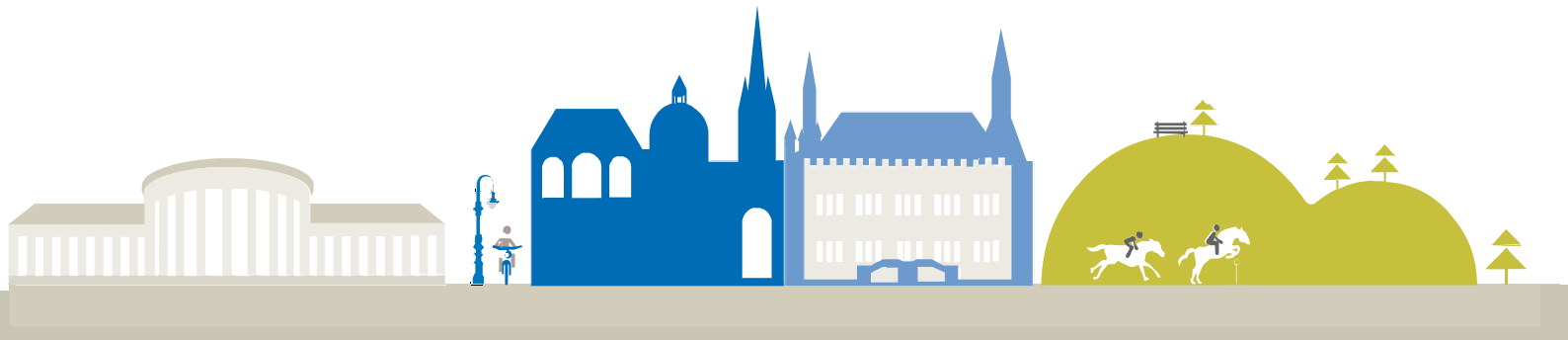
15iFK

FLUID POWER: ENABLING FUTURE TECHNOLOGIES

9TH - 11TH MARCH 2026 | PROCEEDINGS

15TH INTERNATIONAL FLUID POWER CONFERENCE



Editor: Katharina Schmitz
Conference Proceedings, 15th International Fluid Power Conference Aachen
Fluid Power: Enabling Future Technologies

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction in any form, and storage in databases. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Act of 9 September 1965 (Federal Law Gazette I, p. 1273), as last amended by Article 25 of the Act of 23 June 2021 (Federal Law Gazette I, p. 1858), in its current version. Permission for use must always be obtained from HP – Fördervereinigung Fluidtechnik e.V. Violations are liable to prosecution under the German Copyright Act.

This volume constitutes a preprint edition of the conference proceedings and is intended solely for distribution to participants of the conference. The papers included herein have not yet undergone final editorial processing or formal publication. The final, citable versions of the contributions will be published subsequently by the designated publisher.

Germany, Aachen, 2026

WELCOME TO THE 15TH INTERNATIONAL CONFERENCE ON FLUID POWER (IFK)

The current economic and societal environments both present the fluid power community with significant challenges and important opportunities. The mechanical engineering sector faces economic uncertainty, volatile markets, geopolitical complexity, and a persistent shortage of skilled workers. At the same time, climate change and sustainability demands require focused and forward-looking action. Strengthening innovation, competitiveness, and technological progress is therefore essential.

Against this backdrop, the 15th International Fluid Power Conference operates under the motto "Fluid Power: Enabling Future Technologies." As a proven and robust technology, fluid power must be continuously advanced through creative and interdisciplinary approaches to remain a key enabler of future technologies. By bringing together science and industry, the IFK fosters exchange, collaboration, and innovation. In particular, artificial intelligence, machine learning, and digitalization are recognized as decisive drivers for the long-term competitiveness of fluid power in both research and industrial applications.

The IFK is one of the world's largest scientific conferences in fluid power. Since its first edition in 1974 in Aachen by Prof. Wolfgang Backé, it is now a leading forum for high-level scientific exchange and networking across the entire fluid power community. In 2026, the IFK is taking place again in Aachen and will be for the first time jointly organized with the GFPS Ph.D. Symposium, further strengthening international exchange and the integration of young researchers.

This year's program features 4 plenary lectures, 18 overview lectures, and 102 scientific presentations. Speakers from 21 countries underline the strong international character of the conference. Of the contributions, 52% originate from Germany, and 48% from international institutions and companies. In addition, 18 exhibitors will present their latest innovations in the accompanying exhibition.

The conference begins on Monday with the Symposium, focusing on scientific contributions from academia and industrial research. The day concludes with a networking get-together at the accompanying exhibition.

On Tuesday morning, the official opening ceremony will take place, with welcoming addresses and an international overview of current developments in academic fluid power research worldwide. To sharpen the technological look ahead, the following two plenary lectures focus on the role of artificial intelligence and digital twins in fluid power technology and applications. Tuesday's program places particular emphasis on industrial contributions and application-oriented topics, and concludes with a gala dinner in the Coronation Hall of Aachen City Hall, offering a distinguished setting for further exchange.
On Wednesday, the conference continues with additional technical sessions and concludes with a farewell event and the presentation of the Best Paper Awards for both the IFK and the GFPS Ph.D. Symposium.

Finally, we express our sincere gratitude to all committee members, speakers, authors, reviewers, session chairs, and exhibitors for their commitment and support. We wish all participants an inspiring and rewarding 15th IFK in Aachen and hope that this conference will provide valuable impulses for future research, innovation, and collaboration in fluid power technology.





Univ.-Prof. Dr.-Ing. Katharina Schmitz

PROGRAM COMMITTEE

Bauer, F.	Dr.-Ing., Hydac Technology GmbH, Sulzbach/Saar
Beck, M.	Dr.-Ing., Emerson Automation Solutions AVENTICS GmbH, Laatzen
Brahmer, B.	Dipl.-Ing., J.M. Voith SE & Co. KG I VTHL, Stuttgart
Brinkschulte, L.	Dr.-Ing., ARGO-HYTOS GmbH, Kraichtal
Boes, C.	Dr.-Ing., MOOG GmbH, Böblingen
Faß, U.	Dr.-Ing., Volvo Construction Equipment Germany GmbH, Konz
Fiedler, M.	Dr.-Ing., J. Schmalz GmbH, Glatten
Finzel, R.	Dr.-Ing., Wacker Neuson Linz GmbH, Hörsching, AT
Flaig, A.	Dr.-Ing., Bosch Rexroth AG, Lohr am Main
Fuchs, M.	Dipl.-Ing., Bürkert Werke GmbH, Ingelfingen
Gauchel, W.	Dr.-Ing., Festo SE & Co. KG, Esslingen
Geis, C.	Dr.-Ing., VDMA, Frankfurt a.M.
Hunger, J.	Dr.-Ing., Walter Hunger GmbH & Co. KG
Kohmäscher, T.	Dr.-Ing., Claas KGaA mbH, Harsewinkel
Krieg, M.	Dr.-Ing., Bosch Rexroth AG, Lohr am Main
Leifeld, R.	Dr.-Ing., Maschinenfabrik Bernard KRONE GmbH & Co. KG, Spelle
Lüüs, H.	Dipl.-Ing., Bucher Hydraulics GmbH, Klettgau
Merkelbach, S.	Dr.-Ing., Norgren GmbH, Alpen
Rahmfeld, R.	Dr.-Ing., Danfoss Power Solutions GmbH & Co. OHG, Hamburg
Roosen, K.	Dr.-Ing., Parker Hannifin GmbH, Kaarst
Schmitz, K.	Prof. Dr.-Ing., RWTH Aachen
Schultz, A.	Dr.-Ing., Magnet-Schultz GmbH & Co. KG, Memmingen
Schulze Schencking, D.	Dr.-Ing., Hauhinco Maschinenfabrik GmbH & Co. KG, Sprockhövel
Von Dombrowski, R.	Dr.-Ing., G. Siempelkamp GmbH & Co. KG, Krefeld
Wechsel, T.	Dipl.-Ing., HAWE Hydraulik SE, München
Weber, J.	Prof. Dr.-Ing., TU Dresden
Zähe, B.	Dr.-Ing., Sun Hydraulics LLC, Erkelenz

INTERNATIONAL ADVISORY COMMITTEE

Professor Bideaux, Eric	Laboratoire Ampère, INSA Lyon, Villeurbanne, FR
Professor Chiang, Mao-Hsiung	Energy Research Center, National Taiwan University, TWN
Professor De Negri, Victor J.	Federal University of Santa Catarina, Trindade Florianópolis, BR
Professor Ericson, Liselott	Fluid and Mechatronic Systems (FLUMES), Linköping University, Linköping, SE
Professor Frosina, Emma	Department of Engineering (DING), Universita Degli Studi Del Sannio, Benevento, IT
Professor Geimer, Marcus	Institut für Fahrzeugsystemtechnik Institutsteil Mobile Arbeitsmaschinen, Karlsruher Institut für Technologie, Karlsruhe, DE
Professor Handroos, Heikki	School of Energy Systems, Lappeenranta University of Technology, Lappeenranta, FI
Professor Krus, Petter	Division of Fluid and Mechatronic Systems, Linköping University, SE
Professor Kumar, Niranjan	Power Hydraulics & Pneumatics Lab, Indian Institute of Technology, IN
Professor Minav, Tatiana	Faculty of Engineering and Natural Sciences, Tampere University, Tampere, FI
Professor Plummer, Andrew	Department of Mechanical Engineering, University of Bath, Bath, UK
Professor Sanada, Kazushi	Faculty of Engineering Division of Systems Research, Yokohama National University, Yokohama, JP
Professor Schmidt, Lasse	Fluid Power and Mechatronic Systems, Aalborg University, Aalborg, DK
Professor Schmitz, Katharina	Institut für Fluidtechnische Antriebe und Systeme, RWTH Aachen University, Aachen DE
Professor Senatore, Adolfo	Universita Degli Studi Di Salerno, Fisciano, IT
Professor Songjing, Li	Provincial Joint Lab, Harbin Institute of Technology, Harbin, CN
Professor Stelson, Kim	Center for Compact and Efficient Fluid Power, University of Minnesota, Minneapolis, US
Professor Vacca, Andrea	Maha Fluid Power Research Center, Purdue University, Indiana, US
Professor Weber, Jürgen	Professur für Fluid-Mechatronische Systemtechnik, Technische Universität Dresden, Dresden, DE
Professor Yang, Huayong	School of Mechanical Engineering, Zhejiang University, Hangzhou, CN

CONFERENCE ORGANISATION

Sprink, Johannes	Institut für Fluidtechnische Antriebe und Systeme, RWTH Aachen University, Aachen DE
Schlegel, Felix	Institut für Fluidtechnische Antriebe und Systeme, RWTH Aachen University, Aachen DE
Zacharias, Jutta	Institut für Fluidtechnische Antriebe und Systeme, RWTH Aachen University, Aachen DE

PAPER OVERVIEW

DAY 1 09:30-10:30

SYMPOSIUM: OPENING SPEECH

Prof. Dr. Huayong Yang	Development of inspection and replacement robots for excavation tools of Tunnel Boring Machine (TBM)	19

M1 - SYMPOSIUM: MOBILE SYSTEMS

Zezhou Wang	Automated Steering Control for Articulated Off-Highway Machines	35
Dr. Matteo Bertoli	Design and Application of a Prediction Model of the Loads and Mission Profile for Hydraulic Steering System in Heavy-Duty Off-Road Vehicles	55
Jiaming Wu	Power optimization strategy design for electric wheel loader work function with variable speed and pump displacement control	75

M2 - SYMPOSIUM: CONTROLS

Simo Käki	A PID Controller Gain Adjustment Mechanism to Balance EHA Tracking and Health	93
Amy Rankka	Machine Learning Based Controller for Energy-Optimized Operation of Hydraulic Loader Cranes	110
Dr. Junjie Mi	Multidimensional Neural Networks Based RISE Control for Hydraulic Manipulators with Disturbances Compensation	128

M3 - SYMPOSIUM: TRIBOLOGY

Marius Hofmeister	Determination of Gas Solubility for Technical and Flammable fluids	145
Dr. Ana Trajkovski	Evaluating sliding properties of selective laser melted maraging steel against aluminium alloys in water and hydraulic oil	163
Pascal Moor	Performance evaluation of a wear test rig for positive displacement pumps	183

PAPER OVERVIEW

DAY 1 11:00-12:40

M4 - SYMPOSIUM: COMPONENTS PUMPS I

Roman Ivantysyn	Electrification-Ready Tribology:Experimental and Simulation Insights into Wear and Performance of Surface Optimized Valve Plates	197
Matthias Vogt	Analysis of Structure-Borne Noise Emissions of Axial Piston Pumps Using Blocked Forces and Virtual Point Transformation	225
Thomas Heeger	Review of Single-Sided and Double-Sided Axial Piston Machine Characteristics	240
Ishan Suvarna	Optimization of the Cylinder Block-Valve Plate Interface Considering Coupled Interaction from Piston Dynamics and Shaft Deformation	256
Dr. Rongfu Tang	New Prototype of High-Speed External Gear Pump with Filling Oil from Tooth Root	277

M5 - SYMPOSIUM: MOBILE APPLICATIONS

Dr. Lulu Gao	Duty-based Dumping Control for Underground Load-Haul-Dump Vehicle Based on Reinforcement Learning	295
Hongyun Mu	Energy efficiency evaluation of hybrid distributed-centralized load sensing system for electrified excavator	311
Paulo Leonel Teixeira	Innovative fan drive solution for mobile machines	327
Yihan Qiao	An Electro Hydraulic Hybrid System to Downsize Motor Power of Electric Wheel Loader Working Function	344
Yangxiu Xia	Motion Control of Multi-DOF Hydraulic Manipulator Driven by Independent Metering System with Friction Compensation	363

M6 - GFPS SYMPOSIUM: SYSTEMS & COMPONENTS

Marcel Steidle	Direct model reference adaptive control (MRAC) augmented frameworks for load pressure and position control of servo-valve controlled hydraulic cylinders	379
Seshan Calapatti Suresh	Design Overview and Performance Assessment of an Integrated Electro-Hydraulic Actuator (iEHA)	397
Lukas Boden	Non-destructive Detection of Fatigue Cracks in Hydraulic Components	415
Fabian Blatter	Investigation of Fatigue Crack Growth Under High Pressure Gradients with Fluid-Structure Interaction Model	433

PAPER OVERVIEW

DAY 1 13:40-15:00

M7 - SYMPOSIUM: SYSTEMS & ENERGY I

Vili Vironmäki	Feasibility of Electro-Hydrostatic Actuators in Wheel Loader Application: A Case Study Focusing on Energy Efficiency, Controllability and Thermal Behaviour	451
Dr. Bobo Helian	Automation Framework for Constrained Optimal Trajectory Planning and Adaptive Robust Motion Control of Electrohydraulic Mobile Machines	469
Artur Tozzi de Cantuaria Gama	Exploratory Analysis of Optimisation Results for the Multi-Pump System Using Data-Driven Methods	485
Ruiyao Tao	Analysis of High-Pressure Pneumatic Stage Separation System for Reusable Launch Vehicles Based on Real Gas Equation of State	499

M8 - SYMPOSIUM: COMPONENTS VALVES

Yuhang Zhang	Radial Flow Force Reduction for Hydraulic Spool Valve Core using Balanced Orifice Design	515
Krishna Bhaskar	Modelling and Analysis of the Performance for a Pressure Compensated Flow Control Valve	533
Junhao Chen	Thermal–fluid coupling analysis of the Deflector Jet Servo Valve: effects of internal air region, oil viscosity, and coil heat source	549
Timo Rothhardt	Model-Based Valve Design: Potential of CFD Simulations as a Database for Prediction Models	565

M9 - GFPS SYMPOSIUM: PUMPS & TRIBOLOGY

Felix Schlegel	Experimental Comparison of Polymer and Non-Ferrous Metal Slippers for Oil-Hydraulic Axial Piston Machines	583
Avae Béatrice Carla Berger	Modeling and Simulation of Oil Film Influence on Hydraulic Piston Pump Vibrations	601
Yuhan Song	Grain Microstructural Evolution of Friction Layers in Piston Pump Friction Pair	619
Dingchong Lyu	Physics-Informed Neural Network-Based Prediction of Mixed Lubrication Characteristics in Cylinder Block/Valve Plate Interfaces via Mechanism-Data Integration	635

PAPER OVERVIEW

DAY 1 15:30-17:30

M10 - SYMPOSIUM: SYSTEMS & ENERGY II

Harsh Kulkarni	Multi-Objective Optimization of Kfloid-Type Gerotor Pumps for High-Pressure Applications	653
Tobias Vonderbank	Beyond Pilot Control: A Prototype for Direct Actuation of Large Industrial Proportional Valves	669
Dr. Nejc Novak	Innovative Approach to Cleaning and Cleanliness Control of Newly Manufactured Hydraulic Assemblies	687
Christa Düsing	Automated Processing of FMEA-Induced Fault-Injection Simulation for Safety Functions in Dangerous Machinery	707
Dr. Qianpeng Chen	Hydraulic Spool Valve Contamination Susceptibility Test and Visualization Study of Clearance Particle Ingress-Aggregation Phenomenon	725

M11 - SYMPOSIUM: COMPONENTS PUMPS II

Prof. Dr. Gudrun Mikota	Performance analysis of rotary mechanical pulsation compensators for different types of pumps	741
Alfonso Rosario Apuzzo	Enhancing Volumetric Efficiency and Productivity of External Gear Pumps through Optimized Run-In Strategies.	757
Svenja Horn	Influence of surface structuring on the hydrostatic and hydrodynamic pressure distribution of slipper sealing lands	775
Liangyu Song	Effects of contamination on the wear process of piston/cylinder interface in axial piston machines considering three-body abrasive wear model	795
Dr. Komsan Ngamkham	Hydraulic Oil Temperature in Bridge Circuit and Hydraulic Bent Axis Piston as Load	811

M12 - GFPS SYMPOSIUM: MOBILE APPLICATIONS

Mohamed Allam	High-Resolution Force Control in a Multi-Pressure Hydraulic System for Excavator Work Functions	823
Dr. Zhixin Yan	Research on the Performance of a Heavy-Duty Robotic Arm Featuring Pump-Valve Coordination and Independent Metering Control with Hydraulic-Pneumatic Energy Storage	841
Bernhard Sender	Operator effect on the performance of a wheeled excavator	859
Róger Mateus Sehnem	Optimal preliminary design of electro-mechanical actuators over drive-cycle for mobile machinery	875
Daniel Kriegl	Universal and Highly Efficient Speed Control of the Hydraulic Motor in a Wire Saw Unit for a Hybrid Constant Flow and Constant Pressure System	891

PLENARY LECTURES

DAY 2 09:00 - 12:30

Prof. Dr. Katharina Schmitz (ifas), Dr. Ralf Otten (Mayor Aachen), Prof. Dr. Wolfgang Schröder (Dean RWTH), Dr. Christian Geis (VDMA)	Opening	
Prof. Dr. Huayong Yang (ZJU), Prof. Dr. Andrea Vacca (Purdue), Prof. Dr. Marcus Geimer (KIT)	Fluid Power Research Worldwide	
Dr. Steffen Haack (Bosch Rexroth)	AI in Fluid Power Technology	
Prof. Dr. Katharina Schmitz (ifas)	Digital Twins and AI in Fluid Power: Improving Systems by Connecting Real and Virtual Worlds	907

PAPER OVERVIEW

DAY 2 13:30-14:40

T1 - CONFERENCE: CONTROLS

Dr. Dirk Wroblewski	Next generation of intelligent electro-hydraulic control valve actuation	925
Filip Vlach	Push-pull electromagnet in hydraulics	939
Dr. Siegfried Rotthäuser	Recuperation on a transmission test bench	949

T2 - CONFERENCE: COMPONENTS PUMPS

Patrick Fischer	Next Generation Fixed Displacement Pumps for Electrified Mobile Applications	959
Alexandre Vucemilovic	Experimental study of the influence of the channel geometry of regenerative pumps on their internal flow	977
Nandakumar P S	Navigating Electrification Challenges: Low speed behavior of external gear units	995

T3 - CONFERENCE: ARTIFICIAL INTELLIGENCE

Faras Brumand-Poor	Geometry-Aware Physics-Informed Neural Network: A Framework for Solving the Reynolds Equation with Starvation for Lubricated Contacts	1013
Vaishnavi Kulkarni	Information Extraction from Various Formats for Engineering Decisions	1029
Maximilian Romeser	Enabling Vibroacoustic End-of-Line Testing of APUs Across Varying Fluid Temperatures through ML-based Conditional Anomaly Detection	1047

PAPER OVERVIEW

DAY 2 15:30-17:00

T4 - CONFERENCE: MOBILE SYSTEMS

Fabian Kreutmayr	Case Study on Independent Metering Hydraulics for Energy-Efficient Battery-Electric Telescopic Loaders	1065
Dr. Marcel Neu	Efficient holistic system simulation method for the analysis of thermal-management concepts for electrified construction machinery through inverse hydraulic modelling	1083
Prof. Dr. Lasse Schmidt	On Displacement Controlled Drive Networks	1099

T5 - CONFERENCE: COMPONENTS VALVES

Dr. Kristof Schlemmer	Development of a flow-optimised slip-in servo cartridge valve for high-performance applications	1115
Dr. Seiei Masuda	Modeling and Simulation of Damping Orifices for Pressure Control Valves (2nd Report)	1131
Keitaro Kamata	The optimal design of hydraulic pump performance and reliability through lubrication, mechanism, and structural coupling analysis	1149
Harsh Maniar	Performance Improvement of Pressure Compensated Flow Control Valve Using Virtual Engineering	1159

T6 - CONFERENCE: FLUIDS

Wolfgang Bock	Fluid Solutions (Innovative and sustainable hydraulic fluid concepts)	1171
Klaus Ellenrieder	Evaluating Hydraulic Fluids for Performance and Sustainability: Bosch Rexroth's Approach	1179
Sebastian Deuster	Influence of Fluid Mechanical Properties of Hydraulic Oils with Different Chemical Structur on the Efficiency of Hydraulic System Components	1189
Dr. Dmitriy Shakhvorostov	Advanced Hydraulic Fluids Make Mobile Machinery Not Only More Sustainable, but Also More Economical	1205

PAPER OVERVIEW

DAY 3 09:00-10:30

W1 - CONFERENCE: DIGITAL TRANSFORMATION

Dr. Mark Krieg	Digitalization & Standardization: Enablers for the Future of Industrial Hydraulics	1217
Dr. Vinay Partap Singh	Machine Learning- Based Health Monitoring of High-Power Electro-Mechanical Actuator for Non-Road Mobile Machinery Using Experimental Data	1225
Moritz Neltner	Digitaliztion in Sustainability for Fluid Power Components and Systems	1243
Simon Knoll	Predicting leakage and operating conditions in axial piston pumps using vibration-based machine learning	1257

W2 - CONFERENCE: MOBILE APPLICATIONS I

Thorsten Stiller	Improved System Efficiency thorugh Drained Case Operation of Axial Piston Units	1275
Dr. Fabio Scolari	Real time fault detection in hydraulic system for wheel loader application	1287
Sai Kishore Ayalasomayajula	Enhancing Energy Efficiency in Drilling: Innovative Pipe Handling with Hydraulic Energy Recovery	1305
Dr. Matteo Pellegri	Dextreme MAX: A Digital Displacement® Excavator System for Energy Recovery and Reduced Losses	1323

W3 - CONFERENCE: PNEUMATICS I

Dr. Wolfgang Gauchel	Environmental Sustainable Pneumatics	1341
Vladimir Boyko	Real-Time Fault Detection in Pneumatic Drive Applications Through Pattern Classification	1365
Prof. Dr. Olivier Reinertz	Quasi-Isobaric Compressed Air Reservoirs	1381
Prof. Dr. Zeljko Situm	Integration of Pneumatic Systems in Training Devices for Combat Sports	1399

PAPER OVERVIEW

DAY 3 11:00-12:30

W4 - CONFERENCE: NET ZERO

Prof. Dr. Lasse Schmidt	Estimating the Global Emissions by Fluid Power Machinery	1411
Dr. Andreas Schumacher	Decarbonization in mobile machinery – from reducing the carbon emissions to emission-free drive-trains	1427
Felix Ruthmann	Calculating use-phase greenhouse gas emissions of hydraulic components	1439
Heiko Steinbach	A Holistic Approach Towards Energy Efficient Excavators	1457

W5 - CONFERENCE: INDUSTRIAL

Dr. Johannes Gattinger	Innovative self-sufficient clamping system for machine tools	1469
Dr. Zhiqiang He	Digital Architecture Design and Implementation for Performance Optimization of Electromechanical-Hydraulic Systems: A Case Study on Injection Molding Machines	1479
Dr. Mark-Patrick Muehlhausen	Downsizing of hydraulic power units by accurate prediction of disperse flows	1497
Reiner Stammberger	Sensor-Efficient Health Monitoring for Brownfield Industrial Hydraulic Systems	1507

W6 - CONFERENCE: PNEUMATICS II

Dr. Harald Kuolt	Energy-efficient handling through passive vacuum generation and biomimetic functional principles	1525
Philipp Hüpper	Development and Experimental Validation of a Novel Origami-Inspired Pneumatic Actuator Concept with Progressive Force-Stroke Characteristics	1541
Micha Koebele	AAS-Orchestrated Production – End-to-End Digital Integration from Machine to Visualization	1557
Cong Liu	CFD-Based Optimization of a Supersonic Vacuum Ejector Using Combined 2D–3D Simulations and Latin Hypercube Sampling	1573

PAPER OVERVIEW

DAY 3 13:30-15:00

W7 - CONFERENCE: FLUID 4.0

Paul Remde	Digital fluid technology in the data space: From semantic component search to the development of mobile hydraulic systems using AAS	1591
Fabian Geibel	Vertical Integration of Energy Data - an Architectural Evaluation of Asset Administration Shell (AAS)	1613
Dr. Heiko Baum	System Simulation Meets Asset Administration Shell – Potentials and Challenges of a Standardized, Model-Based Development of Fluid Power Systems	1631
Barkha Sharma	Unlocking Semantic Interoperability in Industry with Large Language Models	1649

W8 - CONFERENCE: MOBILE APPLICATIONS II

Heiko Laffrenzen	Best of both worlds – load dependent and load independent hydrostatic driving switchable on-demand with innovative hybrid electro-proportional control	1667
Ralph Wassermann	High efficient Hydraulic Actuator Valve Sensor Unit without dedicated hydraulic supply	1683
Felix Figge	EHAs in mobile machinery: Experimental validation of a high-speed EHA with recuperation and load-holding capability	1699

W9 - CONFERENCE: SEALS

Dr. Niklas Bauer	Beyond PFAS: Challenges and Opportunities in Material Innovation for replacing PTFE in Hydraulic Seals	1717
Jiehao Wang	Percolation Analysis of Reciprocating Sealing Surfaces Based on the Statistical Evolution of HPD and PSD Induced by Wear	1735
Diego Manuel Chamorro	Wear modeling and sensitivity analysis for seals in hydraulic cylinders.	1753

PAPER OVERVIEW

DAY 3 15:30-17:00

W10 - CONFERENCE: DATA & SENSORS

Markus Fritton	Novel ultrasonic position sensor for hydraulic clamping cylinders	1771
Dr. Benedikt Müller	From Pressure Pulses to Noise Sources – NVH-Oriented Design of Hydraulic Systems	1783
Olaf Kissing	Current developments in magnetostrictive position sensors	1801

W11 - CONFERENCE: ADDITIVE MANUFACTURING & MATERIALS

Zita Tappeiner	Additive Manufacturing in Hydraulics: Design Methodologies for Manifolds	1811
Jan Bartolj	3D-Printed Polymer and Ceramic Spools as Alternatives to Steel for Hydraulic Valves	1829
Johannes Klotz	Hydraulic Substitute Base for Demolition Hammers	1847

W12 - CONFERENCE: SPECIAL APPLICATIONS

Prof. Dr. Mathias Niebergall	Hydraulic Artificial Muscles for Exoskeleton Actuation Solutions	1863
Dr. Christian Lessmeier	New hydrogen fuel lines from carbon steel tubes and their qualification for high-pressure hydrogen applications	1881
Dr. Enrico Pasquini	From Turbulence to Tranquillity: A Multidisciplinary Blueprint for Optimising Liquid Cooling in Precision Machinery	1891
Weihang Xi	Research of the cavitation in the rear brake chamber of hydraulic rock drill	1907

THE OPTIONAL PEER REVIEW

Many public institutions that support research projects require a regular publication of their results. To ensure that these results are in fact of scientific value, the assessment by a third party is often required. This process is known as peer-review and due to its demanding logistics is increasingly rare. The Institute for Fluid Power Drives and Systems (ifas) and organizers of the 15th IFK in Aachen wanted scientists to have the opportunity to have their work reviewed accordingly, so as to demonstrate its scientific value.
By tradition, the IFK unites scientists with industry in an international forum to exchange their knowledge. Therefore, a peer-review only makes sense for some of the authors. At this juncture, ifas would like to point out that the peer-review is intended in no way to classify the papers, but only to support the need for review.
The reviewed papers and the corresponding presentations are tagged with the sign below in the Scientific Programe.



Development of inspection and replacement robots for excavation tools of Tunnel Boring Machine (TBM)

Huayong Yang[1,2], *Cheng Wang*[1,2], *Haibo Xie*[1,2]

1 State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou, China

2 Institute of Advanced Machines, Zhejiang University, Hangzhou, China

Abstract.

The cutterhead of a tunnel boring machine (TBM) operates in a confined, high-pressure, high-humidity environment where manual inspection and replacement of excavation tools is risky and time-consuming. To enable robotised maintenance, we developed a TBM inspection-and-replacement system composed of a snake-like inspection robot and a heavy-load disc-cutter replacement robot. For inspection, a compact cable-driven snake robot is designed to work at pressures up to 2 MPa and temperatures up to 65 C, integrating cleaning, lighting, and zoom imaging to achieve high-precision tool wear assessment. For replacement, a 6-DOF PRRPRR manipulator driven by electro-hydrostatic actuators (EHAs) is developed for heavy-load, constrained-space operation. A position-pressure feedback control (PPFC) method is proposed to improve the system response and position accuracy. Field deployment on the 'Pioneer II' TBM in the Shenzhen Airport metro project validates the robustness and effectiveness of the system.

Keywords. TBM maintenance robots; snake-like inspection robot; disc cutter replacement; electro-hydrostatic actuator; trajectory planning; adaptive control.

1. INTRODUCTION

Urban rail transit expansion and large-diameter tunnel construction have led to a sharp increase in TBM usage. Disc cutters on the cutterhead suffer severe wear and damage due to rock-breaking loads, and their condition directly affects excavation efficiency and safety.

In deep-buried and high-water-pressure projects, manual inspection and replacement require working under high pressure, poor visibility, and muddy conditions, which introduces significant safety risks and long downtime. Engineering statistics report that tool inspection and replacement can occupy a substantial fraction of the excavation schedule and a large share of project cost, while a large proportion of TBM-related accidents are associated with manual cutter maintenance.

Robotised maintenance is therefore essential to reduce human exposure and improve machine utilization. A complete system must provide reliable inspection, accurate wear assessment, and safe replacement in a confined environment, while coordinating with the TBM control architecture.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Recently, research institutes and companies worldwide have started to develop robotic systems for cutterhead inspection and tool replacement to overcome the limitations of manual in-chamber operations and to move toward robots instead of humans. Herrenknecht has developed an auxiliary cutter inspection robot [1]. The main body uses a folding arm, with lighting, a camera, and a high-pressure water device at the end, plus an industrial endoscope that can pass through the cutterhead to observe tools; the system is still under development. OC Robotics built a snake-like inspection prototype, JetSnake [2], and tested it in the Port of Miami and the Hong Kong International Airport projects. Public information indicates that their tests were performed only at atmospheric pressure, without reliability validation under high-pressure and high-humidity conditions [3]. The University of Tokyo designed a cutter inspection robot for a 5 m class TBM, WInBot [4], with a telescopic structure [5] that can extend 2.7x vertically and 1.5x horizontally, enabling operation in narrow spaces around the cutterhead. However, its limited DOF allows inspection in only one direction, and no field application has been reported. CREG invented a shield machine inspection robot system [6] using a telescopic design to wash and inspect cutters. CSIC Haiwei Zhengzhou High-Tech Co. proposed a snake robot for shield inspection [7] that remains at the proof-of-concept stage. Beijing Institute of Petrochemical Technology developed a folding-arm cutterhead cleaning robot [8] that can clean cutters with a high-pressure water gun and observe the excavation chamber, but it did not address the complex soil chamber environment and has limited obstacle avoidance, with no follow-up development. Shanghai Jiao Tong University studied a snake-arm inspection robot [9], but the work remains at simulation level and does not consider the high-pressure and high-humidity TBM soil chamber.

The earliest exploration of robotic cutter replacement was by Bouygues, which launched the TELEMACH project in 2007 with funding from the French National Research Agency. After years of development, TELEMACH was demonstrated in June 2015 during the Tuen Mun to Chek Lap Kok tunnel project in Hong Kong using a 17.6 m diameter Herrenknecht TBM. The TELEMACH replacement robot is based on a KUKA KR FORTEC industrial robot with 500 kg payload and a dedicated end effector for cutter replacement. The main contribution was a new cutter system for robotic replacement and supporting tooling for cutter assembly and disassembly. However, because a standard industrial robot was used, the system is large and poorly suited to confined spaces; it can only be installed on large-diameter TBMs and is not applicable to small-diameter machines [10-11]. Shanghai Hongrun Construction Co. and the Shanghai University Robotics Institute developed a multi-DOF manipulator in which a six-DOF arm is mounted on a three-DOF linear track. All joints are motor-driven and the end payload exceeds 200 kg [12-13].

This paper consolidates the design and validation of a complete inspection-and-replacement robot system for TBM cutterhead maintenance. The system integrates a snake-like inspection robot for cleaning and visual inspection, and a heavy-load replacement robot for cutter swapping. We present the mechanical design, workspace planning, kinematics and dynamics modeling, EHA-driven control strategies, and laboratory/field verification.

2. SYSTEM ARCHITECTURE

The TBM robotised maintenance system consists of: (1) a snake-like inspection robot responsible for cleaning and imaging cutter tools, and (2) a heavy-load replacement robot

for removal and installation of disc cutters. The inspection robot provides real-time assessment of cutter wear and damage, supporting maintenance decision-making, while the replacement robot executes planned trajectories to swap cutters through the access gate.

The workflow of the TBM robotised maintenance system is shown in Fig. 1. Embedded tool sensors monitor cutter condition online and provide preliminary fault locations. The supervisory computer decides whether in-chamber inspection is needed. If required, the inspection robot enters the chamber, moves near the target cutter, cleans it, and performs visual inspection. Based on the visual results, the system decides whether replacement is necessary; if damage is confirmed, the replacement robot enters the chamber to swap the cutter.

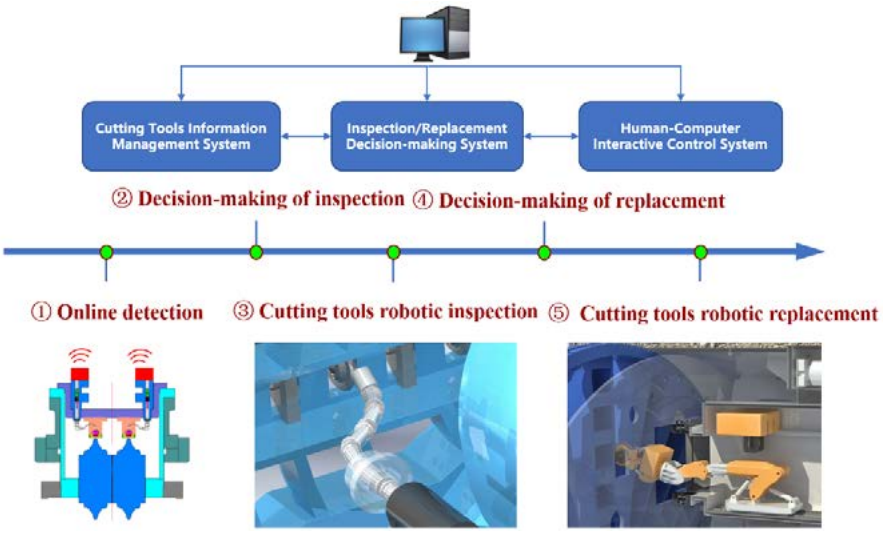


Fig. 1 Workflow for TBM cutter inspection and replacement

3. SNAKE-LIKE INSPECTION ROBOT

The snake-like inspection robot is cable-driven and segmented to enable long reach and high dexterity. Actuators are located in the rear module to avoid direct exposure to high pressure and water. The front section integrates a cleaning module, lighting, and an industrial zoom camera for in-situ inspection.

3.1. Inspection Coverage and Key Parameter Design

To support installation and testing on a TBM, we cooperated with CREG to design an inspection-robot prototype for a 6 m diameter EPB shield. Because various devices are installed behind the bulkhead of a typical 6 m EPB TBM, there is no extra space for a robot cabin. After discussions with CREG engineers, the available cabin space was determined to be a cylinder 600 mm in diameter and 2600 mm in length, as shown in Fig.2. The cabin center is 2207 mm from the TBM center, and the line connecting the two centers forms a 31 deg angle with the horizontal.

4

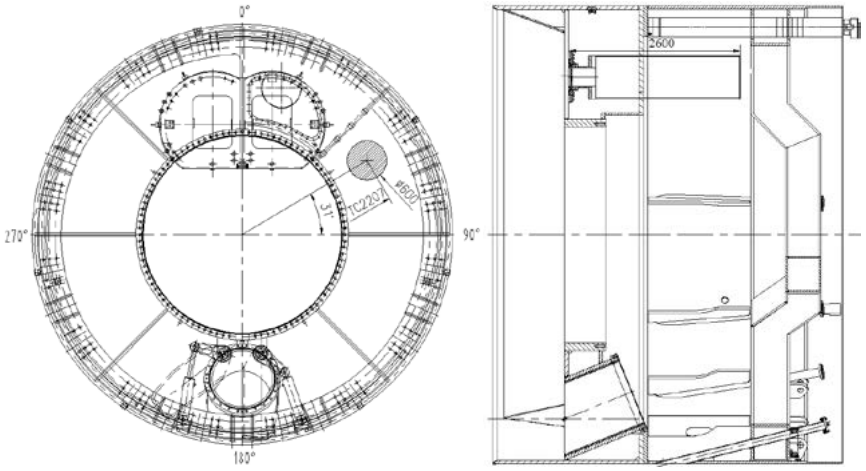


Fig. 2 Installation space of the inspection robot cabin in a 6 m class TBM

The snake-like inspection robot consists of a manipulator section and a drive module. The arm length determines the inspection reach, while the drive module length limits the maximum cable displacement and thus the joint angle range. To maximize inspection coverage within the limited installation space, we analyzed the robot's motion envelope inside the sealed chamber, obtained the maximum cable displacements in the drive space, and determined the screw stroke accordingly. This process finalized the key parameters of the snake robot.

Each joint of the snake robot has two degrees of freedom and can bend in arbitrary directions. Therefore, its workspace is symmetric. By analyzing planar motion, the maximum inspection range inside the TBM can be determined. Fig. 3 illustrates the workspace analysis of the snake robot in the XOY plane. The origin O is at the cabin door center and is 2207 mm from the TBM center. The Y-axis is aligned with the robot's centerline and points toward the cutterhead; the X-axis points to the cutterhead center. The distance from the cutter back edge to the bulkhead (along the X-axis) is 1300 mm. The wrist joint is denoted as a_n and the end link as a_{n+1}. The robot must pass through a narrow hatch between the robot cabin and the sealed chamber; this hatch is installed in the bulkhead and can be designed as an electric gate or replaced by a large-diameter ball valve. As shown in Fig. 3 the cabin outlet has a size D, and the robot must avoid interference with this opening during inspection. For cleaning and inspection, the robot should adopt a posture behind the cutterhead; an ideal pose is with the end link normal to the cutterhead and 300 mm from the cutter back edge. The inverse kinematics is solved to keep the end link horizontal (aligned with the Y-axis) while approaching edge and center cutters as close as possible to the 300 mm standoff. If a perpendicular end pose is infeasible, the end joint is tilted by a certain angle to observe the cutters. Using the optimal inverse-kinematics method, the maximum inspection range in the XOY plane is obtained.

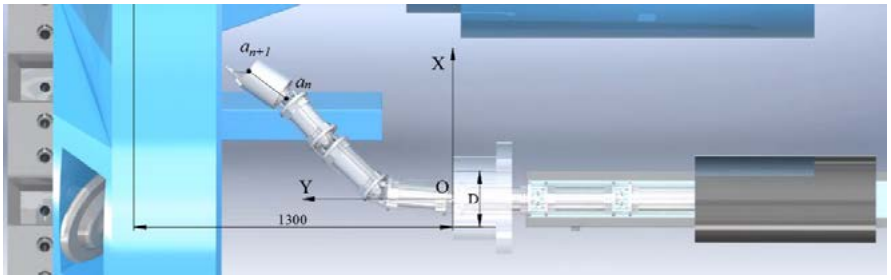


Fig. 3 Inspection range of the snake robot in the XOY plane

Equation (1) gives the inverse-kinematics model. The total length of the robot arm section is set to 1900 mm. The objective function drives the end-effector to stay as close as possible to the line Y = 1000, with constraints that prevent any link from interfering with the hatch during exit. The initial slope vector of the end link is set to k = [0 1 0]^T When the inverse-solution algorithm cannot continue, *k* is adjusted so that the end link adopts a tilted pose to observe the cutters.

$$\begin{aligned} \min \quad & \left\| y_{a_{n+1}} - 1000 \right\| \\ s.t. \quad & \begin{cases} q_i \le \max JointAngle, i = 1,...,n \\ x_{a_1} = 0 \,\&\, z_{a_i} = 0, i = 1,...,n+1 \\ -1030 \le x_{a_{n+1}} \le 2200 \\ \left\| a_i - a_{i+1} \right\| = L_i, i = 1,...,n \\ a_{n+1} = a_n + k \cdot L_n \\ if \;\; y'_{a_i} < 0, then \;\; \dfrac{x_{a_{i+1}} - x_{a_i}}{y_{a_{i+1}} - y_{a_i}} \cdot \left| y_{a_i} \right| + x_{a_i} \le \dfrac{D}{2}, i = 1,...,n \end{cases} \end{aligned} \tag{1}$$

Using the optimal inverse-kinematics method, the inspection range in the XOY plane is obtained as shown in Fig. 4 The robot can keep the end link perpendicular to the cutterhead and 300 mm from the cutter back edge within x = -1030 to x = 1300 mm,this is defined as the complete inspection range. When x > 1300 mm, the end link can be tilted by a certain angle for cleaning and inspection; this is defined as the tilted inspection range. At the limit, the distance between the end link and the back edge of the center cutter is 600 mm.

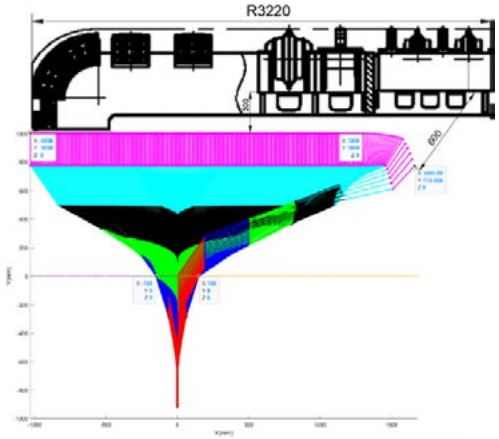


Fig. 4 Inspection range of the snake robot in the XOY plane for a 6 m class TBM

The maximum displacement of the drive cable is 150 mm. Considering a 30 mm slider length in the screw-nut pair and a safety margin, the screw length is set to 200 mm. The motor and coupling length is 180 mm; the fixed link length is 140 mm (equal to the hatch thickness); and a 150 mm cable space is reserved behind the robot. The total arm length is thus 1920 mm. The minimum end-link length is 220 mm, so the remaining five links are each 340 mm long. The overall dimensions are shown in Fig. 5.

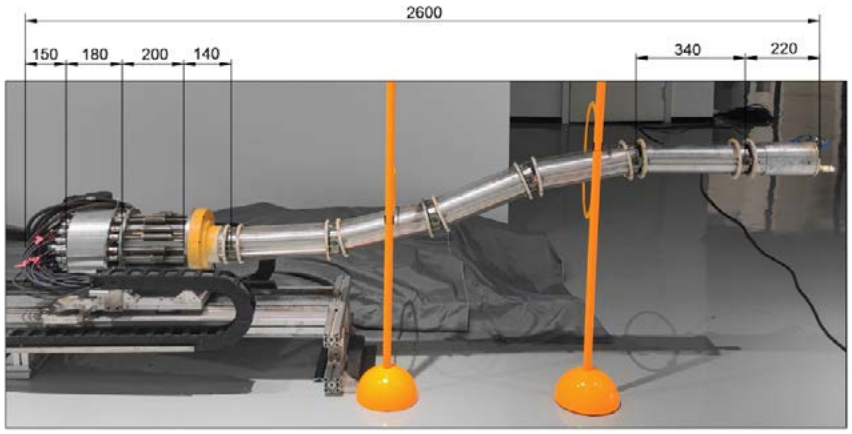
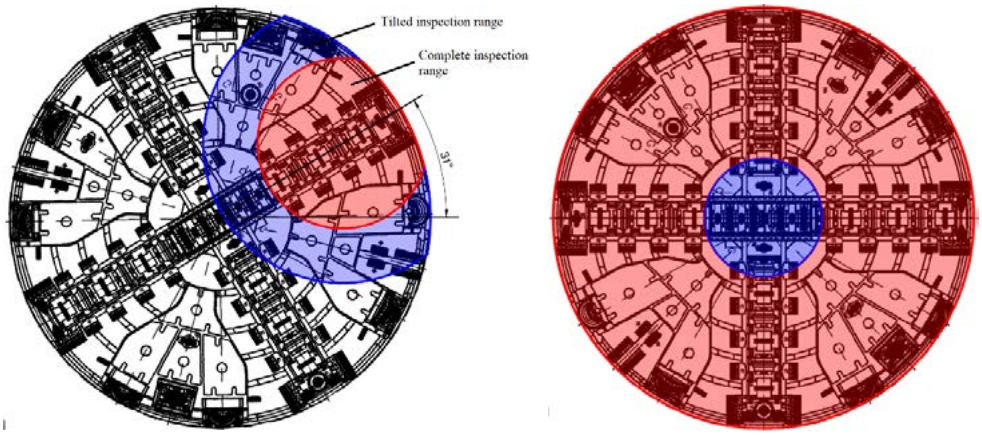


Fig. 5 Dimensions of the TBM snake-like inspection robot

After determining the design parameters, the cutterhead inspection coverage is calculated using this method, as shown in Fig. 6 (a), where the red area is the complete inspection range and the blue area is the tilted inspection range. In practice, cutterhead rotation can be coordinated with the robot to inspect more cutters; the resulting coverage is shown in Fig. 6 (b).By computing the ratio of the inspection area to the total cutterhead area under different conditions, Table 1 is obtained. The results show that full cutterhead coverage can be achieved when inspection is coordinated with cutterhead rotation.



(a) Inspection range with cutterhead stationary (b) Inspection range with cutterhead rotation

Fig. 6 Cutterhead inspection coverage of the snake robot

	Complete inspection range	Tilted inspection range	Total inspection range
Cutterhead stationary	15.11%	16.82%	31.93%
Cutterhead rotating	92.19%	7.81%	100%

Table 1 Cutterhead inspection coverage for a 6 m class TBM

3.2. End-effector cleaning and sensing

To ensure reliable inspection in muddy conditions, the end-effector integrates a self-balancing high-pressure nozzle (35 MPa) with one forward jet and three backward jets (120 deg spacing). The nozzle diameter is 1.0 mm with an orifice diameter of 0.68 mm, delivering about 30 L/min while keeping recoil below 10 N.

High-power lighting, a zoom industrial camera, and an air-knife/laser module are used to clean and illuminate the target area before image capture. The combined cleaning and imaging arrangement allows accurate detection of wear, cracks, and missing cutter components.

3.3. Protection Design in Harsh Environments

The harsh-environment protection design of the TBM snake inspection robot is shown in Fig. 7. The industrial camera at the end effector uses a pressure housing rated to 2 MPa so the camera operates at normal pressure and humidity. The magnetic rotary encoder boards inside the joints are potted, isolating them from high pressure and humidity without compromising measurement accuracy. In addition, the arm is enclosed by a pressure-resistant bellows that runs through the entire manipulator. A double-seal design is used at the base: the inner seal clamps the bellows root, and the outer seal connects to the base, further ensuring safe and reliable operation of precision joint components and magnetic rotary encoders.

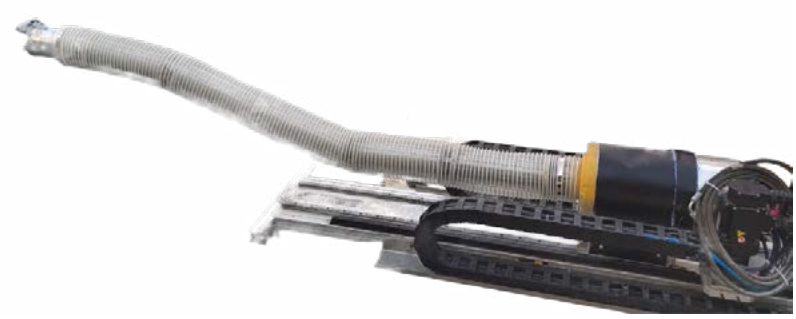


Fig. 7 Protective design of the snake robot for harsh environments

4. HEAVY-LOAD DISC CUTTER REPLACEMENT ROBOT

The replacement robot is a 6-DOF PRRPRR serial manipulator, as shown in Fig.8. Joint 1 is a linear slide for horizontal positioning; joint 4 is a telescopic arm for depth adjustment; joints 2 and 3 are hydraulic cylinders for boom angle and reach; and joints 5 and 6 are rotary swing cylinders for end-effector orientation. The design provides high payload capacity and sufficient stiffness for cutter handling.

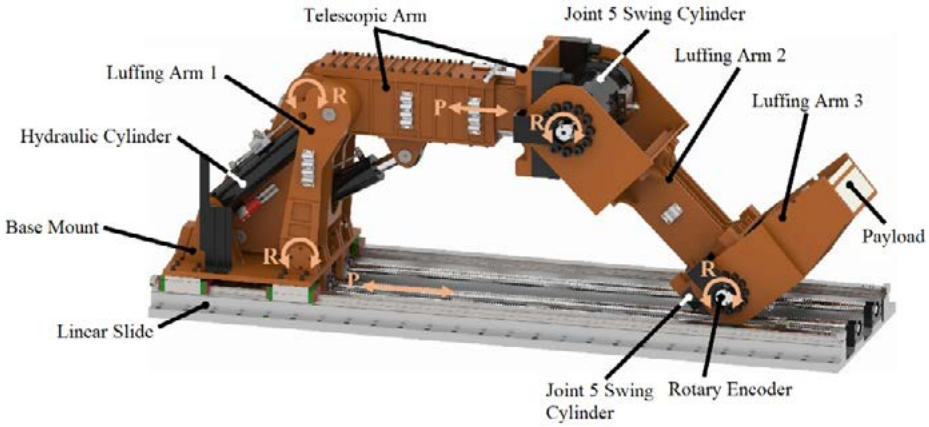


Fig. 8 Mechanical configuration of the replacement robot (joint layout)

4.1. Installation space and access

The robot is designed for an installation space of approximately 1 m x 1 m x 3 m and passes through a 0.65 m diameter, 0.39 m long access gate, as shown in Fig.9. The layout of the TBM chamber, partitions, and access door constrains robot motion. Manual maintenance under these conditions is hazardous, motivating robotised replacement.

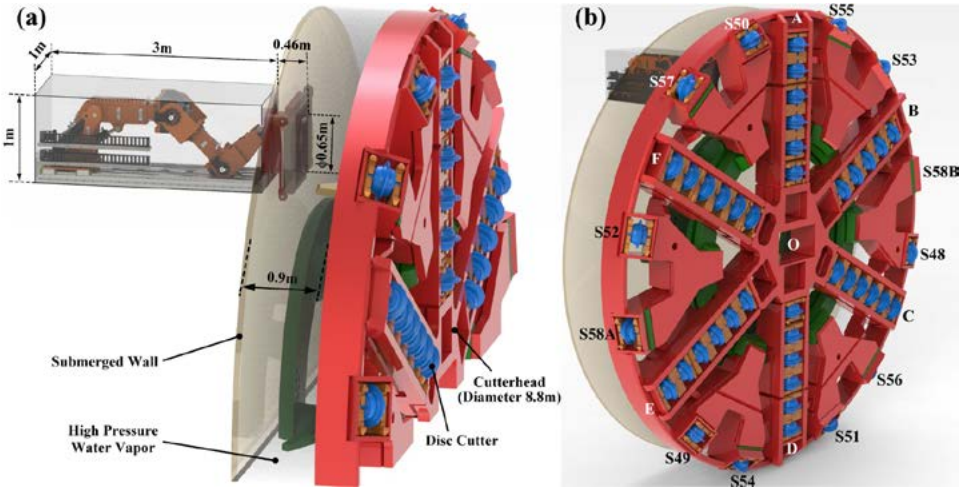


Fig. 9 TBM chamber layout and access positions for cutter replacement

4.2. Kinematics and dynamics modeling

A D-H based kinematic model is established to compute forward and inverse kinematics, as shown in Fig.10. And the D-H parameters table for the replacement robot is shown in Table 2. Redundancy is resolved by constraining joints 2 and 3 to reduce energy consumption and avoid collisions in the confined workspace. The cutterhead of an 8.8 m EPB TBM includes 35 radial cutters and 12 edge cutters, all of which must be reachable.

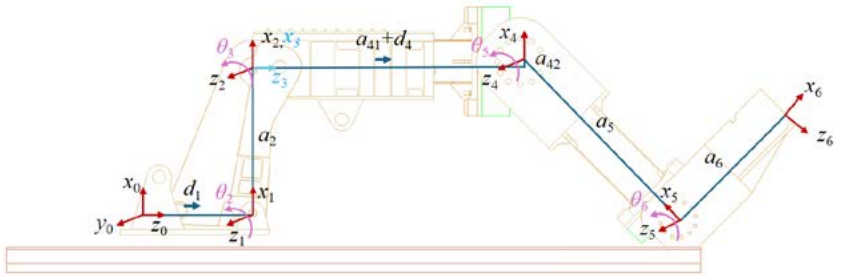


Fig. 10. Coordinate frames and kinematic parameters of the replacement robot

	θ_i	d_i	a_i	α_i
Joint 1	0	d_1	0	-90°
Joint 2	θ_2	0	a_2	0
Joint 3	θ_3	0	0	90°
Joint 4	0	$a_{41}+d_4$	a_{42}	-90°
Joint 5	45°+θ_5	0	-a_5	0
Joint 6	-90°+θ_6	0	a_6	90°

Table 2. D-H parameters table for the replacement robot

4.3. Constrained-space trajectory planning

A region-based joint-space planning strategy is proposed. Cutters are divided into middle, lower, and upper regions. For each region, a small number of key poses (initial, one or two intermediate poses, and target pose) are selected to minimize joint reversals, especially for joints 2 and 3. Trajectories are generated using cubic spline interpolation to ensure smooth velocity and acceleration profiles. Fig.11 shows cutter replacement process for multiple tool locations.

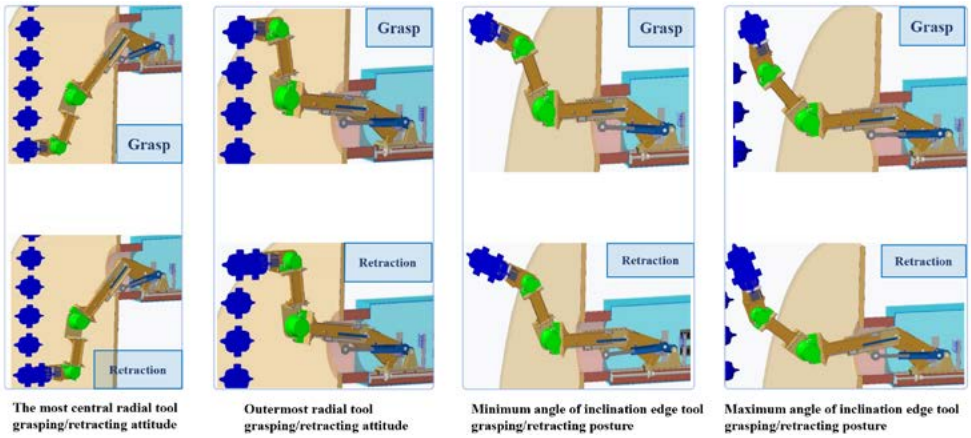


Fig. 11 Demonstration of cutter replacement process for multiple tool locations

4.4. EHA Modeling and Control

Electro-hydrostatic actuators (EHAs) drive joints 2 to 6. As shown in Fig. 12，it includes two major parts, the first part is mainly comprised of a servomotor, an EHA pump, a pressurized tank and a series of cartridge valves, mounted on the manifold block. And the second part is an integrated cylinder, which includes an internal magnetostrictive displacement sensor and an external dual pilot operated check valve. The position control strategy using both position feedback and pressure feedback is proposed. Here, the high gain pressure loop only comes into effect at the beginning of reversing. Hence the pressure in the contained volumes can be raised as soon as possible, which is beneficial for increasing the response speed of the actuator.

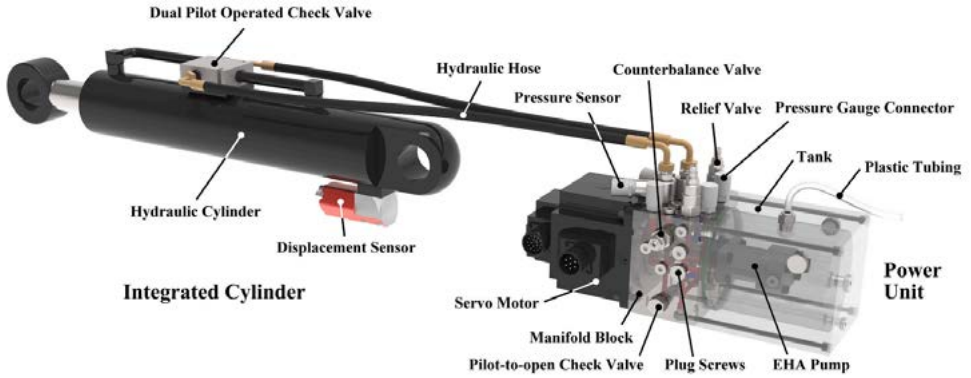


Fig. 12 Global 3D view of the electro-hydrostatic actuator (EHA)

Fig. 13 is the schematic diagram of the presented EHA system, and the corresponding 3D structure diagram is shown in Fig. 12. In this paper, the integrated cylinder is separated from the power unit, which is connected by hydraulic hoses. The integrated hydraulic cylinder is installed on the manipulator and is located in its installation space, while the power unit of the EHA is installed outside the manipulator's installation space.

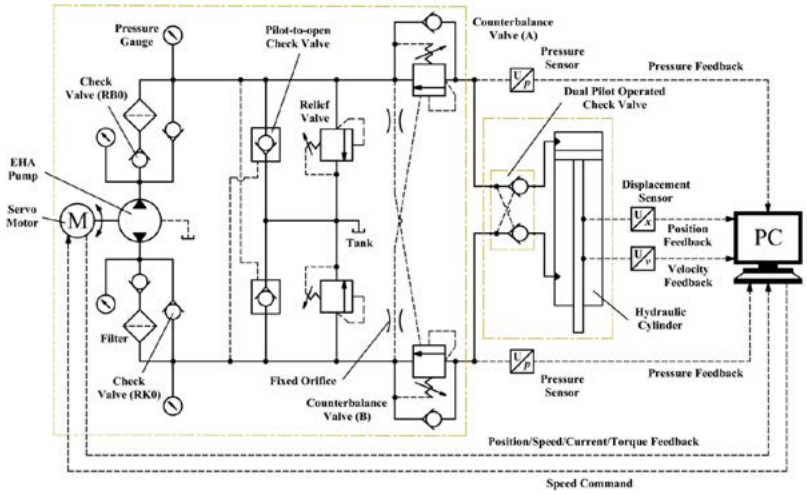


Fig. 13 Schematic diagram of the EHA

The control strategy combines two control loops, a high-gain pres-sure loop and a traditional position loop, as shown in Fig. 14. The goal of the high-gain pressure loop is to increase the speed of system response by increasing the pressure build-up speed of the working chamber when the piston rod moves in reverse, where the pressure controller is a high gain P controller. The position loop is used to realize the precise position control of the actuator, where the position controller is a PD controller. Adopting position-pressure feedback control (PPFC) strategy is to achieve both high positioning accuracy and high dynamic response of the pump control system.

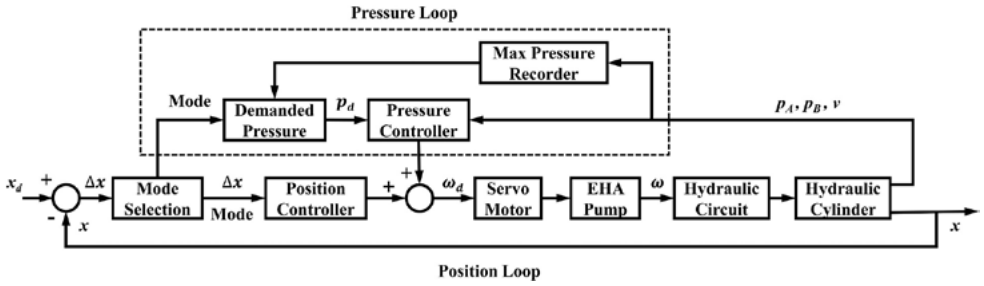


Fig. 14 Block diagram of the control strategy for the EHA

5. FIELD APPLICATION

The inspection-and-replacement system was deployed on the 'Pioneer II' TBM in the Shenzhen Airport metro tunnel project. The TBM operated in mixed geology. The overburden depth along the interval tunnel ranged from 10.13 to 265.81 m, with a maximum longitudinal gradient of 28%. The minimum horizontal curve radius was 568 m. The tunnel section passed through 3965 m of moderately to slightly weathered granite (87% of the alignment), while the remaining sections were gravelly cohesive soil and completely to strongly weathered granite, crossing seven fault zones and four densely jointed zones.

The engineering application of the inspection/replacement robots was first evaluated at the Foshan assembly plant to verify installation inside the TBM and to conduct cutter inspection tests, shown in Fig.15.

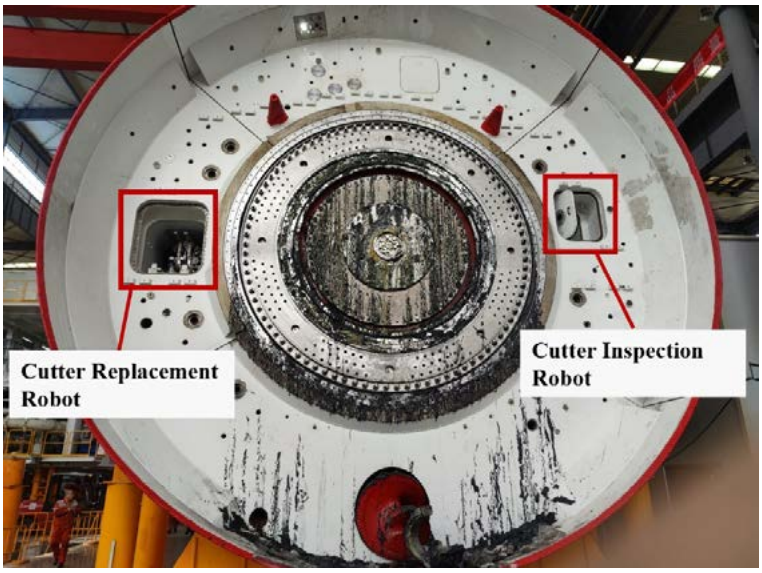


Fig. 15 The inspection-and-replacement system deployed on the 'Pioneer II' TBM

After installation, the snake robot was controlled via a virtual-environment human-machine interaction platform to conduct cutter inspection tests shown in Fig. 16. The platform includes a 3D cutterhead model; the cutterhead status displayed in software together with the robot's real-time posture was used to determine the relative position between the robot and the cutter. A serpentine motion-control method combined with end-camera images enabled cutter inspection.

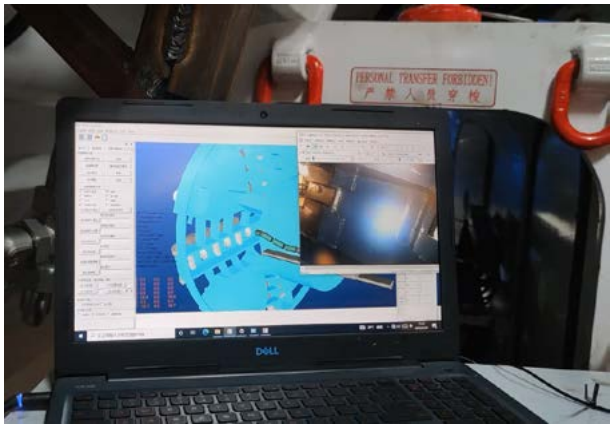


Fig. 16 Cutter inspection using the virtual-environment human-machine interaction platform

During installation and commissioning, the in-TBM installation issue of the snake robot was resolved. As shown in Fig. 17, a complete out-of-cabin inspection and retraction test was performed, preliminarily verifying the robot's operational capability inside the TBM.



Fig. 17 The snake robot exiting the cabin for cutter inspection.

During field operation, the snake robot completed cleaning and inspection of front cutters, and the replacement robot validated safe access for disc-cutter replacement. Fig. 18 shows the snake robot performing cutter cleaning at the tunneling site.



Fig. 18 The snake robot performing cutter cleaning at the tunneling site

6. DISCUSSION AND FUTURE WORK

While the proposed system demonstrates reliable inspection and replacement in harsh environments, several challenges remain: (1) current work is limited to visual image examination, future work should integrate specialized equipment such as depth-sensing cameras and laser profilometers to achieve precise wear measurement of the cutters, (2) the intense vibrations during TBM tunneling can easily cause failures in the cutter inspection and replacement robot, future work should consider implementing vibration-resistant design

to enhance the robot's reliability. Future work will address these issues and extend the system to larger TBM diameters and deeper burial depths.

7. CONCLUSION

A TBM robotised maintenance system combining a snake-like inspection robot and a heavy-load cutter replacement robot has been developed and validated. The inspection robot integrates cleaning and visual sensing for high-precision wear assessment under extreme conditions, while the replacement robot employs region-based trajectory planning and EHA control to achieve safe, accurate cutter swapping. Field application confirm reliability of the inspection robot in harsh environments, as well as its effectiveness in cutter cleaning and inspection.

8. REFERENCES

[1] Edelmann T. Monitoring of cutting tools in TBM[R].2013.

[2] OCRobotics. OC Robotics successfully delivers new JetSnake system to Dra gages/Bouygues joint venture in Hong Kong[EB/OL]. https://www.ocrobotics. com/news-en/oc-robotics-successfully-delivers-new-jetsnake-system-to-dragagesb ouygues-joint-venture-in-hong-kong/.

[3] Bruno COMBE E L. Innovations in TBM Tunneling[R].2019.

[4] Yamada Y, Fukui R, Warisawa S, Morioka E, Uetake M, Terada S. WInBot: A Disc Cutter Wear Inspection Robot for a Tunnel Boring Machine[C]//: IEEE, 2019: 1837-1843

[5] TSUBAKIMOTO. Zip Chain transmission device [EB/OL]. https://tt-net.tsuba kimoto.co.jp/tecs/pdct/sad/pdct_ZCA.asp.

[6] Liu Feixiang, Shao Jizhou, Zeng Hua, Zhu Chen, Ma Jianbing, He Tao, Huang Chunxia. Inspection robot system for shield machine and its inspection method [P]. 2020-08-07.

[7] Chen G. Design and Test of Intelligent Inspection and Replacement System of TBM Excavation Tools[C]//: IEEE, 2019: 219-222

[8] Xue Long, Cao Yingyu, Wang Kun, Liu Feixiang, Cheng Yongliang. Shield machine cutterhead cleaning robot system and cleaning method [P]. 2018-06-12.

[9] Xiong Zhilin. Research on key technologies of a snake-arm robot for shield cutter inspection [D]. Shanghai Jiao Tong University, 2018.

[10] https://vimeo.com/118895178. [EB/OL], 2015

[11] Rubrecht S. Contributions to the control of constrained robots [D]. Universite Pierre et Marie Curie-Paris VI, 2011.

[12] Du L., Yuan J., Bao S., et al. Robotic replacement for disc cutters in tunnel boring ma-chines[J]. Automation in Construction, 2022, 140: 104369.

[13] Yuan J., Guan R., Du J. Design and implementation of disc cutter changing robot for tun-nel boring machine (TBM)[C]//2019 IEEE International Conference on Robotics and Bio-mimetics (ROBIO). IEEE, 2019: 2402-2407.

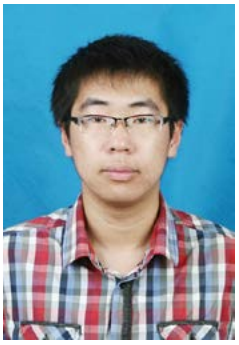
Biographies



Huayong Yang received the Ph.D. degree in fluid power transmission and control from the University of Bath, Bath, U.K., in 1988. He is an Academician of the Chinese Academy of Engineering. His current research interests include motion control and energy saving of mechatronic systems, microfluidic devices and systems, and R&D of fluid power components.



Haibo Xie received the Ph.D. degree in mechatronic engineering from Zhejiang University, Hangzhou, China, in 2004. He is a Professor with the State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University. His research interests include redundant robotics, mobile hydraulic control systems and components, and tunnel boring machine driving technique.



Cheng Wang received the Ph.D. degree in mechatronic engineering from Zhejiang University, Hangzhou, China, in 2023. He currently works as a researcher at Institute of Advanced Machines, Zhejiang University. His research interests include cable-driven mechanisms, redundant manipulators, mechanics-based modeling, and control.

Automated Steering Control for Articulated Off-Highway Machines

Zezhou Wang[1]*, Magdalena Janocha[1], Lukas Michiels, Benjamin Kazenwadel[1], Marcus Geimer[1]

[1]*Institute of Mobile Machines, Karlsruhe Institute of Technology*

** Corresponding author: Zezhou Wang (zezhou.wang@kit.edu)*

Abstract.

Automated steering control is a more and more important function for Articulated Off-Highway Machines (AOHMs) operating in construction, mining, and forestry. Their articulation-based kinematics enable high maneuverability but also introduce challenges for lateral control, particularly when tracking paths with large curvature or frequent directional changes. In the forestry industry, AOHMs must achieve precise and stable path tracking in unstructured environments, where disturbances and dynamic path updates are common. This paper investigates a combined control approach that combines a Pure Pursuit (PP) and a Stanley (ST) controller into a feedforward–feedback framework tailored to AOHMs. As an example application in an AOHM, a forwarder is used in simulations on four representative paths—circular, sinusoidal, straight, and real-world trajectories—in both forward and backward driving. The results show that the PP-ST controller consistently outperforms the individual controllers, achieving maximum lateral errors below 0.2 m across all test paths. Robustness analysis under dynamic steering and velocity noise, payload variation, and initial offsets demonstrates convergence in all cases, with steady-state lateral errors remaining within 0.1 m. Even for extreme orientation deviations at the initial position, the controller eventually corrects the trajectory and returns to the reference path. Overall, the PP–ST controller provides accurate, stable, and robust lateral control for articulated off-highway machines in representative forestry scenarios.

Keywords. Assistance systems, articulated vehicle, lateral control, mobile machine, Pure Pursuit, Stanley.

1. INTRODUCTION

Nowadays, off-highway mobile working machines play a critical role in industries such as construction, mining, and forestry, where operations are performed in unstructured and often challenging environments [1]. Automated driving technology for these machines has the potential to enhance safety, reduce operator fatigue, and improve overall productivity [2]. By enabling precise navigation and efficient task execution, these systems can also help optimize fuel usage and minimize environmental impact [3].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Within this category, Articulated Off-Highway Machines (AOHMs)—comprising multiple rigid units connected via articulation joints—are widely adopted due to their exceptional maneuverability in confined spaces [4]. The articulated steering mechanism enables tight turning radii and improved mobility on complex terrain. However, the kinematic complexity introduced by articulation makes lateral control more challenging than for rigid-frame vehicles, requiring control strategies capable of handling large steering angles and rapid changes in path curvature.

AOHMs are frequently deployed in the forestry industry, in machines like forwarders and harvesters, where precise lateral control is essential for safe and efficient operation. Reliable environmental perception—such as detecting and parameterizing individual trees—is a prerequisite for navigating the narrow and irregular lanes found in forest environments, often on uneven and unstructured terrain [5]. In such conditions, the planned path is continuously updated in response to changes in the surrounding environment and operational demands. Consequently, the lateral control system must not only achieve high tracking accuracy on paths with large curvature and frequent directional changes but also adapt in real time to sudden adjustments while maintaining vehicle stability.

In addition to the single-joint configuration that dominates current forestry machines, multi-articulated vehicle architectures have been investigated in the broader field of heavy-duty vehicle dynamics and control. Examples include multi-unit articulated systems comprising dual multi-axle trailers with coordinated steering for motion control [6]. Further studies have analyzed articulated multi-body vehicle systems that combine articulated joints and other joint types to capture complex interactions between interconnected units [7]. This paper focuses on AOHMs with a single articulation joint between the front and rear units, which remains the prevailing design in forestry applications such as forwarders and harvesters.

Among the existing lateral control methods, geometric approaches such as the Pure Pursuit (PP) and Stanley (ST) controllers are well known for their robustness, simplicity, and suitability for real-time implementation [8]. This work adapts and integrates these two controllers into a feedforward–feedback framework tailored to the kinematics of articulated off-highway machines, aiming to achieve robust and precise path tracking under diverse and dynamic operating conditions, including those encountered in forestry operations.

2. Lateral Control Methods for All-Wheel-Driven Articulated Off-Highway Machines

In this section, the kinematics of AOHMs are described, followed by the introduction of two lateral control methods: The Pure Pursuit and the Stanley controller.

Advanced control approaches such as Model Predictive Control (MPC) and nonlinear MPC require accurate dynamic vehicle models, reliable state estimation and considerable real-time computational resources. For articulated vehicles, tube-based robust MPC for tracking control explicitly relies on a detailed vehicle model and online optimization of the control sequence [9]. Nonlinear MPC for underground articulated vehicles is based on a high-fidelity dynamic model and nonlinear optimization at each control step [10], and multi-time-scale predictive control for underground articulated vehicles further illustrates the

computational effort associated with predictive control in this class of machines [11]. Robust path-following control for articulated heavy-duty vehicles addresses model uncertainty but still requires substantial tuning and validation effort to guarantee performance across the operating range [12]. In contrast, geometric controllers such as PP and ST controller operate with a kinematic model and low computational cost, which makes them suitable for real-time implementation on AOHMs. These characteristics motivate the use of the combined PP and ST controller in this work.

2.1. Kinematics of the Articulated Off-Highway Machine

Figure 2.1 illustrates the geometry of an AOHM. The steering angle δ can be derived from the orientation angles θ_i of the vehicle units [13]:

$$\delta = \theta_2 - \theta_1 \qquad (2.1)$$

A (virtual) axle can be defined between the two wheels of each vehicle unit. From the midpoint of this axle to the articulation joint, the lengths l_i and the corresponding radii r_i to a virtual center of rotation center o— about which the AOHM moves during a steering maneuver — can be determined. The radii are given by [13]:

$$r_1 = \frac{l_1 \cdot cos(\delta) + l_2}{sin(\delta)} \qquad (2.2)$$

$$r_2 = \frac{l_2 \cdot cos(\delta) + l_1}{sin(\delta)} \qquad (2.3)$$

To standardize the motion orientation of the AOHM, a quasi-static planar kinematic model is considered. When the machine moves forward, the pose reference point is defined at the front axle center; when moving backward, it is defined at the rear axle center. Under this kinematic assumption (negligible tire slip and compliance), r_1 and r_2 in Eq. (2.2) and Eq. (2.3) denote the instantaneous curvature radii of the chosen reference point. These radii are geometric quantities only and do not represent the actual turning radius, which in practice is influenced by payload, tire–soil interaction, velocity, and vehicle dynamics.

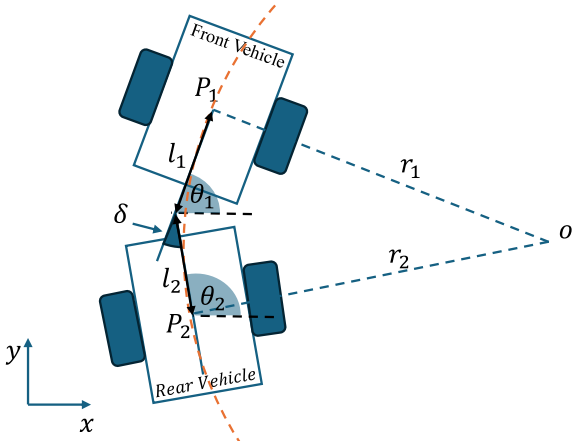


Figure 2.1. Geometry of an Articulated Off-Highway Machine

2.2. *Pure Pursuit Controller for AOHM*

The Pure Pursuit controller steers the vehicle toward a target point located at a specified lookahead distance l_d along the planned path [14, 15]. The lookahead distance is

$$l_d = max\left(l_{d,min}, min\left(k_{PP} \cdot |v|, l_{d,max}\right)\right) \qquad (2.4)$$

where v is the current vehicle speed, k_{PP} is the speed gain, and $l_{d,min}$, $l_{d,max}$ are the allowable bounds. This distance is smoothed to prevent abrupt steering changes.

As illustrated in Figure 2.2, l_d is measured along the reference path from the vehicle's current projected position to the lookahead point. The angle α is the heading from the vehicle to the lookahead point, while $\Delta\psi_1$ and $\Delta\psi_2$ are orientation differences between the Pure Pursuit arc and the reference path at the first and second candidate lookahead points, respectively. An orientation difference check reduces l_d in highly curved segments until $|\Delta\psi|$ falls below a predefined threshold, improving stability in dynamic paths.

Once the lookahead point is determined, the turning radius R can be calculated [16]:

$$R = \frac{l_d}{2 \cdot sin\alpha} \qquad (2.5)$$

By applying the tangent half-angle substitution, where $t = \tan\frac{\delta_{PP}}{2}$, and considering the forward motion case in which the turning radius $R = r_1$, Eq. (2.2) can be transformed into:

$$(l_1 - l_2) \cdot t^2 + 2\, r_1\, t - l_1 - l_2 = 0 \qquad (2.6)$$

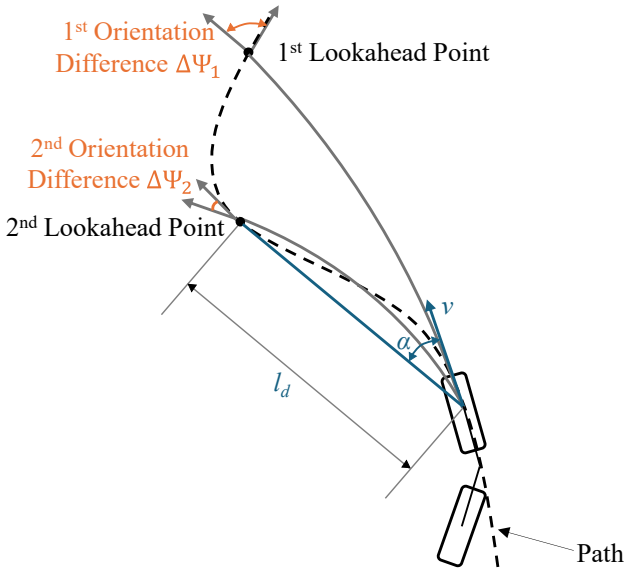


Figure 2.2. Selection of Lookahead Point for Pure Pursuit Controller

The steering angle of the Pure Pursuit controller for AOHM can then be obtained by applying the quadratic formula:

$$\delta_{PP} = 2\left(\tan^{-1}\left(\frac{-r_1 \pm \sqrt{l_1^2 - l_2^2 + r_1^2}}{l_1 - l_2}\right) + \pi n\right), \quad n \in Z \tag{2.7}$$

The same applies to the backward motion case, where r_1 in Eq. (2.6) and Eq. (2.7) is replaced by r_2 from Eq. (2.3), and n denotes the periodicity of the tangent function.

2.3. Stanley Controller for AOHM

The Stanley controller, similar to the Pure Pursuit controller, is employed for the lateral control of vehicles. However, instead of relying on a fixed lookahead point as in the Pure Pursuit controller, the Stanley controller determines the steering command based on the path point that is closest to the vehicle along its heading direction. To prevent oscillatory steering behavior when the reference point is too close, a lookahead distance is introduced in practice. Compared to the Pure Pursuit controller, the Stanley controller is generally more effective in suppressing lateral errors and achieving higher path-tracking accuracy, particularly in low-speed or sharp-turn scenarios [17].

The Stanley controller is a geometric method for lateral vehicle control that simultaneously corrects the heading error and the lateral position error to align the vehicle with the planned path. As shown in Figure 2.3, (c_x, c_y) denotes the closest point on the path to the AOHM's current reference point, e_l is the signed lateral error measured perpendicularly from this closest point to the vehicle reference point, and θ_h is the heading error between the longitudinal axis of AOHM's front vehicle and the tangent direction of the path at (c_x, c_y). The front vehicle's longitudinal speed at the reference point is denoted by v, and δ_{ST} is the steering command generated by the controller.

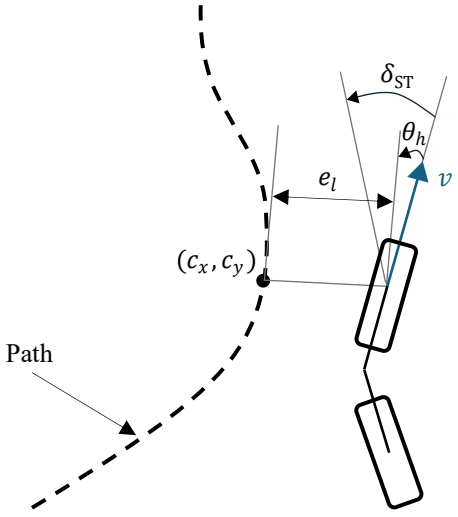


Figure 2.3. Stanley Controller for AOHM

In this work , based on the original Stanley controller, as proposed by Hoffmann et al. [18], a small positive constant k_v is integrated in the denominator to suppress excessive steering

at very low speeds and to avoid division by zero. The Stanley steering angle can be computed as

$$\delta_{ST} = \theta_h + \arctan\left(\frac{k_l e_l}{v + k_v}\right) \tag{2.8}$$

where the lateral correction term scales inversely with the vehicle's speed v, and k_l is the lateral error gain that scales the correction based on the magnitude of e_l. The first term in the equation corrects the heading to match the path's tangent direction, while the second term drives the lateral error toward zero. This combination allows the controller to achieve accurate path tracking across a wide range of speeds and curvature conditions.

3. Overall Framework of Automated Steering Control for Articulated Off-Highway Machines

The feasibility of combining the Pure Pursuit (PP) and Stanley (ST) controllers has been proved for robust lateral path tracking of indoor mobile robots [19]. Building upon this concept, the PP–ST combination will be adapted to the lateral control of AOHMs.

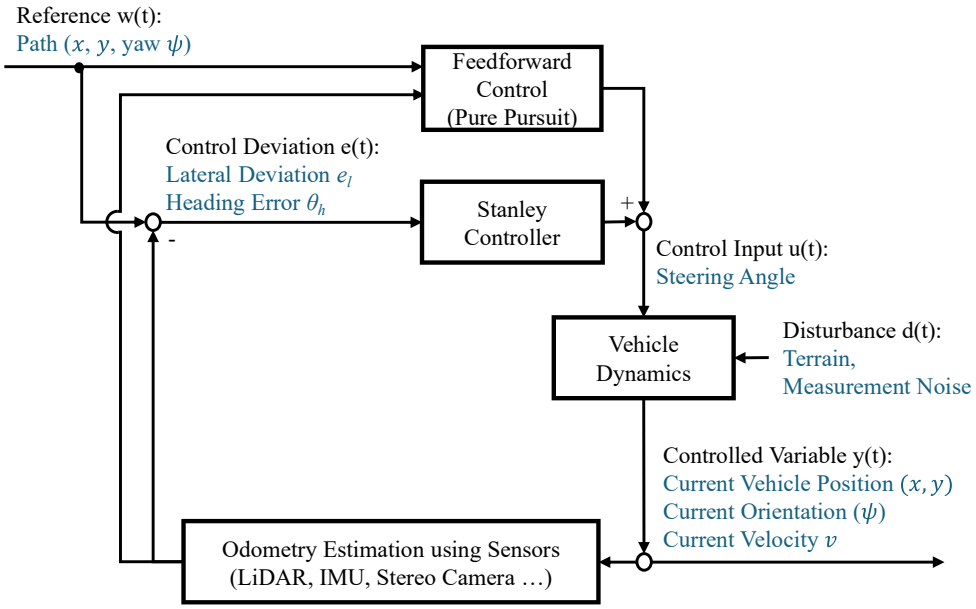


Figure 3.1. Overall Framework of Automated Steering Control for AOHMs

The overall framework is illustrated in Figure 3.1. In this framework, the planned path is available from a higher-level planning module, which generates a sequence of target poses (x, y, ψ) based on environmental perception and task objectives. At each control step, the vehicle's current pose and velocity are obtained from the odometry estimation module. A feedforward–feedback strategy is applied to generate the steering command: the PP module

determines the curvature required to reach a look-ahead point on the planned path, while the ST controller compensates for both lateral displacement and heading misalignment relative to the nearest path point. The outputs of both controllers are summed to form the steering angle command, which is sent to the articulated steering joint via the vehicle's actuation interface. The resulting vehicle motion is continuously monitored, closing the control loop and enabling precise path tracking under the kinematic constraints of articulated steering.

4. SIMULATION SETUPS

In this paper, a Forwarder is used as an exemplary application of the proposed control schematics.

Table 1: Technical Specifications of the Test Vehicle

Technical Specifications			
Total length (m)	10.96	Curb Weight (kg)	18,500
Width (m)	3.0	Maximum Payload (t)	14
Height (m)	3.79	Weight Distribution	60/40 (Front/Rear)
Front Axle Length l_1 (m)	1.799	Tire Diameter (m)	1.312
Rear Axle Length l_2 (m)	3.480	Tire Width (m)	1.42
Maximum Articulation Angle δ_{max} (°)	30	Ground Friction (rolling / static)	0.35 / 0.45

The Forwarder will be simulated in MATLAB Simulink with Simscape Multibody, where the forwarder is represented as a multi-body system including the articulated steering joint and axle constraints listed in Table 1. Tire–ground interaction is modeled with constant rolling and static friction coefficients (0.35 / 0.45). A variable-step solver with adaptive step sizes is applied to ensure numerical stability of multi-body dynamics. The simulation horizon is adjusted to allow traversal of each path. Outputs such as pose, steering angle, and lateral error are sampled at every time step.

In the simulations, the vehicle pose and velocity are provided as ideal ground-truth states from the multibody model without sensor noise. This assumption isolates the controller performance from perception uncertainties and allows a focused evaluation of the lateral control scheme. The initial forward velocity is set to a baseline value of 2 m/s, which corresponds to typical low-speed operation of forestry machinery and is used as the nominal input for all simulated scenarios.

The Forwarder will be tested on four paths: a circular path, a sinusoidal path, a straight line, and a real-world path collected from the Forwarder in the forest.

The formulas of the first three paths are defined in Table 2, where x starts from 0 m and path points are sampled every 0.5 m, resulting in a total of 150 points forming the final path. On each path, the Forwarder is tested once in forward and once in backward direction, i.e.,

evaluating the lateral control performance when either the front or the rear vehicle is leading. In addition, to investigate the robustness of the lateral controllers, further simulations are conducted on the sinusoidal path by introducing various disturbances, including dynamic steering and velocity noise, payload variations, and initial offsets.

Table 2: Equations of the Test Path

Path	Equation
Circular Path	$y_{circular}(x) = -12 \pm \sqrt{12^2 - x^2}$
Sinusoidal Path	$y_{sin}(x) = 5 \cdot \sin(x/10)$
Straight Line	$y_{straight}(x) = 0$

For the PP and ST controllers, the relevant parameters in Eq. (2.4) and Eq. (2.8) are listed in Table 3. These parameters are determined based on a sinusoidal path defined by $f_{\sin 2}(x) = 20 \cdot \sin(x/20)$, which will be detailed in subsequent sections.

Table 3: Parameters of Lateral Controllers

Parameter	Value
Maximum Lookahead Distance $l_{d,max}$ (m)	15.0
Minimum Lookahead Distance $l_{d,min}$ (m)	2.0
Speed Gain of PP controller k_{pp} (s)	3.0
Orientation Difference $\Delta\Psi$ (°)	20
Step Size for Orientation-Difference $\Delta\Psi$ Adaptation (m)	0.5
Velocity Offset k_v (m/s)	0.1
Lateral Error Gain k_l (s^{-1})	5.0

5. RESULTS AND ANALYSIS

In the following sections, simulations based on the setups described in Section 4 are conducted to evaluate the performance of the PP-ST controller introduced in Section 3. First, the optimal ST controller parameters, such as lateral error gain k_l, are determined through parameter variation. Next, the path-tracking performance of the PP-ST controller is compared with that of the standalone PP and ST controllers. Finally, different disturbance scenarios are introduced in the simulations to investigate the robustness of the PP-ST controller.

5.1. Parameter Tuning of the PP–ST Controller

In this paper, the controller parameters are tuned using a trial-and-error–based, simulation-driven approach. To illustrate the tuning procedure, the lateral error gain k_l of the ST controller is selected as a representative example.

As shown in Eq. (2.8), the lateral error gain k_l is a key parameter influencing the performance of the ST controller. The following simulations are conducted on the previously defined $f_{sin2}(x)$ path using different values of k_l to determine the optimal setting.

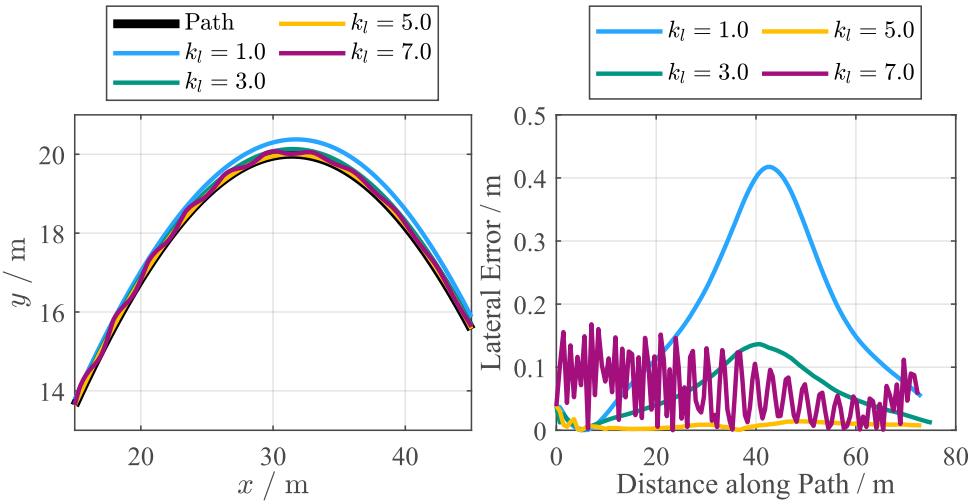


Figure 5.1. Optimization of Stanley Controller's Lateral Error Gain

As illustrated in Figure 5.1, four cases are compared: $k_l = 1.0$, $k_l = 3.0$, $k_l = 5.0$, and $k_l = 7.0$. The left-hand side of the figure presents the path-tracking results of the ST controller under different k_l values, while the right-hand side shows the corresponding lateral errors relative to the reference path. It can be observed that when $k_l = 1.0$, the lateral error is relatively large, particularly in sections with higher curvature. As k_l increases, for example to 3.0, the overall lateral error decreases. However, when $k_l = 7.0$, the lateral error exhibits strong oscillations. Consequently, a compromise value of 5.0, achieves relatively low and stable lateral errors without noticeable oscillations.

Table 4: Performance of ST Controller's Lateral Error Gains on Lateral Error

Lateral Error Gain k_l (s^{-1})	Mean / m	Max / m	Median / m
1.0	0.192	0.418	0.167
3.0	0.063	0.137	0.063
5.0	**0.008**	**0.038**	**0.008**
7.0	0.062	0.168	0.062

The lateral error metrics in Table 4, namely the mean, maximum, and median values, further confirm that the ST controller with $k_l = 5.0$ achieves the best performance across all

indicators. Based on this tuning procedure, the value of k_l is fixed to 5.0 in the subsequent simulations. Following the same trial-and-error–based tuning strategy, the remaining PP–ST controller parameters listed in Table 3 are determined analogously.

5.2. Comparison of Lateral Controllers

Subsequently, the PP, ST, and the PP-ST controller are tested and compared on the four paths described in Section 4. The first scenario considers driving forward.

As shown in Figure 5.1, all three controllers achieve close adherence to the reference circular path in terms of path tracking, as show on the left. However, the lateral error plots on the right indicate that, although the errors of the PP and ST controllers are not large and tend to stabilize after initial oscillations, they still exceed those of the PP-ST controller.

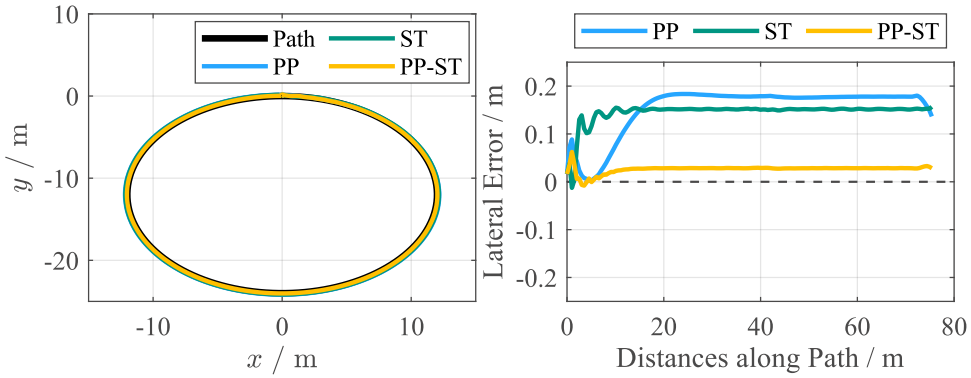


Figure 5.1. Lateral Controller Comparison during Forward Driving on a Circular Path

For the sinusoidal path in Figure 5.2, all three controllers exhibit noticeable sinusoidal oscillations in lateral error. However, the PP-ST controller shows the smallest amplitude, whereas the PP controller produces the largest. Nevertheless, the maximum absolute lateral error of all three controllers remains below 0.2 m.

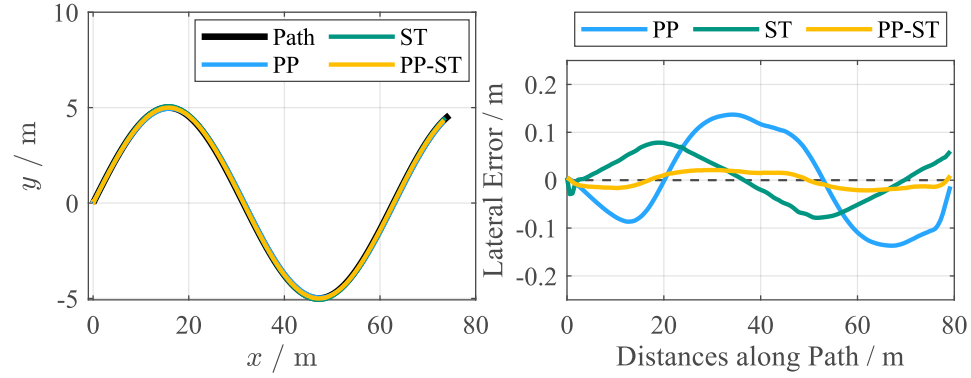


Figure 5.2. Lateral Controller Comparison during Forward Driving on a Sinusoidal Path

In the straight-line simulation shown in Figure 5.3, all three controllers exhibit strong oscillations at the beginning, with the ST controller even showing lateral swaying. However, in the later stage, all three converge to stable behavior and achieve a high degree of accuracy in straight-line tracking, with their lateral errors approaching 0 m.

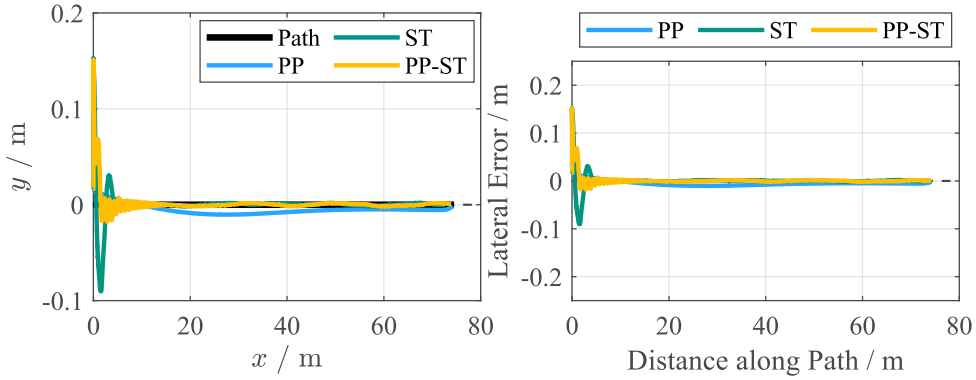


Figure 5.3. Lateral Controller Comparison during Forward Driving on a Straight Line

Figure 5.4 illustrates a real-world path. As a result, all three controllers exhibit high oscillation frequencies. Nevertheless, their overall performance remains satisfactory, with the PP-ST controller once again demonstrating the best results.

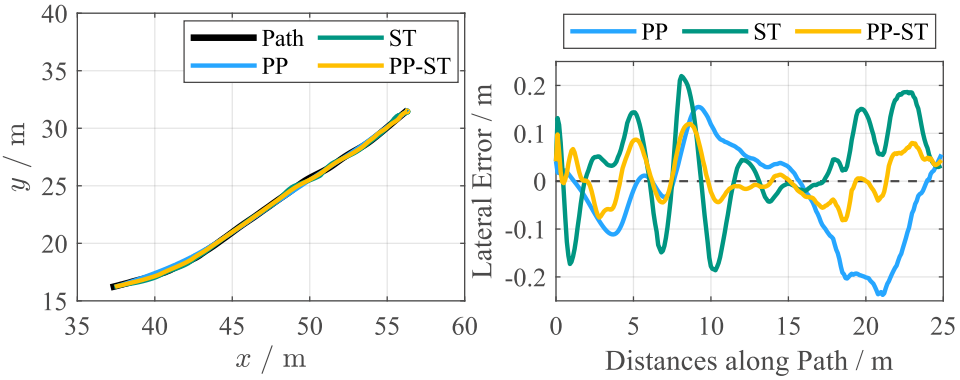


Figure 5.4. Lateral Controller Comparison during Forward Driving on a Real Path

Table 4 summarizes the controller performance in forward and backward driving across the four paths, based on identical simulation settings to ensure comparability. Although in certain cases, such as the straight-line scenario, the ST controller achieves marginally lower mean and median errors (e.g., 0.005 m and 0.001 m, respectively), the PP-ST controller introduced in this paper consistently delivers superior results in most test cases.

For forward circular and sinusoidal paths, the PP-ST controller reduces the mean error by more than 70% compared to PP and ST individually. Specifically, the mean error is 0.027 m compared to 0.153 m and 0.145 m for the circular path, and 0.014 m compared to 0.084 m and 0.042 m for the sinusoidal path. On the real path, it again halves the error (0.039 m) compared to PP and ST. This demonstrates its strong ability to track curved trajectories with high precision.

Table 4: Controller Performance on Absolute Lateral Error

Scenario Path	Lateral Controller	Mean / m	Max / m	Median / m
Driving Forward Circular Path	PP	0.153	0.184	0.178
	ST	0.145	0.157	0.152
	PP-ST	**0.027**	**0.063**	**0.028**
Driving Forward Sinusoidal Path	PP	0.084	0.137	0.087
	ST	0.042	0.078	0.041
	PP-ST	**0.014**	**0.040**	**0.015**
Driving Forward Straight Line	PP	0.007	**0.023**	0.006
	ST	**0.005**	0.153	**0.001**
	PP-ST	**0.005**	0.151	**0.001**
Driving Forward Real Path	PP	0.080	0.239	0.059
	ST	0.078	0.221	0.053
	PP-ST	**0.039**	**0.120**	**0.037**
Driving Backward Circular Path	PP	0.286	0.365	0.318
	ST	0.182	0.442	0.187
	PP-ST	**0.064**	**0.202**	**0.071**
Driving Backward Sinusoidal Path	PP	0.101	0.192	0.114
	ST	0.053	0.130	0.051
	PP-ST	**0.022**	**0.110**	**0.020**
Driving Backward Straight Line	PP	0.024	0.030	0.026
	ST	**0.004**	**0.191**	**0.001**
	PP-ST	0.005	0.201	0.003
Driving Backward Real Path	PP	0.035	0.127	**0.010**
	ST	0.028	0.127	0.013
	PP-ST	**0.021**	**0.109**	0.011

The backward driving cases further highlight the robustness of PP-ST. In the circular and sinusoidal paths, its mean error is reduced to 0.064 m and 0.022 m, respectively, significantly lower than the pure PP or ST controllers. Even in the more challenging real-path scenario, PP-ST achieves the smallest mean error (0.021 m) while maintaining stable maximum and median errors, underscoring its ability to generalize across both motion directions.

Overall, despite isolated cases where ST exhibits competitive performance, particularly in straight-line driving, the PP-ST controller demonstrates superior accuracy and stability across diverse paths and driving directions, validating its effectiveness as a more reliable solution for articulated off-highway machine lateral control.

Figure 5.5 presents boxplots of the lateral errors for all cases. The PP-ST controller achieves a maximum lateral error of less than 0.2 m, which is clearly lower than that of the other two controllers, whose performances are relatively similar to each other.

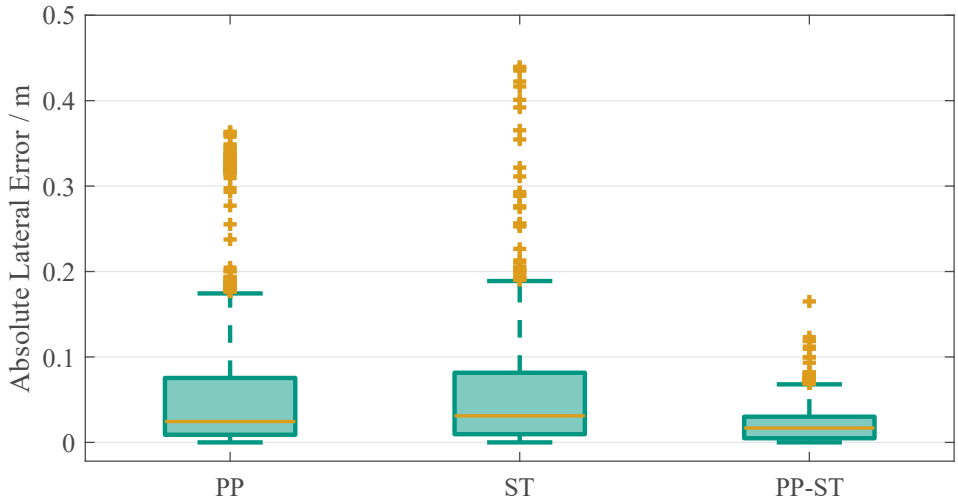


Figure 5.5. Overall Lateral Error Distributions of Absolute Lateral Controllers

5.3. Robustness Analysis of PP-ST Controller

Next, different disturbances are introduced on the sinusoidal path in Figure 5.2 to evaluate the robustness of the PP-ST controller. These disturbances include adding a noise term $2\sin(3t)$ to both the steering angle $\delta(t)$ and the velocity $v(t)$, where t denotes time; applying an additional load of 14 t to the rear vehicle of the forwarder; introducing an extra orientation offset of 45° at the initial position; and starting path tracking from a location 2 m away from the initial position.

Figure 5.6 illustrates the path-tracking performance and the corresponding signed lateral errors of the PP-ST controller under different disturbance scenarios. In the upper plot, the trajectories remain close to the reference path in most cases, with only the scenario starting with an additional orientation offset of 45° showing a visible deviation at the beginning before converging.

The lower plot shows the lateral error. When sinusoidal noise $2\sin(3t)$ is added to either the steering angle $\delta(t)$ or the velocity $v(t)$, the error exhibits oscillations but remains within 0.1 m. The additional rear-vehicle load of 14 t and the initial orientation offset of 45° cause slightly larger fluctuations than the normal case, yet the errors converge to near zero. The initial position offset of 2 m has only a minor transient effect. Overall, the controller demonstrates strong robustness under all disturbances.

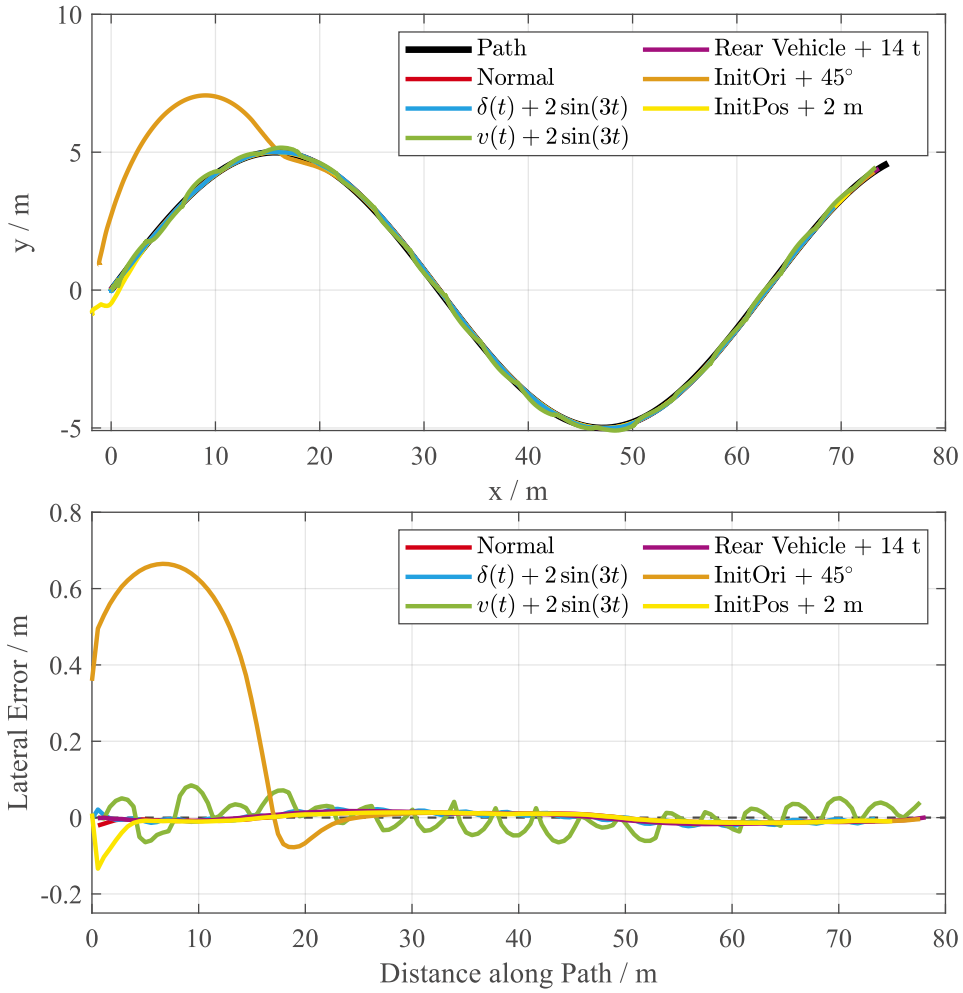


Figure 5.6. Performance of the PP-ST Controller under Different Disturbance Scenarios

The quantitative metrics in Table 5 confirm the observations in Figure 5.6. Under normal conditions (with no noise), the mean, maximum, and median lateral errors are 0.023 m, 0.049 m, and 0.024 m. Adding sinusoidal noise to the steering angle $\delta(t)$ causes only a marginal increase, with errors nearly identical to the baseline. Velocity disturbance results in higher values, with a mean of 0.068 m and a maximum error of 0.215 m. An additional rear-vehicle load of 14 t has almost no effect on the controller's performance. By contrast, the initial orientation offset of 45° produces the most significant degradation, with the maximum error rising to 3.622 m, while the mean remains relatively moderate at 0.588 m due to eventual recovery. Starting from an initial position offset of 2 m leads to a transient maximum error of 0.361 m, but the mean and median remain close to the normal case, indicating fast convergence.

Overall, both the trajectory plots and the error metrics demonstrate that the PP-ST controller is robust against most dynamic and static disturbances. Only the extreme initial orientation

offset leads to a substantial temporary degradation, while in all other cases the steady-state lateral errors remain within 0.1 m.

Table 5: Metrics of PP-ST Controller's Absolute Lateral Error under Disturbance

Scenario	Mean / m	Max / m	Median / m
Normal	0.023	0.049	0.024
$\delta(t) + 2sin(3t)$	0.024	0.054	0.024
$v(t) + 2sin(3t)$	0.068	0.215	0.060
Rear Vehicle + 14 t	0.024	0.040	0.025
InitOri + 45°	0.588	3.622	0.027
InitPos + 2 m	0.029	0.361	0.023

6. CONCLUSIONS

This paper presented a lateral control framework for Articulated Off-Highway Machines (AOHMs) by combining Pure Pursuit (PP) and Stanley (ST) controllers. The lateral error gain of the ST controller for a typical AOHM, a Forwarder, is first simulated on circular, sinusoidal, straight, and real paths demonstrated that the proposed PP–ST controller consistently outperformed standalone PP and ST controllers, achieving maximum lateral errors below 0.2 m in most cases. Robustness analysis under disturbances such as dynamic steering and velocity noise, payload variations, and initial offsets confirmed stable convergence, with steady-state lateral errors remaining within 0.1 m except for extreme orientation offsets, which caused only transient degradation. Overall, the results show that the PP–ST controller combines the complementary strengths of both methods, providing accurate, stable, and robust automated steering control for AOHMs.

7. REFERENCES

[1] M. Geimer, 'Mobile Working Machines', 2020, doi: 10.4271/9780768094329.

[2] D. Kaur, 'The impact of autonomous vehicles on mining operations: Enhancing safety and productivity through technological advancements', Scholarly Review Journal, 2024, doi: 10.70121/001c.124875.

[3] Z. Sun, 'Optimization and Evaluation of Energy Savings for Connected and Autonomous Off-Road Vehicles', 2024, doi: 10.2172/2476184.

[4] L. Gao, Y. Dong, and J. Zhao, 'Dynamic Modeling and Characteristic Analysis of Articulated Steering Vehicles', Applied Sciences, 2023, doi: 10.3390/app13085099.

[5] Z. Wang, et al., 'LiDAR-Based Tree Detection and Parameterization for SLAM in Autonomous Forestry Machinery, 2025 5th International Conference on Computer,

Control and Robotics (ICCCR), pp. 1–7, Hangzhou, China, May 16-18, 2025, doi: 10.1109/ICCCR65461.2025.11072654.

[6] X. Li, et al., 'A Modelling Method for All-Axle Steered Multi-Articulated Vehicles', 2024 UKACC 14th International Conference on Control (CONTROL), pp. 236–241, Winchester, United Kingdom, 2024, doi: 10.1109/CONTROL60310.2024.10532122.

[7] A. Ajorkar and Y. He, 'Design of autonomous driving controls for multi-trailer articulated heavy vehicles', Journal of Vibration and Control, 2025, doi: 10.1177/10775463251313657.

[8] L. Michiels, C. Geiger, and M. Geimer, 'Autonomous Driving on Skid Tracks for Forestry Machines', 2025, doi: 10.5445/IR/1000179648.

[9] D. Jeong and S. B. Choi, 'Tube-Based Robust Model Predictive Control for Tracking Control of Autonomous Articulated Vehicles', IEEE Trans. Intell. Veh., 2024, doi: 10.1109/TIV.2023.3320795.

[10] N. Sun, W. Zhang, and J. Yang, 'Integrated Path Tracking Controller of Underground Articulated Vehicle Based on Nonlinear Model Predictive Control', Applied Sciences, 2023, doi: 10.3390/app13095340.

[11] L. Liu, et al., 'Path Tracking Control for Underground Articulated Vehicles with Multi-Timescale Predictive Modeling', Actuators, 2025, doi: 10.3390/act14100477.

[12] F. M. Barbosa, et al., 'Robust path-following control for articulated heavy-duty vehicles', Control Engineering Practice, 2019, doi: 10.1016/j.conengprac.2019.01.017.

[13] P. I. Corke and P. Ridley, 'Steering kinematics for a center-articulated mobile robot', IEEE Trans. Robot. Automat., 2001, doi: 10.1109/70.928568.

[14] R. C. Conlter, 'Implementation of the Pure Pursuit Path Tracking Algorithm', Robot. Inst., Tech. Rep. CMU-RI-TR-92-01, Pittsburgh, PA, USA, 1992.

[15] J. M. Snider, 'Automatic Steering Methods for Autonomous Automobile Path Tracking', Robot. Inst., Tech. Rep. CMU-RI-TR-09-08, Pittsburgh, PA, USA, 2009.

[16] J. Morales, et al., 'Pure-Pursuit Reactive Path Tracking for Nonholonomic Mobile Robots with a 2D Laser Scanner', EURASIP J. Adv. Signal Process., 2009, doi: 10.1155/2009/935237.

[17] A. E. Bousskoul, I. Ouachtouk, and A. Ait Elmahjoub, 'Control strategies for autonomous vehicle path tracking: A comparative study of PID, Pure-Pursuit, and Stanley methods', EPJ Web Conf., 2025, doi: 10.1051/epjconf/202533006001.

[18] G. M. Hoffmann, et al., 'Autonomous Automobile Trajectory Tracking for Off-Road Driving: Controller Design, Experimental Validation and Racing', 2007 American Control Conference, pp. 2296–2301, New York, NY, USA, 2007, doi: 10.1109/ACC.2007.4282788.

[19] Y. Cao, et al., 'PP-ST: An Indoor Mobile Robot Path Tracking Algorithm', IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3326247.

Biographies



Zezhou Wang received his B. Sc. in Mechanical Engineering in 2022 and his M. Sc. in Automotive Engineering in 2024, both from RWTH Aachen University, Germany. He is currently working as a Research Associate at the Institute of Mobile Machines (Mobima) at the Karlsruhe Institute of Technology (KIT). His research focuses on localization, intelligent control and Sim2Real technologies of highly automated and autonomous forestry machines.



Magdalena Janocha received her B. Sc. in Mechatronics and Information Technology in 2025 and is currently pursuing her M. Sc. in Mechatronics and Information Technology at the Karlsruhe Institute of Technology (KIT), Germany. Her research interests include intelligent control of articulated vehicles.



Lukas Michiels received his B. Sc. in Mechanical Engineering in 2017 and his M. Sc. in Mechanical Engineering in 2019, both from the Karlsruhe Institute of Technology (KIT), Germany. In 2024, he obtained his Dr.-Ing. (Ph.D.) degree from KIT. His doctoral research focused on the development of driving functions and environment perception using artificial intelligence for the (semi-)autonomous operation of forestry machines, as well as the simulation of coupled fluid–structure interaction in the drive systems of mobile working machines.



Benjamin Kazenwadel received his B. Sc. in Mechanical Engineering in 2019 and his M. Sc. in Mechanical Engineering in 2021, both from the Karlsruhe Institute of Technology (KIT), Germany. He also earned the title Magister en Energia y Ambiente (Master in Energy and Environment) from the Instituto Tecnológico de Buenos Aires (ITBA), Argentina. He is currently working as a Research Associate at the Institute of Mobile Machines (Mobima) at KIT. His research focuses on the development of assistance systems to improve the efficiency of agricultural machinery.



Marcus Geimer received the Diploma degree in mechanical engineering from the RWTH Aachen University, Germany, in 1990, and the Ph.D. degree from the Institute of Hydraulics and Pneumatics, today named the Institute for Fluid Power Drives and Systems, RWTH Aachen University. Since 2005, he has been a Full Professor and the Director of the Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), Germany. He started his industrial career in 1995 in the field of construction at the company Krupp Berco Bautechnik GmbH, Germany, where he was the Leader of the research group for hydraulic breakers. In 2000, he was with Bucher Hydraulics GmbH, Germany, where he led the construction and customer development for mobile hydraulics.

Design and Application of a Prediction Model of the Loads and Mission Profile for Hydraulic Steering System in Heavy-Duty Off-Road Vehicles

M. Bertoli, B. Zardin, M. Borghi

Enzo Ferrari Department of Engineering, University of Modena and Reggio Emilia, Modena, Italy 41125
matteo.bertoli@unimore.it, barbara.zardin@unimore.it, massimo.borghi@unimore.it.

Abstract.

Centralized hydraulic circuits are typically employed to actuate the steering system and other different users in off-road vehicles, particularly in agricultural tractors. These solutions, while robust and widely adopted, lead to high energy demands and significant contributions to overall fuel consumption and greenhouse gas emissions. This study forms part of a wider academic research project aimed at developing an experimental platform integrated with virtual models to study alternative architectures, with the objective of improving energy efficiency and sustainability of heavy-duty off-road vehicles. In particular, this work is dedicated to developing a mathematical model that predicts the loads acting on the hydraulic steering actuator during working cycles. Indeed, measuring the loads acting on a tractor's steering system during real-world operation is challenging and costly. Field tests require the vehicle to be equipped with sensors and to undergo controlled work cycles on different terrains, which demand considerable time, resources and logistical complexity. Predictive models thus offer a clear advantage, addressing the lack of experimental data by generating virtual but realistic steering loads. The novel contribution of this paper is a simulation framework that, accounting for the vehicle's front axle load, soil type, and tire sinkage, creates realistic duty cycles to feed a hardware-in-the-loop test bench. These parameters, in fact, critically affect the forces transmitted to the hydraulic actuator. The model is based on a detailed kinematic study of the front axle with core terramechanics concepts to define the tire–soil interaction. The mathematical model was corroborated by comparing the results with those of a multibody simulation model of the front axle. The virtual assembly includes the main rigid bodies and a double-rod hydraulic steering cylinder. This simulation shows the mechanical behavior under both static and dynamic loads, providing force estimates across different realistic operating scenarios. We derived the modeling assumptions and boundary conditions from relevant literature on heavy-duty vehicle dynamics. As a future application, the results obtained from this validated model will be employed to actuate the solenoid valves of a load generating system in a test bench, thereby exerting the desired working conditions on the hydraulic power steering system.

Keywords. Hydraulic, steering, heavy duty vehicles, test rig, multibody simulation

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Hydraulic steering systems in heavy-duty off-road vehicles are typically driven by centralized load sensing circuits. While robust and widely adopted, these architectures result in high energy demands and significant hydraulic losses. Recent studies highlight the impact of these systems on fuel consumption and greenhouse gas emissions, especially during multi-actuator operations where overload pressure and compensator-induced dissipation are prevalent [1, 2]. Driven by growing industrial interest in sustainability and energy efficiency in the agricultural sector, research focuses on alternative hydraulic layouts. These include independent metering valve systems, electronically controlled variable displacement pumps and electro-hydraulic power steering systems with solenoid valves [3, 4, 5, 6]. Such configurations improve control flexibility and decrease power losses, particularly when integrated with digital control strategies that interact with mechanical models [7, 8].

However, the performance evaluation of hydraulic steering systems in realistic operating conditions is still a complicated and resource-intensive process. Field testing requires advanced instrumentation, controlled cycles on different terrains and considerable logistical effort [9, 10, 11]. In this context, predictive modeling proves to be an effective method for estimating steering loads, thereby avoiding the necessity for extensive experimental testing. Prior research validated the effectiveness of multibody simulations and terramechanics-based models for tire–soil interactions [12, 13, 14, 15, 16]. Angelucci et al. [17] explore the applicability of terra-mechanical models to optimize soil-tire interaction and highlight a gain on performance. Analytical foundations for predictive models are based on multidisciplinary contributions: Hegendus introduces dimensional analysis and plate penetration techniques to derive soil parameters [18], while Bekker's model estimates ground deformation, sinking and shear stresses induced by vehicular loads [19]. Thus, Taborek provided the theory for studying kinematic behavior on vehicle steering mechanics, while Terzaghi's theoretical soil mechanics further clarified the role of cohesion, internal friction, and effective stress in soil failure [20, 21]. These cross-disciplinary notions are used in several engineering fields, e.g. robotics: Jia et al. [22] and Ishigami et al. [23] built a predictive model for a rover focused on tire-soil interaction, which can be integrated into multibody simulations. Instead, Patel et al. [24] applied Hegendus terramechanics to shovel-ground interaction in excavators. In conclusion, Cao et al. [25] pointed out the importance of axle friction and alignment forces, during pivot steering alongside terra-mechanical contributions. Despite the extensive literature on terramechanics and vehicle dynamics, a significant research gap remains in the availability of reliable, experimental-like load data for the validation of steering systems without relying on costly field campaigns. Drawing inspiration from rover studies, this research adapts the methodology to the agricultural sector. The novelty lies in the integration of these analytical foundations into a comprehensive multibody co-simulation approach. The result is a 'virtual sensor' capable of generating reliable duty cycles for hardware-in-the-loop validation.

2. EXPERIMENTAL TEST BENCH LAYOUT

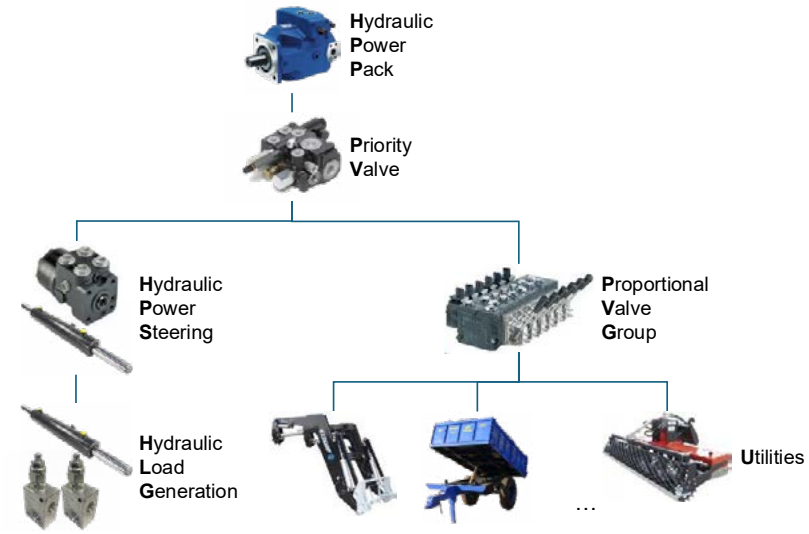


Figure 2.1. Baseline configuration layout

The design of an experimental bench for testing hydraulic and electro-hydraulic components, typically used in heavy duty vehicles, needs a method to replicate a duty cycle on the primary user, i.e. the steering system. Therefore, the primary purpose of the platform is to exert on the hydraulic power steering (HPS) system the same forces and displacements that would be observed during actual field operation. Although the platform does not include a physical front axle, this apparent hardware limitation is in fact advantageous: the platform core is a digital twin that replicates the front axle and its kinematics through the hydraulic load generation (HLG) system. Different axle geometries, kinematics or tractor dimensions can be simulated by updating the model, without mechanical reconfiguration.

The layout also includes secondary users, such as lifting actuators and orbital motors, or in general users that can be mounted on a tractor. Hydraulic power distribution is managed by a priority valve. The described layout is known as the baseline configuration and is illustrated in Figure 2.1. Obviously, the platform's modular design allows for the configuration of alternative hydraulic architectures. For example, it can operate without a priority valve and with users powered independently. Nevertheless, the focus of this work is the steering function. Therefore, alternative circuit configurations and secondary users, including the priority valve, are not considered in this context. The analysis addresses HPS and HLG subsystems.

2.1. HPS Main Components

Figure 2.2 illustrates the hydraulic layout of the steering branch of the experimental platform. HPS replicates the steering architecture typically utilized in utility tractor. We chose two representative components:

- Danfoss OSPC80LS steering unit (4 in Fig. 2.2) [26]

- Ognibene 57/35 × 190 mm double acting steering cylinder with through rod (3A) [27]

The OSPC80LS is connected to the steering wheel and adjusts the flow to the cylinder based on the pilot's manual input. The cylinder is selected with a through rod, similar to many of those installed on front axles.

2.2. HLG Main Components

The innovative core of the test bench is the HLG system. This unit provides an autonomous short-circuited load system, unlike conventional rigs, which include safety valves to release oil in case of unexpected pressure spikes. This represents an effective solution for avoiding power saturation of the bench control unit. The main components of the HLG are:

- Ognibene 57/35 × 190 mm double-acting steering cylinder with through rod (3B) [27]
- Hydraforce TS12-27 two-way proportional solenoid valves (5A and 5B) [28]
- Anti-cavitation and anti-shock safety valves (6)

Note that we chose the same actuator as the steering one, as it provides the same area ratios on both sides. However, the platform is not limited to this configuration. In fact, if different cylinders are used, a correction factor derived from the chamber area ratio could be considered. The proportional pressure relief valves operate alternately depending on the steering direction. Their opening is controlled by a PWM signal generated by the data acquisition and control system (DAQ). However, a key consideration for this configuration is the need for high-frequency acquisition to quickly control valve opening, thus ensuring stability within the simulated duty cycle.

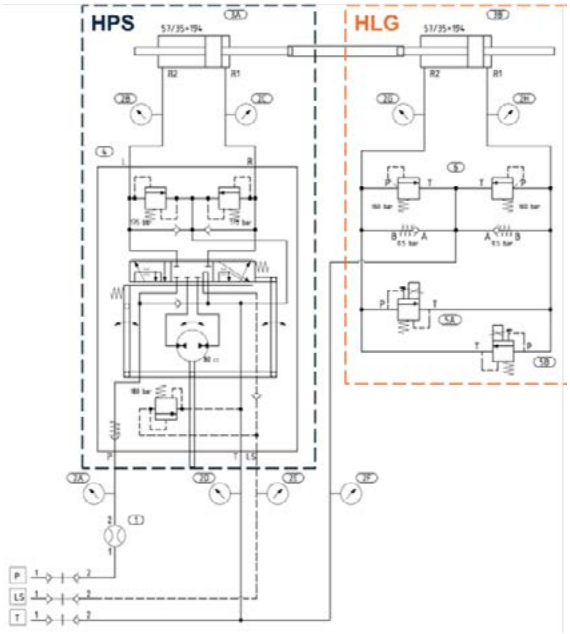


Figure 2.2. Hydraulic circuit of power steering (HPS) and load generation (HLG) systems

2.3. DAQ Design

The platform's data acquisition and control system is implemented on a National Instruments CompactRIO (cRIO), a real-time embedded controller integrating a real-time processor, an FPGA, and modular I/O. Acquisition is handled by the NI Scan Engine, which performs a deterministic I/O scan cycle over a user-defined scan period (typically 1 ms, up to ~1 kHz), ensuring synchronous sampling of multiple signals, such as hydraulic pressures, flow rates, angular steering position, linear displacement, and load cell measurements. Scan Engine logic is easier to program because you can access I/O variables directly in LabVIEW Real-Time as shared variables, but the maximum cycle speed is limited by the scan speed. PWM signals for Hydraforce T12-27 are also generated through the Scan Engine. High-frequency FPGA acquisition is not required in this context.

Based on this hardware setup, three operating scenarios are supported. The first uses fixed pressure thresholds on proportional valves, simplifying the identification of steady-state behavior. This mode is used for preliminary hardware calibration and does not rely on the simulation model. The second scenario uses a look-up table that correlates predefined steering resistances with wheel angular positions. The validated model presented in this paper is specifically aimed at generating the data for this second scenario. The last scenario, representing the next step of the research work, provides continuous co-simulation during the experimental test. In this setup, the DAQ integrates real-time simulation outputs to dynamically adjust the valve opening, effectively connecting the physical steering system with the digital twin.

3. ANALYTIC APPROACH

This chapter presents the analytical approach used to validate the virtual model of the front steering axle and to estimate the soil-induced resistance on the vehicle. Modeling is broken down into two problems: the kinematics and dynamics of the steering axle, which provides how steering inputs turn into wheel motion and reaction forces; and the terra-mechanical model, which estimates ground-tire forces in relation to terrain characteristics and tire/load conditions. The aims of the analytical formulation are to obtain a reduced-order dynamic system of the front axle and to define a mathematical model that correlates the lateral forces on the tire surface to the steering cylinder based on the steering angle.

3.1. Assumptions and Reference Frames

The steering axle and components are modeled as rigid bodies, the vehicle is considered stationary, and friction in the kinematics is assumed negligible due to lack of literature data. Furthermore, a limiting scenario is considered in which the minimal imprint that the tire can create in the soil corresponds to the square of its thickness.

We considered both a fixed O_{xyz} and a rotating $O_{x'y'z'}$ reference system. The O_{xyz} frame is located along the cylindrical joint connecting the hub and the axle, and it is aligned perpendicularly to the ground plane. The mobile reference system $O_{x'y'z'}$ is obtained from the former by a γ-rotation around the z-axis, which is the steering angle, and a δ-rotation around the x-axis, which is the kingpin angle. Naming O the fixed frame O_{xyz} and O' the rotating frame $O_{x'y'z'}$, Equation 3.1 provides the rotation matrix to pass from O to O'.

$$R_{O\to O'} = R_z(\gamma)R_x(\delta) = \begin{bmatrix} \cos\gamma & -\sin\gamma\cos\delta & \sin\gamma\sin\delta \\ \sin\gamma & \cos\gamma\cos\delta & -\cos\gamma\sin\delta \\ 0 & \sin\delta & \cos\delta \end{bmatrix} \quad (3.1)$$

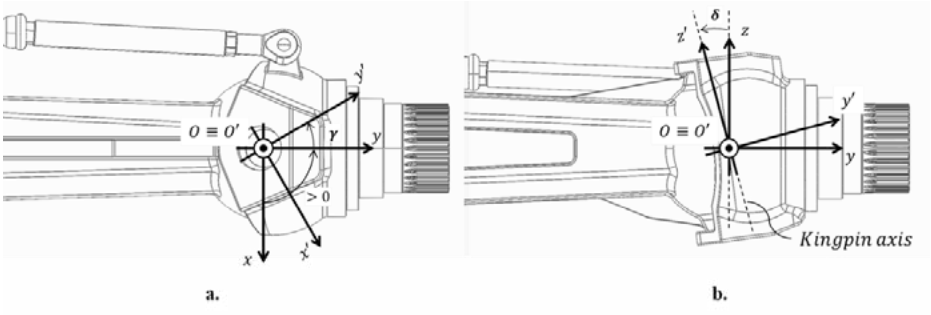


Figure 3.1. Fixed reference frame O_{xyz} and mobile $O_{x'y'z'}$ on xy-plane (a) and yz-plane (b)

3.2. Front Axle Analysis: Kinematics and Dynamics

The steering axle is approximated as a crank mechanism with deflected thrust, as Figure 3.2 depicts. For the sake of clarity, we use the same reference systems as common methods. Note that α, named the crank angle, can be written as $\alpha(\gamma)$ if $\widehat{AOG} = \pi/2$, where G is the tire center of gravity. Furthermore, the domain of $\alpha \notin [0; 2\pi]$ but is limited due to the kinematics of the axis. Eq. 3.2 represents the OAB vector closing triangle in terms of its coordinates.

$$A = \begin{cases} a\cos\alpha \\ a\sin\alpha \end{cases} \quad B = \begin{cases} s \\ h \end{cases} \quad (3.2)$$

where h is a fixed dimension that depends on the axle geometries, s is the cylinder stroke. Using Eq. 3.3 we obtain the length of the rod AB.

$$b = \sqrt{(s - a\cos\alpha)^2 + (h - a\sin\alpha)^2} \quad (3.3)$$

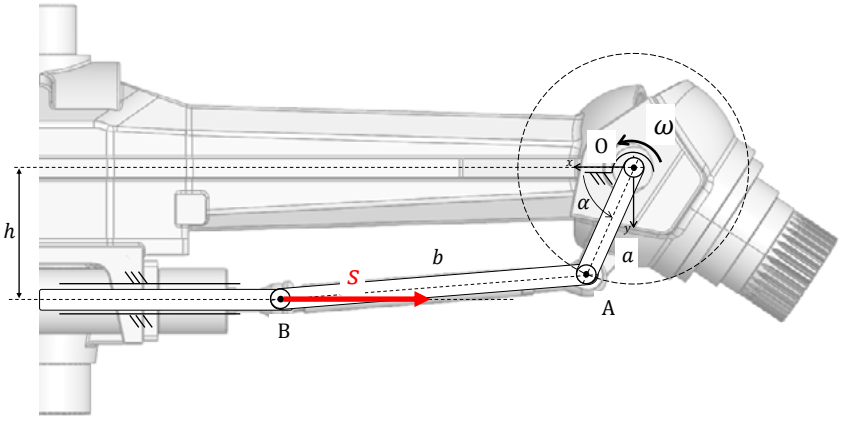


Figure 3.2. Kinematics right steered analysis

This condition results in a quadratic equation in s from which the solution, described in Eq. 3.4, is obtained.

$$s(\alpha) = a\cos\alpha + \sqrt{b^2 - (h - a\sin\alpha)^2} \tag{3.4}$$

It is now possible to derive Eq. 3.4 to obtain the speed of the steering actuator. Similarly, acceleration can be obtained by deriving the velocity.

Subsequently, we focus on dynamic analysis, with the system being modeled again as a crank mechanism with deflected thrust. The aim of this formulation is to guarantee that M_{res}, denoting the torque at hub, is balanced from the cylinder force. We present the right steering scenario, illustrated in Figure 3.3a.

However, this approach can also be applied to left steering. Therefore, the rod angle β can be expressed in Eq. 3.5.

$$\beta(\alpha) = \cos^{-1}\left(\frac{h - a\sin\alpha}{b}\right) \tag{3.5}$$

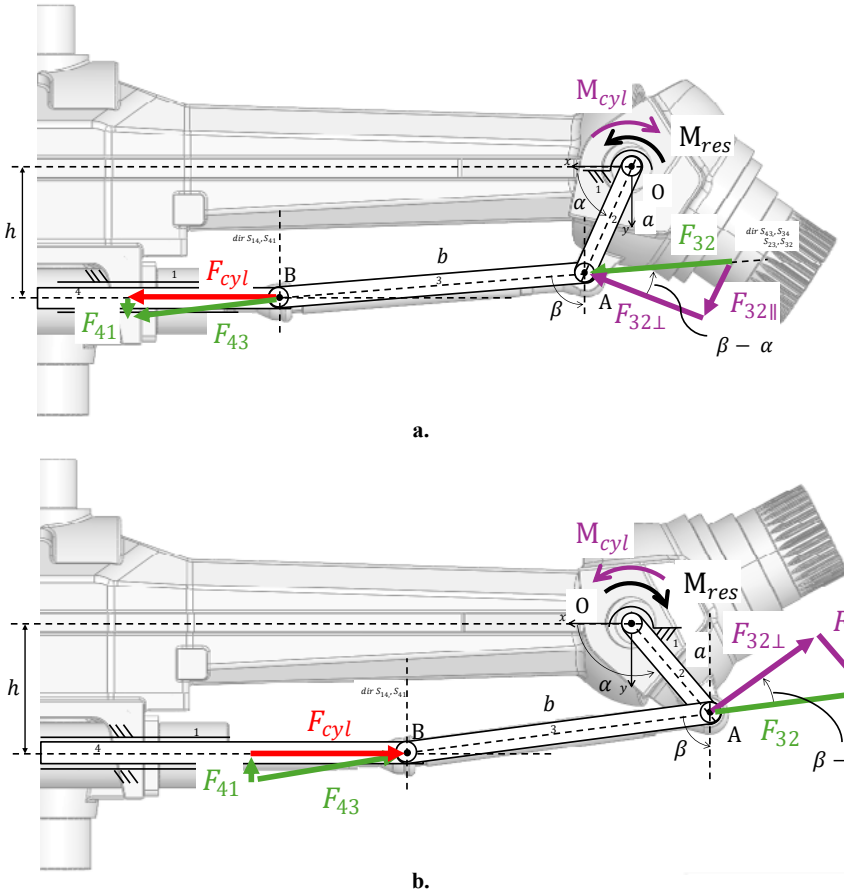


Figure 3.3. Dynamic right (a) and left (b) steered analysis

F_{cyl} represents the force applied by the cylinder in the direction of the guide. The axial force acting on the connecting rod AB is obtained by projecting it along the slider's axis, while the perpendicular component to the crank AO produces the resisting torque at the hub M_{res} around point O. Equation 3.6 illustrates the computation of M_{res}.

$$M_{res} = aF_{32,\perp} = aF_{32}\cos(\beta - \alpha) = a\frac{F_{cyl}}{\sin\beta}\cos(\beta - \alpha) \tag{3.6}$$

In conclusion, M_{res} is moved from the fixed frame to the mobile frame $O_{x'y'z'}$ by means of the rotation matrix $R_{O\to O'}$, to represent the hub torque on kingpin axis. Note that this model refers to a specific axle configuration and, in the future, with a library of axles available, the same method can be applied to other configurations.

3.3. Kingpin Torque Calculation and Terramechanics Analysis

In this subchapter, the resistance torque acting on the steering hub is estimated by combining different theoretical frameworks from kingpin analysis to terramechanics-based tire–soil interaction models. We chose a customized analytical approach rather than commercial "black-box" models in order to ensure flexibility and explicit control over critical variables, particularly tire sinkage (h). This allows for rapid reconfiguration of soil-vehicle parameters and direct definition of sinking levels, which are the main determinants of lateral loads on the steering. The current approach assumes a stationary vehicle and a pivot steering; therefore, future developments will extend the model to include vehicle motion, allowing the simulation of dynamic efforts.

3.3.1. Normal Contact Pressure and Sinkage

According to Bekker's pressure–sinkage relationship, the normal contact pressure $\sigma(h)$ between tire and soil is expressed as Eq. 3.7.

$$\sigma(h) = \left(\frac{k_c}{b} + k_\varphi\right)h^n \tag{3.7}$$

where k_c, k_φ, n are soil parameters, h is sinkage, and b is the tire width. The wheel sinkage $h(\theta)$ at arbitrary angular position θ is given geometrically by Eq. 3.8.

$$h(\theta) = r\,(\cos\theta - \cos\theta_s) \tag{3.8}$$

The contact angle θ_s is derived from the vertical load balance,

$$W = \int_{-\theta_s}^{\theta_s} \sigma(\theta)\, br\, \cos\theta\, d\theta \tag{3.9}$$

as W is the known tractor load on the wheel. The static sinkage is then evaluated using the geometric relation described in Eq. 3.10.

$$h_s = r(1 - \cos\theta_s) \tag{3.10}$$

where r is the tire radius and θ_s is the contact angle. Finally, the local pressure distribution can be represented as a function of θ, as shown in Eq. 3.11.

$$\sigma(\theta) = \left(\frac{k_c}{b} + k_\varphi\right)[r\,(\cos\theta - \cos\theta_s)]^n \tag{3.11}$$

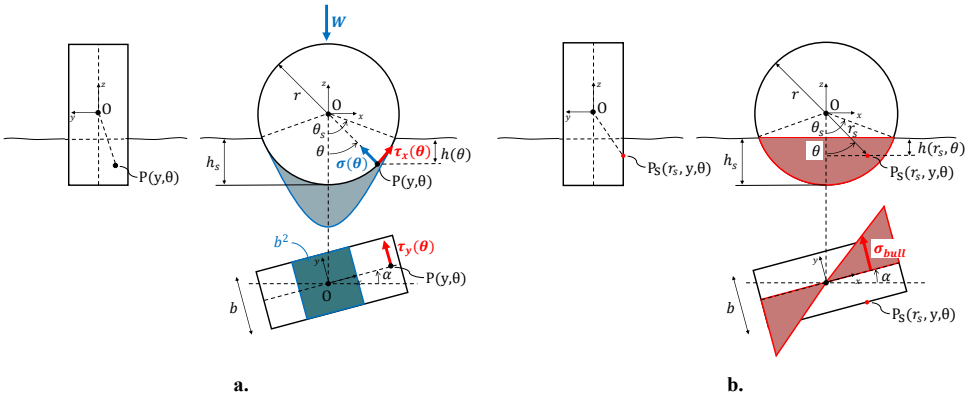


Figure 3.4. Representation of the contributions of normal pressure in $P(y,\theta)$ (a) and bulldozing in $P_s(r_s,y,\theta)$ (b) on the wheel

3.3.2. Shear Stress

The maximum shear stress that can be transferred between the ground and the tire is limited by the weakest contact, determined by the tire-to-ground friction and the internal cut resistance of the ground (Mohr-Coulomb). This relation is described in Eq. 3.12.

$$\tau_{max} = min\ (\mu_s\sigma, c + \sigma \tan\phi) \tag{3.12}$$

where μ_s is the sliding friction coefficient, c is the soil cohesion, and φ is the soil internal friction angle. For the integration of this contribution, we consider the point $P(y,\theta)$ as

$$P(y,\theta) = [r\ \sin\theta \quad y \quad -r\cos\theta]^T \tag{3.13}$$

In the stationary vehicle assumption ($v_x = 0$), slip ratio and slip angle lose physical meaning, so the Janosi–Hanamoto formulation is excluded. It will be introduced in future model extensions to account for longitudinal motion.

3.3.3. Bulldozing Resistance

In addition to shear stress, the soil exerts resistance on the lateral sinkage tire surface. We consider the point $P_s(r_s,y,\theta)$ as Figure 3.4 (a) depicts.

$$P_s(r_s,\ y,\theta) = [r_s\ \sin\theta \quad y \quad -r_s\cos\theta]^T \tag{3.14}$$

Thus, the relative sinkage is evaluated using Eq. 3.15.

$$h(r_s,\theta) = r_s\ \cos\theta - r\ \cos\theta_s \tag{3.15}$$

Following Hegedus model, the stress per unit width is evaluated using Eq. 3.16.

$$\sigma_{bull}(r_s,y,\theta) = D_1\left[c\ h(r_s,\theta) + D_2\frac{\rho_d h(r_s,\theta)^2}{2}\right] \tag{3.16}$$

In the equation above $D_1$and D_2 represent geometric coefficients derived from the soil internal friction angle, while ρ_d is the soil density.

3.3.4. Steering Aligning Torque

Due to kingpin inclination, steering must overcome soil resistance and generates a restoring torque. The steering aligning torque is evaluated using Eq. 3.17.

$$M_\theta = -W[e + (r - h_s)\tan\delta]\sin(2\delta)\sin\left(\frac{\gamma}{2}\right) \tag{3.17}$$

where e is the kingpin offset and δ the kingpin angle. This torque tends to realign the wheel toward the neutral position.

3.3.5. Resulting Force and Moments

Lateral (Y-direction) friction forces are computed by integrating over the cylindrical surface element $rd\theta dy$, which characterizes the interface between the wheel and soil (Figure 3.5).

$$F_y = \int_{-b/2}^{b/2}\int_{-\theta_s}^{\theta_s}\left(r\cdot\tau_y(\theta)\right)dyd\theta \tag{3.18}$$

An additional contribution from soil bulldozing must be considered, integrated over the lateral surface increment $r_s\ d\theta dr_s$ within the sink region of the tire.

$$F_{bull} = \int_{r-h(\theta)}^{r}\int_{-\theta_s}^{\theta_s}\sigma_{bull}(\theta)dr_s d\theta \tag{3.19}$$

The corresponding resisting torque is then evaluated using Eq. 3.20.

$$M_{res} = M_z = \int r\ sin\theta\ dF_y + \int r_s\ sin\theta\ dF_{bull} + M_\theta \tag{3.20}$$

In conclusion, M_{res} is moved from the fixed frame to the mobile frame $O_{x'y'z'}$ by means of the rotation matrix $R_{O\to O'}$ (Eq. 3.1).

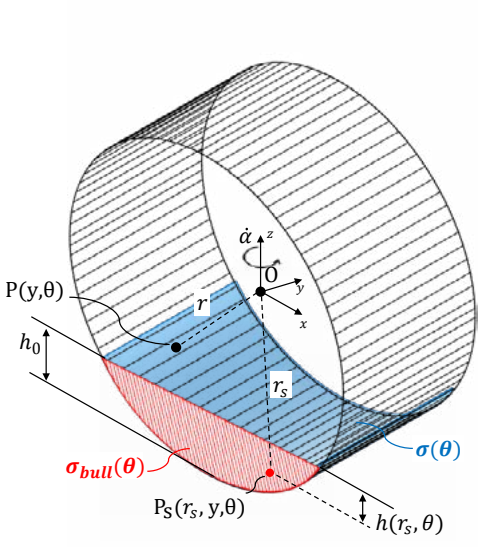


Figure 3.5. Integrals graphical representation of sigma and shear stresses on a tire

4. MULTIBODY APPROACH

The analytical formulation is then translated into a virtual prototype developed in a multi-body environment. In the present work, the modeling strategy is divided into two parts: the development of the steering axle model in ADAMS [29] and the implementation of the control logic in Simulink [30], which together forms the basis for co-simulating the complete steering system.

4.1. Front Axle Analysis: ADAMS model

The steering geometry was first designed in CAD and then imported into ADAMS for the definition of kinematic constraints and dynamic interactions. Some rigid components were merged according to their relative motion to maintain computational efficiency. For instance, the steering knuckle and hub were combined into one unit, while the cylinder of the hydraulic tie-rod assembly was incorporated into the housing.

Kinematic constraints were implemented to replicate the physical linkages, such as revolute joints at the kingpins, translation motion for the actuator rods, and spherical joints at the ball connections between the hubs and brackets.

The model dynamic behavior is introduced by applying a torque on the hubs joint aligned with the kingpin axis. This resistant torque, calculated in Simulink according to the selected operating scenario (for example clay soil with the New Holland T4.75 vehicle without cabin), reproduces the terra-mechanical resistance on the steering system. At the same time, the force exerted by the hydraulic actuator along its sliding axis is measured as the fundamental output of the multi-body model.

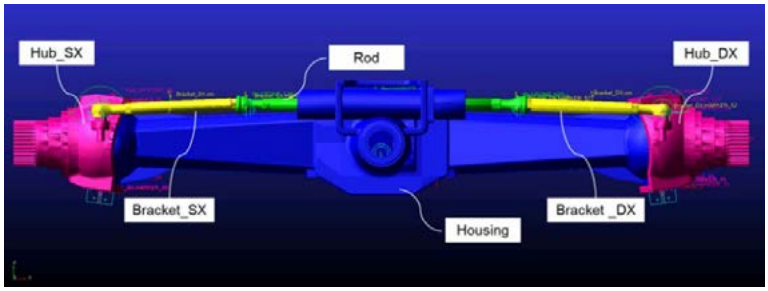


Figure 4.1. CAD model imported to ADAMS and creating constraints

4.2. Simulink Control Loop

The control subsystem was implemented in MATLAB/Simulink to drive the multibody model developed in ADAMS. The Simulink model is structured around different functional blocks, which together reproduce the dynamics of the steering bench and the input control to the hydraulic actuator.

The key parameters for the selected vehicle–ground configuration are read from .csv files, based on the user choices provided through the drop-down menu selections (1 in Fig. 4.2). In Tables 4.1 and 4.2, we provide parameters that characterize respectively soils and tractors, which relate to the results of the simulations discussed in the following chapter. It is important to note that we have assumed a vehicle weight distribution of 45% front and 55% rear. Also, the vehicle was evaluated without ballast and tools.

The control input of the system is the steering wheel angular position. Currently, a lock-to-lock maneuver is simulated; however, this signal will be logged in real-time concerning the angular position of the platform's steering wheel (2). These data and the geometric properties of the bench components are used to derive the displacement of the steering cylinder.

Blocks (3) and (4) contain MATLAB functions for computing the wheel hub-resistant torque $M(\gamma, k_{tractor}, k_{soil})$ and the actuator position $s(\theta_w, V_c)$, derived from the analytical model described in the previous sections.

The interface between Simulink and ADAMS was created using Adams Control/Plants (5), which exports the multibody model to a block that can be directly used in Simulink. Furthermore, a closed-loop is implemented to dynamically use the actual steering angle in the terramechanics function (3).

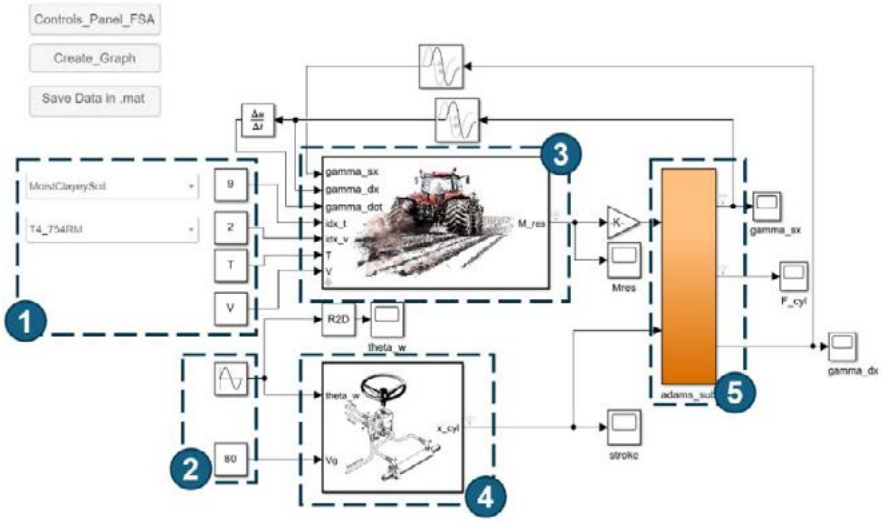


Figure 4.2. Simulink model with block for co-simulation with ADAMS

Table 4.1
Parameters for soil characterization [20] [21] [31].

Terrain Name	μ	Moist	ρ_d	n	k_c	k_φ	c	ϕ
Unit.		%	$\frac{kN}{m^3}$		$\frac{kN}{m^{n+1}}$	$\frac{kN}{m^{n+2}}$	Pa	deg
Asphalt (A)	0.75	0	0	0	0	0	0	0
Standard Sandy Loam (SSL)	0.45	15	16	0.7	5.27	1515.04	1720	29
Fertile Loam (FL)	0.45	24	14	1.01	0.06	5880	3100	29.8
Moist Heavy Clay (MHC)	0.3	55	20	0.11	1.87	103.27	20690	6

Table 4.2
Parameters for vehicle characterization [32] [33].

Tractor Name	Weight on front	b	r	e	δ
Unit.	N	mm	mm	mm	deg
T4.75.4RM	8829	280	450	225	7.5
T6.180	27755	480	670	225	7.5

4.3. Co-simulation Setup

To realize the interfacing between the multibody model in ADAMS and the control model in Simulink, the ADAMS/Controls plugin is used. This tool allows the export of the ADAMS model as a block "plant", i.e. a template that defines the co-simulation inputs and outputs, into Simulink model. This block is a dynamic system that receives commands and returns kinematic/dynamic measurements to the control loop. From Simulink to ADAMS, the following state variables are provided:

- Torque corrections derived from the feedback law, M_{res}.
- Actuator rod displacement, s.

From ADAMS to Simulink, the following state variables are provided:

- Actual steering angle of the hubs, γ.
- Actuator force, F_{cyl}.

ADAMS simulates one time step while maintaining the received inputs constant. Then, Simulink simulates for the step, keeping the ADAMS outputs constant. The new values are then exchanged, and this loop continues.

5. RESULTS

The first set of results, shown in Figure 5.1 (a–b), presents a comparison between the analytical formulation and the multibody simulation. A lock-to-lock maneuver was simulated that corresponded to four complete turns of the steering wheel. In this manner, the steering actuator executes its complete stroke within the bench configuration.

Note that, due to the modeled mechanism, the full stroke of the actuator corresponds to a steering sweep of approximately 60° on the hub. Therefore, the steering angle asymmetry shown in Figure 5.1b is inherent to the steering axle kinematics. Future developments may include testing different CAD configurations. Finally Figure 5.1 (c–d) illustrates the steering cylinder force and the resisting hub torque, respectively, throughout the simulated duty cycle. The clear overlap between the curves indicates a strong correlation between the analytical and multibody models, with negligible deviations, confirming the accuracy of the kinematic and dynamic implementation of the virtual model.

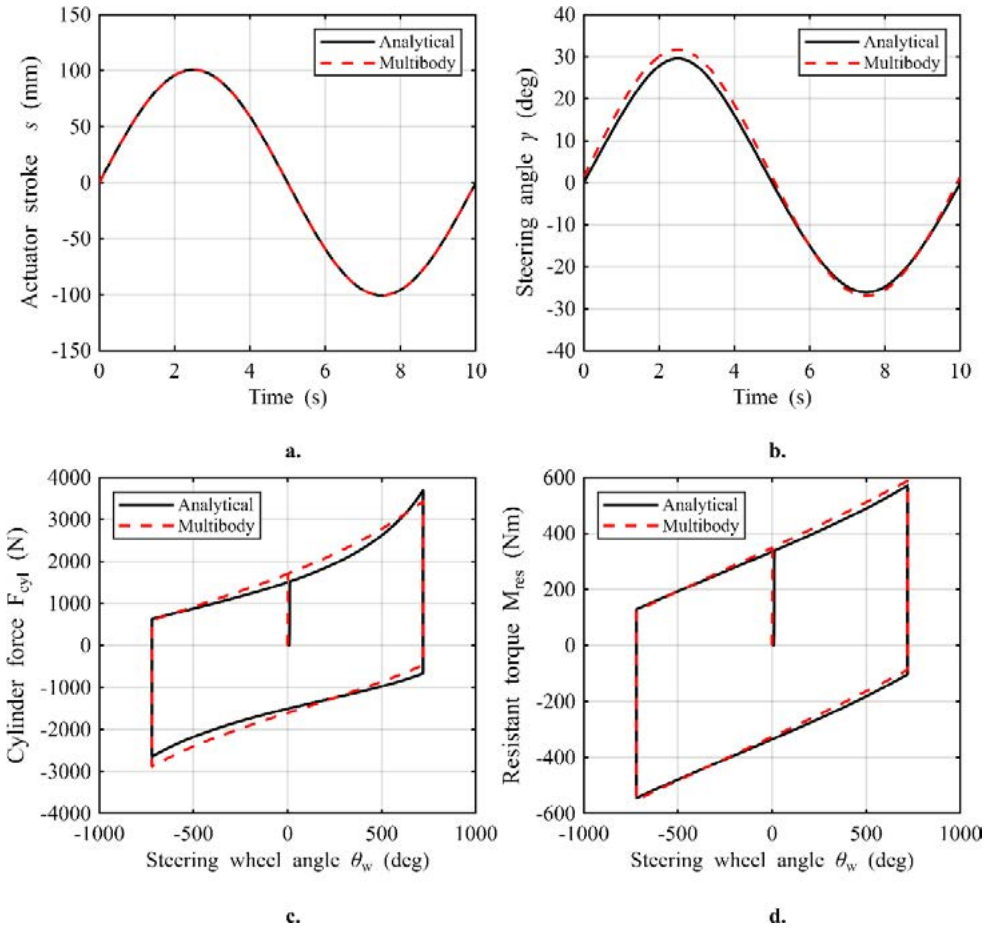


Figure 5.1. Simulation results for T4.75.4RM on Fertile Loam

Figure 5.2 illustrates a comparative analysis between the lighter New Holland T4.75 and the heavier T6.180 across four simulated soil conditions: Asphalt (A), Standard Sandy Loam (SSL), Fertile Loam (FL), and Moist Heavy Clay (MHC). For the New Holland T4.75.4RM (Figure 5.2 a-b), the impact of soil type on steering loads is minimal: the resulting resistance torques is similar in different scenarios. In fact, with a low vertical load, the deformation of the contact area and the corresponding soil shear stress are minimal, thereby reducing the variability of the steering resistance.

In contrast, in the case of the heavier vehicle (Figure 5.2 c-d), the influence of soil characteristics becomes significant. The heavier load on the front axle increases the tire-soil contact pressure, resulting in greater static sinkage (approximately 10 cm). As a result, more cohesive or compact soils generate the strongest torques, while softer soils allow easier lateral movement of tires. A distinct behavior can be observed on clayey soil (MHC): this soil provides low surface resistance, which is at most balanced by the self-aligning torque due to the kingpin's geometry during the rod's return movement. It is important to highlight that this kind of soil provided the greatest resistance when the tire sinkage is severe (for instance, during overload or "stuck" scenarios).

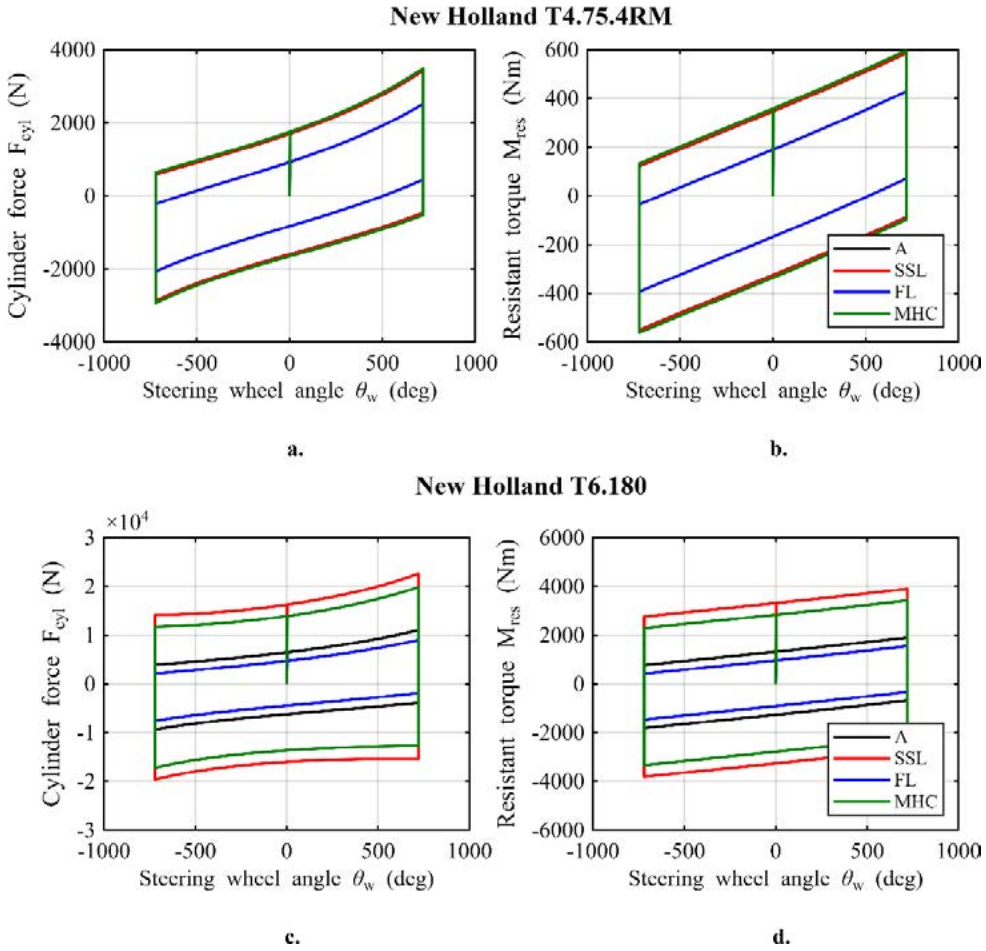


Figure 5.2. Comparison of the simulation results for New Holland T4.75.4RM (a-b) and New Holland T6.180 (c-d) on different soil

6. CONCLUSIONS

This work supports the performance evaluation of hydraulic steering systems in heavy vehicles, highlighting the role of digitalization in laboratory reproduction of operational scenarios through a digital twin approach. The proposed layout represents modular and scalable architecture, facilitating the acquisition of hydraulic and mechanical data that are often estimated during vehicle tests. This model represents the preliminary step of a forward-looking Hardware-in-the-Loop (HIL) approach. Despite the simplifying hypotheses adopted, the results appear robust and physically consistent, supported by the tuning of the terra-mechanical parameters according to the analytic approach.

Driven by growing industrial interest in improving the efficiency of on-board hydraulic systems, especially steering, future research will aim to simulate different front axle configurations. In addition, future developments will extend the model to include vehicle dynamics, influence of tire inflation pressure and non-stationary steering maneuvers.

Finally, the implementation of a Digital Twin, characterized by real-time bidirectional data exchange and synchronization between the digital model and the physical platform, represents the goal of this research project. Future works will focus on this final integration step and on the experimental validation of the complete co-simulation loop. This approach lays the groundwork for future advancements in virtual testing methodologies, contributing to more sustainable and efficient development of off-road vehicle systems.

ACKNOWLEDGMENTS

The authors would like to acknowledge CNH Industrial for the collaboration in the modeling and research activity. The authors would also like to thank Walvoil SpA and Ognibene SpA for providing their support in terms of know–how and materials. This research has been funded by PR-FESR EMILIA ROMAGNA 2021-2027 in the context of the project FACT Future of Agricultural Tractors (ID 38034, CUP E37G22000510007).

7. REFERENCES

[1] M. Borghi, B. Zardin, F. Pintore, F. Belluzzi, 'Energy Saving in the Hydraulic Circuit of Agricultural Tractor', Energy Procedia 45, 2014, https://doi.org/10.1016/j.egypro.2014.01.038.

[2] D. Padovani, P. Dimitriou, T. Minav, 'Challenges and solutions for designing Energy-Efficient and Low-Pollutant Machines in Off-Road hydraulics', Energy Conversion and Management, 2024, https://doi.org/10.1016/j.ecmx.2024.100526.

[3] A. Gaiola, B. Zardin, P. Casoli, et. al., 'The Hydraulic Power Generation and Transmission on Agricultural Tractors: feasible architectures to reduce dissipation and fuel consumption – Part 1', E3S Web of Conferences 197, 2020, https://doi.org/10.1051/e3sconf/202019707009.

[4] P. Casoli, B. Zardin, et. al., 'The Hydraulic Power Generation and Transmission on Agricultural Tractors: feasible architectures to reduce dissipation and fuel consumption – Part 2', E3S Web of Conferences 197, 2020, https://doi.org/10.1051/e3sconf/202019707010.

[5] H. Du, et. al., 'The design, simulation, and experiment of high-accuracy multi-axle electro-hydraulic control servo steering system', Advances in Mechanical Engineering, 2016, https://doi.org/10.1177/1687814016674383.

[6] B. Zardin, M. Borghi, F. Gherardini, N. Zanasi, 'Modelling and Simulation of a Hydrostatic Steering System for Agricultural Tractors', Energies, 2018, https://doi.org/10.3390/en11010230.

[7] Y. Gültekin, et. al., 'Modelling and Simulation of Power Steering System for Agricultural Tractors', International Journal of Advances on Automotive and Technology, 2016.

[8] G. Panetta, F. Mancarella, M. Borghi, et. al., 'Dynamic Modelling of an Off-Road Vehicle for the Design of a Semiactive, Hydropneumatic Spring-Damper System', ASME International Mechanical Engineering Congress and Exposition Proceedings, 2015.

[9] I. Schröder, T. Schmidt, J. Krusborg, C. Woernle, 'Parametrization of a Real-Time Vehicle Model from Driving Tests for HiL Testing of Hydraulic Steering Systems', New Trends in Mechanism and Machine Science, EuCoMeS, 2020.

[10] J.A. Pytka, et. at., 'An instrumented vehicle for offroad dynamics testing', Journal of Terramechanics 48, 2011, https://doi.org/10.1016/j.jterra.2011.06.003.

[11] L. Angelucci, M. Mattetti, ‘The development of reference working cycles for agricultural tractors’, Biosystems Engineering, 2024, https://doi.org/10.1016/j.biosystemseng.2024.04.004.
[12] M. Martelli, S. Gessi, P. Marani, et. al., ‘Comprehensive lumped parameter and multibody approach for the dynamic simulation of agricultural tractors with tyre–soil interaction’, IET Cyber-Systems and Robotics, 2023, https://doi.org/10.1049/csy2.12092.
[13] F. Mocera, A. Somà, A. Nicolini, ‘Grousers Effect in Tracked Vehicle Multibody Dynamics with Deformable Terrain Contact Model’, Appl. Sci., 2020, https://doi.org/10.3390/app10186581.
[14] J. Rekem, et. al.,‘Multi-Body Model of Agricultural Tractor for Vibration Transmission Investigation’, Appl. Sci., 2024, https://doi.org/10.3390/app14188451.
[15] H. Yarmohamadi, V. Berbyuk, ‘Kinematic and dynamic analysis of a heavy truck with individual front suspension’, Vehicle System Dynamics, 2013.
[16] S. Laughery et. al., ‘Off-Road Vehicle Locomotion Using Bekker's Model’, SPIE, Orlando, 2000.
[17] L. Angelucci, M. Varani, F. Pinet, et. al., ‘The role of tyres and soil conditions in enhancing the efficiency of agricultural tractors’, Soil and Tillage Research, 2025.
[18] E. Hegedus, ‘A simplified method for the determination of bulldozing resistance’, Land Locomotion Research, Laboratory, Army Tank Automotive Command Report, 61, 1960.
[19] M.G. Bekker, ‘Introduction to Terrain-Vehicle Systems’, University of Michigan Press, Ann Arbor, MI, 1969.
[20] J. J. Taborek, ‘Mechanics of Vehicles’, Penton Publishing Co., Cleveland, 1957.
[21] K. Terzaghi, ‘Theoretical Soil Mechanics’, John Wiley and Son Inc, USA, 1943.
[22] Z. Jia, et. al., ‘Terramechanics-based wheel–terrain interaction model and its applications to off-road wheeled mobile robots’, Robotica, 30(3), 2012, https://doi.org/10.1017/S0263574711000798.
[23] G. Ishigami, et. al., Terramechanics-Based Model for Steering Maneuver of Planetary Exploration Rovers on Loose Soil, Journal of Field Robotics, 2007, https://doi.org/10.1002/rob.20187.
[24] B. P. Patel and J. M. Prajapati, ‘Evaluation of Resistive Force using Principle of Soil Mechanics for Mini Hydraulic Backhoe Excavator’, International Journal of Machine Learning and Computing, Vol. 2, No. 4, 2012.
[25] D. Cao, et. al., ‘Study on Low-Speed Steering Resistance Torque of Vehicles Considering Friction between Tire and Pavement’, Appl. Sci., 2019, https://doi.org/10.3390/app9051015.
[26] Danfoss, Hydraulic Power Steering Systems Catalogue, Danfoss Power Solutions.
[27] Ognibene Power, Hydraulic Steering Cylinders Catalogue, Ognibene S.p.A.
[28] HydraForce, TS12-27 Solenoid Valve Technical Catalogue, HydraForce Inc.
[29] MSC Software Corporation, Adams 2023: Multibody Dynamics Simulation Software.
[30] MathWorks Inc., Simulink: Dynamic System Simulation for MATLAB, Release R2024b.
[31] J.Y. Wong, ‘Theory of Ground Vehicles’, 3rd ed., New York, Wiley, 1997.
[32] New Holland Agriculture, T4 Series Tractor Technical Datasheet, CNH Industrial.
[33] New Holland Agriculture, T6 Series Tractor Technical Datasheet, CNH Industrial.

Biographies



Matteo Bertoli obtained a bachelor's degree in mechanical engineering from University of Ferrara in 2021 and master's degree in mechanical engineering at the "Enzo Ferrari" Engineering Department, University of Modena and Reggio Emilia, in 2024. After one year as a research fellow, he decided to pursue a PhD, driven by growing interest in fluid power applications and hydraulic systems.



Barbara Zardin is an Associate Professor of Fluid Machines and Energy Systems. Her research focuses especially on evaluating and improving the efficiency of fluid power systems and components for mobile applications. She has written over 40 research papers and articles, and is currently the local principal investigator for the University of Modena and Reggio Emilia on the European project H2REF-DEMO (co-funded by the European Union's "Horizon. Europe" programme under the "Clean Hydrogen Partnership", grant agreement no. 101101517), and the PR-FESR Emilia-Romagna 2021–2027 project, "FACT: Future of Agricultural Tractors" (ID 38034, CUP E37G22000510007).



Massimo Borghi is a full professor of Fluid Machines and Energy Systems. His research interests span internal combustion engines and fluid power systems and components, including hydrogen applications. He has written over 90 research articles and papers, making significant contributions to the analysis, modelling and simulation of hydraulic pumps, motors and systems for mobile applications (in both the automotive and off-highway sectors).

Power optimization strategy design for electric wheel loader work function with variable speed and pump displacement control

Jiaming Wu[1], Yihan Qiao[1], Feng Wang[1*], Bing Xu[1]

[1] *State Key Laboratory of Fluid Power Components and Mechatronic Systems in the School of Mechanical Engineering, Zhejiang University, Hangzhou, Zhejiang, 310027, China*

[*] *Corresponding author. E-mail address: dieter@zju.edu.cn*

Abstract.

Powertrain electrification is key to decarbonizing mobile machinery like wheel loaders. However, their work functions still rely on centralized valve control, causing significant energy losses. Electro-hydrostatic actuators (EHA) offer a decentralized, efficient alternative but often use fixed-displacement pumps, leading to poor efficiency under high-torque, low-speed conditions. Existing cooperative speed and displacement control are rule-based and neglect system dynamics, limiting optimal effect. This paper proposes a power optimization strategy for electric wheel loader work function with variable speed and pump displacement cooperative control. Offline dynamic programming (DP) first determines globally optimal operating points under typical conditions. Then a real-time optimization strategy is proposed based on the optimal results from DP algorithm. Results show the proposed system achieves a 9.12% reduction in energy consumption compared with the EHA system under the familiar cycle (9.76% for DP system). Under another unfamiliar cycle, the proposed system still has a 4.9% energy consumption reduction, demonstrating its energy-saving potential across various operating conditions.

Keywords. Electro-hydraulic actuator, variable speed and pump displacement control, dynamic programming, distributed hydraulic system, mobile machine.

1. INTRODUCTION

Powertrain electrification has emerged as a pivotal technological pathway for reducing carbon emissions in mobile machinery and represents a major focus of current research [1]. Among various types of electric mobile equipment, wheel loaders have achieved the highest market penetration [2]. Conventional electric wheel loaders typically utilize two electric motors—one for wheel drive and another for the working functions, and the latter consumes a substantial portion of the total battery energy. Although these systems employ high-efficiency electric motors, the working functions often continue to rely on centralized

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

hydraulic valve control systems [3]. This results in significant throttling losses, reduced overall system efficiency, and limited operational duration per charge [4].

Decentralized hydraulic systems have gained increasing attention due to their potential for higher energy efficiency and superior control performance compared to centralized configurations [5, 6]. A prominent solution in this category is the electro-hydrostatic actuator (EHA), which uses a variable-speed electric motor as the prime mover for each hydraulic actuator. Many studies have been conducted on that solution. Qu applied an electro-hydrostatic variable speed drive system in a skid steer loader, achieving a steady-state energy efficiency of up to 59.9% [7]. Ge proposed a EHA driving by an asymmetric pump and the energy consumption during the entire boom cylinder loading cycle can be reduced by 75.0 % [8]. Guo proposed a variable speed drive steering system for wheel loaders, which reduced energy consumption by 22.8% compared with centralized systems [9].

However, in most EHA implementations, the hydraulic pump displacement remains fixed. This limitation forces the electric motor's operating points to be dictated directly by the external load, leading to relatively low system efficiency—particularly under high-torque, low-speed conditions. To address this challenge, several cooperative control strategies for simultaneous speed and displacement adjustment have been proposed. Huang proposed an active load-sensitive EHA with a dual control variable system. The output flow of the pump is implemented by adjusting the motor speed and the pump displacement based on the load-sensitive principle. The system efficiency is optimized in real time and the system dynamic performance is improved [10, 11]. Lin developed a combined motor speed and hydraulic pump displacement control strategy, achieving hierarchical system pressure control, reducing energy consumption by up to 70% [12].

Nevertheless, existing methods predominantly rely on simplified rule-based approaches, which fail to achieve optimal system energy efficiency. Moreover, these strategies often neglect system dynamic characteristics, thereby limiting their practical effectiveness. The dynamic programming (DP) algorithm is a well-established method for solving global optimization problems in dynamic systems [13, 14]. However, its high computational burden renders it unsuitable for real-time implementation.

In this context, this paper proposes an energy-optimized control strategy for the work function of electric wheel loaders, based on cooperative variable speed and displacement control. The DP algorithm is first applied offline to derive efficiency-optimal operating parameters under typical working cycles. These results then guide the design of a real-time optimization strategy that incorporates system dynamics. The rest of the paper is organized as follows. Section II presents system mathematical modelling, section III shows DP algorithm and real-time optimization strategy, section IV presents simulation study and section V gives conclusions.

2. MATHEMATICAL MODELLING OF ELECTRO-HYDRAULIC ACTUATOR SYSTEM

2.1. System Overview

The schematic of EHA system with variable speed and pump displacement control for wheel loaders is shown in Figure 2.1. It uses an electric motor (EM) as a primary mover for each

hydraulic working function and is able to optimize the working points of both electric motor and hydraulic pump (HP). The working hydraulic system consists of lift and tilt functions. The hydraulic variable displacement pumps can work in four quadrants. Two 4/3 directional valves (DV) allow for load holding capability. An on-off valve works as a bypass valve (BPV) parallel to the tilt cylinder (TC). The accumulators (ACC) compensate for the differential flow of the cylinders during extension phase and intake excess flow during retraction phase. Two sets of pilot check valves (PCV) control the charging and discharging process of the accumulator.

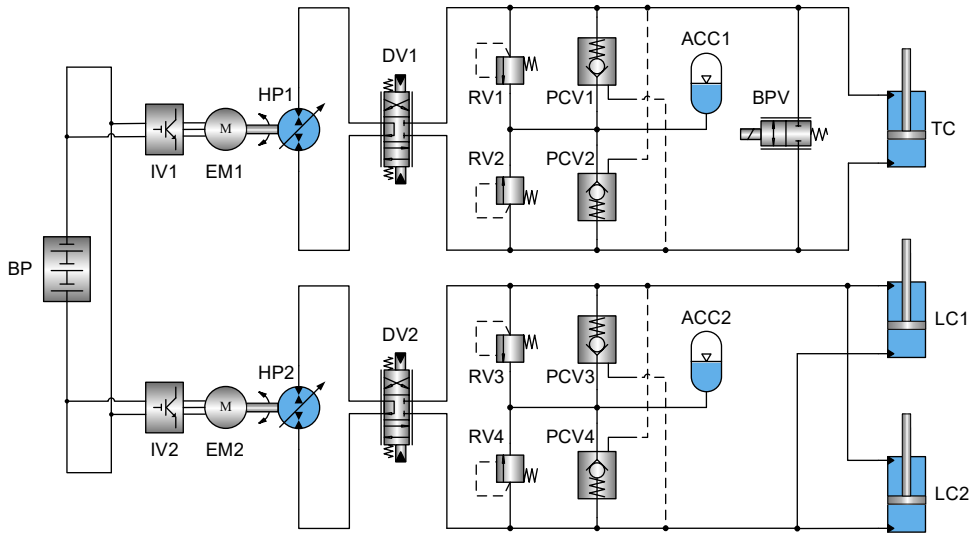


Figure 2.1. Schematic of electro-hydraulic actuator system with variable speed and pump displacement control

2.2. System Modelling

The physical-based mathematic models are established in this study for the following power optimization strategy design, including the models of battery pack, electric motors, hydraulic pumps, and so on. The SI unit is used in the mathematic models.

Battery Pack

A widely used equivalent circuit model composed of a voltage source, U_{oc}, and a resistance, R_b, is used for battery modelling [15,16]. The battery charge/discharge current, I_b, is given by:

$$I_b = \frac{U_{oc} - \sqrt{U_{oc}^2 - 4R_b P_{bout}}}{2R_b} \tag{2.1}$$

where P_{bout} is the output power (electric load power) of battery and can be expressed as:

$$P_{bout} = P_{iv1} + P_{iv2} \tag{2.2}$$

where P_{iv1}, and P_{iv2} are the input power of two inverters respectively.

Electric Motor

The electric motors in this study are permanent magnet synchronous motors (PMSM) that can work either as motors or as generators. A commonly used mathematical model of PMSM is to present electromagnetic vectors in an orthogonal d-q coordinate with d axis coinciding rotor flux linkage direction [17]. The winding stator voltage components in the d-q axis, u_d, u_q are:

$$\begin{cases} u_d = R_s i_d + \frac{d\psi_d}{dt} - \omega_{em}\psi_d \\ u_q = R_s i_q + \frac{d\psi_q}{dt} + \omega_{em}\psi_q \end{cases} \quad (2.3)$$

where R_s is stator winding's equivalent resistance, ω_{em} is the shaft speed of electric motor, i_d and i_q are equivalent currents in the d-q axis, ψ_d and ψ_q are d-q components flux linkage.

The electric motor output shaft torque T_{em} is:

$$T_{em} = \frac{3}{2} n_p i_q \left[\psi_f + (L_d - L_q) i_d\right] \quad (2.4)$$

where n_p is the number of pole pairs of motor stator.

The dynamics of electric motor shaft speed is:

$$(J_{em} + J_{hp})\dot{\omega}_{em} = T_{em} - T_{hp} - k_{vf}\omega_{em} \quad (2.5)$$

where J_{em} and J_{hp} are shaft inertia of electric motor and hydraulic pump, T_{hp} is hydraulic pump torque, k_{vf} is rotary friction coefficient.

Hydraulic Pump

The shaft torque of hydraulic pump, T_{hp}, is:

$$T_{hp} = \frac{D\Delta p}{\eta_{mech}{}^{\beta}} \quad (2.6)$$

where D is pump displacement, Δp is pressure difference between pump outlet and inlet pressure, η_{mech} is pump mechanical efficiency, β is mode factor ($\beta = 1$ in pumping mode and $\beta = -1$ in motoring mode).

The pump output flow rate, q_{hp} is:

$$q_{hp} = D\omega_{em}\eta_{vol}{}^{\beta} \quad (2.7)$$

where η_{vol} is pump volumetric efficiency.

3. DYNAMIC PROGRAMMING ALGORITHM AND REAL-TIME STRATEGY

3.1. Dynamic Programming Algorithm

The Dynamic Programming is a powerful algorithmic technique for solving complex optimization problems by breaking them down into simpler, overlapping subproblems. Its core principle is to solve each subproblem only once, store the solution, and reuse it when needed, thereby avoiding redundant calculations and improving efficiency, as shown in Figure 3.1. This "divide-and-conquer-with-memorization" approach is particularly effective for sequential decision-making processes. The method relies on Bellman's principle of

optimality, which states that an optimal policy consists of optimal sub-policies. By recursively defining the value of a state based on subsequent states, DP constructs an optimal solution from back to front, making it ideal for applications like pathfinding, resource allocation, and optimal control strategies.

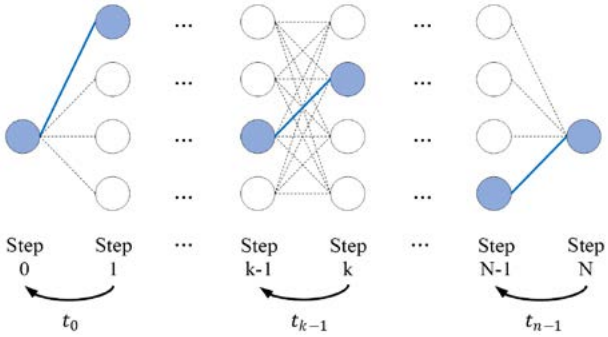


Figure 3.1. Schematic of dynamic programming algorithm

DP is adopted in this paper to obtain the efficiency-optimal operating parameters (rotary speed of electric motor and pump displacement) under typical working conditions of wheel loaders. Considering that the calculation complexity of DP algorithm increases exponentially with the dimensions of state variables and control variables, the pressure dynamics in the pipelines are neglected for simplicity and feasibility.

The flow chat of dynamic programming algorithm is shown in Figure 3.2. After getting the piston velocity and load force from loading cycle, the system operating mode is classified. The operating mode of wheel loaders normally includes two quadrants since the load force direction remains the same.

During extension mode ($v > 0$), the system power losses are given by:

$$P_{pvl} = \Delta p \omega_{em} D(1-\eta_{vol}) = \Delta p \frac{A_{head}}{\eta_{vol}}(1-\eta_{vol})v \quad (2.8)$$

$$P_{pml} = \Delta p \omega_{em} D \frac{1-\eta_{mech}}{\eta_{mech}} = \Delta p \frac{A_{head}}{\eta_{vol}} \frac{1-\eta_{mech}}{\eta_{mech}} v \quad (2.9)$$

$$P_{ml} = \omega_{em} T_{hp} \frac{1-\eta_{em}}{\eta_{em}} = \Delta p \frac{A_{head}}{\eta_{vol}\eta_{mech}} \frac{1-\eta_{em}}{\eta_{em}} v \quad (2.10)$$

$$P_{bl} = I_b^2 R_b \quad (2.11)$$

where P_{pvl} and P_{pml} are volumetric and mechanical power losses of hydraulic pump, P_{ml} is power loss of electric motor and inverter, P_{bl} is power loss of battery.

During retraction mode ($v < 0$), the system operations between lift and tilt functions are different. The lift function recoveries gravitational potential energy of loader working device and its hydraulic pump works in motoring mode. The system power losses are given by:

$$P_{pvl} = \Delta p A_{head}(1-\eta_{vol})v \quad (2.12)$$

$$P_{pml} = \Delta p A_{head}(1-\eta_{mech})\eta_{vol} v \quad (2.13)$$

$$P_{ml} = \Delta p A_{head}\eta_{vol}\eta_{mech}(1-\eta_{em})v \quad (2.14)$$

$$P_{bl} = I_b^2 R_b \tag{2.15}$$

For the tilt function, the bypass valve placed in parallel to the tilt cylinder is open during retraction mode to accelerate the motion process. Since the pressure difference between pump outlet and inlet pressure equals to zero, the system power losses during the bucket retracting are neglected.

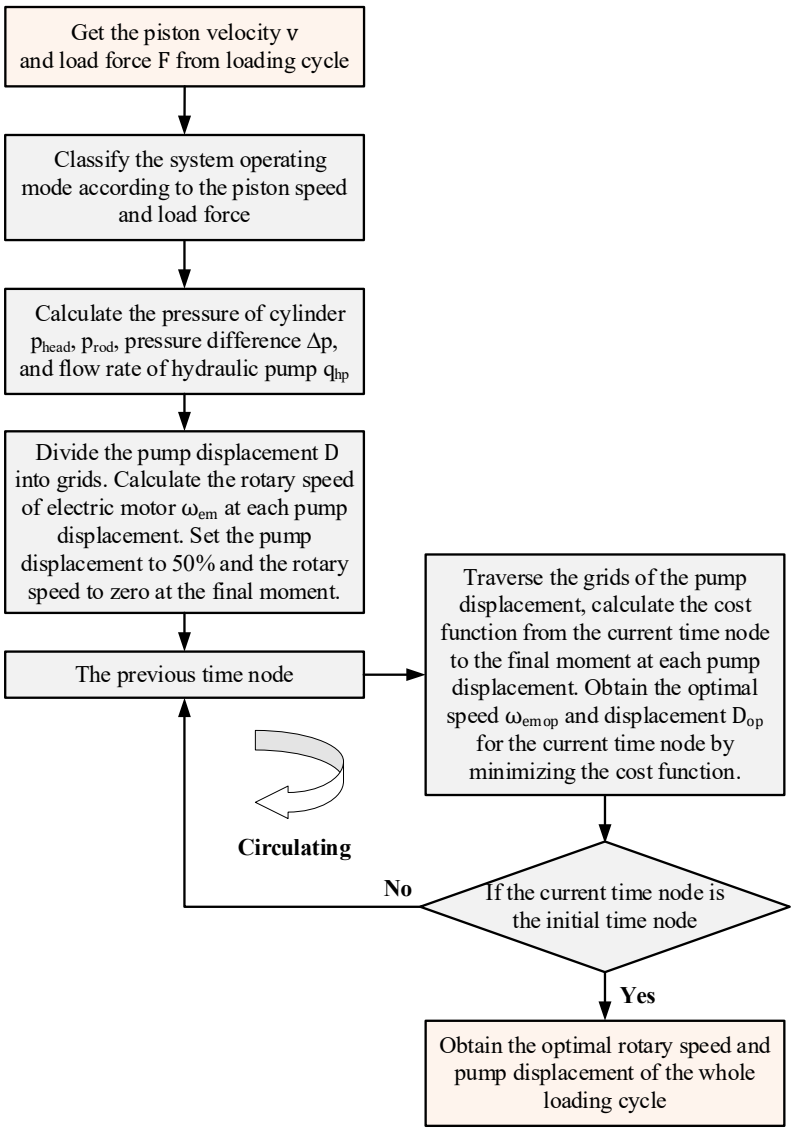


Figure 3.2. Flow chat of dynamic programming algorithm

After calculating the system states, a set of pump displacement is selected from 2.5 cc/rev to 250 cc/rev (100% maximum displacement) with an interval of 2.5 cc/rev. Then the rotary speed of electric motor at each pump displacement is calculated according to Eq. (2.7). To

facilitate the DP algorithm to find the optimal displacement more quickly, the pump displacement at the final moment is set to 50%. The rotary speed at the final moment is set to zero.

During the dynamic programming process, the rotary speed, acceleration of electric motor and the rate of change of the pump displacement should be restricted within proper range according to the real dynamics of electric motor and hydraulic pump. Therefore, a penalty function is added to the optimization problem to discard the state exceeding physical limitations in the optimization process, since a larger value from penalty function will be added besides cost function.

By setting the weight coefficient of all constraints. The expression of the penalty function is given by:

$$\begin{aligned} f_{pen} = k_{penalty}\big[&H(\omega_{em} - \omega_{emmax}) + H(\omega_{em} - \omega_{emmin}) + H(\dot{\omega}_{em} - \dot{\omega}_{emmax}) + \\ &H(\dot{\omega}_{em} - \dot{\omega}_{emmin}) + H\big(\dot{D} - \dot{D}_{max}\big) + H\big(\dot{D} - \dot{D}_{min}\big)\big] \end{aligned} \tag{2.16}$$

$$\dot{\omega}_{emmax} = \frac{T_{emmax}\text{-}T_{hp}\text{-}k_{vf}\omega_{em}}{J_{em}+J_{hp}} \tag{2.17}$$

$$\dot{\omega}_{emmin} = \frac{T_{emmin}\text{-}T_{hp}\text{-}k_{vf}\omega_{em}}{J_{em}+J_{hp}} \tag{2.18}$$

where $k_{penalty}$ is weight coefficient of penalty function, ω_{emmax} and ω_{emmin} are the maximum and minimum rotary speed of electric motor, $\dot{\omega}_{emmax}$ and $\dot{\omega}_{emmin}$ are the maximum and minimum rotary acceleration of electric motor, T_{emmax} and T_{emmin} (reverse) are the maximum and minimum output shaft torque of electric motor, $\dot{D}_{max}$ and $\dot{D}_{min}$ are the maximum and minimum rate of change of the pump displacement obtained from test results.

The cost function from the current time node to the final moment at each pump displacement is given by:

$$J_k = \int_{kT_s}^{(k+1)T_s} P_{pvl} + P_{pml} + P_{ml} + P_{bl}dt + f_{pen,k} + J_{k+1} \tag{2.19}$$

where J_k, J_{k+1} are cost functions of time nodes k and $k+1$, T_s is discrete-time timestep.

By traversing the grids of the pump displacement, the optimal pump displacement for the current time node is obtained by minimizing the cost function:

$$D_{op,k} = \underset{D \in \Omega}{\operatorname{argmin}}\, J_k \tag{2.20}$$

The nested loops are used that start at the end and the optimal pump displacement is computed for every node recursively.

3.2. Real-Time Rotary Speed and Pump Displacement Optimization Strategy

Although DP is an effective method to solve global optimization problems, its substantial computational demands make real-time optimization infeasible. Therefore, in this paper, a real-time rotary speed and pump displacement optimization strategy is designed based on the optimized results from dynamic programming algorithm.

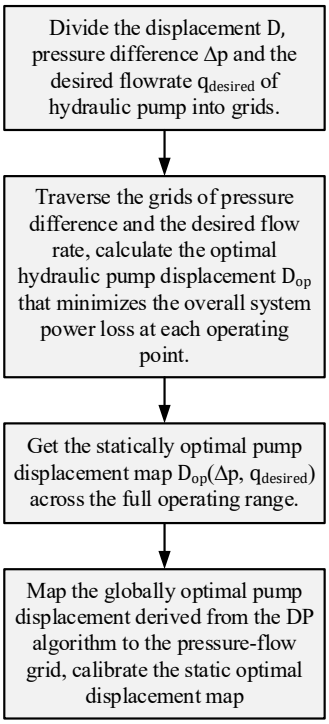


Figure 3.3. Flow chat of real-time optimization strategy

The flow chat of real-time optimization strategy is shown in Figure 3.3. The real-time optimization strategy is a two-stage methodology for determining the globally optimal displacement of a hydraulic pump. The first stage is to obtain the statically optimal pump displacement map $D_{op}(\Delta p, q_{desired})$ across the full operating range. Key operational parameters (displacement D, pressure difference Δp, and desired flow rate $q_{desired}$ of hydraulic pump) are discretized into grids in the beginning. For each discrete operating point within this grid, it calculates the specific pump displacement that minimizes the overall system power loss. This exhaustive search yields a preliminary, statically optimal displacement map, which defines the most efficient pump setting for any given pressure and flow requirement within the predefined operating range.

In the final and crucial step, the results from DP algorithm, which computes a sequence of optimal decisions over time, are mapped onto the same pressure-flow grid. This integration allows the static map to be calibrated and refined with the globally optimal trajectory from the DP solution, thereby enhancing its accuracy and ensuring system-wide optimality.

4. SIMULATION STUDY

4.1. Simulation Parameters

The dynamic simulation model of the proposed system is developed in Amesim. The simulation models target a 12 tons electric wheel loader. The major parameters are shown in Table 4.1. The nominal voltage of battery is 648 V. The accumulator volume is calculated to compensate for the differential volume of the cylinders during extension phase. The

displacement of the hydraulic pumps determines the shaft speed limits of electric motors. In the simulation, two 22 L accumulators are selected for the tilt and lift functions.

The efficiency map of electric motor is obtained from the experimental data as shown in Figure 4.1. The flowrate and shaft torque of hydraulic pump/motor is shown in Figure 4.2 and Figure 4.3. The first loading cycle used in this study is 60 s long and includes six phases: preparing, bucket filling, load holding, boom lifting, dumping and boom lowering.

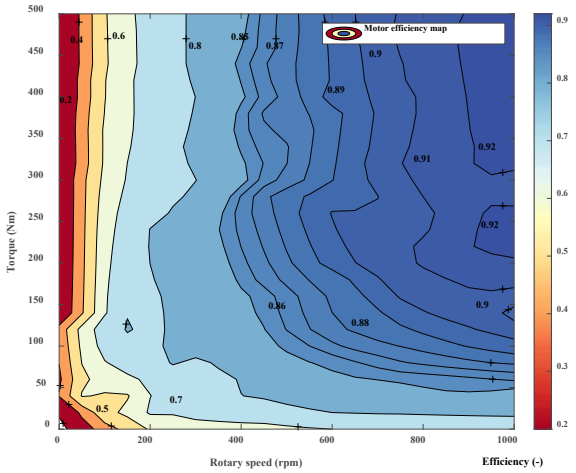


Figure 4.1. Efficiency map of the electric motor

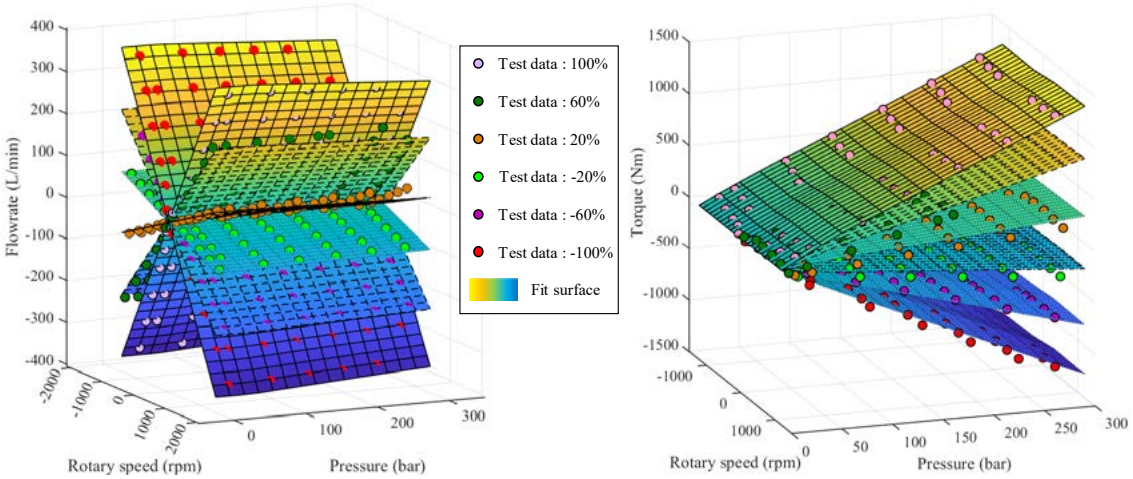


Figure 4.2. Flowrate and shaft torque of hydraulic pump/motor

Table 4.1. Main simulation parameters

Property		**Value**	**Unit**
Battery	Nominal voltage	648	V
Electric motor	Maximum rotary speed	2000	rpm
	Maximum torque	648	Nm
Hydraulic pump	Geometric displacement	250	cc/rev

Accumulator	Volume	22.0	L
	Precharge pressure	8	bar
Direction valve	Maximum throttling area	256.12	mm2
	Natural frequency	20.0	Hz
	Damping ratio	1.2	-
Tilt/Lift cylinder	Piston stroke	350.0/450.0	mm
	Piston diameter	180.0/150.0	mm
	Rod diameter	80.0/90.0	mm

4.2. Simulation Results

There are three hydraulic system configurations considered in the simulation study: normal EHA system, variable speed and pump displacement control system with DP algorithm (VSVD-DP), variable speed and pump displacement control system with real-time optimization strategy (VSVD-RT).

In the EHA system, each actuator is driven by a stand-alone electric motor with a fixed displacement hydraulic pump and the actuator speed is controlled through variable speed drive of electric motor. The 4/3 directional valve is fully open when the actuator moves and closes in standby mode. The comparison of piston displacement between three systems is shown in Figure 4.3. The piston displacements of the three systems are essentially consistent, except for the discrepancy during the bucket lowering phase. This is due to the drastic changes in piston velocity during the bucket lowering phase, which poses challenges to the control accuracy of variable speed and displacement system.

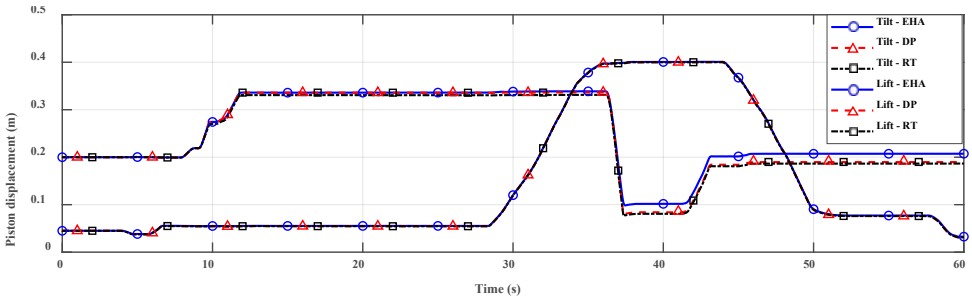


Figure 4.3. Comparison of piston displacement between three systems

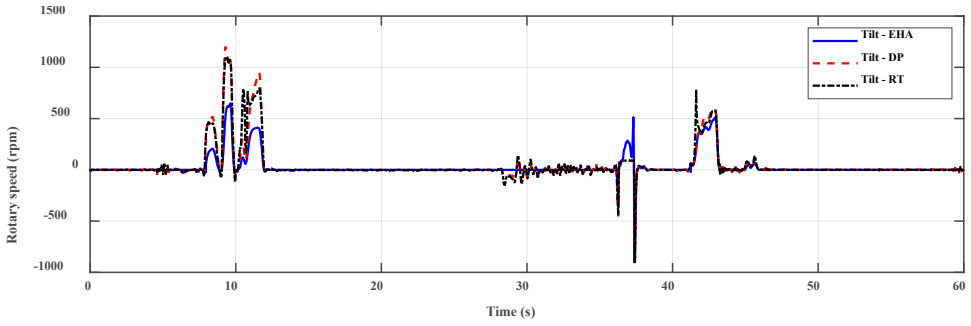


Figure 4.4. Comparison of tilt rotary speed between three systems

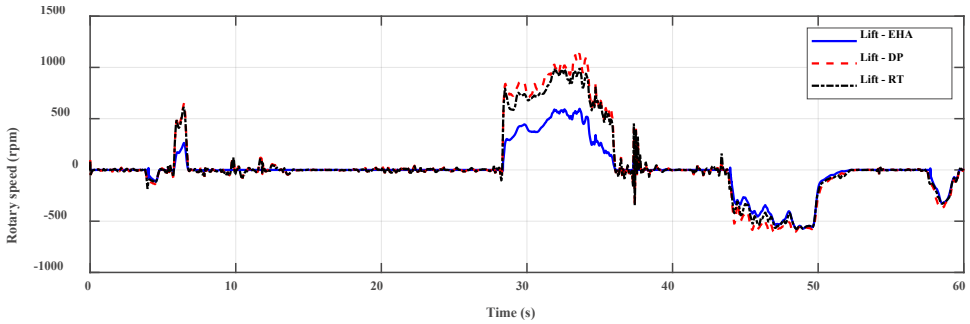


Figure 4.5. Comparison of lift rotary speed between three systems

The comparison of rotary speed between three systems is shown in Figure 4.4 and Figure 4.5. As the 4/3 directional valve opens and closes during system status switching process, there are some pressure shocks leading to the peak fluctuations of motor shaft torque in the EHA system. Compared with the EHA system, the shaft speeds of DP and RT systems are much higher during bucket filling and boom lifting phase.

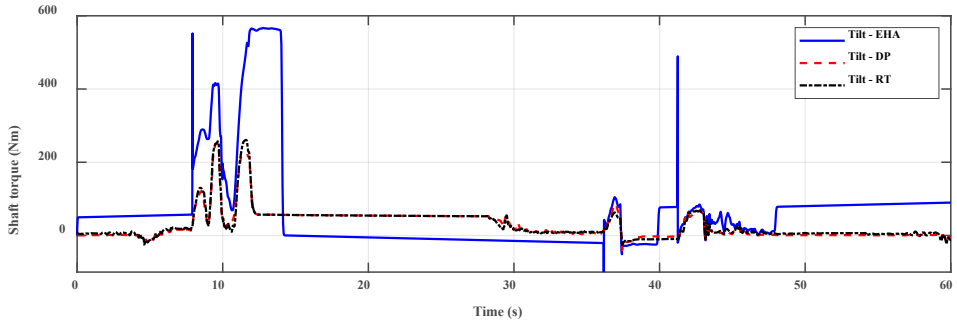


Figure 4.6. Comparison of tilt shaft torque between three systems

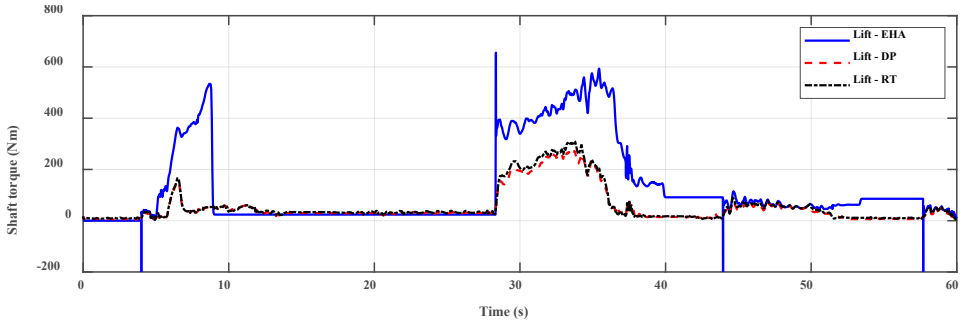


Figure 4.7. Comparison of lift shaft torque between three systems

The shaft torque results are shown in Figure 4.6 and Figure 4.7. Due to the reduction in hydraulic pump displacement, the motor output torque of the DP and RT systems in operating mode decreases significantly compared with EHA system. The pump displacement results between three systems are shown in Figure 4.8 and Figure 4.9. The pump displacement in the EHA system is always maintained at 100%, while the pump

displacement in the DP and RT systems is adjusted in real time according to operating conditions. By adjusting the pump displacement, the operating points of the motors in the DP and RT systems shift toward the high-efficiency region characterized by high speed and low torque. The 4/3 directional valves in the DP and RT systems are fully open from the beginning to the end. In the standby mode, both DP and real-time strategies keep the pump displacement close to zero to reduce the power loss of electric motor.

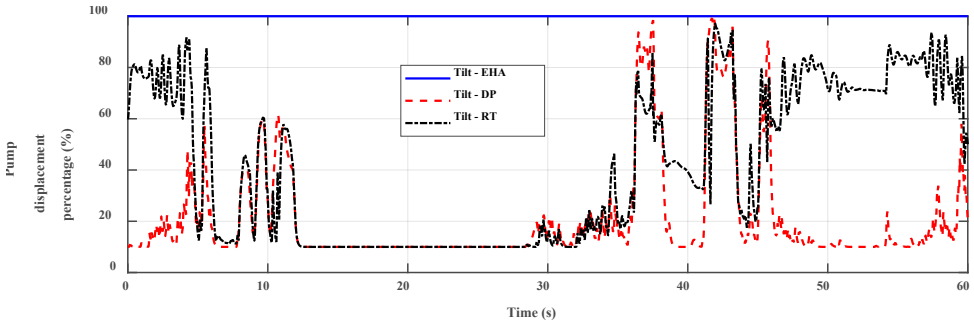


Figure 4.8. Comparison of tilt pump displacement between three systems

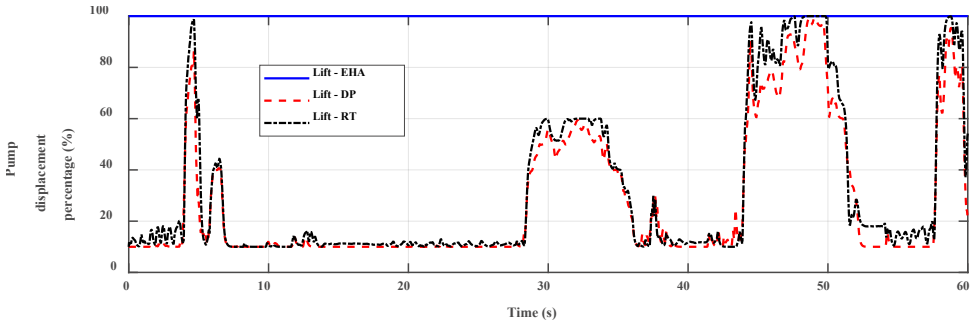


Figure 4.9. Comparison of lift pump displacement between three systems

The pump displacement percentage curves between DP and RT systems still have slight differences. Since the DP strategy accounts for dynamic constraints of the system, the obtained globally optimal displacement percentage exhibits smoother variations. In contrast, the real-time optimization strategy leads to more fluctuations. However, owing to the calibration of the real-time strategy based on DP results, the pump displacement from RT system is close to the results from DP with much smaller computational complexity.

The comparison of energy consumption between EHA, DP, optimized and non- optimized RT systems is shown in Figure 4.10. The total energy consumptions of these system are 218.3 kJ, 197.0 kJ, 198.4 and 202.9 kJ respectively. Compared with the EHA system, the energy consumptions of VSVD systems are reduced, primarily due to the decrease in electric motor energy loss. The DP system achieves the lowest energy consumption through offline global optimization. Compared to the EHA system, the non-optimized RT system saves 7.05% in energy consumption, while a 9.12% reduction is achieved in the optimized RT system, approaching the optimal result of the DP system.

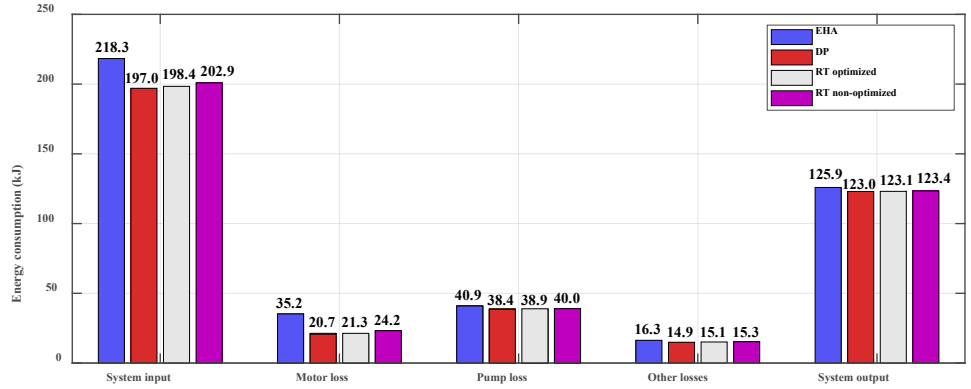


Figure 4.10. Comparison of energy consumption between systems

The aforementioned results verify the energy-saving effectiveness of the real-time optimization strategy under specific operating conditions. To investigate the influence of different operating conditions on the proposed method, an unfamiliar loading cycle lasting 32 s is selected as the system input. The hydraulic cylinder piston displacement curve under the new cycle is shown in Figure 4.11. Compared with the previous cycle, the piston displacement variation is greater, and the intervals between movements are shorter.

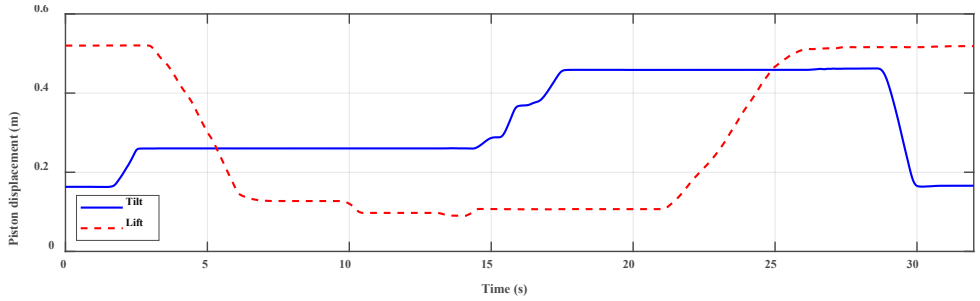


Figure 4.11. Piston displacement under 32 s long loading cycle

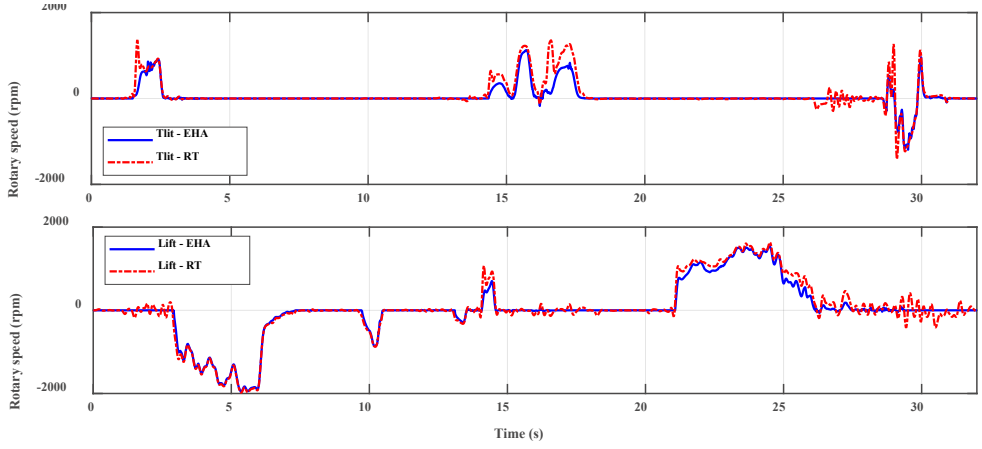


Figure 4.12. Comparison of rotary speed between EHA and RT systems

The comparison of rotary speed between EHA and RT systems is shown in Figure 4.12. The shaft speed of RT system is slightly higher than that of the EHA system. From the results of lift working function, it can be observed that when the cycle requires high flowrate, the motor rotary speeds of the EHA and RT systems are essentially the same, indicating that the hydraulic pump operates at full displacement under such conditions.

The comparison of shaft torque between EHA and RT systems is shown in Figure 4.13. The mean motor shaft torque of the RT system is significantly lower than that of the EHA system. The RT optimization strategy can rapidly reduce motor torque by adjusting the hydraulic pump displacement after the operation is completed, thereby minimizing motor energy loss.

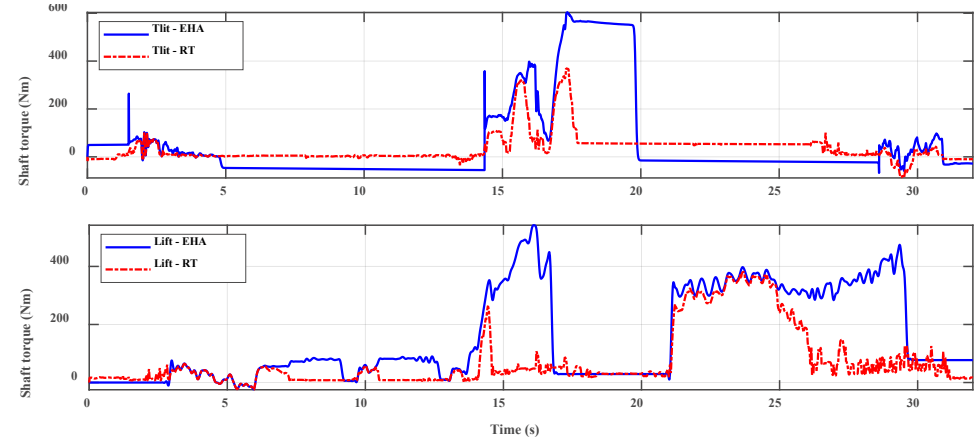


Figure 4.13. Comparison of shaft torque between EHA and RT systems

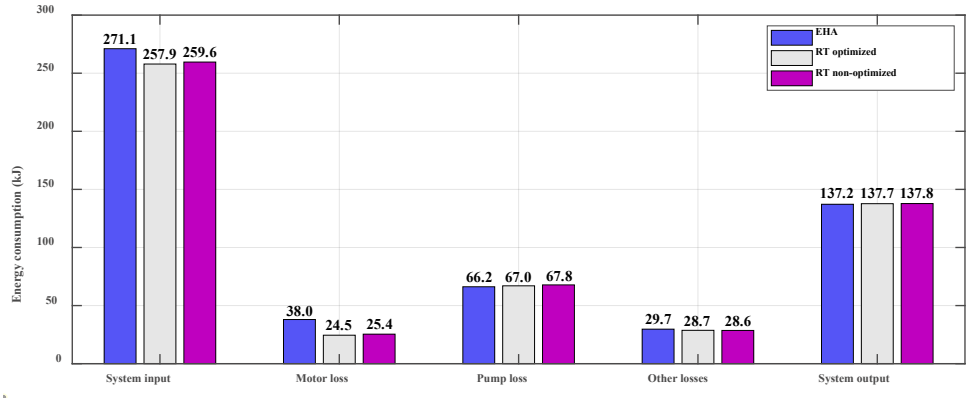


Figure 4.14. Comparison of energy consumption between EHA and RT systems

The comparison of energy consumption between EHA, optimized and non- optimized RT systems is shown in Figure 4.14. The total energy consumptions of these system are 271.1 kJ, 257.9 kJ and 259.6 kJ respectively. Both VSVD systems have smaller energy consumptions than that of the EHA system under an unfamiliar loading cycle. Due to the increased power output required for short loading cycle, the operating time under full hydraulic pump displacement is extended in the VSVD systems, resulting in a reduced margin of energy consumption reduction compared to the EHA system. Although the new

cycle shortens the overall time with an increased piston displacement change, the motion pattern of the working functions remains similar to that of the long cycle. Consequently, the optimized RT strategy achieves additional system energy savings compared to the non-optimized one.

5. CONCLUSIONS

In this paper, a power optimization strategy for electric wheel loader work function with variable speed and pump displacement cooperative control is proposed. The dynamic programming algorithm is applied offline to obtain the efficiency-optimal operating parameters under typical working conditions of wheel loaders. Based on the optimized results, a real-time rotary speed and pump displacement optimization strategy is designed.

The dynamic simulation models of three hydraulic systems (EHA, VSVD-DP, VSVD-RT) are developed in Amesim. Two loading cycles are used in this study. Under the familiar cycle, VSVD-RT system achieves a 9.12% reduction in total energy consumption compared with the EHA system (9.76% for DP system). Under another unfamiliar cycle, the proposed system still has a 4.9% energy consumption reduction. The results indicate that the real-time optimization strategy possesses energy-saving potential across various operating conditions. In practical applications, the real-time model can be calibrated using DP global optimal solutions derived from multiple frequently-used loading cycle, thereby enhancing its adaptability to different operating scenarios.

6. REFERENCES

[1] H. Assaf, S. Sarode, A. Vacca A, et al, ‘Electric machine sizing consideration for ePumps in mobile hydraulics’, Energy science & engineering, vol.12, no. 3, pp.793-809, 2024.

[2] Z. Quan, L. Ge, Z. Wei, et al, ‘A Survey of Powertrain Technologies for Energy-Efficient Heavy-Duty Machinery’, Proceedings of the IEEE, vol. 109, no. 3, pp. 279-308, 2021.

[3] B. Xu, M. Cheng, ‘Motion control of multi-actuator hydraulic systems for mobile machineries: Recent advancements and future trends’, Front Mech Eng, vol. 13, no. 2, pp. 151-166, 2018.

[4] D. B. Beck, D. E. Fischer, D. G. Kolks, et al, ‘Novel system Architectures by individual drives’, In: 10th Int Fluid Power Conf, pp. 29-62, 2016.

[5] J. Lodewyks, P. Zurbrügg, ‘Decentralized energy-saving hydraulic concepts for mobile working machines’, In: 10th Int Fluid Power Conf, pp. 79-90, 2016.

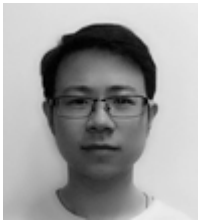
[6] X. He, H. Liu, S. He, et al, ‘Research on the energy efficiency of energy regeneration systems for a battery-powered hydrostatic vehicle’, Energy, vol. 178, pp. 400-418, 2019.

[7] S. Qu, D. Fassbender, A. Vacca, et al, ‘A high-efficient solution for electro-hydraulic actuators with energy regeneration capability’, Energy, vol. 216, pp. 119291, 2021.

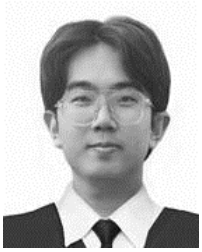
[8] L. Ge, L. Quan, Y. Li, et al, ‘A novel hydraulic excavator boom driving system with high efficiency and potential energy regeneration capability’, Energy Convers Manag, vol. 166, pp. 308–17, 2018.

[9] T. Guo, B. Wu, T. Lin, et al, ‘Closed-Circuit Pump-Controlled Electro-Hydraulic Steering System for Pure Electric Wheel Loader’, Applied Sciences-Basel, vol. 12, no. 11, pp. 5740, 2022.
[10] L. Huang, T. Yu, Z. Jiao, et al, ‘Research on Power Matching and Energy Optimal Control of Active Load-Sensitive Electro-Hydrostatic Actuator’, IEEE Access, vol. 9, pp. 51121-51133, 2021.
[11] L. Huang, T. Yu, Z. Jiao, et al, ‘Active Load-Sensitive Electro-Hydrostatic Actuator for More Electric Aircraft’, Applied Sciences, vol. 10, no. 19, pp. 6978-21, 2020.
[12] T. Lin, Y. Lin, H. Ren, et al, ‘A double variable control load sensing system for electric hydraulic excavator’, Energy (Oxford), vol. 223, pp. 119999, 2021.
[13] J. Nurmi, J. Mattila, ‘Global energy-optimised redundancy resolution in hydraulic manipulators using dynamic programming’, Automation in Construction, vol. 73, pp. 120-134, 2017.
[14] O. Sundstrom, D. Ambiihl, L. Guzzella, ‘On implementation of dynamic programming for optimal control problems with final state constraints’, Oil & Gas Science and Technology, vol. 65, no. 1, pp. 91-102, 2010.
[15] H. Zhang, F. Wang, B. Xu, et al, ‘Extending battery lifetime for electric wheel loaders with electric-hydraulic hybrid powertrain’, Energy, vol. 261, pp. 125190, 2022.
[16] F. Wang, Z. Wu, B. Xu, et al, ‘A Mode-driven Control Strategy to Reduce Electric Drive Peak Power of Hybrid Wheel Loader Propulsion System’, IEEE Transactions on Vehicular Technology, vol. 72, no. 5, pp. 5948-5961, 2023.
[17] S. H. Kim, ‘Electric Motor Control: DC, AC, and BLDC Motors’, 2017.

Biographies



Jiaming Wu received B.S. degree in mechanical engineering from Zhejiang University, Hangzhou, China, in 2019. He is currently pursuing his Ph.D. degree in mechanical engineering at Zhejiang University, Hangzhou, China. His main research focuses on modeling and control of electro-hydrostatic actuator system for off-road vehicles.



Yihan Qiao received B.S. degree in mechanical engineering from Zhejiang University, Hangzhou, China, in 2022. He is currently pursuing his Ph.D. degree in mechanical engineering at Zhejiang University, Hangzhou, China. His main research focuses on electro-hydraulic hybrid variable speed drive system for wheel loader.



Feng Wang received B.S., M.S., and Ph.D. degrees in mechanical engineering from Zhejiang University, Hangzhou, China, in 2003, 2005, and 2009.

In 2009 he became a Postdoctoral Associate in the Department of Mechanical Engineering at University of Minnesota, Minneapolis, MN, USA, where he joined the NSF Engineering Research Center for Compact and Efficient Fluid Power (CCEFP). In 2016 he became an Assistant Professor in the School of Mechanical Engineering at Zhejiang University, Hangzhou, China, where he joined the State Key Laboratory of Fluid Power and Mechatronics Systems. His research interests include modeling and control of intelligent hydraulic powertrains for heavy-duty vehicles and large utility wind turbines.



Bing Xu received Ph.D. degree in mechanical engineering from Zhejiang University, Hangzhou, China, in 2001. He is currently a Professor in the school of mechanical engineering and also the Director of the State Key Laboratory of Fluid Power Transmission and Control at Zhejiang University.

He has authored or co-authored more than 200 journal and conference papers, and authorized 49 patents. His research interests include fluid power components and systems, mechatronic systems design and motion control for mobile machinery. Prof. Xu is a Chair Professor of the Yangtze River Scholars Program and a science and technology innovation leader of the Ten Thousand Talent Program.

A PID Controller Gain Adjustment Mechanism to Balance EHA Tracking and Health

Simo Käki, Mikko Huova, Tatiana Minav

Innovative Hydraulics and Automation, Tampere University, Tampere, Finland, firstname.lastname@tuni.fi

Abstract.

Electro-hydraulic systems power much of today's working machinery, which often operate in remote or harsh locations where maintenance is difficult. Given the cost and safety concerns of unplanned outages, strategies that prevent breakdowns between service intervals are crucial. One prominent approach is to adjust the parameters of existing controllers in a health-aware manner to potentially extend the remaining useful life (RUL) of the components.

Consequently, this paper presents a health-aware adjustment method for a PID-controlled electro-hydrostatic actuator that tunes the controller gain online to balance tracking and health. The method estimates a hidden degradation state using an extended Kalman filter and models the short-horizon degradation as a Wiener process. It derives an inverse-Gaussian RUL distribution from the degradation estimates and fits a ridge model that links control effort to degradation rate. Moreover, a supervisory cost function weighs an integral-squared error against a soft health constraint and selects gain updates accordingly.

In accelerated simulation studies, an update of the controller gain slowed degradation and thus extended the RUL. The system met a 95 percent survival target while keeping the control error modest. This approach adds improved life-extension potential with minimal architectural change, and it can run as a cloud-based supervisory layer that updates parameters using recorded data.

Keywords. Electro-hydrostatic actuator, external gear pump, health-aware control, PID-control, remaining useful life

1. INTRODUCTION

Maintaining working machinery is crucial, as the cost of unplanned outages can be significant. However, this is often challenging in practice, since machines frequently operate in remote and harsh environments. Accurate remaining useful life (RUL) prediction can provide key information for maintenance scheduling. In recent years, estimating the RUL of various components in electro-hydraulic systems has seen considerable research interest. Bejaoui et al. [1] and Dahal et al. [2] investigated the RUL of rotating machines using data-

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

driven methods. On the hydraulic side, the pump has been the focus of multiple papers utilizing physics-based approaches [3, 4], data-based approaches [5–7], or a combination of these, i.e., hybrid approaches [8, 9], to estimate the RUL.

Previous research has tended to focus on RUL prediction and fault-tolerant control rather than on failure prevention [10–13]. However, there has been growing interest in life-extending control or health-aware control (HAC) for systems that are difficult to maintain [14–16]. A recent survey by Zagorowska et al. [17] highlights the potential of HAC and underlines the need for more research that integrates degradation-dependent models in simulations when validating optimization and scheduling layers.

A regulatory control system requires accurate knowledge of the system behavior over short timescales. In contrast, degradation is often a very slow process and depends on the way in which the system is operated as well as on external factors. Thus, integrating both the performance and degradation mitigation in the same control loop leads inevitably to computational problems [18]. This problem naturally leads to a multi-layer approach where a long-run RUL problem is solved separately from the lower-level regulatory problem.

The aim of this paper is to develop an adjustment method for a proportional-integral-derivative (PID) controller in an electro-hydrostatic actuator (EHA) system. The EHA is selected, as it combines tight servo performance with compact, self-contained hydraulics. Moreover, the adoption of EHAs is anticipated to increase substantially with the growing introduction of electrification of working machinery and aerospace applications [19].To enable RUL-extending capabilities, measurements of oil flow, pressures, and pump rotational speed are required. This approach provides an alternative to complex HAC systems by adjusting the existing controller rather than designing degradation-aware control from scratch. This has the added benefit that the adjustment mechanism can be executed as a cloud-based service. Additionally, minimal assumptions are made about the hidden degradation model of the gear pump in the degradation estimation step.

The remainder of this paper is structured into four sections. Section 2 outlines the methodology applied throughout the study. Section 3 introduces the simulation model and explains the techniques. Section 4 discusses the main results in detail, and Section 5 concludes with a summary of the key findings and their relevance to fluid power applications.

2. METHODOLOGY

2.1. Degradation Model

In this paper, the RUL optimization is called the adjustment mechanism, which updates the parameters of the regulatory controller. To simulate the behavior of an adjustment mechanism, accurate knowledge of the degradation is needed. A single-input single-output (SISO) system with degrading actuator is considered. Here, the time-varying and hidden degradation variable is defined $\phi(t)$. To model the evolution of the degradation, a nonlinear Wiener process model is employed [19–21] with the following equation:

$$\phi(t) = \phi_0 + \int_0^t \varphi(\tau, \boldsymbol{\theta}, \dots) d\tau + \sigma B(t), \tag{1}$$

where ϕ_0 is the initial value, $\varphi(\tau, \boldsymbol{\theta}, \dots)$ is the deterministic nonlinear function of τ with the parameter vector $\boldsymbol{\theta}$ and possible other inputs that describe the degradation rate, σ is the diffusion parameter which reflects the degradation variation, and $B(t)$ is the standard Brownian motion.

The deterministic part of Equation (1) is often coupled with the state, inputs, and/or outputs of the actuator, making this a factor-based model. This coupling is heavily content dependent, and a considerable amount of research has been conducted to investigate suitable formulations [17].

The degradation variable represents the strength of an item. The item fails when the process reaches an adverse threshold state for the first time. In this case, the lifetime in component-level prognostics is usually defined as the first hitting time (FHT) of the degradation process reaching a failure threshold ϕ_f [23], which is denoted as

$$t_f = \inf\{t \geq 0 : \ \phi(t) \geq \phi_f \mid \phi(0) < \phi_f\}, \tag{2}$$

where t_f is the failure time. Here, it is worth noting that the above definition is only applicable to a case with an increasing degradation process. Additionally, the choice of failure threshold is not obvious. Thus, careful engineering knowledge of the system requirements should be used to determine a sufficient value.

Differentiating Equation (1) yields

$$\frac{d}{dt}\phi(t) = \varphi(t, \boldsymbol{\theta}, \dots) + \sigma w(t), \tag{3}$$

where $w(t) \sim \mathcal{N}(0,1)$ is white noise. The degradation rate $\varphi(\dots)$ is often quite slow, so it is safe to assume that an average rate gives a reasonable estimate of the immediate future degradation behavior. This assumption holds when the variation in the degradation rate over time is negligible, making the average rate particularly suitable for short-term prediction. Otherwise, a time-dependent rate should be considered for improved accuracy. Following this assumption, the nonlinear degradation model (1) can be estimated with the linear Wiener process model

$$\phi(t) \approx \phi_0 + \mu t + \sigma B(t), \tag{4}$$

where μ is the (constant) degradation rate. This simplification can be improved with historical knowledge of degradation progression.

2.2. Extended Kalman Filter

An extended Kalman filter (EKF) is a nonlinear state estimation technique that extends the classical Kalman filter to systems with nonlinear dynamics and/or measurement models [24]. By linearizing the system around the current estimate with first-order Taylor expansion, the EKF enables recursive estimation of the system state in real time. This makes the EKF an effective choice for estimating the current degradation rate in addition to the current degradation. To achieve this, Equation (4) is written as

$$\begin{bmatrix} \dot{\mu} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \phi \end{bmatrix} + w, \tag{5}$$

where $w \sim \mathcal{N}(0, \Sigma_w)$ is the process noise with a zero-mean Gaussian distribution and $\boldsymbol{x}_{kf} = [\mu \quad \phi]^{\mathrm{T}}$ is the state vector. Equation (5) should be supplemented with (possibly) the nonlinear measurement function

$$\boldsymbol{y}_{kf} = h(\boldsymbol{x}_{kf}, \boldsymbol{u}_{kf}) + v, \tag{6}$$

where $v \sim \mathcal{N}(0, \Sigma_v)$ is the measurement noise with a zero-mean Gaussian distribution that relates the degradation $\phi(t)$ to the measured values $\boldsymbol{y}_{kf}$ and inputs $\boldsymbol{u}_{kf}$.

The EKF algorithm consists of a prediction step and an update step. In the prediction step, the state and covariance are propagated forward in time using the system model. This *a priori* information is denoted by a superscript minus sign ($^-$). In the update step *a posteriori* information is denoted by a superscript plus sign ($^+$).

The EKF algorithm requires the Jacobians of the system dynamics and measurement functions. For the case of Equation (5), the discrete time Jacobians are

$$F = \begin{bmatrix} 1 & 0 \\ t_{s,ekf} & 1 \end{bmatrix}, \tag{7}$$

$$H = \frac{\partial}{\partial \boldsymbol{x}_{kf}} h(\boldsymbol{x}_{kf}, \boldsymbol{u}_{kf}), \tag{8}$$

where $t_{s,ekf}$ is the sampling time of the filter. These are then utilized in the prediction step to update the state estimate and covariance with equations

$$\hat{\boldsymbol{x}}_{kf}^{-}[k] = F\hat{\boldsymbol{x}}_{kf}^{+}[k-1], \tag{9}$$

$$\Sigma_{\boldsymbol{x}}^{-}[k] = F\Sigma_{\boldsymbol{x}}^{+}[k-1]F^{\mathrm{T}} + \Sigma_w, \tag{10}$$

where k is the current sampling timestep.

The update step starts with calculating the *a priori* measurements

$$\hat{\boldsymbol{y}}_{kf}^{-}[k] = h(\hat{\boldsymbol{x}}_{kf}^{-}[k], \boldsymbol{u}_{kf}[k]), \tag{11}$$

and the Kalman gain

$$K = \Sigma_{\boldsymbol{x}}^{-}[k]H^{\mathrm{T}}(H\Sigma_{\boldsymbol{x}}^{-}[k]H^{\mathrm{T}} + \Sigma_v)^{-1}. \tag{12}$$

The *a priori* measurements and the Kalman gain are then used to update the state estimate

$$\hat{\boldsymbol{x}}_{kf}^{+}[k] = \hat{\boldsymbol{x}}_{kf}^{-}[k] + K(\boldsymbol{y}_{kf}[k] - \hat{\boldsymbol{y}}_{kf}^{-}[k]). \tag{13}$$

Finally, the covariance matrix is updated with the Joseph form

$$\Sigma_{\boldsymbol{x}}^{+}[k] = (I - KH)\Sigma_{\boldsymbol{x}}^{-}[k](I - KH)^{\mathrm{T}} + K\Sigma_v K^{\mathrm{T}}, \tag{14}$$

where I is the identity matrix of a necessary size. The use of the Joseph form prevents the loss of positive definiteness in the covariance matrix due to numerical errors. This formulation is crucial when dealing with small numerical updates in long-term estimation tasks. Moreover, this formulation of the EKF enables an estimate of the degradation and the drift of a wide range of components. However, there are limits, which are outside the scope of this paper, as to the extent to which EKF can be utilized in this manner.

2.3. RUL Estimation

To estimate the RUL of the component, the remaining degradation margin is calculated with

$$L = \phi_f - \hat{\phi}[k], \tag{15}$$

where $\hat{\phi}[k]$ is the current degradation estimate from the EKF. For a meaningful prognostic horizon, $L > 0$ is required. Utilizing the slow degradation assumption in Equation (4), the RUL for this process becomes a classical FHT problem. For Brownian motion with a positive drift, the FHT to a fixed level has an inverse Gaussian (IG) distribution [23]. This link between drifted Brownian motion and the IG law is a standard result in stochastic process theory and is widely used in degradation-based reliability modeling.

In particular, if L is the distance to the failure threshold, then

$$\hat{t}_{RUL} \sim \mathrm{IG}\left(\frac{L}{\mu}, \frac{L^2}{\sigma^2}\right). \tag{16}$$

The corresponding cumulative distribution function (CDF) can be written using the standard normal distribution $\Phi(\cdot)$

$$F_{IG}(t) = \Phi\left(\frac{\hat{\mu}t - L}{\hat{\sigma}\sqrt{t}}\right) + \exp\left(\frac{2\hat{\mu}L}{\hat{\sigma}^2}\right)\Phi\left(-\frac{\hat{\mu}t + L}{\hat{\sigma}\sqrt{t}}\right), \tag{17}$$

where $\hat{\mu}$ is the current drift estimate and $\hat{\sigma}$ is the current diffusion estimate.

For a stable estimate of RUL, an evaluation time with a moving average can be utilized. As the EKF estimates the degradation with a low sampling time, a variety of unwanted dynamics and noise can negatively impact the calculations. At every evaluation time t_{eval}, the average from a recent data window for $\hat{\mu}$ can be calculated. Accordingly, the current diffusion estimate measures the random variability of degradation that is not explained by the drift estimate. The maximum likelihood estimate for the diffusion can be calculated with

$$\hat{\sigma}^2 = \frac{1}{\sum t_s}\sum_k \frac{(\boldsymbol{\phi}_W[k] - \boldsymbol{\phi}_W[k-1] - \hat{\mu}t_s)^2}{t_s}, \tag{18}$$

where $\boldsymbol{\phi}_W$ is the vector of degradation estimates from the EKF with the length of the moving window. For numerical stability, it is recommended to normalize $\hat{\phi}$ and $\hat{\mu}$ by dividing them with the failure threshold.

Finally, the mean RUL can be calculated with

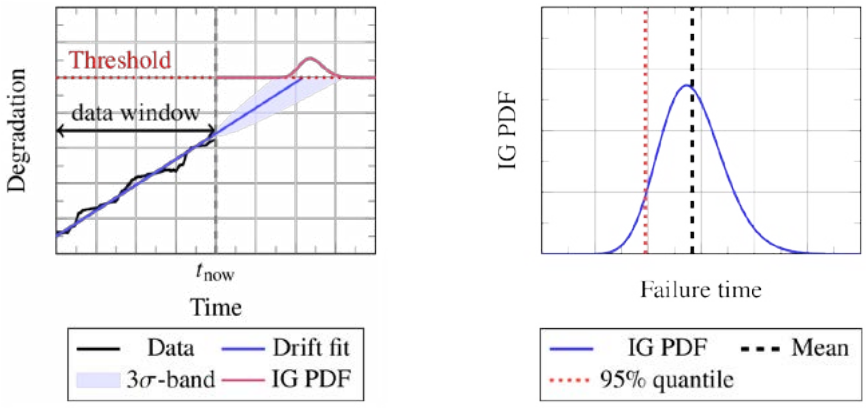
$$\hat{t}_{RUL} = \frac{L}{\hat{\mu}}. \tag{19}$$

For a more conservative estimate, a survival percentage can be chosen (i.e., 95%). Then, the RUL must be solved from

$$F_{IG}(\hat{t}_{RUL,p}) = 1 - p, \tag{20}$$

where p is the survival probability.

The RUL estimate is visualized in Figure 2.1. The Figure 2.1 (a) shows the recent degradation estimate date from the EKF and the failure threshold. Then, the model from Equation (4) is fitted to this data and plotted forward in time with a 3σ-band to show the uncertainty. The IG probability density function (PDF) from Figure 2.1 (b) is also plotted in Figure 2.1 (a). In Figure 2.1 (b), the mean RUL and the more conservative 95% survival RUL are plotted with vertical lines.



(a) **(b)**

Figure 2.1: Visualization of the RUL estimation from buffered data. (a) Drift fit to recent degradation data. (b) IG PDF with $\hat{t}_{RUL}$ and $\hat{t}_{RUL,95\%}$.

2.4. Adjustment Mechanism

The adjustment mechanism requires information on the effect of the control signal $u(t)$ on the degradation speed. A simple linear model

$$\hat{\mu} \approx X\boldsymbol{\beta}, \tag{21}$$

where X is the feature matrix and $\boldsymbol{\beta}$ is the fitted coefficients of the model, can be fitted to buffered degradation rate estimates $\hat{\mu}$ from the EKF. The fitting can be achieved using ridge regression [25]. In this study, the features are chosen as constant, rotational speed, and rotational acceleration. The corresponding drift term becomes $\hat{\mu}(t) \approx \beta_0 + \beta_1 u(t) + \beta_2 \dot{u}(t)$. These features were selected to be consistent with the deterministic degradation contributions in later Equations (34) and (35), and because $u(t)$ is the only variable the controller can directly influence, making it the most relevant regressor for analyzing how control actions correlate with degradation speed in this framework.

More generally, degradation can also depend on exogenous and operating-condition variables such as pressure, temperature, and contamination level. Such variables can be appended as additional columns in X when reliable measurements are available. However, they are not included here because the present study focuses on a simplified simulation setting and isolates the influence of the control action.

It is assumed that the controlled system can be presented as a SISO discrete time linear state-space model

$$\boldsymbol{x}_p[k+1] = A_d\boldsymbol{x}_p[k] + B_d u[k], \tag{22}$$

$$y[k] = C_d\boldsymbol{x}_p[k] + D_d u[k], \tag{23}$$

where $\boldsymbol{x}_p$ is the state vector, A_d is the dynamics matrix, B_d is the input matrix, C_d is the output matrix, and D_d is the feedthrough matrix. The system output y is controlled by discrete 2DOF-PID controller of the form [26]

$$x_{f,2}[k] = q_1 x_{f,2}[k-1] + q_2\Big((cr[k] - y[k]) - x_{f,1}[k-1]\Big), \tag{24}$$

$$x_{f,1}[k] = x_{f,1}[k-1] + t_s x_{f,2}[k], \tag{25}$$

$$u[k] = k_p(br[k] - y[k]) + u_i[k] - k_p t_d x_{f,2}[k], \tag{26}$$

$$u_i[k+1] = u_i[k] + \frac{k_p t_s}{t_i}(r[k] - y[k]), \tag{27}$$

where r is the target value for y, $x_{f,1}$ and $x_{f,2}$ are the states of a lowpass filter, k_p is the controller gain, t_i is the integral time, t_d is the derivative time, t_s is the sampling time, b is the proportional reference weight, and c is the derivative reference weight. The internal gains of the lowpass filter are calculated with

$$q_1 = \frac{\tau_f^2}{\tau_f^2 + 2t_s\tau_f + 2t_s^2}, \tag{28}$$

$$q_2 = \frac{2t_s}{\tau_f^2 + 2t_s\tau_f + 2t_s^2}, \tag{29}$$

where τ_f is the filter time constant.

Finally, the adjustment mechanism can be formulated as

$$\min_{k_p} J(k_p) \text{ such that } k_{p,min} < k_p < k_{p,max}, \tag{30}$$

where $J(k_p)$ is the cost function that relates the performance and RUL of the system to the controller gain k_p. To ensure dimensional consistency and allow tuning of priorities, the cost function is formulated as

$$J(k_p) = w_1 \cdot ISE(k_p) + w_2 \cdot HEALTH(k_p), \tag{31}$$

where w_1 and w_2 are weighting factors that balance tracking and health preservation, $ISE(k_p)$ is the integral squared error that can be calculated from the past reference values and simulated with Equations (22)–(27) and $HEALTH(k_p)$ is the health constraint. The health constraint is formulated as a soft constraint

$$HEALTH(k_p) = \begin{cases} 0 & \text{if } \hat{t}_{RUL,95\%}(\hat{\mu}(k_p)) \geq t_{RUL,req} \\ 10^6\left(t_{RUL,req} - \hat{t}_{RUL,95\%}(\hat{\mu}(k_p))\right)^2 & \text{otherwise} \end{cases}, \tag{32}$$

where $t_{RUL,req}$ is the remaining mission time, $\hat{t}_{RUL,95\%}(\hat{\mu})$ is the 95% quantile RUL from Equation (20), and $\hat{\mu}(k_p)$ is the degradation rate from Equation (21), where the feature matrix can be constructed from the simulation.

The complete framework for the adjustment mechanism is presented in Figure 2.2. The extended Kalman filter is utilized to estimate the degradation and degradation rate from the rotation speed, generated flow, and port pressures of the pump. From these, a RUL estimate can be made. The degradation rate is used in the feature extraction block to estimate how the rotational speed affects the degradation speed. Finally, the optimizer employs this information to find the best value for the controller gain.

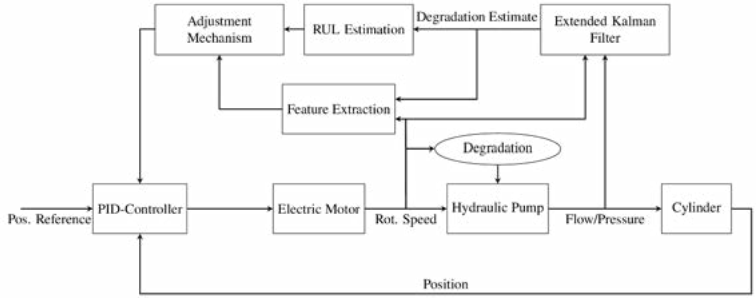


Figure 2.2: PID-controlled EHA with prognostics and health monitoring system connected to the adjustment mechanism to schedule the PID parameters.

3. SIMULATION MODEL

A Simulink model of an EHA system is used to evaluate the effectiveness of the adjustment mechanism. The whole simulation model is presented in more detail in a previous publication [27]. This section focuses on the main differences in the simulation model to the previous usage. The hydraulic diagram of an EHA is shown in Figure 3.1.

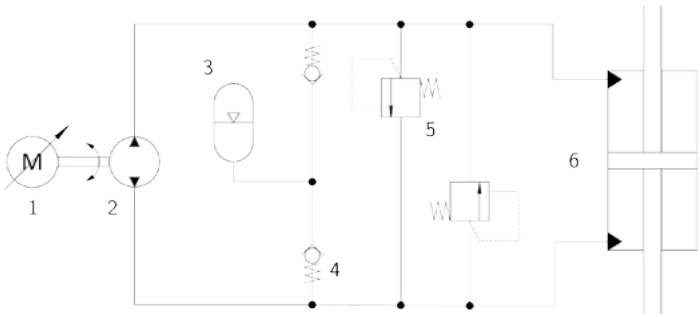


Figure 3.1: EHA hydraulic diagram. 1. Electric motor. 2. Hydraulic pump. 3. Accumulator. 4. Check-valves. 5. Pressure-relief valves. 6. Double-acting cylinder.

The hydraulic pump is modeled as an external gear pump. The outflow of the pump is modeled with Schlösser's loss coefficient model [28]

$$Q_A = V_p n - C_{sl}\frac{V_p(p_{h,A} - p_{h,B})}{2\pi\nu} - C_{st}V_p^{\frac{2}{3}}\sqrt{\frac{2(p_{h,A} - p_{h,B})}{\rho}} - C_{se}p_{h,A}, \quad (33)$$

where V_p is the volumetric displacement, n is the rotational speed, C_{sl} is the laminar slip coefficient, $p_{h,A}$ and $p_{h,B}$ are the hose pressures, ν is the dynamic viscosity, C_{st} is the turbulent slip coefficient, ρ is the oil density, and C_{se} is the external leakage coefficient.

During the system operation, the performance of the actuator gradually deteriorates. This degradation can be represented by certain actuator parameters. In this study, the slip coefficients C_{sl} and C_{st} are utilized as hidden variables to characterize the degradation of the pump. This is motivated by the fact that internal leakage increases over time [29]. This choice is illustrative as other parameters could also be used. Based on Equation (1), the hidden degradation process is modelled as

$$C_{sl}(t) = C_{sl}(0) + \int_0^t \lambda_1|n(\tau)| + \lambda_2\left|\frac{d}{d\tau}n(\tau)\right| d\tau + \sigma B_1(t), \quad (34)$$

$$C_{st}(t) = C_{st}(0) + \int_0^t \lambda_1|n(\tau)| + \lambda_2\left|\frac{d}{d\tau}n(\tau)\right| d\tau + \sigma B_2(t), \quad (35)$$

where λ_i is the degradation rate coefficient. Here, it is assumed that both slip coefficients increase by the same proportion, but the Brownian motion is different. It is also assumed that the amount of degradation per unit of time is proportional to the amplitude of the rotational speed and the rotational acceleration. This assumption reflects the fact that stress is usage-dependent and increases with higher rotational speed and fast changes between operation points. Therefore, increasing controller aggressiveness can increase peak acceleration, but may still reduce cumulative degradation if it shortens transients and reduces the time spent at elevated rotational speed. The resulting effect is not necessarily monotonic in k_p and depends on the duty cycle and the coefficients of the speed- and acceleration-dependent terms.

In the EKF, a lumped parameter model for internal leakage

$$Q_A = V_p n - \hat{\phi}(p_{h,A} - p_{h,B}) - C_{se}p_{h,A}, \quad (36)$$

was utilized as the measurement function in Equation (6). This formulation allows for a straightforward definition of the failure threshold based on the operation requirements of the cylinder. For the simulations, the failure threshold was set to $\phi_f = 4 \cdot 10^{-12}\ (m^3/s)/Pa$. It is assumed that the volumetric displacement and the external leakage coefficient are known, while flow, rotational speed, and pressures are measured.

The simulation model parameters are summarized in

Table 1. The table presents only the parameters that differ from the previous publication [27]. The position reference and load forces are also identical to the previous publication. These were repeated until the degradation reached the specified failure threshold in the simulation.

Table 1: EHA simulation model parameters compared to previous publication [27]

Parameter	Symbol	Value
Pump displacement	V_p	$1.2732 \cdot 10^{-6}\ m^3/rad$
Initial laminar slip coeff.	$C_{sl}(0)$	$5 \cdot 10^{-9}$
Initial turbulent slip coeff.	$C_{st}(0)$	10^{-11}
External leakage coeff.	C_{se}	10^{-13}
Rot. speed degradation rate coeff.	λ_1	$2 \cdot 10^{-11}$
Rot. accel. degradation rate coeff.	λ_2	$8 \cdot 10^{-12}$
Diffusion coefficient	σ	10^{-9}
Dynamic viscosity	ν	$0.03\ Pa \cdot s$
Bulk modulus (cylinder)	β_{cyl}	$1200 \cdot 10^6\ Pa$
Bulk modulus (hose)	β_{hose}	$400 \cdot 10^6\ Pa$

The EHA position is controlled by a 2DOF-PID controller. The control signal is calculated according to Equations (24)–(27). The parameters of the controller are presented in Table 2. The controller gain is adjusted with the previously described adjustment mechanism. The parameters of the prognostics and health monitoring system are presented in Table 3.

Table 2: PID controller parameters

Parameter	Symbol	Value
Nominal gain	k_p	$4000\ (rad/s)/m$
Integral time	t_i	$1.6000\ s$
Derivative time	t_d	$0.3750\ s$
Filter time constant	τ_f	$0.0250\ s$
Proportional reference weight	b	0.5
Derivative reference weight	c	0
Controller sampling time	t_s	$0.01\ s$

Table 3: Prognostics and health monitoring system parameters

Parameter	Symbol	Value
EKF sampling time	$t_{s,ekf}$	$0.05\ s$
EKF state covariance	Σ_w	$\begin{bmatrix} 0.0006 & 0.0179 \\ 0.0179 & 0.7170 \end{bmatrix} \cdot 10^{-16}$

EKF measurement covariance	Σ_v	$2 \cdot 10^{-5}$
Buffer time	t_{buf}	$200\ s$
Adj. mech. evaluation step	t_{eval}	$400\ s$
Controller gain, lower bound	$k_{p,min}$	$1000\ (rad/s)/m$
Controller gain, upper bound	$k_{p,max}$	$10000\ (rad/s)/m$
Cost function tracking weight	w_1	$1/m^2 s$
Cost function health weight	w_2	$1/s^2$

4. RESULTS

The deterioration of the gear pump in an EHA is inevitable but can be delayed by adjusting the controller parameters. To illustrate this application of the predicted RUL and adjustment mechanism, the proposed method is applied to the EHA position control system. The simulations were conducted with MATLAB Simulink in accelerated conditions, i.e., the coefficients of Equations (34) and (35) are greatly exaggerated. This leads to the pump failing in about 30 minutes rather than thousands of hours. In the simulations the EHA is supposed to operate for longer than $t_{RUL,req} = 1850\ s$ with the probability $p = 95\%$. The adjustment mechanism evaluation times were set $400\ s$ apart.

Figure 4.1 shows the degradation evolution for the baseline system and the system with the adjustment mechanism enabled. The points where the controller gain is adjusted are highlighted with dashed grey lines. After the first adjustment at $t = 400\ s$, it is evident that the degradation rate slows down compared to the baseline. Although the gain increase can raise peak acceleration during reference transitions, the net degradation slows because Equations (34) and (35) depend on the integrated rotational speed/acceleration. Under this duty cycle the updated gain reduces the duration and average level of the stress terms. In the baseline simulation, $t_{RUL,req}$ was broken with 25 seconds. The adjustment mechanism successfully extended the RUL of the system by 46 seconds, which fulfilled the operation time requirement.

As an additional reference, a fixed-gain simulation was conducted by setting the proportional gain k_p to $8000\ (rad/s)/m$. This was done to maximize useful life for this specific duty cycle and model parameterization. In this case, the achieved useful life increased to approximately $2060\ s$, compared with $1825\ s$ for the nominal fixed-gain baseline and $1871\ s$ for the gain-updated case. This result should be interpreted as scenario specific as the optimal fixed gain depends on the duty cycle, load, and modeling uncertainty, and therefore serves as an upper bound rather than a generally applicable tuning.

In Figure 4.2, the adjusted controller gain of the PID controller is shown with the constant gain of the baseline simulation. The effect of changing the controller gain to the squared control error is illustrated in Figure 4.3. At the first adjustment time, there is a small increase in the control error caused by the sudden change of controller gain. Figure 4.3 also shows that the peaks of the control error increase over time. This is caused by the deterioration of the pump, which causes it to produce less flow and in turn makes it more difficult to follow

the changing reference points. The error peaks of the adjusted controller are slightly smaller, caused by the cost function (31) including the ISE in the optimization.

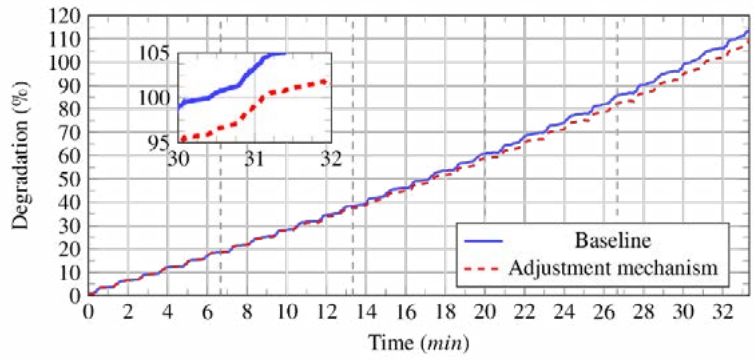


Figure 4.1: Evolution of estimated degradation in accelerated simulation conditions.

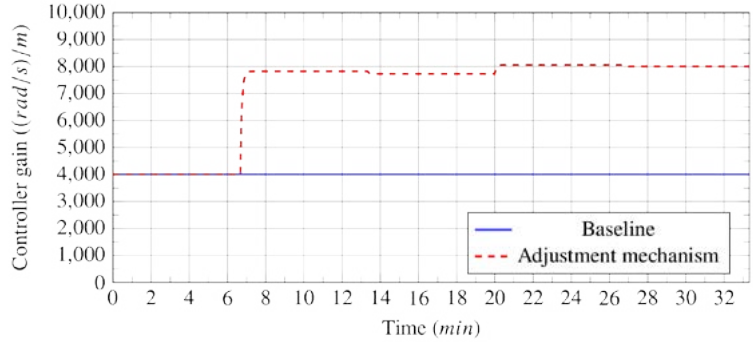


Figure 4.2: Proportional gain in the baseline test and under the adjustment mechanism.

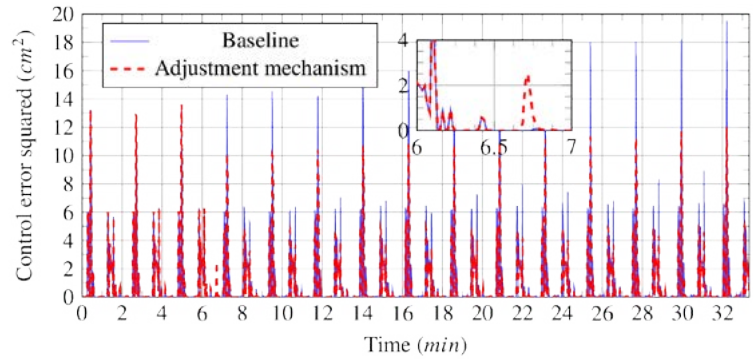


Figure 4.3: Squared control error for the two scenarios with zoomed in section of the first adjustment time.

From Figure 4.2, it is evident that after the first adjustment, there is no need for the subsequent adjustments. This is mostly due to the assumption of linear degradation evolution holding very well for the whole simulation, which is caused by two main factors. Firstly, the cyclic nature of the position reference and load signals utilized in the simulation cause the degradation to increase almost the same amount per cycle. This is somewhat offset by the increased leakage flow for which the controller must compensate by increasing the rotational speed, which in turn accelerates the degradation towards the end of the simulation. Secondly, the chosen degradation models, i.e., Equations (34) and (35), increase the degradation speed independently from the age/degradation of the component. In real systems this might not be true, as the degradation speed of the component can be considerably faster near the end of its life.

5. CONCLUSION

This paper introduced a degradation-aware adjustment method for a PID-controller in an EHA system. An extended Kalman filter-based degradation estimator was combined with an inverse-Gaussian RUL formulation and a simple ridge model linking the control effort to the degradation rate. These were used in a cost function that trades tracking performance against a soft health constraint. The approach targets systems where maintenance is challenging and controller retrofits are preferred over a full health-aware control redesign.

Simulations with an EHA model under accelerated degradation showed that the proposed mechanism met the mission-time requirement at the specified survival-probability level. The nominal fixed-gain baseline achieved a useful life of $1825\ s$, whereas the system with the proposed adjustment mechanism increased the useful life to $1871\ s$ with only a modest transient increase in control error. As an additional benchmark, selecting the near scenario optimal fixed gain increased the useful life to approximately $2060\ s$. This gap indicates that, for a static and known duty cycle, offline tuning can outperform delayed online adjustment. Accordingly, the proposed mechanism is primarily aimed at settings where the best tuning is not known beforehand and/or operating conditions evolve over time.

The practical implication is that existing PID-controlled electro-hydraulic installations can gain significant life-extension potential with minimal architectural change. The method can be implemented as a supervisory layer (potentially cloud-based) that periodically re-tunes a small set of controller parameters using rotational speed, flow, and pressure data. While speed feedback and pressure are commonly available, direct flow sensing typically is not and may require added instrumentation or a validated virtual-sensing alternative. For practitioners, this provides a low-intrusion pathway to improve availability and maintenance planning by aligning controller aggressiveness with the remaining mission time, provided the required measurements are available.

This study has limitations worth noting. The results are based on simulations with accelerated degradation and simplified Wiener-type models whose parameters and threshold selection impact RUL accuracy. The controller adjustment tuned only the controller gain, and other PID parameters were kept fixed. The simulated duty cycle was cyclic, which favored near-linear degradation over the horizon, and may not generalize to all duty cycles or late-life behaviors where degradation can accelerate.

Future work should include experimental validation on hardware and the identification of more detailed degradation mechanisms and thresholds to improve model fidelity. In addition, the supervisory strategy should be evaluated under dynamic duty cycles, including varying loads and reference trajectories, to quantify robustness and to determine when online supervision provides a clear advantage over a priori fixed-gain tuning. Robust estimation under changing operating conditions, adaptive evaluation schedules, and integration with maintenance planning tools are also recommended to broaden applicability and reliability.

ACKNOWLEDGMENT

We acknowledge the financial support of the Finnish Ministry of Education and Culture through the Intelligent Work Machines Doctoral Education Pilot Program (IWM VN/3137/2024-OKM-4).

6. REFERENCES

[1] I. Bejaoui, D. Bruneo, and M. G. Xibilia, "A Data-Driven Prognostics Technique and RUL Prediction of Rotating Machines Using an Exponential Degradation Model," in *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, June 2020, pp. 703–708. doi: 10.1109/CoDIT49905.2020.9263930.

[2] A. Dahal, Y. Li, S. Khanal, S. G. Niazi, and H.-Z. Huang, "Remaining useful life prediction of induction motor based on hybrid temporal convolution neural network," in *14th International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE 2024)*, July 2024, pp. 1280–1285. doi: 10.1049/icp.2024.3616.

[3] Cun Shi, Shaoping Wang, Xingjian Wang, and Jun Wang, "Remaining useful life prediction based on particle filtering for high-speed pump in load sensing electro-hydrostatic actuator," in *CSAA/IET International Conference on Aircraft Utility Systems (AUS 2018)*, Guiyang, China: Institution of Engineering and Technology, 2018, p. 238 (6 pp.)-238 (6 pp.). doi: 10.1049/cp.2018.0312.

[4] Y. Zhang, S. Wang, C. Zhang, and R. Chen, "Physics-Based Machine Learning Hydraulic Piston Pump Performance Degradation and Prognostics," in *2023 7th International Conference on System Reliability and Safety (ICSRS)*, Bologna, Italy: IEEE, Nov. 2023, pp. 389–394. doi: 10.1109/ICSRS59833.2023.10381233.

[5] M.-S. Lee, T. A. Shifat, and J.-W. Hur, "Kalman Filter Assisted Deep Feature Learning for RUL Prediction of Hydraulic Gear Pump," *IEEE Sens. J.*, vol. 22, no. 11, pp. 11088–11097, June 2022, doi: 10.1109/JSEN.2022.3167926.

[6] R. Guo, Y. Li, L. Zhao, J. Zhao, and D. Gao, "Remaining Useful Life Prediction Based on the Bayesian Regularized Radial Basis Function Neural Network for an External Gear Pump," *IEEE Access*, vol. 8, pp. 107498–107509, 2020, doi: 10.1109/ACCESS.2020.3001013.

[7] P. Liu, S. He, K. Gao, G. Yang, J. Zou, and L. Ye, "Prediction of hydraulic pumps remaining useful life based on LSTM and Transform with dual self-attention," in *2023 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big*

Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Wuhan, China: IEEE, Dec. 2023, pp. 824–831. doi: 10.1109/ISPA-BDCloud-SocialCom-SustainCom59178.2023.00140.

[8] C. Li, J. Wang, X. Li, X. Li, J. Xu, and D. Mba, "PGMTKD: A Physics-Guided Multi-Teacher Knowledge Distillation network for external gear pump onboard fault diagnosis with missing modality," *Mech. Syst. Signal Process.*, vol. 239, p. 113296, Oct. 2025, doi: 10.1016/j.ymssp.2025.113296.

[9] W. Xu, Z. Wang, Z. Zhou, C. Sun, R. Yan, and X. Chen, "Wear state assessment of external gear pump based on system-level hybrid digital twin," *Mech. Syst. Signal Process.*, vol. 209, p. 111123, Mar. 2024, doi: 10.1016/j.ymssp.2024.111123.

[10] T. Escobet, V. Puig, and F. Nejjari, "Health Aware Control and model-based Prognosis," presented at the 2012 20th Mediterranean Conference on Control and Automation, MED 2012 - Conference Proceedings, 2012, pp. 691–696. doi: 10.1109/MED.2012.6265718.

[11] S. A. Nahian, D. Q. Truong, P. Chowdhury, D. Das, and K. K. Ahn, "Modeling and fault tolerant control of an electro-hydraulic actuator," *Int. J. Precis. Eng. Manuf.*, vol. 17, no. 10, pp. 1285–1297, Oct. 2016, doi: 10.1007/s12541-016-0153-2.

[12] L. Liao and F. Köttig, "Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction," *IEEE Trans. Reliab.*, vol. 63, no. 1, pp. 191–207, Mar. 2014, doi: 10.1109/TR.2014.2299152.

[13] Y. Hu, S. Liu, H. Lu, and H. Zhang, "Remaining Useful Life Model and Assessment of Mechanical Products: A Brief Review and a Note on the State Space Model Method," *Chin. J. Mech. Eng.*, vol. 32, no. 1, p. 15, Feb. 2019, doi: 10.1186/s10033-019-0317-y.

[14] J. Thuillier, M. S. Jha, M. Galeotta, and D. Theilliol, "Control reconfiguration strategies for Remaining Useful Life extension," *IFAC-Pap.*, vol. 55, no. 34, pp. 114–119, Jan. 2022, doi: 10.1016/j.ifacol.2022.11.317.

[15] J. Matias and J. Jäschke, "Balancing production maximization and equipment degradation: Experimental application of a health-aware controller," *Control Eng. Pract.*, vol. 156, p. 106193, Mar. 2025, doi: 10.1016/j.conengprac.2024.106193.

[16] E. Kipchirchir, M. H. Do, J. G. Njiri, and D. Söffker, "Prognostics-based adaptive control strategy for lifetime control of wind turbines," *Wind Energy Sci.*, vol. 8, no. 4, pp. 575–588, 2023, doi: 10.5194/wes-8-575-2023.

[17] M. Zagorowska, O. Wu, J. R. Ottewill, M. Reble, and N. F. Thornhill, "A survey of models of degradation for control applications," *Annu. Rev. Control*, vol. 50, pp. 150–173, Jan. 2020, doi: 10.1016/j.arcontrol.2020.08.002.

[18] R. D. de Oliveira and J. Jäschke, "Integration of time scales in health-aware control," presented at the IFAC-PapersOnLine, 2024, pp. 222–227. doi: 10.1016/j.ifacol.2024.08.340.

[19] M. Kumar, “A survey on electro hydrostatic actuator: Architecture and way ahead,” *Mater. Today Proc.*, vol. 45, pp. 6057–6063, Jan. 2021, doi: 10.1016/j.matpr.2020.10.049.

[20] Z. Zhang, X. Si, C. Hu, and Y. Lei, “Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods,” *Eur. J. Oper. Res.*, vol. 271, no. 3, pp. 775–796, Dec. 2018, doi: 10.1016/j.ejor.2018.02.033.

[21] H. Wang, X. Ma, and Y. Zhao, “An improved Wiener process model with adaptive drift and diffusion for online remaining useful life prediction,” *Mech. Syst. Signal Process.*, vol. 127, pp. 370–387, July 2019, doi: 10.1016/j.ymssp.2019.03.019.

[22] X. Si, Z. Ren, X. Hu, C. Hu, and Q. Shi, “A Novel Degradation Modeling and Prognostic Framework for Closed-Loop Systems With Degrading Actuator,” *IEEE Trans. Ind. Electron.*, vol. 67, no. 11, pp. 9635–9647, Nov. 2020, doi: 10.1109/TIE.2019.2952828.

[23] M.-L. T. Lee and G. A. Whitmore, “Threshold Regression for Survival Analysis: Modeling Event Times by a Stochastic Process Reaching a Boundary,” *Stat. Sci.*, vol. 21, no. 4, Nov. 2006, doi: 10.1214/088342306000000330.

[24] S. Särkkä, *Bayesian filtering and smoothing*. in Institute of Mathematical Statistics textbooks, no. 3. Cambridge, U.K. ; New York: Cambridge University Press, 2013.

[25] A. K. M. E. Saleh, M. Arashi, and B. M. G. Kibria, *Theory of ridge regression estimation with applications*. Hoboken, NJ: John Wiley & Sons, 2019.

[26] K. J. Åström and T. Hägglund, *Advanced PID control*. Research Triangle Park, NC: ISA - The Instrumentation, Systems, and Automation Soc, 2006.

[27] S. Käki, V. P. Singh, M. Huova, and T. Minav, “Health Aware Control Design for Extending Remaining Useful Life of Electro-Hydrostatic Actuator Based Steering,” in *The 19th Scandinavian International Conference on Fluid Power, SICFP’25*, Linköping University, Sweden: River Publishers, 2025, pp. 176–186. doi: 10.13052/rp-9788743808251A17.

[28] M. Torrent, P. J. Gamez-Montero, and E. Codina, “Parameterization, Modeling, and Validation in Real Conditions of an External Gear Pump,” *Sustainability*, vol. 13, no. 6, p. 3089, Mar. 2021, doi: 10.3390/su13063089.

[29] Y. Duensing, A. Merkel, and K. Schmitz, “Degradation Identification of an EHA Piston Pump by Analysis of Load-Holding States,” *Int. J. Fluid Power*, pp. 273–290, July 2024, doi: 10.13052/ijfp1439-9776.2528.

Biographies



Simo Käki graduated with the M.Sc. (Tech.) degree in automation technology from Tampere University, Finland, in 2024. Currently he is a Doctoral Researcher at Innovative Hydraulics and Automation (IHA), Tampere University, Finland. His research focuses on advanced control of electro-hydraulic systems and extension of remaining useful life of components with intelligent control.



Mikko Huova received the doctoral degree at Tampere University of Technology, Finland, in 2015. Currently he is a Senior Research Fellow at IHA lab of Tampere University, Finland. His research focuses on advanced control, multi-pressure systems, and energy efficient digital hydraulic systems of heavy-duty mobile machines.



Tatiana Minav has over 18 years of experience improving the efficiency and sustainability of off-road mobile working machines and innovatively combining signal processing, electrical and mechanical engineering. Since 2019 she has been an associate professor (tenured) in the largest Finnish fluid power laboratory at Tampere University, which specializes in off-road machinery, efficiency, powertrain, fluid power and automation-related topics. Professor Minav has authored ca 180 international scientific publications.

Machine Learning Based Controller for Energy-Optimized Operation of Hydraulic Loader Cranes

Amy Rankka [1,2], Alessandro Dell'Amico [1]

1: Division of Fluid and Mechatronic Systems, Linköping University, Linköping, Sweden, 2: Hiab AB amy.rankka@liu.se, alessandro.dellamico@liu.se

Abstract.

As energy efficiency becomes an increasingly important property of hydraulic working machines, new ways of improving the system need to be explored. This paper explores the potential of improving the energy efficiency of the hydraulic system by how it is operated. With a kinematically redundant system, such as the boom system of a loader crane, there is a possibility of reducing the energy consumption of a given task by optimizing how the different actuators are operated to form the crane tip motion. This work considers the semi-automatic function Crane Tip Control (CTC) where the operator controls the velocity of the crane.

A new, energy efficient, CTC controller with two parts is developed. The first part of the controller decides the strategy of the crane tip motion by setting the speed command for one of the boom actuators. The second part handles the reference velocity following by calculating the required speed commands for the other two boom actuators. The input to the controller is the reference velocity, both speed and direction, of the crane tip together with the state of the crane boom cylinders. In its current state, the controller is optimized for a single payload, but in future versions, the payload can be estimated or measured and set as input.

The strategy part of the controller is represented by a neural network, that is trained by a Reinforcement Learning approach with a hydraulic energy minimization objective. The training is performed on a simulation model of the crane, which includes geometrical relations for the crane booms and cylinders, and an energy consumption estimation.

The performance of the new CTC controller is benchmarked with regards to two other CTC controllers. The first one is a near-optimal offline controller implemented by a dynamic programming optimization. The second one is a a rule-based controller with a strategy that always holds one of the actuators still. An energy flow analysis has revealed great losses due to simultaneous operation of several actuators. Although the rule-based controller is not optimized with respect to the crane's operation, it is an important comparison since it provides a less development heavy solution.

Keywords. loader cranes, energy efficiency, semi-autonomous operation, neural network controller, redundancy resolution

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Although, improving the energy efficiency of hydraulic working machines has been widely investigated for a long time with many interesting concepts being presented, very few solutions has been implemented by the industry due to the associated cost and complexity with changing to a new architecture. This study explores a way of reducing the energy consumption by automating the movement of the machine, and thus, does not require any hardware changes. The work is applied to a loader crane, which is a kinematically redundant system, meaning it has an extra degree of freedom which can be used for optimizing the movements with regards to some objective. Loader cranes are not yet fully automated, but there is a commercially available semi-automatic feature by several manufacturers, see for example [12], [14] and [13], where the motion of the crane tip is controlled by the operator and the corresponding motions of the hydraulic cylinders are controlled by the control system. The feature can be denoted Crane Tip Control, CTC, and in this paper an approach to minimize the energy consumption during CTC operation is investigated.

1.1 Problem Formulation

The CTC feature which is used as reference here is provided by crane manufacturer Hiab [12]. It is optimized for lifting capacity, and in order to investigate how large impact an optimized automated feature can have on the energy consumption, the aim of this study is to develop a new CTC controller that is optimized for energy efficiency. A complete CTC controller should be able to take any valid operator command, which is given in the form of a speed command in the horizontal direction of the crane tip and a speed command in the vertical direction, and provide a set of cylinder speed commands that result in this motion. The CTC feature considered here works in a two-dimensional plan and controls the three boom cylinders of the crane, the slewing function is excluded. The scope of this study is limited to investigate movements in the horizontal direction with a fixed positive speed. CTC is a realtime feature which must be executed on the embedded control system of the crane. For this reason, traditional offline optimization methods are excluded. Instead, a machine learning approach, where part of the CTC controller is set up as a neural network, is explored. A simple rule-based controller is also developed to form a baseline. As a best case benchmark, near-optimal solutions to a limited number of test cases are calculated by an offline Dynamic Programming algorithm developed in previous work by the authors [8].

1.2 Delimitations

The scope of this paper is limited to investigating one type of neural network structure, feed forward neural network, as well as using one Reinforcement Learning approach, policy optimization, as training method. Further, the CTC controller is limited to working with a fixed speed and payload in the positive x-direction of the working area of the crane. The main focus will be on presenting results from one such controller, although results for training controllers with different speed and payload will be briefly presented to show the potential for extending the method in future work.

1.3 System Description

The system used as reference application in this work is a medium sized loader crane with three hydraulic cylinders used by the CTC feature,the first boom, "1b", cylinder, the second boom, "2b", cylinder and the extension, "ext", cylinder. The first and second boom cylinders control revolute joints with angles denoted α_{1b} and α_{2b} respectively. The extension cylinder works as a prismatic joint with a position denoted l_{ext}. The position of the three joints together with the dimensions of the crane defines the crane pose. The

kinematic redundancy gives that there can be an infinite amount of different crane poses for a single crane tip position. The crane tip position is given in Cartesian coordinates in the plane with axes x and y.

The cylinders of the crane are connected to a pressure compensated load sensing hydraulic system with a variable displacement pump. The pressure compensation gives that the speed commands to each cylinder, denoted u_{1b}, u_{2b} and u_{ext}, can be directly translated to spool position commands to the directional valve.

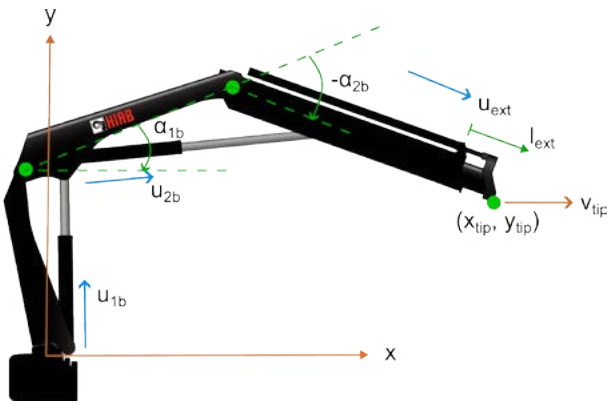


Figure 1: Crane with cylinders, joints and crane tip position.

1.4 Prior Art

The problem for a CTC controller or similar is to translate a desired motion in the task space (crane tip movement in the CTC case) to a motion in the joint space (cylinder velocities in the CTC case). To go from tip velocity to joint velocities the inverse of the so called jacobian matrix has to be somehow calculated. For a redundant system, the exact inverse of the jacobian does not exist. A common way to solve the problem is to compute the pseudo-inverse of the matrix, which gives the solution with the smallest joint velocity magnitude. This method is described for example in [4] and [3]. The drawback with this approach is that only locally optimal solutions can be found and that minimizing the joint velocity magnitude is the only objective. In [5] it is shown that the energy consumption performance of the pseudo-inverse method is significantly worse than for a dynamic programming algorithm.

In [11], the pseudo-inverse solution is used for precision tracking of a three degree of freedom robotic arm and combined with a minimization of the kinetic energy by a velocity vector pointing in the direction where the kinetic energy integral is minimized for a chosen trajectory. The optimal control is predicted for a window and continuously updated as the simulation is rolled out. The resulting controller is shown to perform pretty close to the global optimal solutions on trajectories with a length of 10 s. However, to only minimize the kinetic energy is too much a simplification for the crane with a load sensing hydraulic system, where kinetic energy cannot be directly translated to input energy to the system.

The possibility of using a neural network as CTC controller has been presented by [10]. In this work, the CTC controller is optimized with the objective of moving the crane tip as fast as possible. The same methodology has been applied to the crane used as application in this study with good results considering the speed and precision of the crane tip movement.The promising results gives an incentive for exploring a similar methodology and controller structure but with an energy minimization objective. This work was started in [8] where an energy efficient CTC controller was trained for a very limited part of the working area of the crane with good results.

A neural network controller has also been implemented for controlling the automatic log collecting of a forwarder in [1]. The network is trained with a reinforcement learning approach in a simulation environment and shows promising results with regards to energy efficiency. In another recent publication, [2], a controller for autonomous operation of a forwarder is also developed by reinforcement learning. The controller is trained in a simulation environment and later deployed on a real forwarder with good results.

2. Baseline and Benchmark

The energy consumption of moving the crane tip according to some specified trajectory with different strategies for moving the cylinders differ mainly due to three properties of the system:

Total oil need The total traveled distance of the cylinders can differ and affects the average flow demand of the pump.

Cylinder pressures Different cylinders require different pressure levels for lifting the payload in different crane poses. The cylinder pressures depend on the load force, friction and cylinder dimensions. The highest cylinder pressure (or more accurately the highest load pressure sensed by the directional valve) sets the pump pressure.

Simultaneous losses In a load sensing system with a single pump, the flow to all cylinders active simultaneously is supplied at the same pressure. If a cylinder requires a lower pressure, the excess pressure is wasted at the compensator valve.

For an optimized system, all these three properties and how they together form the flow and pressure demand of the pump at every instant must be taken into account, as well as the total energy of a movement of some length of time. A Dynamic Programming, DP, algorithm that can make this optimization for a discretized version of the system has been developed in previous work [8]. Results from this algorithm will serve as a benchmark for the new controller in Section 5.

Without any optimization, a simple rule-based controller that tries to minimize the third property, simultaneous losses, can be developed. The strategy of this controller is to always keep one cylinder still and to use the other two to make the necessary motion. A previous study, [7], of the reference loader crane shows that simultaneous losses account for about 30% of the losses in the hydraulic system. At the start of a motion, the controller tries to keep the 1st boom still, as this is often the cylinder that has to carry the most weight. If this leads to an illegal solution, due to violation of maximum cylinder positions or speeds, it tries to keep the 2nd boom cylinder or extension cylinder still. During the motion, it then keeps the strategy of the previous timestamp as long as possible, in order to achieve smooth motions. Results from this controller will serve as a baseline for the new controller in Section 5.

3. Controller Structure

The new CTC controller optimized for energy consumption consists of two parts, a strategy controller and a velocity controller, and its layout is presented in Figure 2. The inputs to the controller are the crane state, represented by the joint angles and extension position, and the reference tip velocity. The outputs are the speed commands to the three boom cylinders. Development of the controller was started in the work presented in [8], and the method has now been extended to training a controller that is able to control the crane tip motion in the whole working area of the crane.

The strategy controller is implemented as a feed forward neural network with three hidden layers with 32 nodes each. Adding more nodes than that does not improve the performance of the network.

LeakyRelu is used as activation function for all layers except the output layer, where instead HardTanh is used in order to get the output saturated to [-1,1]. -1 corresponds to maximum negative speed command, and 1 to maximum positive speed command. LeakyReLU stands for leaky rectified linear unit and introduces non-linearity to the network with reduced risk of de-activating neurons compared to simple ReLU. LeakyReLU is defined according to eq. (1), where α is set to 0.01.

$$LeakyReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{if } x \leq 0 \end{cases} \tag{1}$$

The task of the strategy controller is to select a strategy that minimizes the required pump energy to move the crane some distance with CTC. It should be noted that the objective is not to minimize the instant pump power, but to look ahead some distance and minimize the energy consumption of a movement that goes on for some time. An assumption then has to be made regarding how the crane tip reference will be in the near future. In this work, the energy is optimized on trajectories of a length of 13 s with a constant reference speed of 200 mm/s. With this speed, the crane can cover about a third to half of the working area in the horizontal direction during a trajectory, depending of the vertical position of the crane tip. This distance is considered a reasonable assumption of how far the crane is moved along a constant direction during normal work. The selection of the time horizon over which the accumulated cost during optimization is calculated can also be seen as a trade-off between promoting long term gains while still being able to adapt to changes in the input.

The output from the strategy controller is the speed command to the extension cylinder. It could just as well have been the command to the 1st or 2nd boom, the key is that the strategy is set by deciding the motion for a single one of the cylinders since the system only has one extra degree of freedom.

Once the output to the extension cylinder is given, the problem of finding the output to the 1st and 2nd boom has a unique solution. The velocity controller thus takes the output from the strategy controller and calculates the output to the 1st and 2nd boom while making sure that the reference crane tip trajectory is followed by using the inverse kinematics equations of the boom system.

If no valid combination of cylinder positions can be found, the velocity controller sets all outputs to zero. It thus functions as a layer between the neural network and the crane that can also act as a safety layer that can stop the crane, or reduce the speed, in unsafe situations when deployed on a real crane.

4. Training Environment and Simulation Model

The network is trained by Reinforcement learning policy optimization in a simulation environment implemented in Python, where the network functions as the policy approximator. The PyTorch framework is used for setting up the network and running the training loop. The crane is implemented as a quasi-static geometry model. Since the network only is trained to provide a high level motion strategy, no dynamics is needed in the crane model. It is however, important to have a good energy consumption model, since the strategy decisions are made based on this. The simulation model boundary is set at the outlet of the variable pump, so the signals of interest are the hydraulic pump pressure and pump flow.

The flow leakage in the crane is low and the pump flow is thus modelled directly as a summation of the oil required to move each cylinder according to the motion within the

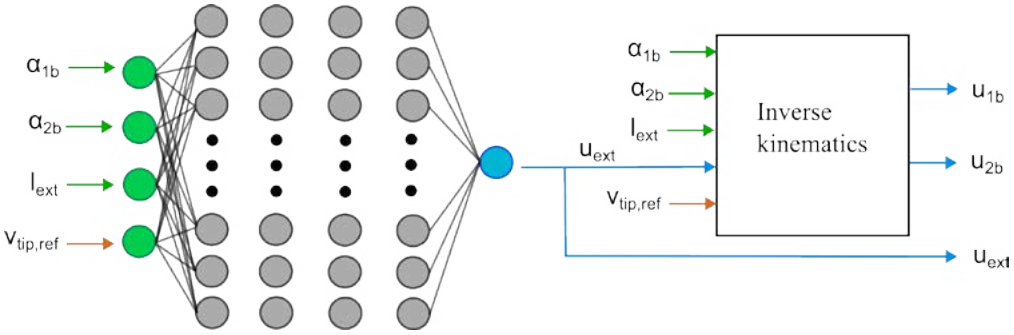


Figure 2: Layout of the neural network for the strategy controller and its connection to the velocity controller.

timestep, according to Equation 5. The hydraulic system is a load-sensing system, which gives that the pump pressure can be calculated as the maximum of the load pressures plus a pressure margin, according to Equation 6, where $p_{func,cyl}$ are the active side cylinder pressures, Δp_{func} are the pressure drops between the valve and the cylinders and Δp_{pump} is the pump pressure margin.

The cylinder pressure level required to move each cylinder depends on the force on the cylinder induced by the payload and the dead weight of the crane, on the counter pressure and on the friction in the system. The force can be calculated based on the geometry model, but given that it is hard to physically model the friction, especially in the extension system, the rest of the contribution to the cylinder pressure is modelled by a black-box model, with cylinder flow and force as inputs, fitted to measurement data.

Initially, a machine learning approach with gaussian process models, presented in [9], was tested. Although the fit of the model was very good in most working points, it was sensible to the distribution of the data points and its interpolation scheme could produce outputs that were not in line with the expected behavior. For example, if the force is kept constant, the pressure of a cylinder chamber is expected to increase with speed, but for some specific force values, the pressure could decrease for a small range of increasing speed.

Inconsistencies like this are easily exploited by the dynamic programming optimization, but more difficult for the gradient based optimization of the network to find, creating differences in the results that might not hold on the real system. For this reason, polynomials of the second degree are used instead, and the coefficients are fitted with with the least square minimization criterion. For the extension cylinder, different polynomial models are created depending on direction of movement and gravitational force, and the currently active pressure, P_{active}, can be calculated according to equations 3 to 4, where $P_{pist,static}$ and $P_{rod,static}$ are the contributions from the static force on the cylinder to the piston and rod side chamber pressure, and $P_{friction}$ is the contribution from the friction force given by the blackbox model. Figure 3 show the measurement data for the friction induced pressure as a function of cylinder flow for one of these cases, retraction of the cylinder and gravitation force pulling the cylinder in the extension direction, together with the fitted model. Figure 4 show the resulting fit of the active pressure to validation data from a test cycle when all four extension models are combined.

$$P_{pist,static} = \cdot \frac{F_{ext}}{A_{pist}} \tag{2}$$

$$P_{rod,static} = \cdot \frac{F_{ext}}{A_{rod}} \tag{3}$$

$$P_{active} = \begin{cases} |P_{rod,static}| + P_{friction,F-,v-} & \text{if } F_{ext} \leq 0 \text{ and } q \leq 0 \\ -|P_{pist,static}| + P_{friction,F-,v+} & \text{if } F_{ext} \leq 0 \text{ and } q > 0 \\ -|P_{rod,static}| + P_{friction,F+,v-} & \text{if } F_{ext} > 0 \text{ and } q \leq 0 \\ |P_{pist,static}| + P_{friction,F+,v+} & \text{if } F_{ext} > 0 \text{ and } q > 0 \end{cases} \tag{4}$$

The pressure drop between the valve and the cylinder is modelled as a simple flow dependent pressure drop model with coefficients fitted to measurement data together with load holding valve models. These models are presented in [6].

$$q_{pump} = \frac{\Delta 1b_{clen}}{T_s} * A_{1b,cyl} + \frac{\Delta 2b_{clen}}{T_s} * A_{2b,cyl} + \frac{\Delta ext_{clen}}{T_s} * A_{ext,cyl} \tag{5}$$

$$\begin{aligned} p_{pump} = &\max[p_{1b,cyl} + \Delta p_{1b}, p_{2b,cyl} + \Delta p_{2b}, p_{ext,cyl} + \Delta p_{ext}] + \\ &+ \Delta p_{pump} \end{aligned} \tag{6}$$

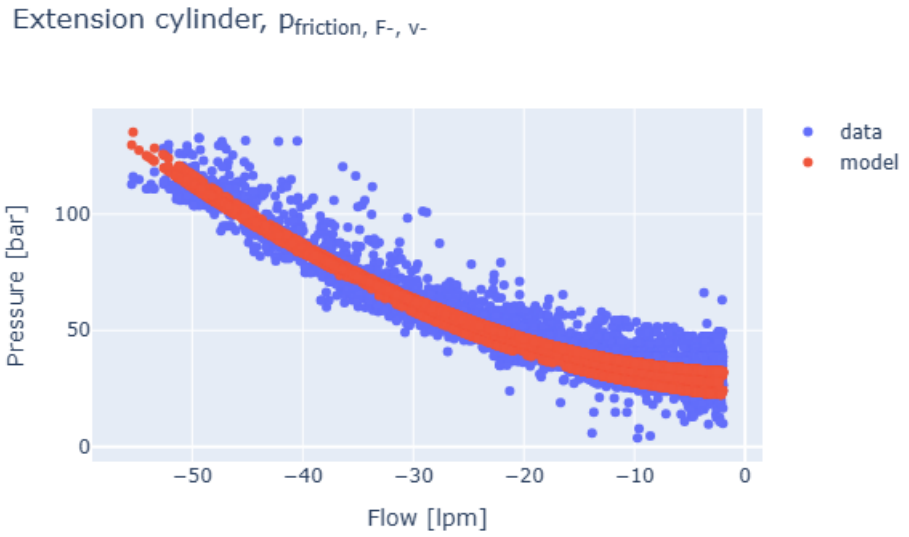


Figure 3: Model of extension friction based pressure as a function of flow. The variation for a fixed flow comes from the small impact of the static force.

During training of the network, a batch of 1000 cylinder positions that corresponds to crane tip positions within the working area of the crane, see Figure 5, is sampled. A simulation with a fixed timestep on a fixed horizon is rolled out on the batch for a number of episodes. Given that the objective of the optimization of the network is to minimize the energy consumption, the cost function is formulated according to eq. (7), where P_i is the pump power required during the current step and H is the time horizon of the simulation.

$$C = \sum_{i=1}^{H} P_i \tag{7}$$

$$P_i = q_{pump,i} * p_{pump,i} \tag{8}$$

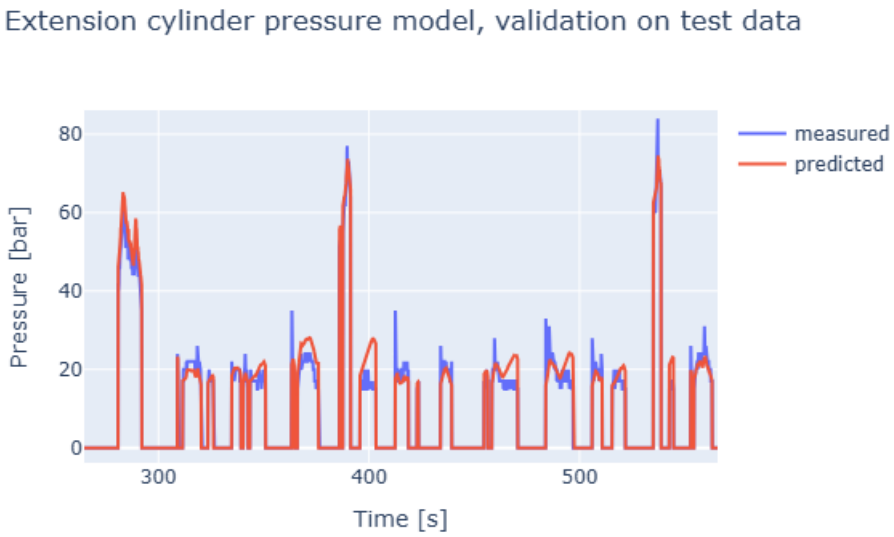


Figure 4: Validation of model of active extension cylinder pressure.

The policy, that is the network during training, is probabilistic. In order to allow the system to explore new solutions, the current best guess is not always selected as output from the policy. Instead, the network is set to produce a mean value and a standard deviation that can form a distribution from which the output (command to the extension cylinder) is sampled. When the controller later is deployed, the sampling from the distribution is replaced by directly taking the mean output from the network.

The network is trained until convergence. New batches of initial positions are sampled at regular intervals to ensure that the solution is generalized. The built-in gradient-descent optimizer AdamW of Pytorch is used to update the weights of the network after each episode based on the gradients of the average cost of all the trajectories in the batch with respect to the weights, or by using the Reinforcement learning terminology, the policy is optimized with respect to the (negative) reward it was given for the actions taken during the episode.

5. Results

The new controller is evaluated by comparing to the rule-based controller baseline and the dynamic programming benchmark on a number of pre-defined test cases. The test cases are divided on four different crane tip trajectories located at different parts of the working area of the crane. For each trajectory there are four different test cases formed by four different starting crane poses. The four trajectories, with one of their starting crane poses each, are displayed in Figure 6. The graphs visualize the motions of the crane when controlled by the new controller. The performance of the new controller, and its similarities and differences to the dynamic programming solutions, on each trajectory is presented in Sections 5.1 to 5.4. A summary of the energy performance comparing both with the dynamic programming benchmark and the rule-based controller baseline, is found in Section 5.5.

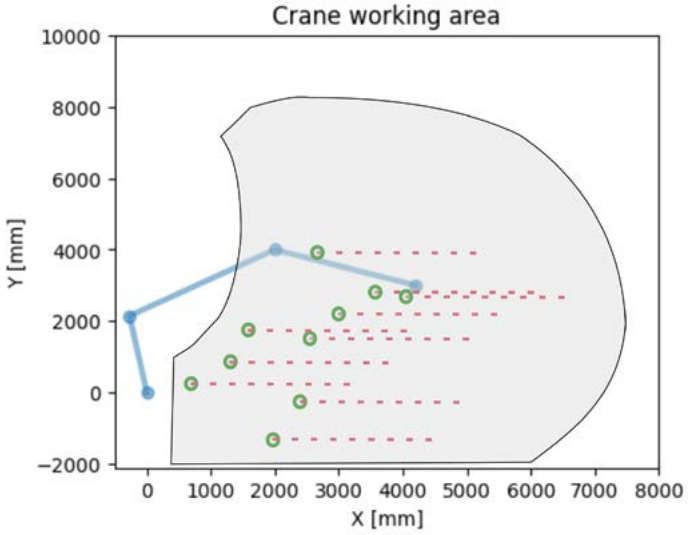


Figure 5: Working area of the crane, on which the neural network controller is trained. The figure also shows 10 example trajectories, with green circles for the starting points and dashed red lines for the trajectories, from a batch on which the controller is evaluated on during each episode.

5.1 Trajectory 1

A comparison between the cylinder commands, active side cylinder pressures, pump flow and pump pressure, for the dynamic programming, DP, controller and the new, neural network based, NN, controller is displayed in Figure 7 for one of the test cases for trajectory 1. At a quick glance, the cylinder commands are rather similar, but while the dynamic programming controller holds the extension still at most times, the new controller constantly runs it at a low speed. This creates a big difference in pump pressure, as the extension cylinder sets the highest pressure during the first half of the movements for the new controller. As the flow demand is similar for both controllers, the much higher pressure demand for the new controller results in a significantly bigger power demand.

The new controller is not able to find the solution where the extension is held still during training due to that the possibility for a neural network to output exactly zero is extremely small. It is however, possible to make adjustments after the training is completed. If the command to the extension is lower than a certain limit, it is plausible to assume that the power demand will be similar, or lower, if the command is set to zero. To keep the crane tip reference following, the 1st and 2nd boom have to be adjusted accordingly, which is handled by the velocity controller. If the adjusted outputs, or resulting cylinder positions, turn out to be illegal, the controller reverts back to the original solution. If this adjustment is applied to the test case 1.3, the pump power is now much more similar to that the DP controller, see Figure 8. The other three test cases of trajectory 1 follows a similar pattern.

5.2 Trajectory 2

Trajectory 2 shows a slightly different scenario, displayed in Figure 9 for one of the test cases. The dynamic programming controller once again keeps the extension still for a large part of the movement, but contrary to during trajectory 1 the extension does not set the highest pressure, so the pump pressure levels are rather similar for both controllers. The

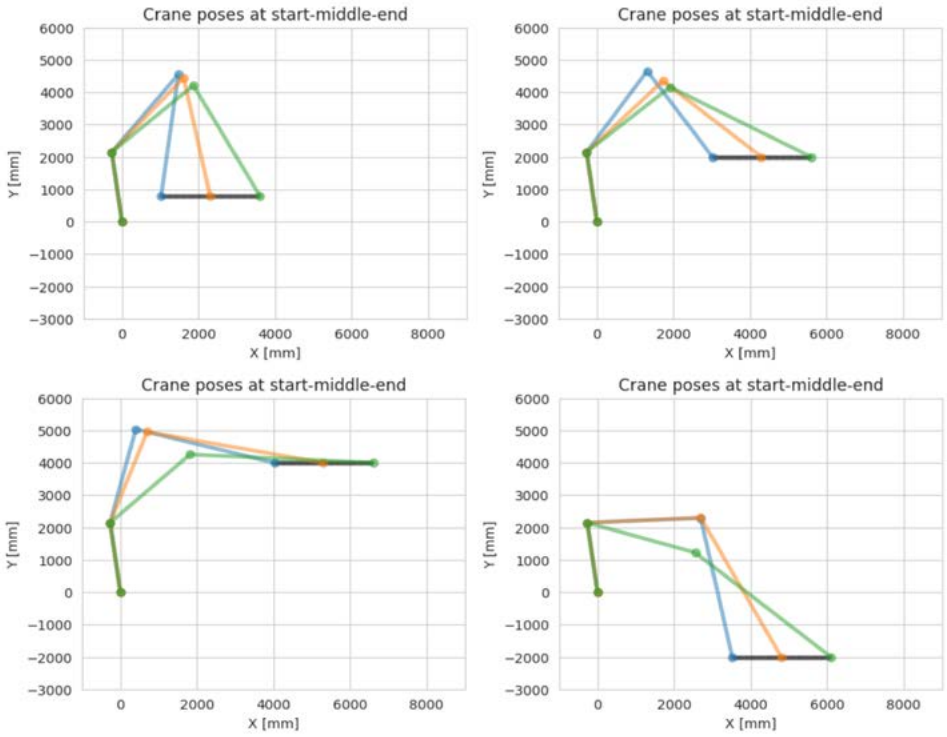


Figure 6: Four different trajectories on which the new controller will be evaluated. One of the four starting crane poses (blue lines), and the corresponding motion (orange lines for mid-time pose and green lines for end pose) of the crane while controlled by the new controller are displayed for each trajectory.

dynamic programming solution does, however, reduce the flow demand compared to the new controller.

5.3 Trajectory 3

The third trajectory is placed in the upper part of the working area of the crane. For all four test cases, the dynamic programming algorithm is able to find a smart solution where the 2nd boom is moved towards a horizontal position during the first part of the trajectory, allowing the extension to be the only function moved in the last part of the trajectory, exemplified by test case 3.2 in Figure 10. Finding exactly the state where the 2nd boom is horizontal is too difficult for the new controller, and consequently it suffers from a high pump pressure being set by moving the 2nd boom upwards during the whole trajectory. The difference between the 2nd boom and the extension pressure can be seen in the plot.

5.4 Trajectory 4

The last trajectory in the comparison is placed in the lower part of the working area of the crane. On this trajectory it is once again highlighted what a big difference to the energy consumption it can make between moving a function slowly or stopping it completely and compensating the movement by the other two functions. The new controller has found solutions where the 1st boom is initially driven slowly upwards, causing the pump pressure to be much higher than for the dynamic programming controller solution. By implementing a similar adjustment as for the extension, that is, a dead zone is placed for small positive

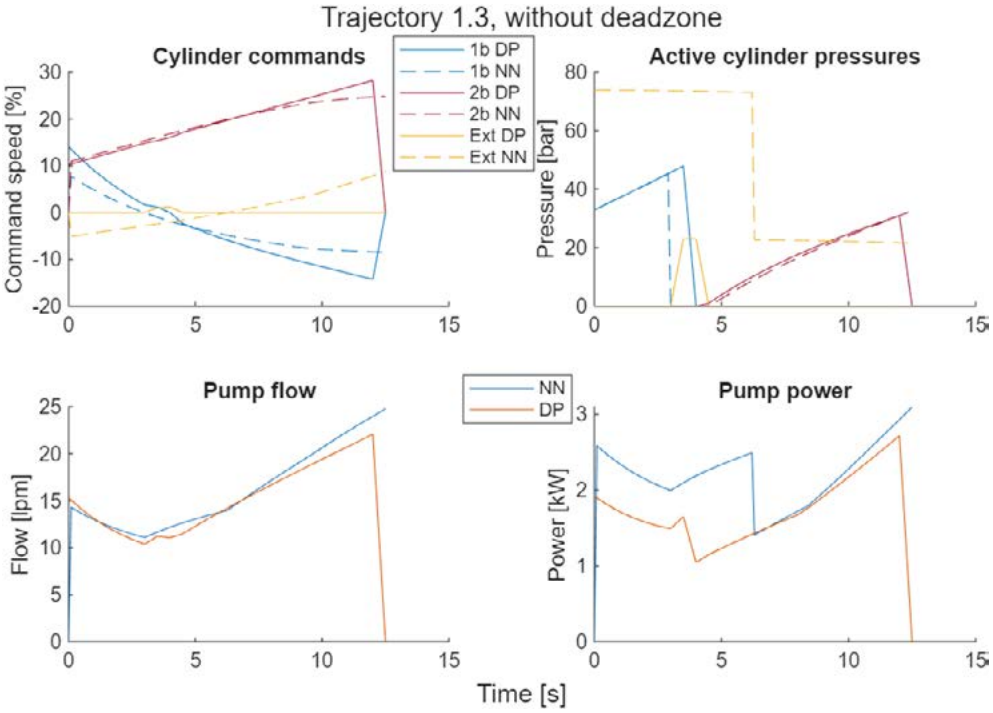


Figure 7: Test case 1.3 with original outputs from the neural network based controller. Note the high pressure set by the extension during the first half of the movement for the NN controller.

commands to the 1st boom which sets the 1st boom command to zero if the adjusted solution is legal. The resulting cylinder commands, as well as active cylinder pressure, pump flow and pump power, are displayed in Figure 11 together with the dynamic programming controller outputs. Even though there are some differences regarding when and how the extension is activated, the resulting mean power consumption is rather similar for both controller, since the 2nd boom is operated in the same direction and sets the highest pressure during the main part of the trajectory.

5.5 Power Comparison

The mean power consumption for each of the 16 test cases is displayed for the three different controllers, the dynamic programming controller (DP), the rule-based controller (r-b) and the neural network based controller (NN), in Table 1. The increase in average power consumption between the dynamic programming controller and the neural network controller varies between 5 % and 34 %, with a mean value of 15 %. In order to be able to draw any conclusions about the quality of these figures, the neural network controller is also compared to the rule-based controller. The average power consumption is most often lower for the neural network controller, with a maximum reduction of 22 %, but for some test cases the consumption is up to 10 % higher for the new controller. The mean value of the reduction is 5.6 %.

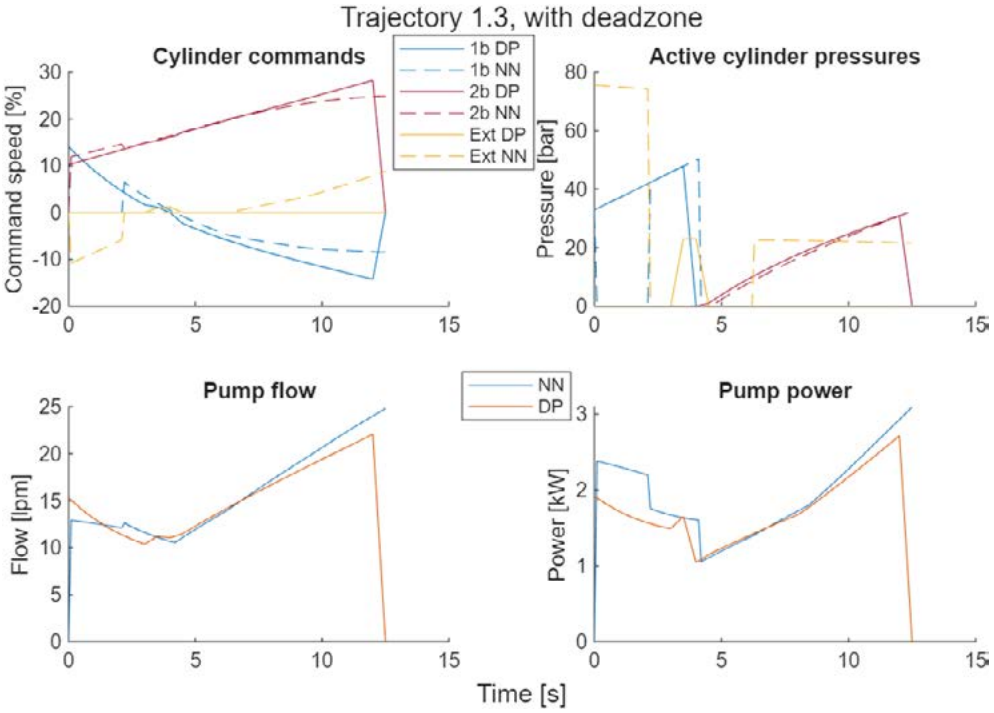


Figure 8: Test case 1.3 with adjusted outputs from the neural network based controller. Note the similar pump power for the two controllers.

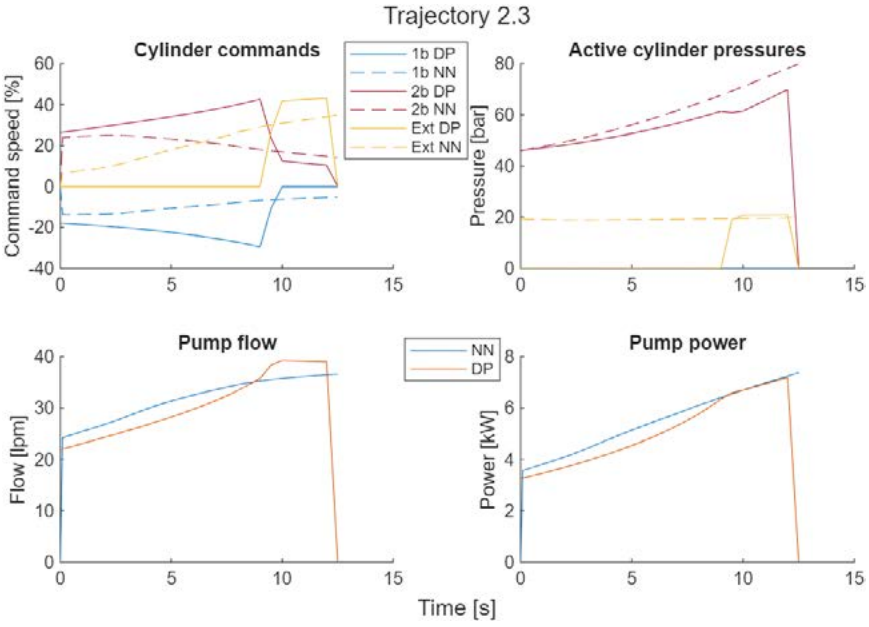


Figure 9: Test case 2.3. Note the lower pump flow for the dynamic programming controller.

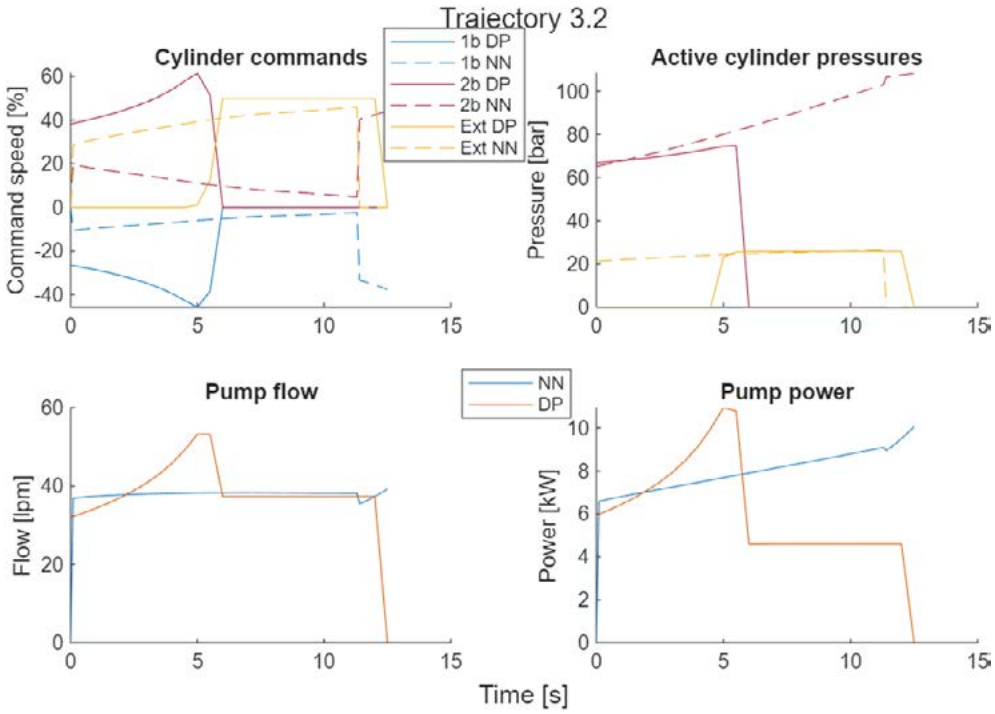


Figure 10: Test case 3.2. Note how only the extension is moved in the second half of the movement for the DP controller.

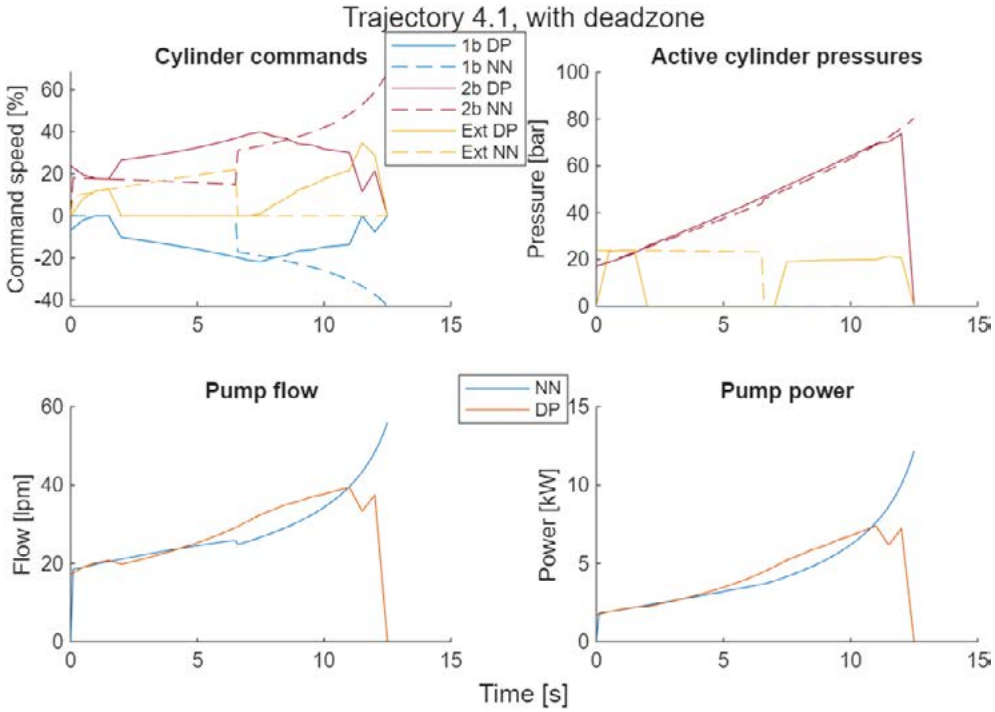


Figure 11: Test case 4.1. Note how the 2nd boom sets the highest pressure during the main part of the trajectory and how the power difference only comes from the flow.

The controller presented above has been optimized for the tip speed of 200 mm/s and the payload of 1000 kg. Two other controllers, one with 100 mm/s tip speed and 1000 kg payload, and one with 200 mm/s tip speed and 2000 kg payload, have also been optimized and evaluated on the same trajectories. The first one has an average power consumption that is 6.5 % higher than for its corresponding dynamic programming solution. The main reason for the smaller difference than for the faster test cases is that the optimizer more often outputs small cylinder commands that can be set to zero during deployment and then be more similar to the dynamic programming solutions. The one with a higher payload has an average power consumption that is 22 % higher than for its corresponding dynamic programming solution. The higher payload gives a larger relative difference in pump pressure between solutions from the dynamic programming where one of the cylinders is often kept still and solutions from the neural network controller where this rarely happens.

6. Discussion

For some of the tested trajectories, the neural network controller finds similar solutions as the near-optimal dynamic programming algorithm. These trajectories are characterized in that the cylinder commands have continuous like shapes that seems to be easier for the network to find than solutions where the commands vary more abruptly or take very specific values. Smooth motions of the cylinders are probably more easy for a real crane to follow without any noticeable small stops or small jerks of the crane tip, so in that perspective the new controller provides good solutions.

However, the results here show that energy efficiency of the neural network controller is rather low compared to the dynamic programming solution. The neural network controller only reduces the energy consumption about 6 % compared to a simple rule-based controller. When taking into account that a complete CTC controller needs to be able to control the crane tip in every possible direction, not just vertical or horizontal, the benefits of the neural network controller will be even smaller. It would be possible to train three more networks, horizontal inwards and vertical up and down, and then combine them with the network presented here for horizontal outwards. But whenever the crane tip should move along a diagonal, there is no guarantee that the energy minimization objective will hold. So in total, a lower energy reduction than 6 % compared to the rule-based controller can be expected.

If a semi-automatic function is required for the autonomy itself, rather than being a means for energy efficiency, it should be noted that there is a rather big difference in energy consumption between an optimized controller and more simple controller. If the semi-automatic function is considered as a means for energy efficiency, the results of the controller evaluated in this work must be put in relation to manual operation of the crane. In previous work, [8], it was shown that a CTC controller optimized with dynamic programming consumed about 14% less energy on a drive cycle than manual operation. Given that the results in this paper indicate that a realtime CTC consumes more than 15% more than the dynamic programming solution, automation through CTC is not considered an effective means of energy reduction. It should be noted that during the manual operation of the drive cycle in [8], the crane tip was not forced to follow straight lines. So automation could still potentially be a means of energy reduction, if given the freedom to move arbitrarily between waypoints.

Why is the trained network controller not better? The network structure and training method with a gradient based optimizer was chosen for his study due to its success in teaching a neural network to control the crane to be fast during CTC operation, proven in a previous study [10]. The kinematic equations of the crane are differentiable and worked well with the gradient based optimizer. Switching to energy consumption as objective introduced some properties that proved to be difficult for the selected method to handle.

Table 1: Comparison of average power demand [kW] for the three different controllers for the 16 test cases (four different trajectories with four different starting positions each).

	$P_{avg,DP}$	$P_{avg,NN}$	$P_{avg,r-b}$	Diff DP-NN	Diff r-b - NN
TC 1.1	1.52	1.68	2.16	+11%	-22%
TC 1.2	2.06	2.48	2.44	+20%	+1.7%
TC 1.3	1.68	1.93	2.36	+15%	-18%
TC 1.4	1.89	2.32	2.46	+23%	-5.6%
Average TC 1				+17%	-11%
TC 2.1	4.98	5.42	6.11	+8.8%	-11%
TC 2.2	4.60	5.35	5.50	+16%	-2.7%
TC 2.3	4.89	5.47	5.95	+12%	-8.2%
TC 2.4	4.80	5.59	6.21	+16%	-9.9%
Average TC 2				+13%	-8.0%
TC 3.1	6.70	8.03	8.35	+20%	-3.9%
TC 3.2	6.04	7.94	8.40	+31%	-5.6%
TC 3.3	7.01	7.76	7.62	+11%	+1.9%
TC 3.4	6.02	8.08	8.38	+34%	-3.7%
Average TC 3				+24%	-2.8%
TC 4.1	4.18	4.42	4.86	+5.9%	-9.0%
TC 4.2	4.21	4.46	4.54	+6.0%	-1.8%
TC 4.3	4.42	4.68	4.74	+5.8%	-1.3%
TC 4.4	4.10	4.31	3.95	+5.2%	+9.4%
Average TC 4				+5.7%	-0.7%
Average all				+15%	-5.6%

A linear combination such as the output from the feed forward neural network has an inherent difficulty of taking on very specific values, such as exactly zero or the command the moves the extension cylinder to a very specific position. Small reference command outputs from the controller can be adjusted to exactly zero during deployment of the controller, which was shown to significantly improve the performance. It is however not possible to apply this deadband during training as it is a non-differentiable function that hinder the gradient based optimizer to track the gradients of the loss function. Another property of the system that it is difficult for the optimizer to track is the non-linearity of the hydraulic system. For a load sensing system, the highest load pressure of the currently active cylinders sets the pump pressure. When the setup of active cylinder changes, the pump pressure can make a large step. So, a neural network could be a good choice for controller if a deadband is applied during training, and if an optimizer that does not rely of gradients of the loss function is used. Non-gradient based optimizers are for example population based methods such as genetic algorithms and particle swarm optimization.

7. Conclusion

Training a neural network to learn an optimal control strategy with regards to energy efficiency that can control a large part of the working envelop of the crane has proven to be a difficult task with the investigated optimization method. On average, the network controller consumes 15 % more energy than the near-optimal solution on the test cases presented here, and only consumes 5.6 % less energy than a simple rule based controller.

During the work of this paper it has become clear that the properties of the system to be optimized, that is the energy consumption of the crane and its load sensing hydraulic system, is not very well suited for gradient-based optimization. For optimizing the speed of the crane tip, gradient-based optimization worked well. But when considering the energy consumption of the hydraulic system a significant non-linearity is introduced - the pump pressure can make a large step when the reference to a cylinder goes from zero to a non-zero value, if the load on that cylinder is higher than the previously active cylinders. Future work will be focused on investigating if other types of optimizers that does not need to calculate gradients, such as different population based methods, are better suited for training neural networks to act as controllers for a load sensing hydraulic system.

Acknowledgment

The authors would like to thank Hiab AB, and Reza Ghabcheloo at Tampere University for their support during the work with this paper. The work in this paper was sponsored by the Swedish Energy Agency, and is part of STEALTH II - Sustainable, Electrified and Automated Load Handling, App. no 44427-3.

References

[1] Jennifer Andersson, Kenneth Bodin, Daniel Lindmark, Martin Servin, and Erik Wallin. Reinforcement learning control of a forestry crane manipulator. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021.

[2] Rohit Dhakate, Christian Brommer, Christoph Bohm, Harald Gietler, Stephan Weiss, and Jan Steinbrener. Autonomous control of redundant hydraulic manipulator using reinforcement learning with action feedback. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[3] Ignacy Duleba and Michal Opalka. A comparison of jacobian-based methods of inverse kinematics for serial robot manipulators. *Int. J. Appl. Math. Comput. Sci.*, 23:373–382, 2013.

[4] Rickard Nilsson. Inverse kinematics, 2009.

[5] Jarmo Nurmi and Jouni Mattila. Global energy-optimal redundancy resolution of hydraulic manipulators: Experimental results for a forestry manipulator. *Energies*, 10, 2017.

[6] Amy Rankka and Alessandro Dell'Amico. Loss analysis and concept comparison for electrically driven hydraulic loader crane. In *The 17th Scandinavian International Conference on Fluid Power, SICFP'21*, 2021.

[7] Amy Rankka, Alessandro Dell'Amico, and Petter Krus. Drive cycle generation for a hydraulic loader crane. In *The 16th Scandinavian International Conference of Fluid Power*, 2019.

[8] Amy Rankka, Marcus Rösth, and Alessandro Dell'amico. Evaluating the performance of semi-autonomous kinematically redundant loader crane operation. In *Advancements in Fluid Power Technology: Sustainability, Electrification, and Digitalization. Proceedings of the Global Fluid Power Society PhD Symposium 2024*. Springer Nature, 2024.

[9] Abdolreza Taheri, Pelle Gustafsson, Marcus Rosth, Reza Ghabcheloo, and Joni Pajarinen. Nonlinear model learning for compensation and feedforward control of real-world hydraulic actuators using gaussian processes. *IEEE Robotics and Automation Letters*, 7:9525–9532, 10 2022.

[10] Abdolreza Taheri, Amy Rankka, Pelle Gustafsson, Joni Pajarinen, and Reza Ghabcheloo. End-effector cartesian velocity control for redundant loader cranes using reinforcement learning. *IEEE Transactions on Robotics*, 2024.

[11] Alessandro Tringali and Silvio Cocuzza. Finite-horizon kinetic energy optimization of a redundant space manipulator. *Applied Sciences (Switzerland)*, 11:1–15, 3 2021.

[12] Hiab Official Webpage. Ctc – crane tip control taking ease of use to a new level, 2020.

[13] John Deere Official Webpage. Intelligent boom control, 2025.

[14] Palfinger Official Webpage. Smart control, 2025.

Biography



Amy Rankka received the M.Sc. degree in electrical engineering in 2015 from Linköping University, Linköping, Sweden, where she is currently studying towards the Ph.D. degree in mechanical engineering with the division of fluid and mechatronic systems. She works as an R&D engineer at Hiab AB. Her research interests include energy efficient hydraulic systems.



Alessandro Dell'Amico has a MSc degree and a PhD degree in Mechanical Engineering from Linköping University. His research includes aircraft system design, energy management, actuation systems, and energy efficiency in fluid power systems. He currently holds a position as Adjunct Associate Professor at Linköping University.

Multidimensional Neural Networks Based RISE Control for Hydraulic Manipulators with Disturbances Compensation

Junjie Mi[1], Gang Shen[1], Jianyong Yao[2]

1. School of Mechatronics Engineering, Anhui University of Science & Technology, Huainan, Anhui 232001, China, mijunjie@aust.edu.cn, 2023020@aust.edu.cn

2. School of Mechanical Engineering, Nanjing University of Science & Technology, Nanjing, Jiangsu 210094, China, yaojianyong@njust.edu.cn

Abstract.

With the accelerated development of automation and unmanned operations in heavy-load engineering fields, higher demands are placed on the control performance of hydraulically driven heavy-load manipulators. However, hydraulic manipulators suffer from prominent issues such as large load fluctuations, significant dynamic model errors, and strong joint coupling characteristics, which readily lead to degraded control performance. To achieve high-precision pose control, this study proposes a nonlinear disturbances compensation controller. A multidimensional radial basis function neural network (mRBFNN) is used to approximate various uncertain nonlinearities present in the system. Combined with the backstepping method, a nonlinear robust integral of the sign of the error (RISE) controller is designed. Online compensation for uncertain nonlinearities is realized during the control law design process, thereby enhancing system robustness. The Lyapunov stability theory proves that the system is stable and all signals are bounded. Simulation results demonstrate that the proposed control strategy significantly improves joint tracking accuracy and anti-disturbance capability of the hydraulically driven manipulator, fully validating the effectiveness of the control strategy.

Keywords. Hydraulic manipulator, Electro-hydraulic servo system, Multidimensional radial basis function neural network (mRBFNN), Disturbances compensation.

1. INTRODUCTION

In fields such as heavy equipment, hydraulic systems play an irreplaceable role[1]. However, hydraulic systems inherently possess strong nonlinearity, time-varying parameters, and sensitivity to external disturbances, posing significant challenges to achieving high-precision control[2]. In recent years, with the rapid development of nonlinear control theory,

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

combining advanced control strategies with the characteristics of hydraulic systems to achieve high dynamic performance control for multi-degree-of-freedom (DOF) hydraulic manipulators has become a hot topic of common interest in academia and industry[3].

In the field of nonlinear control theory, methods such as sliding mode control, adaptive control, and robust control have been widely applied in hydraulic systems[4][5]. However, most current research focuses on single-DOF systems, leaving room for improvement in handling the coupled nonlinearities of multi-DOF hydraulic manipulators[6].

In recent years, significant progress has been made in the field of fluid power transmission and control regarding hydraulic system modeling and compensation. Liu et al.[7] proposed a model predictive control based on a state observer for large-scale offshore hydraulic systems, laying a foundation for high-precision control of hydraulic systems under complex working conditions. Literature[8][9] published in the journal Chinese Hydraulics & Pneumatics proposed compensation strategies for the nonlinear flow characteristics and pressure dynamics of valve-controlled asymmetric cylinders, respectively. Notably, Xia et al.[10] experimentally verified the good effect of friction compensation on improving the low-speed performance of hydraulic systems. Regarding parameter identification, intelligent algorithms such as particle swarm optimization[11] and deep reinforcement learning[12] have been introduced to improve model accuracy. However, most existing compensation methods rely on accurate system models, and their adaptability to parameter variations in multi-DOF systems needs further enhancement.

Traditional PID control is still widely used in industry due to its simple structure and few adjustable parameters, but its performance is limited when dealing with the nonlinear coupling characteristics of hydraulic manipulators. To address this issue, scholars have proposed various solutions: Yao et al.[13] designed an adaptive robust controller to handle model uncertainties; Chen et al.[14] combined a disturbance observer with the backstepping method to suppress the effects of external disturbances. In the study of manipulator trajectory tracking, Zhu et al.[15] verified the advantage of prescribed performance control in constraining tracking errors, but this study did not consider the dynamic characteristics of hydraulic actuators.

The remaining sections of this study are arranged as follows: Section 1 introduces the nonlinear mathematical model of the hydraulic manipulator; Section 2 details the design method of the multidimensional neural network; Section 3 presents the design process of the nonlinear integral robust controller; Section 4 verifies the effectiveness of the proposed algorithm through simulation examples; Section 5 summarizes the research.

2. NONLINEAR MATHEMATICAL MODEL OF THE HYDRAULIC MANIPULATOR

2.1. Dynamic Model of the Hydraulic Manipulator

Using the Lagrange energy method, the dynamic equation for a multi-DOF manipulator can be formulated as follows

$$\boldsymbol{\tau} = \boldsymbol{M}(\boldsymbol{y})\ddot{\boldsymbol{y}} + \boldsymbol{C}(\boldsymbol{y},\dot{\boldsymbol{y}})\dot{\boldsymbol{y}} + \boldsymbol{G}(\boldsymbol{y}) + \boldsymbol{\Delta}_{t1}(t) \tag{1}$$

where $\boldsymbol{M}(\boldsymbol{y}) \in \mathbb{R}^{n\times n}$ is the manipulator's mass matrix, which is positive definite and symmetric. Its off-diagonal elements $M_{ij} = M_{ji}$ represent force coupling effects between joints. $\boldsymbol{C}(\boldsymbol{y},\dot{\boldsymbol{y}}) \in \mathbb{R}^{n\times n}$ represents the Coriolis and centrifugal force matrix, describing the coupling between joint velocities and forces. $\boldsymbol{G}(\boldsymbol{y}) \in \mathbb{R}^{n}$ is the gravity vector. $\boldsymbol{\Delta}_{t1}(t) \in \mathbb{R}^{n}$ represents unmodeled uncertainties in the manipulator's dynamic model, such as link elasticity, joint clearance, and connection relationships. $\boldsymbol{y} = [y_1,...,y_n]^{\mathrm{T}} \in \mathbb{R}^{n}$ represents the rotation angle of each joint.

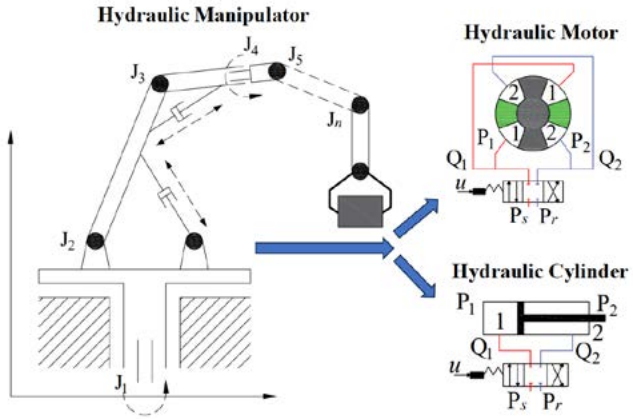


Figure 1 Configuration Diagram and Actuator of m-DOF Hydraulic Manipulator

As shown in Figure 1, hydraulic motors and hydraulic cylinders are responsible for driving different joints. The hydraulic cylinder is a single-rod cylinder, where 1 denotes the rodless chamber and 2 denotes the rod chamber. The hydraulic motor is a rotary actuator, with its position controlled by regulating the flow into and out of its ports via a servovalve.

Assumption 1: The servovalve's orifices are considered symmetric, and its dynamics are simplified as a proportional element.

Assumption 2: Considering the compact layout of the device and oil source, the oil in different chambers is assumed to have the same bulk modulus.

To maintain generality, the pressure dynamics of the electro-hydraulic servo system can be described as:

$$\begin{cases} \dot{p}_{1i} = \beta_e V_{1i}^{-1}\left(q_{1i} - A_{1i}\dot{x}_{si} - C_t\left(\dot{p}_{1i} - \dot{p}_{2i}\right)\right) + d_{1i}(t) \\ \dot{p}_{2i} = \beta_e V_{2i}^{-1}\left(-q_{2i} - A_{2i}\dot{x}_{si} + C_t\left(\dot{p}_{1i} - \dot{p}_{2i}\right)\right) - d_{2i}(t) \end{cases} \tag{2}$$

$$\begin{cases} V_{1i} = V_{01i} + A_{1i}x_{si} \\ V_{2i} = V_{02i} - A_{2i}x_{si} \end{cases} \tag{3}$$

where the subscript i denotes different joint numbers, $i = 1,2,...,n$; β_e represents the bulk modulus of the hydraulic oil, assumed constant; x_{si} represents the displacement of the actuator driving the joint. For a hydraulic motor, this is the joint rotation angle; for a hydraulic cylinder, it is the piston displacement. A_{1i} and A_{2i} are the effective areas of the piston or motor vane; V_{01i} and V_{02i} represent the initial volumes of the two chambers; V_{1i} and V_{2i} represent the actual volumes of the two chambers during motion; p_{1i} and p_{2i} represent the pressures in the two chambers; q_{1i} and q_{2i} denote the flow rates into/out of the two chambers; C_t is the internal leakage coefficient; $d_{1i}(t)$ and $d_{2i}(t)$ represent unmodeled uncertainties and external leakage in the pressure dynamics model.

The servovalve adjusts the orifice opening by changing the spool position, thereby regulating the flow into and out of the actuator chambers. The servovalve flow equation can be expressed as:

$$\begin{cases} q_{1i} = k_{qi}x_{vi}\left(s(x_{vi})\sqrt{p_s - p_{1i}} + s(x_{vi})\sqrt{p_{1i} - p_r}\right) \\ q_{2i} = k_{qi}x_{vi}\left(s(x_{vi})\sqrt{p_{2i} - p_r} + s(x_{vi})\sqrt{p_s - p_{2i}}\right) \end{cases} \tag{4}$$

$$k_{qi} = C_{di}w_i\sqrt{2/\rho_i}, \qquad x_{vi} = k_{ui}u_i \tag{5}$$

where k_{qi} represents the flow gain for the left and right spool displacements of the servovalve; p_s is the supply pressure, p_r is the return pressure; x_{vi} is the servovalve spool displacement; C_{di} is the flow coefficient of the servovalve orifice; w_i is the area gradient of the orifice; ρ is the density of the hydraulic oil. The function $s(*)$ is defined as:

$$s(*) = \begin{cases} 1, & \text{if } * \geq 0 \\ 0, & \text{if } * < 0 \end{cases} \tag{6}$$

Therefore, the flow equation (4) of the electro-hydraulic servo system can be rewritten as:

$$\begin{cases} q_{1i} = k_{ti}R_{1i}u_i \\ q_{2i} = k_{ti}R_{2i}u_i \end{cases} \tag{7}$$

where $k_{ti} = k_{qi}k_{ui}$. Also:

$$\begin{cases} R_{1i} = s(u_i)\sqrt{p_s - p_{1i}} + s(u_i)\sqrt{p_{1i} - p_r} \\ R_{2i} = s(u_i)\sqrt{p_s - p_{1i}} + s(u_i)\sqrt{p_{1i} - p_r} \end{cases} \tag{8}$$

The actual output force/torque of the electro-hydraulic servo system can be expressed as:

2.2. Linear-to-Rotational Nonlinear Transformation of the Actuator

The actuators of the multi-DOF hydraulic manipulator primarily consist of valve-controlled hydraulic cylinders and hydraulic motors. Two hydraulic cylinders are responsible for driving the joints with larger loads (joints 2 and 3), involving a nonlinear transformation from linear to rotational motion, i.e., a structural nonlinear relationship exists.

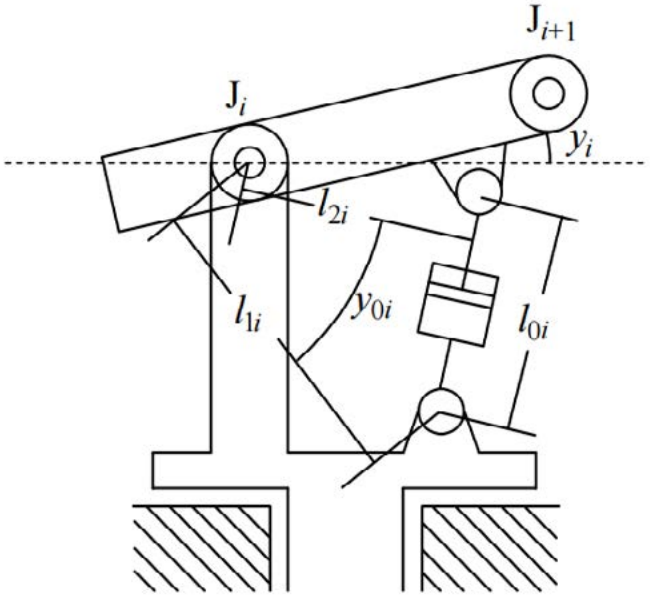


Figure 2 Schematic Diagram of Joint Linear-to-Rotational Nonlinear Transformation

The structural nonlinearity of the hydraulic manipulator is shown in Figure 2, where a hydraulic cylinder pushes a load to rotate about a pivot. Here, y_i represents the initial angle between the driven link and the horizontal plane; y_{0i} represents the initial angle between the two pivot points of the cylinder and the link's rotation axis; l_{1i} is the distance from the fixed pivot of the cylinder to the link's rotation axis; l_{2i} is the distance from the moving pivot of the cylinder to the link's rotation axis; l_{0i} is the initial length of the cylinder.

Based on the geometric relationships and the law of cosines, we obtain:

$$l_i = \sqrt{l_{1i}^2 + l_{2i}^2 - 2l_{1i}l_{2i}\cos\left(y_{0i} + y_i\right)} \tag{9}$$

Let the actual displacement of the cylinder be $x_{pi} = l_i - l_{0i}$. Thus, the actual velocity of the cylinder is:

$$\begin{aligned} v_{pi} = \dot{x}_{pi} &= \frac{\mathrm{d}\left(l_i - l_{0i}\right)}{\mathrm{d}t} = \frac{\mathrm{d}l_i}{\mathrm{d}t} \\ &= \frac{l_{1i}l_{2i}\sin\left(y_{0i} + y_i\right)}{\sqrt{l_{1i}^2 + l_{2i}^2 - 2l_{1i}l_{2i}\cos\left(y_{0i} + y_i\right)}}\dot{y}_i \\ &= S_i\left(y_i\right)\dot{y}_i \end{aligned} \tag{10}$$

where $S_i\left(y_i\right)$ is also the perpendicular distance from the cylinder to the link's rotation axis. Based on the above analysis, x_{si} in Eq. (3) can be expressed as:

$$x_{si} = \begin{cases} y_i & \text{, driven by motor} \\ \sqrt{l_{1i}^2 + l_{2i}^2 - 2l_{1i}l_{2i}\cos\left(y_{0i} + y_i\right)} - l_{0i} & \text{, driven by cylinder} \end{cases} \tag{11}$$

The actuator's own velocity can be obtained from Eq. (11):

$$\dot{x}_{si} = S_i\left(y_i\right)\dot{y}_i \tag{12}$$

$$S_i(y_i)=\begin{cases}1 & \text{,driven by motor}\\ \dfrac{l_{1i}l_{2i}\sin(y_{0i}+y_i)}{\sqrt{l_{1i}^2+l_{2i}^2-2l_{1i}l_{2i}\cos(y_{0i}+y_i)}} & \text{,driven by cylinder}\end{cases} \tag{13}$$

Combining Eq. (9), the actual output torque of the manipulator actuator for each joint can be expressed as:

$$T_i = S_i(y_i)(A_{1i}p_{1i} - A_{2i}p_{2i}) \tag{14}$$

2.3. State-Space Equation of the Hydraulic Manipulator

To facilitate subsequent controller design and analysis, define the state variables $\boldsymbol{x}=[\boldsymbol{x}_1^{\mathrm{T}},\boldsymbol{x}_2^{\mathrm{T}},\boldsymbol{x}_3^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{x}_1=\boldsymbol{y}=[y_1,y_2,\cdots,y_n]^{\mathrm{T}}$, $\boldsymbol{x}_2=\dot{\boldsymbol{x}}_1=[\dot{y}_1,\dot{y}_2,\cdots,\dot{y}_n]^{\mathrm{T}}$, $\boldsymbol{x}_3=\boldsymbol{A}_1\boldsymbol{p}_1-\boldsymbol{A}_2\boldsymbol{p}_2$. Define the manipulator's time-varying rotation angle and angular velocity as y_i and $\dot{y}_i$, respectively. Combining the above equations, the nonlinear dynamic model of the multi-DOF hydraulic manipulator can be expressed in state-space form as:

$$\begin{aligned}\dot{\boldsymbol{x}}_1 &= \boldsymbol{x}_2\\ \dot{\boldsymbol{x}}_2 &= \boldsymbol{M}^{-1}(\boldsymbol{x}_1)\left[\boldsymbol{S}(\boldsymbol{x}_1)\boldsymbol{x}_3-\boldsymbol{G}(\boldsymbol{x}_1)-\boldsymbol{C}(\boldsymbol{x}_1,\boldsymbol{x}_2)\boldsymbol{x}_2\right]\\ &\quad+\boldsymbol{M}^{-1}(\boldsymbol{x}_1)\boldsymbol{\Delta}_{t1}\\ \dot{\boldsymbol{x}}_3 &= \boldsymbol{f}_1\boldsymbol{u}-\boldsymbol{f}_2-\boldsymbol{f}_3+\boldsymbol{\Delta}_{t2}\end{aligned} \tag{15}$$

where the control input vector for each joint's actuator in the hydraulic manipulator system is $\boldsymbol{u}=[u_1,u_2,\cdots,u_n]^{\mathrm{T}}$; expressions for other vectors can be found in [16].

Assumption 3: The commanded trajectory $\boldsymbol{x}_{1d}=\left[x_{11d},x_{12d},\cdots,x_{1nd}\right]^{\mathrm{T}}$ for each joint of the hydraulic manipulator is third-order differentiable, and all its derivatives are bounded.

Assumption 4: The elements of the unmatched uncertainty $\boldsymbol{\Delta}_{t1}=\left[\Delta_{t11},\Delta_{t12},\cdots,\Delta_{t1n}\right]^{\mathrm{T}}$ and the matched uncertainty $\boldsymbol{\Delta}_{t2}=\left[\Delta_{t21},\Delta_{t22},\cdots,\Delta_{t2n}\right]^{\mathrm{T}}$ are sufficiently smooth, and their first and second derivatives exist and are bounded.

3. MULTIDIMENSIONAL NEURAL NETWORK DESIGN

For hydraulic manipulators, complex joint coupling and unmodeled uncertainties are the main factors affecting high-precision position tracking. Therefore, this section employs a multidimensional radial basis function neural networks (mRBFNNs) to approximate and estimate the uncertain joint coupling dynamics and unmodeled disturbances of the hydraulic manipulator, achieving feedforward compensation during the controller design process.

From (16), it can be seen that $\boldsymbol{\Delta}_{t1}$ is related to the position, velocity, and acceleration signals of each joint, while $\boldsymbol{\Delta}_{t2}$ is only related to joint positions. Therefore, $\boldsymbol{\Delta}_{t1}$ and $\boldsymbol{\Delta}_{t2}$ can be designed in the following form:

$$\boldsymbol{\Delta}_{t1}=\left[\boldsymbol{W}_{11}^{\mathrm{T}}\boldsymbol{h}_{11}(\boldsymbol{\chi}_1),\cdots,\boldsymbol{W}_{1n}^{\mathrm{T}}\boldsymbol{h}_{1n}(\boldsymbol{\chi}_n)\right]^{\mathrm{T}}+\boldsymbol{\varepsilon}_1 \tag{16}$$

$$\boldsymbol{\Delta}_{t2}=\left[\boldsymbol{W}_{21}^{\mathrm{T}}\boldsymbol{h}_{21}(y_1),\cdots,\boldsymbol{W}_{2n}^{\mathrm{T}}\boldsymbol{h}_{2n}(y_n)\right]^{\mathrm{T}}+\boldsymbol{\varepsilon}_2 \tag{17}$$

where $\boldsymbol{W}_{1i}=\left[W_{1i1},\cdots,W_{1iH}\right]^{\mathrm{T}}\in\square^{H}$, H represent the number of neurons in the first neural network hidden layer; $\boldsymbol{W}_{2i}=\left[W_{2i1},\cdots,W_{2iL}\right]^{\mathrm{T}}\in\square^{L}$, L represent the number of neurons in the

second neural network hidden layer; $\boldsymbol{\chi}_i = \left[1, y_i, \dot{y}_i, \ddot{y}_i\right]^{\mathrm{T}}$ is the input signal to the first neural network; $\boldsymbol{\varepsilon}_i = \left[\varepsilon_{i1}, \cdots, \varepsilon_{in}\right]^{\mathrm{T}} \in \square^{n}$ represents the function reconstruction error; $\boldsymbol{h}_{1i} = \left[h_{1i1}, \cdots, h_{1iH}\right]^{\mathrm{T}} \in \square^{H}$ and $\boldsymbol{h}_{2i} = \left[h_{2i1}, \cdots, h_{2iL}\right]^{\mathrm{T}} \in \square^{L}$ are the Gaussian radial basis functions for the two different neural networks, expressed as follows:

$$h_{1ij} = \exp\left(-\frac{\left\|\boldsymbol{\chi}_i - \sigma_{1j} I_4\right\|^2}{2\iota_{1j}}\right), \begin{cases} i = 1, 2, \cdots, n \\ j = 1, 2, \cdots, H \end{cases} \tag{18}$$

$$h_{2ij} = \exp\left(-\frac{\left(y_i - \sigma_{2j}\right)^2}{2\iota_{2j}}\right), \begin{cases} i = 1, 2, \cdots, n \\ j = 1, 2, \cdots, L \end{cases} \tag{19}$$

where $\boldsymbol{\sigma}_1 = \left[\sigma_{1j}, \cdots, \sigma_{1H}\right]^{\mathrm{T}}$ and $\boldsymbol{\sigma}_2 = \left[\sigma_{2j}, \cdots, \sigma_{2L}\right]^{\mathrm{T}}$ represent acceptable center ranges; $\boldsymbol{\iota}_1 = \left[\iota_{1j}, \cdots, \iota_{1H}\right]^{\mathrm{T}}$ and $\iota_2 = \left[\iota_{2j}, \cdots, \iota_{2L}\right]^{\mathrm{T}}$ represent the widths of the Gaussian functions.

Assumption 5: The function reconstruction error $\boldsymbol{\varepsilon}_i = \left[\varepsilon_{i1}, \cdots, \varepsilon_{in}\right]^{\mathrm{T}}$ and its second derivative with respect to time are bounded:

$$\begin{cases} \left\|\boldsymbol{\varepsilon}_1\right\| \le \varepsilon_{1d1}, \ \left\|\dot{\boldsymbol{\varepsilon}}_1\right\| \le \varepsilon_{1d2}, \ \left\|\ddot{\boldsymbol{\varepsilon}}_1\right\| \le \varepsilon_{1d3} \\ \left\|\boldsymbol{\varepsilon}_2\right\| \le \varepsilon_{2d1}, \ \left\|\dot{\boldsymbol{\varepsilon}}_2\right\| \le \varepsilon_{2d2}, \ \left\|\ddot{\boldsymbol{\varepsilon}}_2\right\| \le \varepsilon_{2d3} \end{cases} \tag{20}$$

where ε_{id1}, ε_{id2}, and ε_{id3} are positive constants.

Considering that increasing the number of neural network hidden layers and manipulator joints makes tuning the weight matrices $\boldsymbol{W}_{1i}$ and $\boldsymbol{W}_{2i}$ extremely difficult, this study introduces an adaptive method for self-tuning the weight matrices. Based on (17) and (18), their adaptive approximation expressions can be obtained as follows:

$$\hat{\boldsymbol{\Delta}}_{t1} = \left[\hat{\boldsymbol{W}}_{11}^{\mathrm{T}} \boldsymbol{h}_{11}\left(\boldsymbol{\chi}_1\right), \cdots, \hat{\boldsymbol{W}}_{1n}^{\mathrm{T}} \boldsymbol{h}_{1n}\left(\boldsymbol{\chi}_n\right)\right]^{\mathrm{T}} \tag{21}$$

$$\hat{\boldsymbol{\Delta}}_{t2} = \left[\hat{\boldsymbol{W}}_{21}^{\mathrm{T}} \boldsymbol{h}_{21}\left(y_1\right), \cdots, \hat{\boldsymbol{W}}_{2n}^{\mathrm{T}} \boldsymbol{h}_{2n}\left(y_n\right)\right]^{\mathrm{T}} \tag{22}$$

where $\hat{\boldsymbol{W}}_{1i} = \left[\hat{W}_{1i1}, \cdots, \hat{W}_{1iH}\right]^{\mathrm{T}} \in \square^{H}$ and $\hat{\boldsymbol{W}}_{2i} = \left[\hat{W}_{2i1}, \cdots, \hat{W}_{2iL}\right]^{\mathrm{T}} \in \square^{L}$ represent the estimated values of the ideal weights.

The adaptive laws for the neural network weights are as follows:

$$\begin{cases} \dot{\hat{\boldsymbol{W}}}_{1i} = \mathrm{sat}_{\hat{\boldsymbol{W}}_{1iM}}\left(\mathrm{Proj}_{\hat{\boldsymbol{W}}_{1i}}\left(\boldsymbol{\Gamma}_{1i}\boldsymbol{\tau}_{1i}\right)\right) \\ \dot{\hat{\boldsymbol{W}}}_{2i} = \mathrm{sat}_{\hat{\boldsymbol{W}}_{2iM}}\left(\mathrm{Proj}_{\hat{\boldsymbol{W}}_{2i}}\left(\boldsymbol{\Gamma}_{2i}\boldsymbol{\tau}_{2i}\right)\right) \end{cases} \tag{23}$$

$$\begin{cases} \boldsymbol{\tau}_{1i} = \boldsymbol{h}_{1i} z_{2i} \\ \boldsymbol{\tau}_{2i} = \boldsymbol{h}_{2i} z_{3i} \end{cases} \tag{24}$$

where $\boldsymbol{\Gamma}$ is a positive definite matrix determining the adaptive rate of the weights; $\boldsymbol{\tau}$ is defined as the adaptive function for the neural network weights; other function definitions are as shown in [17].

Define the estimation errors as follows:

$$\tilde{\bullet} = \bullet - \hat{\bullet} \tag{25}$$

According to (22), define the following variables:

$$\tilde{\boldsymbol{\Delta}}_{t1} = \left[\tilde{\boldsymbol{W}}_{11}^{\mathrm{T}} \boldsymbol{h}_{11}\left(\boldsymbol{\chi}_1\right), \cdots, \tilde{\boldsymbol{W}}_{1n}^{\mathrm{T}} \boldsymbol{h}_{1n}\left(\boldsymbol{\chi}_n\right)\right]^{\mathrm{T}} \tag{26}$$

$$\tilde{\boldsymbol{\Delta}}_{tw1}=\left[\dot{\tilde{\boldsymbol{W}}}_{11}^{\mathrm{T}}\boldsymbol{h}_{11}(\boldsymbol{\chi}_1),\cdots,\dot{\tilde{\boldsymbol{W}}}_{1n}^{\mathrm{T}}\boldsymbol{h}_{1n}(\boldsymbol{\chi}_n)\right]^{\mathrm{T}} \tag{27}$$

$$\tilde{\boldsymbol{\Delta}}_{th1}=\left[\tilde{\boldsymbol{W}}_{11}^{\mathrm{T}}\dot{\boldsymbol{h}}_{11}(\boldsymbol{\chi}_1),\cdots,\tilde{\boldsymbol{W}}_{1n}^{\mathrm{T}}\dot{\boldsymbol{h}}_{1n}(\boldsymbol{\chi}_n)\right]^{\mathrm{T}} \tag{28}$$

To simplify the proof process, the transformed form of (23) follows the same principle.

4. CONTROLLER DESIGN AND STABILITY ANALYSIS

4.1. Nonlinear Integral Robust Controller Design

Define the error variable vectors as follows:

$$\begin{cases} \boldsymbol{z}_1=\boldsymbol{x}_1-\boldsymbol{x}_{1d} \\ \boldsymbol{z}_2=\dot{\boldsymbol{z}}_1+\boldsymbol{k}_1\boldsymbol{z}_1 \\ \boldsymbol{z}_3=\boldsymbol{x}_3-\boldsymbol{\alpha}_2 \\ \boldsymbol{r}_1=\dot{\boldsymbol{z}}_2+\boldsymbol{k}_2\boldsymbol{z}_2 \\ \boldsymbol{r}_2=\dot{\boldsymbol{z}}_3+\boldsymbol{k}_3\boldsymbol{z}_3 \end{cases} \tag{29}$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are auxiliary error variable vectors, which can help suppress matched and unmatched uncertainties during controller design; $\boldsymbol{k}_1=\mathrm{diag}\{k_{11},k_{12},\cdots,k_{1n}\}$, $\boldsymbol{k}_2=\mathrm{diag}\{k_{21},k_{22},\cdots,k_{2n}\}$, $\boldsymbol{k}_3=\mathrm{diag}\{k_{31},k_{32},\cdots,k_{3n}\}$ represent tunable controller gain matrices.

Based on the above equation, the virtual control law https://media/image111.wmf can be designed as follows:

$$\begin{aligned} \boldsymbol{\alpha}_2&=\boldsymbol{S}^{-1}(\boldsymbol{x}_1)\left[\boldsymbol{\alpha}_{2a}+\boldsymbol{\alpha}_{2s1}+\boldsymbol{\alpha}_{2s2}\right] \\ \boldsymbol{\alpha}_{2a}&=\boldsymbol{G}(\boldsymbol{x}_1)+\boldsymbol{C}(\boldsymbol{x}_1,\boldsymbol{x}_2)\boldsymbol{x}_2-\hat{\boldsymbol{\Delta}}_{t1} \\ &\quad+\boldsymbol{M}(\boldsymbol{x}_1)\ddot{\boldsymbol{x}}_{1d} \\ \boldsymbol{\alpha}_{2s1}&=-(\boldsymbol{k}_{r1}+\boldsymbol{k}_s)\boldsymbol{z}_2-\boldsymbol{M}(\boldsymbol{x}_1)\boldsymbol{k}_1\dot{\boldsymbol{z}}_1-\boldsymbol{M}(\boldsymbol{x}_1)\boldsymbol{k}_2\boldsymbol{z}_2 \\ \boldsymbol{\alpha}_{2s2}&=-\int_0^t\left[(\boldsymbol{k}_{r1}+\boldsymbol{k}_s)\boldsymbol{k}_2\boldsymbol{z}_2+\boldsymbol{v}(\boldsymbol{z}_2)\boldsymbol{\beta}_1\right]\mathrm{d}\tau \end{aligned} \tag{30}$$

where $\mathrm{sign}(*)$ represents the sign function; the sign function matrix $\boldsymbol{v}(\boldsymbol{z}_2)$ is $\boldsymbol{v}(\boldsymbol{z}_2)=\mathrm{diag}\{\mathrm{sign}(z_{21}),\cdots,\mathrm{sign}(z_{2n})\}$; the control gain is $\boldsymbol{k}_{r1}=\mathrm{diag}\{k_{r11},k_{r12},\cdots,k_{r1n}\}$, and all elements in the matrix are positive; $\boldsymbol{\beta}_1=[\beta_{11},\beta_{12},\cdots,\beta_{1n}]^{\mathrm{T}}$ is a gain vector with all positive elements; $\boldsymbol{k}_s=\mathrm{diag}\{k_{s1},k_{s2},\cdots,k_{sn}\}$ is an auxiliary constant variable matrix with all elements being real numbers greater than 0.

The expanded form of the auxiliary error vector $\boldsymbol{r}_1$ is:

$$\boldsymbol{M}(\boldsymbol{x}_1)\dot{\boldsymbol{r}}_1=\boldsymbol{S}(\boldsymbol{x}_1)\dot{\boldsymbol{z}}_3-(\boldsymbol{k}_{r1}+\boldsymbol{k}_s)\boldsymbol{r}_1-\boldsymbol{v}(\boldsymbol{z}_2)\boldsymbol{\beta}_1+\boldsymbol{N}_1 \tag{31}$$

From (31), it is known that $\boldsymbol{N}_1$ is bounded. The bound of $\boldsymbol{N}_1$ can be expressed as:

$$\|\boldsymbol{N}_1\|\le\delta_1,\ \ \|\dot{\boldsymbol{N}}_1\|\le\delta_2 \tag{32}$$

where δ_1 and δ_2 are constants greater than 0.

Based on the expansion of the auxiliary error function $\boldsymbol{r}_2$, the actual control law can be designed as follows:

$$\begin{aligned}\boldsymbol{u} &= \boldsymbol{f}_1^{-1}\left(\boldsymbol{u}_a + \boldsymbol{u}_{s1} + \boldsymbol{u}_{s2}\right)\\ \boldsymbol{u}_a &= \boldsymbol{f}_2 + \boldsymbol{f}_3 + \dot{\boldsymbol{\alpha}}_2 - \hat{\boldsymbol{\Delta}}_{t2}\\ \boldsymbol{u}_{s1} &= -\boldsymbol{k}_{r2}\boldsymbol{z}_3 - \boldsymbol{k}_3\boldsymbol{z}_3\\ \boldsymbol{u}_{s2} &= -\int_0^t\left[\boldsymbol{k}_{r2}\boldsymbol{k}_3\boldsymbol{z}_3 + \boldsymbol{v}\left(z_3\right)\boldsymbol{\beta}_2\right]\mathrm{d}\tau\end{aligned} \tag{33}$$

where $\boldsymbol{\beta}_2 = \left[\beta_{21}, \beta_{22}, \cdots, \beta_{2n}\right]^{\mathrm{T}}$ is a gain vector; $\boldsymbol{u}_a$ is the model compensation term; $\boldsymbol{u}_{s1}$ represents the linear feedback term for system tuning; the sign function matrix $\boldsymbol{v}\left(z_3\right) = \mathrm{diag}\left\{\mathrm{sign}\left(z_{31}\right), \cdots, \mathrm{sign}\left(z_{3n}\right)\right\}$; $\boldsymbol{u}_{s2}$ is the system's RISE feedback term, which can suppress the influence of matched uncertainties on system control performance to some extent.

Combining the specific form of the actual control law in (42), the time derivative of $\boldsymbol{r}_2$ can be expanded as:

$$\dot{\boldsymbol{r}}_2 = -\boldsymbol{k}_{r2}\boldsymbol{r}_2 - \boldsymbol{v}\left(z_3\right)\boldsymbol{\beta}_2 + \boldsymbol{N}_2 \tag{34}$$

From (34), it is known that $\boldsymbol{N}_2$ is bounded. The bound of $\boldsymbol{N}_2$ can be expressed as:

$$\left\|\boldsymbol{N}_2\right\| \le \xi_1, \quad \left\|\dot{\boldsymbol{N}}_2\right\| \le \xi_2 \tag{35}$$

where δ_1 and δ_2 are constants greater than 0.

4.2. Controller Stability Analysis

To analyze the controller's performance and stability, the following theorems are given.

Theorem 1: There always exist real vectors $\boldsymbol{\beta}_1 = \left[\beta_{11}, \beta_{12}, \cdots, \beta_{1n}\right]^{\mathrm{T}}$ and $\boldsymbol{\beta}_2 = \left[\beta_{21}, \beta_{22}, \cdots, \beta_{2n}\right]^{\mathrm{T}}$, whose elements always satisfy the following conditions:

$$\beta_{1i} \ge \delta_{1i} + \frac{1}{k_{2i}}\delta_{2i} \tag{36}$$

$$\beta_{2i} \ge \xi_{1i} + \frac{1}{k_{3i}}\xi_{2i} \tag{37}$$

Here, the functions $P_1(t)$ and $P_2(t)$ are defined as:

$$P_1(t) = \boldsymbol{\beta}_1^{\mathrm{T}}\left|z_2(0)\right| - \boldsymbol{z}_2^{\mathrm{T}}(0)\boldsymbol{N}_1(0) - \int_0^t L_1(\tau)\mathrm{d}\tau \tag{38}$$

$$P_2(t) = \boldsymbol{\beta}_2^{\mathrm{T}}\left|z_3(0)\right| - \boldsymbol{z}_3^{\mathrm{T}}(0)\boldsymbol{N}_2(0) - \int_0^t L_2(\tau)\mathrm{d}\tau \tag{39}$$

$$L_1(t) = \boldsymbol{r}_1^{\mathrm{T}}\left[\boldsymbol{N}_2 - \boldsymbol{v}\left(z_2\right)\boldsymbol{\beta}_1\right] \tag{40}$$

$$L_2(t) = \boldsymbol{r}_2^{\mathrm{T}}\left[\boldsymbol{N}_1 - \boldsymbol{v}\left(z_3\right)\boldsymbol{\beta}_2\right] \tag{41}$$

From the above equations, it can be concluded that the functions $P_1(t)$ and $P_2(t)$ are always positive definite.

Theorem 2: By selecting sufficiently large and appropriate controller gain matrices $\boldsymbol{k}_1$, $\boldsymbol{k}_2$, $\boldsymbol{k}_3$, $\boldsymbol{k}_{\mathrm{r1}}$, and $\boldsymbol{k}_{\mathrm{r2}}$, the following matrix:

$$\Lambda = \begin{bmatrix} \boldsymbol{k}_1 & -\frac{\mathbf{I}_{6\times6}}{2} & \mathbf{0}_{6\times6} & \mathbf{0}_{6\times6} & \mathbf{0}_{6\times6} \\ -\frac{\mathbf{I}_{6\times6}}{2} & \boldsymbol{k}_2 & -\frac{\mathbf{I}_{6\times6}}{2} & \mathbf{0}_{6\times6} & \mathbf{0}_{6\times6} \\ \mathbf{0}_{6\times6} & -\frac{\mathbf{I}_{6\times6}}{2} & \boldsymbol{k}_{\mathrm{r1}} & \mathbf{0}_{6\times6} & \mathbf{0}_{6\times6} \\ \mathbf{0}_{6\times6} & \mathbf{0}_{6\times6} & \mathbf{0}_{6\times6} & \boldsymbol{k}_3 & -\frac{\mathbf{I}_{6\times6}}{2} \\ \mathbf{0}_{6\times6} & \mathbf{0}_{6\times6} & \mathbf{0}_{6\times6} & -\frac{\mathbf{I}_{6\times6}}{2} & \boldsymbol{k}_{\mathrm{r2}} \end{bmatrix} \tag{42}$$

is always positive definite. Then, the designed controller is semi-globally asymptotically stable and ensures that all signals in the closed-loop system are bounded, i.e., $\|z_1\|_\infty = \max\left(|z_{11}|,|z_{12}|,\cdots,|z_{1n}|\right) \to 0$, $t \to \infty$.
The proof process can be referred to in [17].

5. SIMULATION RESULTS

This section verifies the advancement and effectiveness of the designed multidimensional neural network-based integral robust controller for hydraulic manipulators through a simulation example of a three-DOF hydraulic manipulator wrist system. The system parameters are consistent with those in [18]. The structure of the three-DOF hydraulic manipulator wrist system is shown in Figure 3.

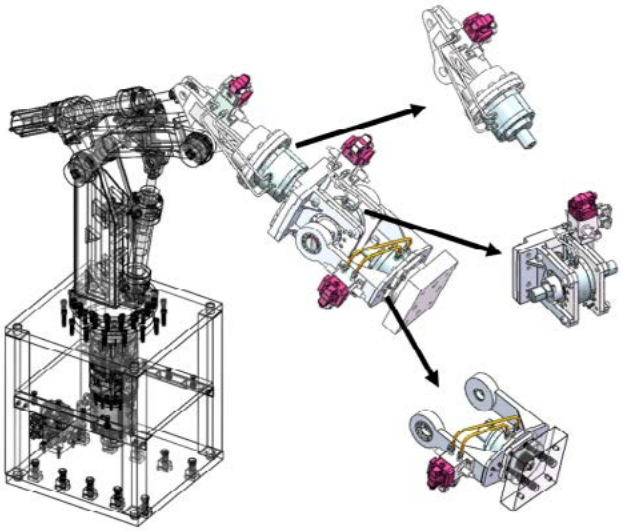


Figure 3 Schematic Diagram of The Hydraulic Manipulator Wrist Structure

To verify the fitting performance of the neural network, matched and unmatched uncertainties are designed as follows:

$$\begin{cases} \boldsymbol{\Delta}_{t1} = \dot{\boldsymbol{x}}_{1d} + \begin{bmatrix} 50 \\ 50 \\ 50 \end{bmatrix} \\ \boldsymbol{\Delta}_{t2} = \begin{bmatrix} 2\times10^4 \sin(0.5t) \\ 1\times10^4 \sin(0.5t) \\ 1\times10^4 \sin(0.5t) \end{bmatrix} \end{cases} \tag{43}$$

To compare the differences in simulation results under different controllers, the controller parameters are set as follows:

(1) Multidimensional Neural Network-Based Integral Robust Controller (C1): This is the controller proposed in this study. Its gains are: $\boldsymbol{k}_1 = \mathrm{diag}\{60,20,20\}$, $\boldsymbol{k}_2 = \mathrm{diag}\{500,800,500\}$, $\boldsymbol{k}_3 = \mathrm{diag}\{60,80,80\}$, $\boldsymbol{k}_{r1} = \mathrm{diag}\{4,5,4\}$, $\boldsymbol{k}_{r2} = \mathrm{diag}\{10,5,10\}$, $\boldsymbol{k}_{\mathrm{s}} = \mathrm{diag}\{10,5,10\}$, $\boldsymbol{\beta}_1 = [8,2,1]^{\mathrm{T}}$, $\boldsymbol{\beta}_2 = [1000,100,150]^{\mathrm{T}}$, $H = 16$, $L = 12$, $\Gamma_{1i} = 25$, $\Gamma_{2i} = 20$.

(2) Feedback Linearization Controller (C2): This is a feedback linearization pose controller. Apart from the parameters not included, other controller gains remain the same as in C1.

The initial tracking command signals for each joint are:

$$\boldsymbol{x}_{1d} = \begin{bmatrix} 0.25\sin(0.5t)\times\left(1-e^{-0.01t^3}\right) \\ 0.20\sin(0.5t)\times\left(1-e^{-0.01t^3}\right) \\ 0.25\sin(0.5t)\times\left(1-e^{-0.01t^3}\right) \end{bmatrix} \tag{44}$$

The simulation results are shown in the following figures. Figure 4 shows the tracking control performance of each joint of the hydraulic manipulator wrist under different controllers. Figure 5 displays the tracking error results for Joint 1 under the two controllers. Figure 6 compares the tracking errors for Joint 2 under different controllers. Figure 7 shows the errors for Joint 3 under the two controllers.

From Figure 4, it can be observed that both controllers, C1 and C2, enable each joint to track the command signal (46) well, with no significant visual difference. However, Figures 5 to 7 more clearly demonstrate the tracking performance of the two controllers. Since the command signal exhibits periodic fluctuations, significant error fluctuations are observed during each hydraulic motor reversal due to the influence of unmodeled errors such as friction and disturbances. Controller C2 is a classic feedback linearization controller. However, lacking disturbance approximation and compensation capabilities, its tracking performance is noticeably inferior to that of the C1 neural network integral robust controller during simulation startup and reversal phases. This verifies the effectiveness of the proposed control strategy.

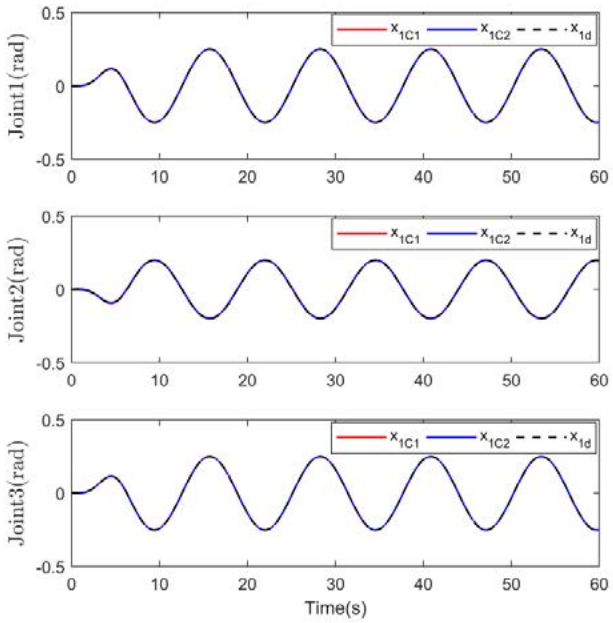


Figure 4 Tracking Effect Diagrams of Each Joint

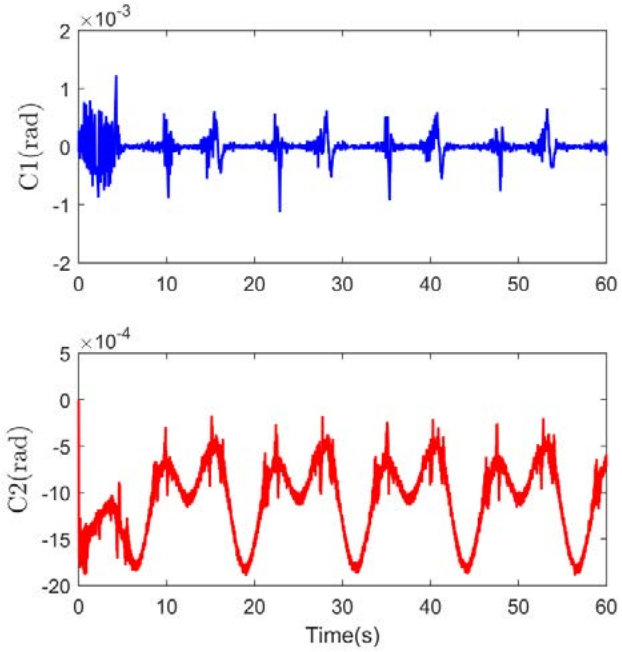


Figure 5 Comparison of Tracking Errors of Joint 1 Under Different Controllers

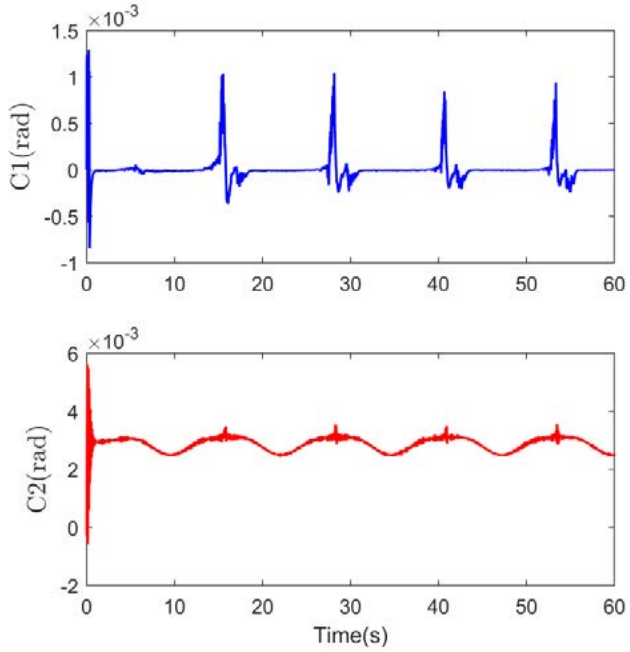


Figure 6 Comparison of Tracking Errors of Joint 2 Under Different Controllers

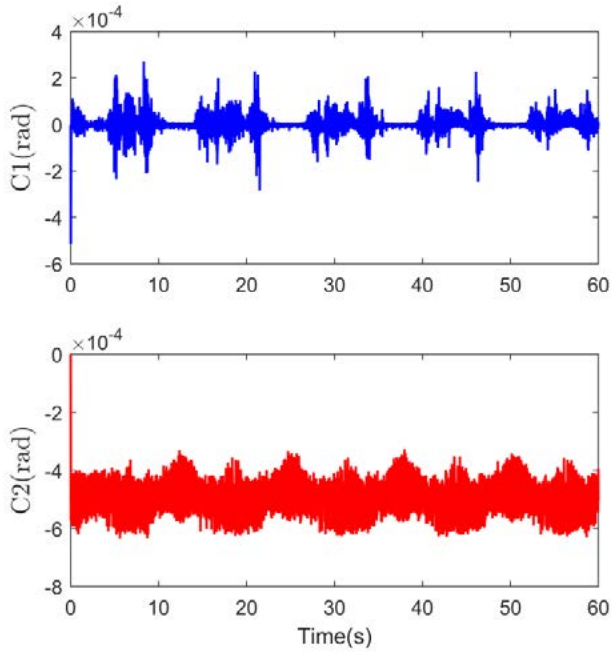


Figure 7 Comparison of Tracking Errors of Joint 3 Under Different Controllers

6. CONCLUSION

With the increasing demand for automation and intelligence in heavy-load operations, higher requirements are placed on the comprehensive control performance of hydraulically driven multi-DOF manipulators. Addressing major problems such as large model errors, high operational risks, and poor pose positioning accuracy in heavy-load hydraulic manipulator control, this study proposes an integral robust control strategy integrated with neural networks, achieving the approximation and compensation of unmodeled uncertainties. The main conclusions are as follows:

This study provides a robust and dynamically compensated solution for high-precision, high-reliability pose tracking control of heavy-load multi-DOF hydraulic manipulators, offering a new approach for high-performance controller design in complex scenarios.

7. REFERENCES

[1] GUO Qing. Development of Nonlinear Control Technology for Electro-hydraulic Servo System [J]. Chinese Hydraulics & Pneumatics,2018,(3): 1-9.
[2] Merritt H E. Hydraulic control systems[M]. New York: Wiley,1967.
[3] YAO Jianyong, JIAO Zongxia, MA Dawei, et al. High-accuracy tracking control of hydraulic rotary actuators with modeling uncertainties[J]. IEEE/ASME Transactions on Mechatronics, 2014, 19(2): 633-641.
[4] WANG Tianlei, WANG Chenxu, XIN Zengmiao, et al. Research on Reinforcement Learning-based Sliding Mode Control for Hydraulic Position Servo Systems[J]. Chinese Hydraulics & Pneumatics, 2025, 49(11):88-97.
[5] FU Xiang, WANG Ranfeng, ZHAO Yangsheng. Self-adaptive control method of fluid feeding with steady pressure for hydraulic support based on overlapping synergetic logic[J]. Journal of China Coal Society, 2020, 45(5):15-25.
[6] WANG Chengwen, JIAO Zongxia, WU Shuai, et al. Nonlinear adaptive torque control of electro-hydraulic load system with external active motion disturbance[J]. Mechatronics, 2014,24(1): 32-40.
[7] LIU Haodong, LI Shizhen, SHI Wenzhuo. Model Predictive Control of Offshore Gangway Hydraulic System Based on Extended State Observer[J]. Chinese Hydraulics & Pneumatics, 2024, 48(9): 22-29.
[8] MA Yanbin, ZHAO Bin, HAO Yunxiao. Characteristics Analysis of Asymmetric Pump-controlled Single-rod Hydraulic Cylinder System[J]. Chinese Hydraulics & Pneumatics, 2023,47(1): 70-76.
[9] JI Hong, ZHAO Hongke, LIU Yaozhuo, et al. Backflushing Jet and Flow Loss in Distribution Valve of Water Piston Pump [J]. Chinese Hydraulics & Pneumatics, 2022, 46(11): 1-8.
[10] Xia Yangxiu, Qi Manzhi, et al. Advanced Motion Control of Hydraulic Manipulator With Precise Compensation of Dynamic Friction[J]. IEEE Transactions on Industrial Informatics, 2024, 20(7): 9375-9384.
[11] Zhang Han, Liu Yuhan, Zhan Liaoyang, et al. A Human–Machine Cooperative Control Strategy Based on Deep Reinforcement Learning to Enhance Heavy Vehicle Driving Safety[J]. IEEE Transactions on Human-Machine Systems, 2025, 55(6): 1006-1015.

[12] ZHANG Jiarui, GE Lei, HAO Yunxiao. Discrete PID Control of Closed Pump Control Hydraulic-mechanical Linear Actuator[J]. Chinese Hydraulics & Pneumatics, 2024,48(9): 54-60.
[13] YAO Zhikai, LIANG Xianglong, JIANG Guoping, et al. Model-Based Reinforcement Learning Control of Electrohydraulic Position Servo Systems[J]. IEEE/ASME Transactions on Mechatronics, 2023, 28(3): 1446-1455.
[14] ZHANG Luyao, WANG Haolin, XIA Wanrun, et al. Scaled Gain-Based Extended State Observer for Disturbed Systems With Time-Varying Delays[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2026, 73(1): 68-72.
[15] ZHU Yukai, QIAO Jianzhong, GUO Lei. Adaptive sliding mode disturbance observer-based composite control with prescribed performance of space manipulators for target capturing[J]. IEEE Transactions on Industrial Electronics, 2020,67(4): 3083-3093.
[16] ZHENG Yuyan, MI Junjie, SHEN Gang. Dual RISE Control for Multi-Degree-of-Freedom Hydraulic Manipulator with Prescribed Performance Function[J]. Chinese Hydraulics & Pneumatics, 2025, 49(6):102-111.
[17] DENG Wenxiang, ZHOU Hang, ZHOU Jin, et al. Neural Network-Based Adaptive Asymptotic Prescribed Performance Tracking Control of Hydraulic Manipulators. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(1):285-295.
[18] MI Junjie, DENG Wenxiang, YAO Jianyong,et al. Linear-Extended-State-Observer-Based Adaptive RISE Control for the Wrist Joints of Manipulators with Electro-Hydraulic Servo Systems[J]. Electronics, 2024, 13(6).

Biographies



Junjie Mi received the B.Tech degree in vehicle engineering from the College of Economics and Technology, Anhui Agricultural University, Hefei, China, in 2017, and the Ph.D. degree in mechatronic control from the Nanjing University of Science and Technology, Nanjing, China, in 2024. In 2024, he joined the School of Mechatronics Engineering, Anhui University of Science and Technology, Huainan, China, as an Lecturer. In 2025, he joined the State Key Laboratory of Digital Intelligent Technology for Unmanned Coal Mining, Huainan, China. His current research interests include electro-hydraulic servo control, hydraulic manipulators, and disturbance rejection control.



Gang Shen received the M.Sc. and Ph.D. degrees in mechanical engineering from the Harbin Institute of Technology, Harbin, China, in 2007 and 2011, respectively. In 2023, he joined the School of Mechatronics Engineering, Anhui University of Science and Technology, Huainan, China. In 2025, he joined the State Key Laboratory of Digital Intelligent Technology for Unmanned Coal Mining, Huainan, China. He is currently working as a Full Professor**.** Dr. Shen current research interests include servo control of mechatronic systems, parallel robot, and adaptive control.



Jianyong Yao received the B. Tech degree in mechatronics from Tianjin University, Tianjin, China, in 2006, and the Ph.D. degree in mechatronic engineering from Beihang University, Beijing, China, in 2012. He was a visiting exchange student with the School of Mechanical Engineering, Purdue University, from Oct. 2010 to Oct. 2011. In 2012, he joined the School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China, and currently as a Full Professor. Dr. Yao current research interests include servo control of mechatronic systems, adaptive and robust control, fault detection and accommodation of dynamic systems.

Determination of Gas Solubility for Technical and Flammable Fluids

Marius Hofmeister, Lukas Boden, Andris Rambaks, Katharina Schmitz

Institute for Fluid Power Drives and Systems, marius.hofmeister@ifas.rwth-aachen.de

Institute for Fluid Power Drives and Systems, lukas.boden@ifas.rwth-aachen.de

Institute for Fluid Power Drives and Systems, andris.rambaks@ifas.rwth-aachen.de

Institute for Fluid Power Drives and Systems, katharina.schmitz@ifas.rwth-aachen.de

Abstract

Dissolved gas is a critical factor in many technical systems, since it influences fluid properties like compressibility, viscosity or thermal conductivity in a negative manner. Cavitation due to dissolved gas can lead to damage of components and increased wear in hydraulic applications. Therefore, a precise understanding of the dissolving and release processes of gases in liquids is essential for efficient and reliable hydraulic systems.

Despite its relevance for technical applications, literature on the gas solubility of hydraulic oils and lubricants is sparse. This is particularly true for data above ambient pressure at pressures relevant for hydraulic application. To close this gap, a measurement set up using volumetric-manometric approach was designed, which can be used to determine the gas solubility of hydraulic oils and other technical liquids. For the investigation of flammable or hazardous liquid/gas combinations an indirect measurement method was developed. In this contribution the design of the measurement set up is presented as well as results for one hydraulic oil in combination with nitrogen. Based on the obtained results, it is evident that substantial amounts of gas may be released during the operation of hydraulic systems.

1. INTRODUCTION

The ability of a liquid to dissolve gas is of great importance for a variety of technical applications. For example, the formation of bubbles in hydraulic systems is influenced by the amount of dissolved gas in the fluid. Bubbles are generally undesirable in hydraulic systems as they can lead to cavitation damage and have a negative impact on fluid properties such as compressibility or viscosity. Therefore, a precise understanding of the dissolving and release processes of gases in liquids is essential for efficient and reliable hydraulic systems [1, 2].

The same applies to cooling systems where bubbles and foam hinder heat transfer between liquid and environment. This becomes increasingly important in terms of the ongoing growth of e-mobility, as efficient temperature management is a prerequisite for e-motors with high power density and battery systems with large storage capacities. Gas solubility

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

also plays a crucial role for combustion engines. Gaseous fuel or combustion gases can dissolve in the lubricating film on the cylinder wall and be drawn in via the piston and thus accumulate in the engine oil. This is particularly important for alternative fuels such as ammonia or hydrogen, which are either toxic or highly explosive [3].

Although gas solubility is extremely relevant for many technical applications as described, gas solubility values can only be found to a limited extent in the literature. Coefficients for gas solubility are mainly available for short-chain molecules and pure liquids. Especially, information on pressure dependent gas solubility is very sparse, since most measurements were performed at ambient pressure. Solubility data in the range of pressures relevant for hydraulic applications is primarily available for crude oil or bitumen, but not for lubricants and hydraulic oils [4].

For this reason, this article will examine the gas dissolving capacity of various hydraulic oils. To determine the gas solubility, several measurement approaches are available [5, 6]. Traditionally, these can be categorized into chemical and physical methods [5]. For analyzing gas solubility in liquids, physical methods are generally more appropriate [5]. These physical approaches can be further divided into manometric-volumetric, mass spectrometric, gas chromatographic, and radiometric techniques [5, 6]. Currently, there is no standardized procedure for determining the solubility of gases in liquids [5]. However, the literature provides some examples of experimental setups used to investigate gas solubility at elevated pressures [7, 8].

In this contribution a volumetric-manometric approach was utilized for the determination of gas solubility at different temperatures and pressures, which is based on preliminary work by Rambaks et al. [9] and has been extensively revised [3]. In the following, first the basic measuring principle is described. Then the design of the test bench and the measurement process are shown. A particular focus is on the determination of gas solubility for highly reactive or toxic liquid/gas mixtures using an indirect measurement method.

Finally, the measurement results for hydraulic oil will be presented. Where possible, the results are compared with data from the literature. Both pressure- and temperature-dependency of the Bunsen coefficient is evaluated by means of the Krichevsky Kasarnovsky equation and the Henry–Van't Hoff relationship. With help of the corresponding parameters the interpolation and extrapolation of the measured data is possible.

2. METHODOLOGY

In the literature, various definitions and measures are used to describe the ability of liquids to dissolve gases. Therefore, this work first provides a general overview of the relevant variables and definitions employed herein, followed by a description of the measurement procedure and its fundamental functionality. Then, the thermodynamic principles underlying the method are derived, and the applicability of the approach to hazardous or flammable gas–liquid systems is discussed.

2.1. Definition of gas solubility

The most common measures used to represent gas solubility are listed in Table 2.1. The most fundamental quantity is the concentration c_G, which expresses the amount of dissolved gas n_G per unit volume of liquid V_L, typically in $\mathrm{mol/m^3}$. Alternatively, the mole fraction x_G

represents the ratio of dissolved gas moles n_G to the total moles of all components in the liquid phase ($n_G + n_L$). This dimensionless quantity is widely employed in thermodynamic modeling. Henry's law constant k_H establishes a proportional relationship between the equilibrium gas partial pressure p and its concentration c_G in the liquid. It serves as a convenient measure to relate experimentally accessible pressures to dissolved amounts, particularly in dilute solutions where the law holds. The Bunsen coefficient is defined as the volume of gas V_G (referred to standard conditions, e.g. $T_0 = 298.15$ Kand $p_0 = 1$ bar) that dissolves in a unit volume of liquid V_L under a partial pressure p.

Table 2.1 Summary of definitions and interrelations of gas solubility parameters

	Concentration $[\mathrm{mol \cdot m^{-3}}]$	Mole fraction [–]	Henry constant $[\mathrm{Pa \cdot m^3 \cdot mol^{-1}}]$	Bunsen coefficient [–]
	$c_G = \frac{n_G}{V_L}$	$x_G = \frac{n_G}{n_G + n_L}$	$k_H = \frac{p}{c_G}$	$\alpha = \frac{V_G}{V_L}\frac{p_0}{p}$
$f(c_G)$	—	$x_G = c_G \frac{M_L}{\rho_L}$	$k_H = \frac{p}{c_G}$	$\alpha = c_G \frac{RT_0}{p_0}$
$f(x_G)$	$c_G = x_G \frac{\rho_L}{M_L}$	—	$k_H = \frac{M_L p}{\rho_L}\frac{x_G}{1 - x_G}$	$\alpha = x_G \frac{RT_0}{p_0}\frac{\rho_L}{M_L}$
$f(k_H)$	$c_G = \frac{p}{k_H}$	$x_G = \frac{pM_L}{k_H \rho_L + pM_L}$	—	$\alpha = \frac{1}{k_H}\frac{RT_0 p}{p_0}$
$f(\alpha)$	$c_G = \alpha \frac{p_0}{RT_0}$	$x_G = \alpha \frac{p_0}{RT_0}\frac{M_L}{\rho_L}$	$k_H = \frac{1}{\alpha}\frac{RT_0 p}{p_0}$	—

Since the Bunsen coefficient has historically been used to describe the gas solubility of hydraulic oils and many other technical liquids, this parameter will also be used in this paper. In addition, many oils are mixtures with largely unknown compositions, meaning that no molar mass M_L can be determined and the amount of substance n_L cannot be calculated for a given liquid volume V_L either. As shown in Table 2.1, the parameters used to describe the ability of liquids to dissolve gases can be converted into one another if necessary.

2.2. General determination of gas solubility according to the manometric-volumetric principle

In this study, the Bunsen coefficient is determined by using a volumetric-manometric method, in which different states of a defined gas volume in thermodynamic equilibrium are evaluated. The pressure in the gas phase is used to quantify the amount of gas that has dissolved in the liquid as illustrated in **Figure 2.1**. The gas and liquid phases are at the same temperature and pressure. In the initial state (0), the liquid contains no dissolved gas, whereas in state 1, a certain amount of gas has already dissolved in the liquid.

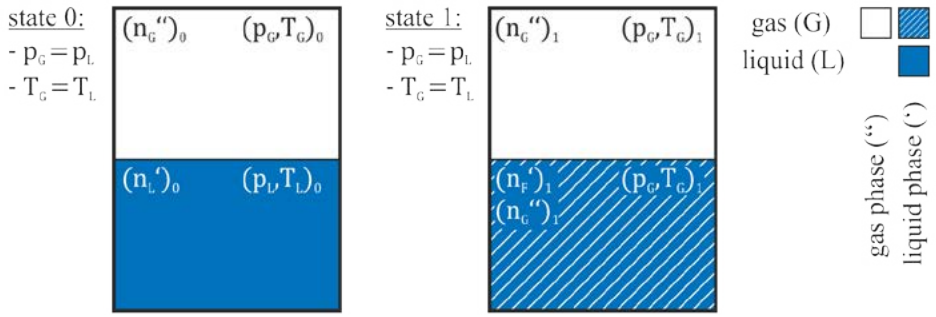


Figure 2.1. Ideal closed system

Since the example system is a closed system, the total amount of gas $(n_G)_i$ is constant and can be represented as the sum of the amount of dissolved gas $(n_G')_i$ and free gas in the gaseous phase $(n_G'')_i$ (2.1).

$$(n_G)_i = (n_G')_i + (n_G'')_i = \text{const.} \tag{2.1}$$

The amount of gas $(n_G'')_i$in the gaseous phase can be described using the real gas law, which is a function of temperature $(T_G'')_i$ and pressure $(p_G'')_i$ as well as the gas volume $(V_G'')_i$, the gas constant R and the compressibility factor $(z_G)_i$.

$$(n_G'')_i = \frac{1}{R}\left(\frac{p_G'' V_G''}{T_G'' z_G}\right)_i \tag{2.2}$$

The amount of dissolved gas $(n_G')_1$ in state 1 corresponds to the difference of the amount of gas in the gaseous phase in state 0 $(n_G'')_0$ and state 1 $(n_G'')_1$. The underlying assumption is that the amount of dissolved gas in state 0 $(n_G'')_0$ is equal to 0.

$$(n_G')_1 - \underbrace{(n_G')_0}_{\approx 0} = (n_G'')_0 - (n_G'')_1 \tag{2.3}$$

As shown in Table 2.1 Bunsen coefficient α_i is the ratio of dissolved gas volume at standard conditions $(V_G')_S$ and the liquid volume $(V_L')_S$. In accordance with the ideal gas law, the gas volume can be also expressed as a function of $(n_G')_S$.

$$\alpha_i = \frac{(V_G')_S}{(V_L')_S} = \frac{R(T)_S}{(V_L')_S (p'')_S}(n_G')_S \tag{2.4}$$

By combining equations (2.2), (2.3) and (2.4), the resulting expression (2.5) is obtained. Finally, the Bunsen absorption coefficient can be determined by comparing the state variables temperature and pressure in both states and by evaluating the volumes $(V_L')_i$ and $(V_G'')_i$.

$$\alpha_i = \frac{(T)_S}{(V_L')_S (p'')_S}\left[\left(\frac{p'' V_G''}{T''}\right)_0 - \left(\frac{p'' V_G''}{T''}\right)_1\right] \tag{2.5}$$

2.3. Measurement setup

For practical implementation of the measurement method described above and to determine the unknown parameters from equation (2.5), the test bench shown in **Figure 2.2** was used. The setup mainly consists of a cylinder which can be filled with liquid and gas. To determine the state variables p'' and T'', corresponding pressure and temperature sensors are located inside the cylinder. The pressure in the gas and liquid phase is equal, so that a single sensor is sufficient to obtain the required pressure data. Since gases and liquids generally heat up at different rates when pressure changes, two separate temperature sensors are used for the respective phases. The volume of the gas phase $(V_G'')_i$ can be calculated using equation (2.6). The piston position x required for this is measured by the installed LVDT sensor.

$$(V_G'')_i = (V_L')_i - [V_0 + A(x_0 - x)] \tag{2.6}$$

The initial volume of the cylinder V_0 was determined in advance by filling it with water. To do this, a container was filled with water and weighed with a precision scale both before and after pouring it into the cylinder. The accuracy of the scale is 0.01 g. The total volume of the cylinder V_0 can then be calculated in accordance with equation (2.7).

$$V_0 = \frac{(m'_{H2O})_0}{(\rho'_{H2O})_0}, \qquad (V_L')_0 = \frac{(m'_L)_0}{(\rho'_L)_0} \tag{2.7}$$

Before filling the test liquid, its mass is also weighed using a precision scale, so that the same relationship as in equation (2.7) can be applied to determine $(V_L')_i$. The required density values are either taken from the literature or determined according to standard DIN 12790 [10]. Changes of $(V_L')_i$ due to compression are neglectable for most liquids at low pressures and are considered for high pressures using a suitable density model.

Adjusting the piston changes the cylinder volume and, thus, the thermodynamic state of the system, enabling the comparison of different states according to equation (2.5).

After the filling process, the cylinder is sealed. To avoid contamination from dissolved gases and the ambient air above the liquid, the cylinder is connected to a vacuum pump. The pump allows the gas phase inside the cylinder to be evacuated, while the resulting pressure reduction promotes the release of any gases previously dissolved in the liquid. During evacuation, care must be taken to ensure that the pressure does not fall below the vapor pressure of the liquid.

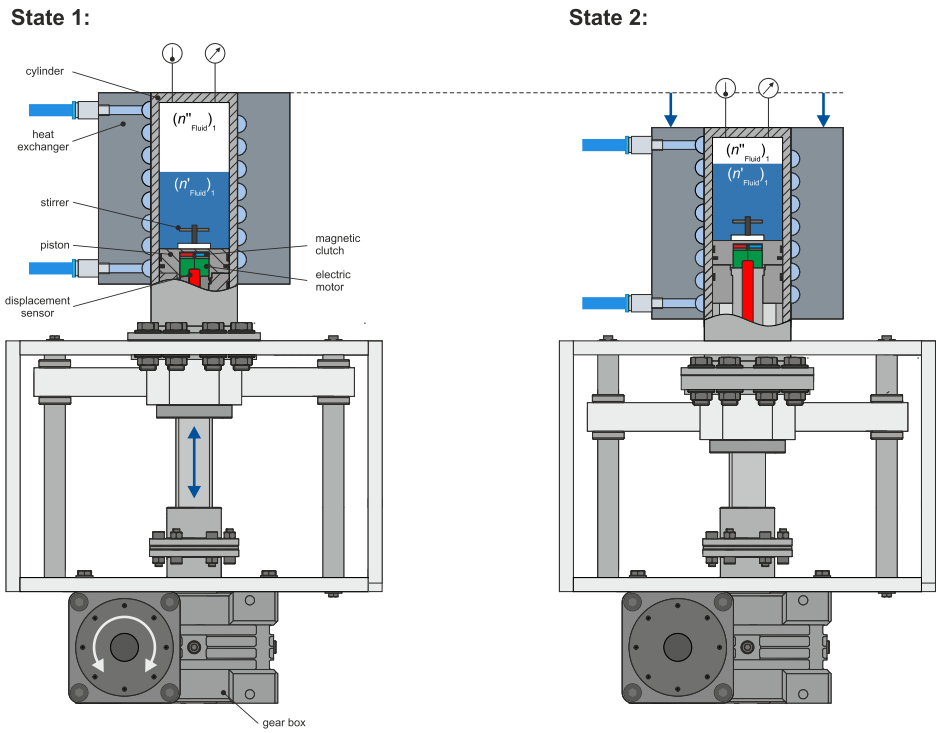


Figure 2.2. Bunsen test bench

After the evacuation of the cylinder, the chamber pressure and volume are adjusted using the piston position. A control loop is implemented to regulate the deviation between the setpoint and the actual pressure in the cylinder, which is translated into a motion of the stepper motor. The stepper motor is mechanically coupled to the piston via a gearbox. In the exemplary measurement shown in **Figure 2.3**, a chamber pressure of 18 bar was set for the first measuring point (1), corresponding to a piston position of approximately 38.8 mm. The increased pressure drives the system out of thermodynamic equilibrium, causing gas to start dissolving into the test liquid. To accelerate the diffusion process, the stirrer mounted on the piston is activated indicated by voltage U. If no change in pressure and temperature outside a small tolerance range can be detected, it is assumed that thermodynamic equilibrium has been reached again. To ensure that the measurement is not influenced by heat generated by the stirrer, it is turned off prior to data acquisition (voltage U: 2 V → 0 V). Thereafter, the quantities temperature T, pressure p, and piston displacement x are recorded for a duration of 10 minutes.

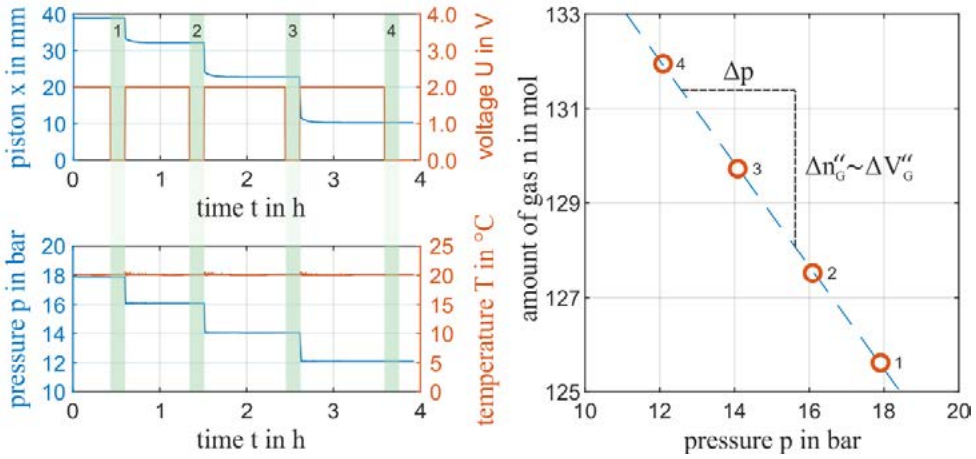


Figure 2.3. Example measurement trace showing the recorded signals

This procedure was repeated for three additional measurement points at 16, 14, and 12 bar. Using the mean values of the recorded quantities, a regression analysis based on equation (2.5) can be performed, as illustrated in the right panel in Figure 1. The slope of the resulting dashed line corresponds to the Bunsen coefficient at the given temperature of 20 °C.

To maintain a constant temperature level, a heat exchanger is mounted on the cylinder. The heat exchanger is connected to a tempering unit controlled by the measuring computer. With help of the temperature signals a closed control loop is feasible. The utilized material for the heat exchanger allows operating temperatures between -20 and +80 °C. In combination with the stepper motor, the described temperature control allows the automated measurement of pressure/temperature maps up to 400 bar. The lower pressure limit is primarily constrained by the vapour pressure of the investigated liquid; in principle, pressures below ambient conditions are attainable.

2.4. Indirect determination of gas solubility for flammable and hazardeous liquid/fuel combinations

The handling of certain mixtures of technical fluids and gases involves significant safety risks. Fuels, as well as hydraulic oils and other lubricants, can ignite upon contact with oxygen. In the context of carbon-free combustion, gaseous fuels such as hydrogen and ammonia, which can be dissolved in considerable amounts in the engine oil of internal combustion engines, are becoming increasingly relevant. In the case of hydrogen, there is a high risk of explosion due to its strong reactivity, whereas ammonia poses serious environmental and health hazards at elevated concentrations. To determine the Bunsen coefficient of such mixtures safely, an indirect measurement method was developed. [3]

Whether a potentially explosive mixture ignites depends on the composition of the gaseous phase. This relationship is illustrated in **Figure 2.4** for a hypothetical ternary system. The a-axis represents the volumetric fraction of vaporized oil or fuel in the gas phase, while the b- and c-axes indicate the respective fractions of nitrogen and oxygen. Mixtures located within the yellow explosion region can potentially ignite.

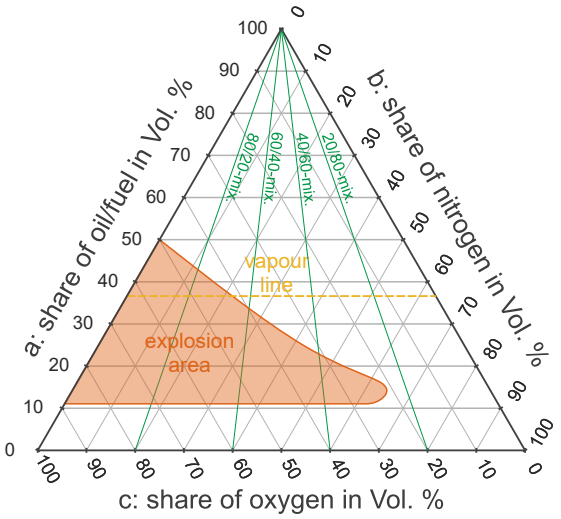


Figure 2.4. flammability diagram for hypothetical oil(fuel)/gas mixture

To estimate the explosion risk of a specific oil or fuel, the molar fraction x_i of the vaporized liquid in the gas phase is calculated according to Raoult's and Henry's law. According to equation (2.8), the molar fraction can be expressed as a function of the total pressure p and the vapor pressure p_i of the liquid.

$$p_{vap} = p \cdot x_i \iff x_{vap} = \frac{p_{vap}}{p} \tag{2.8}$$

Assuming an ideal gas behavior, the molar fraction x_i is equal to the volumetric fraction, and the vapor line (orange) can be plotted in Figure 2.4. For low pressures the volumetric fraction of the vapor increases, while at higher pressures it decreases.

The ratio of the remaining gases, oxygen and nitrogen, is indicated by the green lines in Figure 2.4 and corresponds to the original composition of the gas phase during the filling of the test bench. For an N_2/O_2 ratio of 20/80, the intersection with the vapor line falls within the explosion region. Increasing the fraction of nitrogen shifts the intersection to the right, so that, for example, for synthetic air (N_2/O_2 = 80/20) it lies outside the explosion region. By adjusting the composition of the gas phase in the cylinder, the risk of explosion can thus be eliminated. Similarly, this approach can be used to reduce the toxicity of hazardous gases such as ammonia by mixing them with inert gases.

With the presence of a gas mixture, the Bunsen coefficient of the individual pure components can no longer be determined directly using the method described in Section 2.3. Instead, the Bunsen coefficient of the mixture α_{mix} can be evaluated, based on the individual Bunsen coefficients of the constituent gases and their respective molar fractions (2.9).

$$\alpha_{mix} = \alpha_{N2} \cdot x_{N2} + \alpha_{O2} \cdot x_{O2} \iff \alpha_{O2} = \frac{\alpha_{mix} - \alpha_{N2} \cdot x_{N2}}{x_{O2}} \tag{2.9}$$

By rearranging (2.9), the Bunsen coefficient of oxygen α_{O2} can be determined for the example illustrated in Figure 2.4, based on the corresponding values of the mixture and of pure nitrogen ($\alpha_{mix}, \alpha_{O2}$).

3. RESULTS AND DISCUSSION

Measurements were carried out for hydraulic oil. The results for the temperature and pressure dependent Bunsen coefficient of a HLP 46 oil and nitrogen are shown in **Figure 3.1** [11]. The Bunsen coefficient was determined for pressures between 150 and 350 bar and temperatures ranging from 273.15 K to 323.15 K. The pressure was varied in 50 bar increments and the temperature in increments of 25 K. For each temperature and pressure step two measurements were conducted. The shown results correspond to the mean value obtained for all sensor signals recorded over a duration on 10 minutes with a frequency of 1 Hz. The left panel shows the Bunsen coefficient as a function of pressure, while the right panel illustrates the effect of temperature on gas solubility. The dashed lines indicate model data obtained from a regression analysis. The underlying equations of the analysis will be described later. The measured values for an identical measuring point in some cases differ greatly from one another. This deviation is particularly large for high pressures and low temperatures. Here, the measured Bunsen coefficient for identical boundary conditions differs by a value of 0.005.

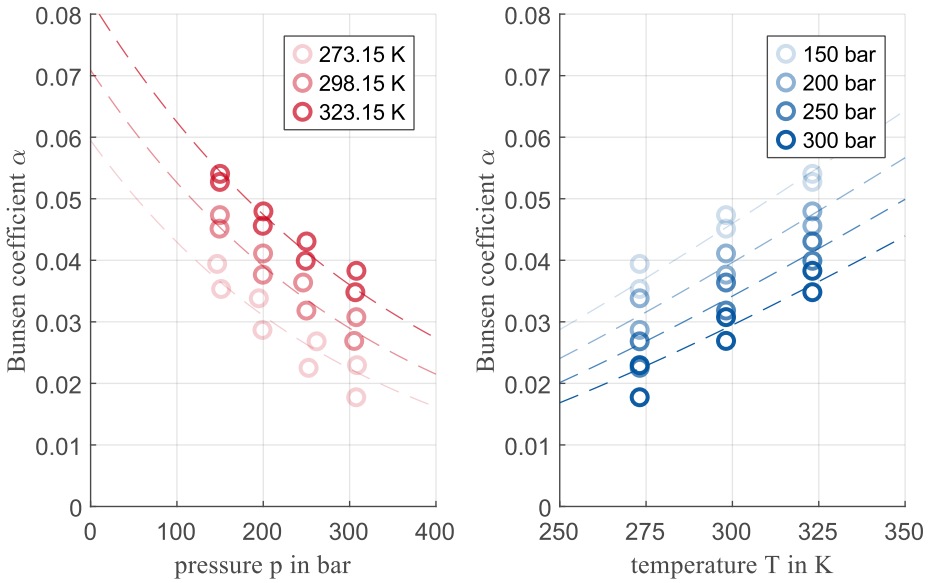


Figure 3.1. Bunsen coefficient of HLP 46 and nitrogen

As can be seen, the Bunsen coefficient decreases for rising pressure values. This trend is consistent with the behavior that emerges from the Krichevsky-Kasarnovsky equation (3.1) for positive values of the partial molar volume ΔV_{solv}. [12]

$$\alpha(p,T) = \alpha(p_0,T) \cdot \exp\left[-\frac{\Delta V_{solv}}{RT}(p - p_S)\right] \tag{3.1}$$

The dependence of gas solubility on temperature can be described using the Henry–Van't Hoff relationship in equation (3.2). According to this, gas solubility follows an exponential curve. For a positive solution enthalpy ΔH_{solv}, increasing solubility is to be expected at rising temperatures, which is also reflected in the experiment. [13]

$$\alpha(p_0, T) = \alpha(p_0, T_0) \cdot \exp\left[\frac{\Delta H_{solv}}{R}\left(\frac{1}{T_S} - \frac{1}{T}\right)\right] \tag{3.2}$$

The nearly proportional relationship between gas solubility and temperature can be explained by the small temperature changes, where the exponential relationship from equation (3.2) can be simplified by a Taylor series expansion and approximately corresponds to the linear expression from (3.3).

$$\alpha(p_0, T) \approx \alpha(p_0, T_0) \cdot \left[1 - \frac{\Delta H_{solv}}{R{T_S}^2}(T - T_S)\right] \tag{3.3}$$

Based on the Krichevsky-Kasarnovsky equation and the Henry–Van't Hoff relationship a regression analysis was performed using a trust-region-reflective algorithm to minimize the sum of squared residuals while accounting for parameter bounds. The derived values for ΔV_{solv}, ΔH_{solv} and $\alpha(p_0, T_0)$ are listed in **Table 3.1** together with the corresponding goodness-of-fit criteria ($R^2, {R^2}_{adj.}, SSE$). Here, the coefficient of determination R^2 quantifies the proportion of variance in the dependent variable that is explained by the model. The adjusted coefficient of determination ${R^2}_{adj.}$ accounts for the number of model parameters and penalizes overfitting. The sum of squared errors (SSE) represents the cumulative squared deviation between observed and predicted values and serves as an absolute measure of the model's residual error. Even though the values for identical measurement points in some cases diverge significantly, the high values for R^2 and ${R^2}_{adj.}$ and low values for SSE indicate that the measurement method is capable of representing the underlying physical laws that determine the pressure- and temperature-dependent gas solubility of oils.

Thermodynamic measures		Goodness-of-fit criteria	
$\alpha(p_0, T_0)$	0.0709	R^2	0.952
ΔV_{solv}	$7.3955\ \frac{m^3}{mol}$	${R^2}_{adj.}$	0.947
ΔH_{solv}	$4756.1\ \frac{J}{mol}$	SSE	$1.015e-4$

Table 3.1 Thermodynamic measures and goodness-of-fit criteria for HLP 46

Since the recorded measurement signals− and consequently the calculated Bunsen coefficient−showed only minor deviations for a single measuring point, the corresponding error bars are not visible in Figure 3.1. Therefore, a detailed analysis of the recorded sensor signals was conducted, as shown in **Figure 3.2**.

The results are shown in Figure 3.2 in the form of box plots, illustrating the deviations from the respective mean value of each measurement. It can be observed that, for the pressure, both the lower and upper quartiles lie within a band of at least ±0.5 bar, indicating a low

pressure variability. Furthermore, the median is located close to the mean value, suggesting a symmetric and evenly distributed temperature signal. The observed deviations can be primarily attributed to uniform noise in the measurement signal, with external influences exerting negligible impact.

For the temperature measurements, the boxes indicate greater variability in both magnitude and position relative to the mean. Nevertheless, the overall deviations are minor, with both the lower and upper quartiles confined to a range of approximately ±0.05 °C.

Deviations in piston position are also minimal, on the order of a few hundredths of a millimeter. Notably, the median consistently lies above the mean value across all measurements, suggesting the presence of a systematic error. Additionally, two individual data points exhibit larger deviations.

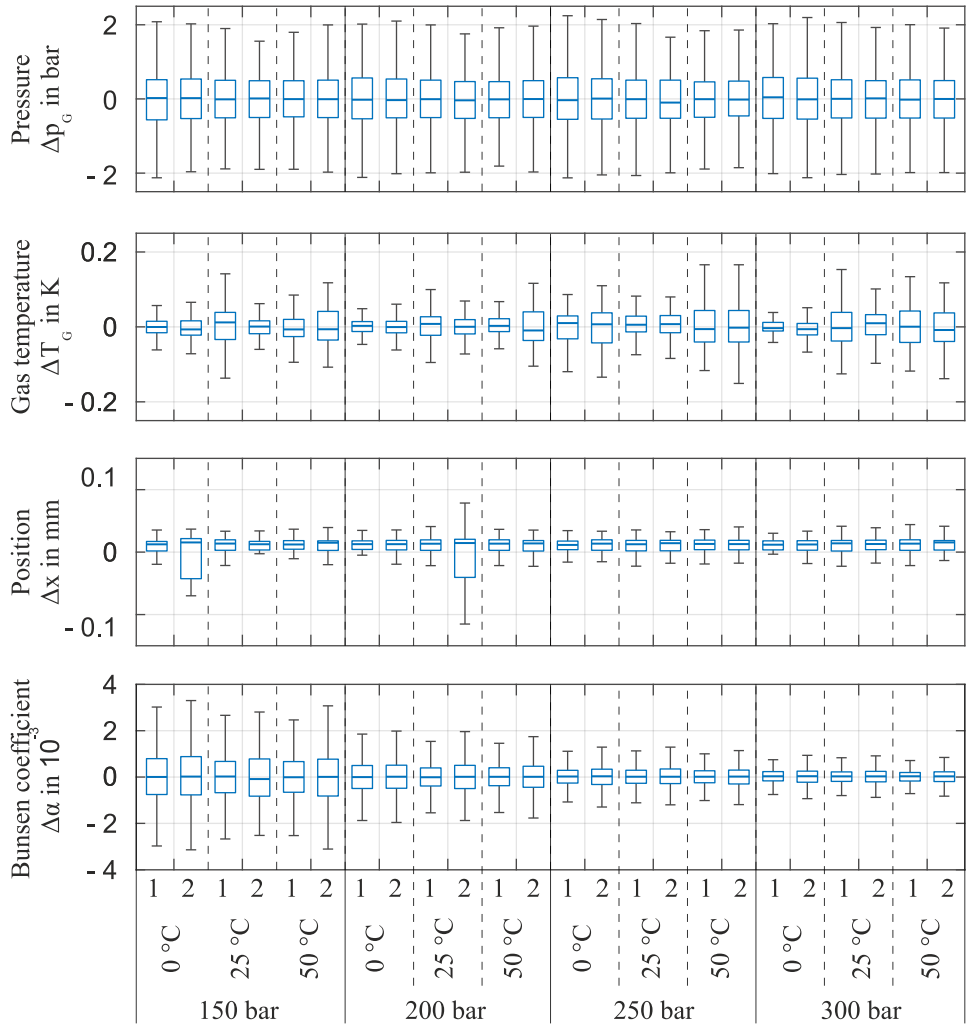


Figure 3.2. Statistical deviations of the sensor signals (gas temperature T_G, pressure p_G, and position x) and the resulting deviations of the Bunsen coefficient α.

Despite the asymmetries observed in the temperature and position data, the Bunsen coefficient demonstrates a uniform distribution around the mean. The bounds of the upper and lower quartiles remain within a narrow range of ±0.001. It is, however, evident that the coefficient values at lower pressures exhibit greater scatter compared to those at higher pressures, whereas temperature appears to have no significant influence on the variation. Consequently, the discrepancies between the two measurement series from Figure 3.1 are unlikely to result from measurement errors. A more plausible explanation is that improper filling of the test rig led to inaccurate assumptions regarding the volumes of liquid and gas and therefore to small deviations between both measurement series.

Since the composition of hydraulic oils is widely unknown, validation with literature data is difficult. The exact composition of the investigated oil is also unknown, so that no calculation of the Bunsen coefficient based on data for pure liquids is possible. However, it can be assumed that HLP oils consist of a mixture of paraffins, naphthenes, and aromatics, allowing for a comprehensive plausibility check [14]. Gas solubility data for the named molecule groups were transferred to the Bunsen coefficient using the interrelations from Table 2.1 and assembled in Figure 3.3 [15]. As can be seen, for most listed hydrocarbons the ability to dissolve nitrogen decreases increasing number of C-atoms. The Bunsen coefficient of the investigated HLP oil is lower than for all n-Alkanes with number of C-atoms lower than 20 and all other hydrocarbons listed. Given that mineral oil-based hydraulic oils within this viscosity class can contain hydrocarbons ranging between 15 and 50 C-atoms, the measured value of 0.0709 seems plausible [16].

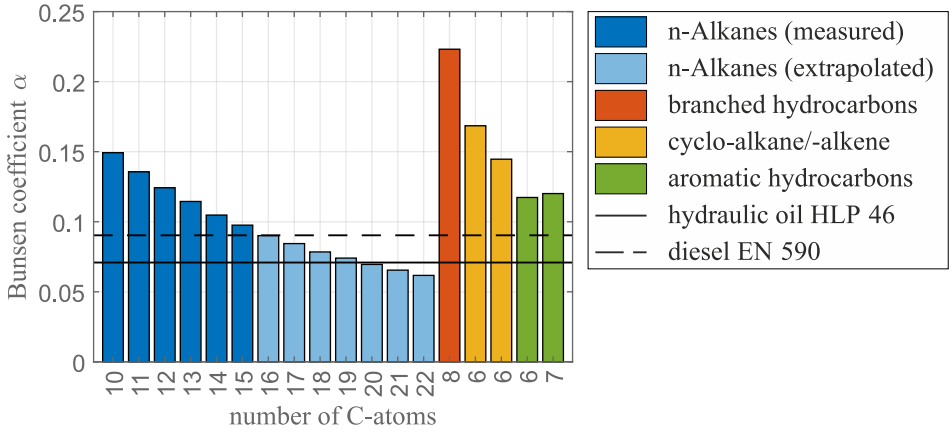


Figure 3.3. Comparison of literature data for the Bunsen coefficient of different hydrocarbons and nitrogen [15]

In a previous study, diesel EN 590 was examined and a Bunsen coefficient of 0.903 was determined [17]. Given that the gas solubility of the hydrocarbons considered decreases with increasing number of C-atoms, the lower values for the comparatively long-chain HLP 46 (Diesel: C13 - C18) are plausible, too [18].

Of particular interest for hydraulic systems is the amount of gas that can be released during operation, as cavitation may damage components and the fluid mechanical properties of the oil can be affected by undissolved gas.

Using the parameters from Table 3.1 and Equations 3.1–3.2, the volume of dissolved gas at standard conditions per unit volume of oil was calculated for pressures and temperatures typical of hydraulic applications. **Figure 3.4** shows that, at high pressures and temperatures, theoretically large amounts of nitrogen can dissolve. For instance, at 400 bar and 80 °C, the gas volume reaches nearly 14 times the oil volume. However, under such extreme conditions, achieving equilibrium is highly unlikely because dissolution processes are slow and gas may immediately re-release at ambient pressure in the tank.

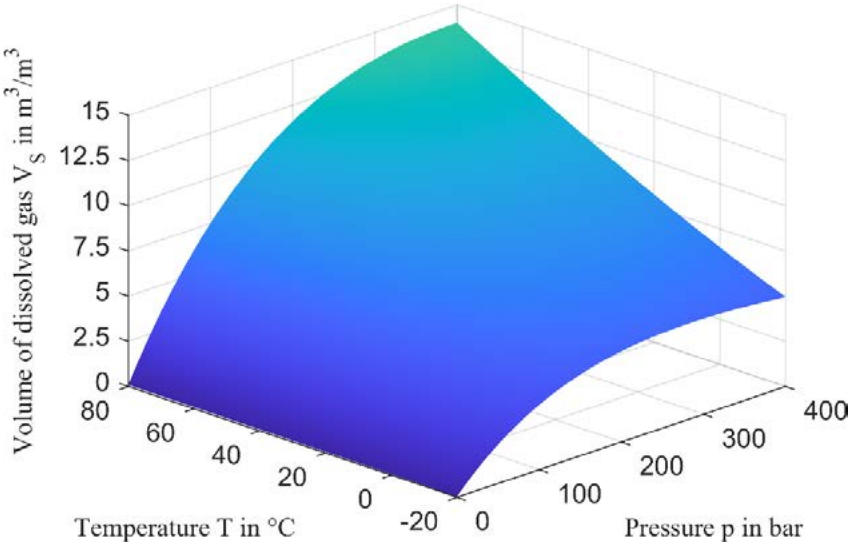


Figure 3.4. Volume of dissolved nitrogen at standard conditions per unit volume of oil

For practical hydraulic operation, the amount of gas that can dissolve under tank pressure and temperature is more relevant. Assuming a system operating at 40 °C and tank pressure equal to ambient, approximately 7.77 % of the oil volume could be saturated with gas. If the pressure drops by 0.1 bar, as might occur in a suction line, the dissolved gas fraction would decrease to 6.99 %, corresponding to about 0.78 % of the oil volume forming bubbles.

Gas release can also be triggered by temperature changes. For example, consider a mobile hydraulic machine operating with a tank temperature of 50 °C. During operation, the oil has sufficient time to reach saturation under these conditions. If the machine is then stored outdoors at −20 °C, and assuming released bubbles remain trapped within pipes, cylinders, or pump housings, the gas volume would occupy approximately 3.18 % of the oil volume after cooling to ambient temperature.

These examples demonstrate that, under normal operating conditions, only a small fraction of the maximum solubility of hydraulic oils is realized. Nevertheless, both operational and standstill conditions can lead to the release of relative gas volumes of a few percent of the oil volume, which may still be significant for system performance. It should be noted that these estimations are based solely on nitrogen values. Accurate determination would require the Bunsen coefficient for oxygen; however, since air consists predominantly of nitrogen, the results are considered to be of the correct order of magnitude.

4. SUMMARY AND OUTLOOK

In this study, a measurement principle for determining the gas solubility of technical liquids was presented, and pressure- and temperature-dependent results for a hydraulic oil (HLP 46) were discussed. Although a direct validation of the results using literature data is challenging, the obtained values were found to be plausible and in good agreement with the underlying thermodynamic models. The obtained gas solubility values indicate that, during hydraulic system operation, dissolved gas can occupy a few percent of the oil volume, potentially causing cavitation and altering the fluid properties of the oil.

In future work, additional oils will be investigated to analyze the influence of oil composition and additives on gas solubility in more detail. Of particular relevance is the study of the solubility behavior of alternative gaseous fuels such as hydrogen and ammonia in engine oils. Understanding how much of these gases can be dissolved, how much may be released to the environment, and how the dissolved gases affect oil degradation is essential for safe and efficient engine operation. Within the framework of the Excellence Cluster *The Fuel Science Center*, a simulation model is currently being developed to describe gas diffusion within the lubricating film between piston ring and cylinder wall. The required material data will be determined experimentally using the measurement method presented in this work.

5. NOMENCLATURE

α	Bunsen coefficient	$[-]$	R^2_{adj}	coefficient of determination (adjusted)	$[-]$
ρ	density	$\left[\frac{kg}{m^3}\right]$	SSE	sum of square due to error	$[-]$
A	piston surface	$[\,m^2\,]$	T	temperature	$[\,K\,]$
c	concentration	$\left[\frac{mol}{m^3}\right]$	SSE	sum of square due to error	$[-]$
ΔH	enthalpy	$[\,J\,]$	t	time	
k_H	Henry's law constant	$\left[\frac{Pa \cdot m^3}{mol}\right]$	U	voltage	
m	mass	$[\,kg\,]$	V	volume	
M	molar mass	$\left[\frac{kg}{mol}\right]$	ΔV_{solv}	partial molar volume	
n	amount of substance	$[\,mol\,]$	x_i	mole fraction	
p	pressure	$[\,bar\,]$	z	compressibility factor	
R^2	coefficient of determination	$[-]$			

$(\square)_G$	gas	$(\square)_1$	state 1
$(\square)_L$	liquid	$(\square)'$	liquid phase
$(\square)_S$	standard conditions	$(\square)''$	gas phase
$(\square)_0$	initial state		

6. ACKNOWLEDGEMENTS

This work was funded by the German Research Foundation (DFG) in the scope of the project „Einfluss der Belastungsgradienten auf das Ermüdungsverhalten hydraulischer Komponenten" with the number 449676223, and under Germany´s Excellence Strategy – Exzellenzcluster 2186 „The Fuel Science Center".The authors would like to thank DFG for its support.

7. REFERENCES

[1] YOUNG, F. Ronald: *Cavitation*. London : Imperial College Press, 1999

[2] NIKOLAJSEN, Jorgen L.: *Viscosity and Density Models for Aerated Oil in Fluid-Film Bearings©*. In: *Tribology Transactions* 42 (1999), Nr. 1, S. 186–191. 10.1080/10402009908982207

[3] MARIUS HOFMEISTER, et al.: Solubility of Ammonia and Hydrogen in Engine Oils. 10.13140/RG.2.2.13368.05127. In: FSC (Hrsg.): *13th International Conference : Fuel Science - From Production to Propulsion, 2025*, 2025

[4] V. P. LOGVINYUK, V. V. MAKARENKOV, V. V. MALYSHEV & G. M. PANCHENKOV: *Solubility of gases in petroleum products*. In: *Chemistry and Technology of Fuels and Oils* (1970), Volume 6, pages 353–355

[5] BATTINO, Rubin, et al.: *The Solubility of Gases in Liquids*. In: *Chemical Reviews* 66 (1966), Nr. 4, S. 395–463. 10.1021/cr60242a003

[6] *The experimental determination of solubilities*. Chichester : Wiley, 2003 (Wiley series in solution chemistry 6)

[7] HUANG, Sam S.-S., et al.: *The phase behavior of two mixtures of methane, carbon dioxide, hydrogen sulfide, and water*. In: *Fluid Phase Equilibria* 19 (1985), 1-2, S. 21–32. 10.1016/0378-3812(85)85033-0

[8] ROUSSEAUX, P., et al.: *A static method for determination of vapour—liquid equilibria and saturated liquid molar volumes at high pressures and temperatures using a new variable-volume cell*. In: *Fluid Phase Equilibria* 11 (1983), Nr. 2, S. 153–168. 10.1016/0378-3812(83)80055-7

[9] RAMBAKS, Andris, et al.: METHOD FOR THE EXPERIMENTAL DETERMINATION OF THE BUNSEN ABSORPTION COEFFICIENT OF HYDRAULIC FLUIDS, FPMC2019. 10.1093/nar/gkn369. In: *Proceedings of the ASME/BATH 2019 Symposium on Fluid Power and Motion Control*, 2019

[10] DIN 12790:2019-12. Dezember 2019. *DIN 12790: Laboratory glassware; hydrometers -Principles of construction and adjustment*

[11] FUCHS LUBRICANTS: *Safety Data Sheet Renolin B 15 VG 46*. 2017

[12] LÜDECKE, Christa, et al.: *Thermodynamik* : *Physikalisch-chemische Grundlagen der thermischen Verfahrenstechnik*. Softcover reprint of the hardcover 1st edition 2000. Berlin, Heidelberg : Springer, 2000 (Springer-Lehrbuch)

[13] CARRERO, J. I.: *Application of the van't Hoff equation to phase equilibria*. In: *ChemTexts* 10 (2024), Nr. 3. 10.1007/s40828-024-00188-x

[14] THEO MANG, et al.: *Lubricants and lubrication*. Third, completely revised and enlarged edition. Weinheim : Wiley-VCH, 2017

[15] BATTINO, Rubin, et al.: *The Solubility of Nitrogen and Air in Liquids*. In: *Journal of Physical and Chemical Reference Data* 13 (1984), Nr. 2, S. 563–600. 10.1063/1.555713

[16] EMMANUEL O. ALUYOR AND MUDIAKEOGHENE ORI-JESU: *Biodegradation of mineral oils – A review*. In: *African Journal of Biotechnology* Vol. 8 (2009), Nr. 6, 915-920

[17] HOFMEISTER, Marius, et al.: A Method for Determining the Bunsen Coefficient of Bio-Hybrid Fuels. 10.4271/2021-01-1187. In: *SAE Technical Paper Series* : SAE International400 Commonwealth Drive, Warrendale, PA, United States, 2021 (SAE Technical Paper Series).

[18] *Fuels and lubricants handbook* : *Technology, properties, performance, and testing*. 2nd edition. West Conshohocken, PA : ASTM International, 2019 (ASTM's manual 37)

Biographies



Marius Hofmeister received his master's degree in mechanical engineering from RWTH Aachen University in 2019. He is currently a research associate at the Institute for Fluid Power Drives and Systems (ifas) and a member of the Fuel Science Center, where his research focuses on alternative fuels, particularly their fluid mechanical properties and material compatibility.



Lukas Boden obtained his master's degree in mechanical engineering at RWTH Aachen University in 2023. Since then, he is working as a research associate at the Institute for Fluid Power Drives and Systems (ifas), focussing on identifying the influence of viscous fluid on fatigue crack propagation in fluid power components.



Andris Rambaks studied mechanical engineering at Riga Technical University from 2011 to 2014 and obtained his master's degree in mechanical engineering from RWTH Aachen University in 2018. From 2018 to 2023, he worked as a research associate at the Institute for Fluid Power Drives and Systems (ifas), where he focused on investigating gas solubility in hydraulic oils and the fatigue behavior of autofrettaged components.

18



Univ.-Prof. Dr.-Ing. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Evaluating sliding properties of selective laser melted maraging steel against aluminium alloys in water and hydraulic oil

Ana Trajkovski, Jan Bartolj, Nejc Novak, Jan Pustavrh, Mitjan Kalin, Franc Majdič

Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva cesta 6, 1000 Ljubljana, Slovenia;

ana.trajkovski@fs.uni-lj.si, jan.bartolj@fs.uni-lj.si , nejc.ovak@fs.uni-lj.si, jan.pustavrh@fs.uni-lj.si , mitjan.kalin@fs.uni-lj.si , franc.majdic@fs.uni-lj.si

Abstract.

Additive manufacturing (AM) of metals enables lightweight, customizable, and high-performance components, with maraging steel (MS1) being widely used for its strength and hardness. While much research exists on AM steels' microstructure and processing, limited studies address their tribological behavior, especially under lubrication. This study investigated the sliding performance of 3D-printed MS1 against aluminium alloys (EN AW 6082) for water hydraulics applications. Aluminium alloy, despite its low hardness, offers advantages in weight reduction and conductivity, and its performance can be improved by surface treatments such as anodization. Reciprocal sliding tests were conducted demineralized water, comparing untreated, conventionally anodized, and hard-anodized aluminium alloy compared to conventional oil lubricated conditions and hardened steel (AISI 440C). Results showed that in oil, sliding speed had little effect, and conventional anodization improved aluminium's wear resistance to nearly match steel. In water, however, sliding speed significantly influenced wear: at low speeds, anodized aluminium performed comparably to steel, but at higher speeds, severe wear occurred for aluminium, even after anodization. The findings suggest aluminium alloys can be potential counterparts to AM maraging steel in water-lubricated systems if sliding speed is carefully controlled.

Keywords. Maraging steel; aluminium alloy; anodization; wear; friction; water; hydraulic oil;

1. INTRODUCTION

Additive manufacturing (AM) of various materials offers unique advantages over conventional manufacturing processes, such as flexibility in geometric design, rapid component production, and product customisation, while reducing product mass through optimised design or cellular structures. In the last decade, AM has reached a technological level suitable for manufacturing high-performance components, and it is predicted that the

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

market will be worth nearly $60 billion by 2030 [1]. Although plastics hold the largest share of the AM market, 3D metal printing enables the production of high-performance, lightweight, and customised metal parts, while eliminating most traditional tooling by allowing integrated cooling channels, lattice structures, and internal features not possible with conventional machining. This approach is highly attractive for the aerospace, defence, robotics, and automotive industries, and also supports efficient material use, rapid prototyping, and replacement of damaged parts. Among the different powder-bed-based approaches for metal printing, selective laser melting (SLM) and selective laser sintering (SLS) are the most common [2]. Extensive fundamental research has been conducted on the microstructure of 3D-printed steel, printing parameters, heat treatments, porosity, manufacturing technology, and post-treatments [2–4]. AM has also been successfully used in hydraulics for manufacturing valves, manifolds, and pump parts [5–8].

There are various metal-based components made from pure metals, alloys, or metal matrices, such as 316L stainless steel, titanium Ti-6Al-4V, aluminium Al 6061, ALSi10MG, Inconel, cobalt-chrome, or maraging steel [1, 2]. There is limited experimental data on the specific wear and friction properties of these materials. Tribological sliding tests are typically conducted under dry conditions, against different steels, carbide, or ceramic counterparts, at room or elevated temperatures. There are fewer studies on the behaviour of these 3D-printed steels under lubricated conditions. Stainless SLM steel 316L has been tested in hydraulic oil compared to brass or nitrided steel, where grinding and polishing [9] and brass counterpart [10] have been suggested for improved performance of SLM steel. Recently, the cavitation erosion resistance of AM materials was found to be lower than that of conventionally produced steel at the beginning of the test [11]. Maraging steel (MS1), also known as 18Ni-300 (1.2709) maraging steel, is a martensitic iron-nickel alloy with low carbon content that has been widely used due to its high yield strength, tensile strength, and hardness, especially in automotive, aerospace, and tooling applications, where high wear resistance is expected. However, most studies focus on microstructure, microhardness, tensile, and impact properties [3, 12–15]. Recently, conventionally and laser bed fusion fabricated MS1 were compared under different normal loads (5–35 N), sliding speeds (1–5 Hz), and temperatures (30–125 °C), but were tested under dry conditions [16] emphasising the importance of testing under specific real-time application conditions. Another study suggested comparable wear but superior corrosion resistance of additively manufactured MS1 components compared to conventional MS1 [17]. In particular, intercritical heat treatment was shown to enable better tribological properties of MS1 under dry conditions [14].

In our recent study, we analysed the tribological properties of 3D-printed maraging steel (MS1) in sliding contact with materials commonly used in hydraulics, particularly water hydraulics: nitrided steel AISI 440C, bronze, and aluminium alloy [18]. The aluminium alloy demonstrated results comparable to conventional nitrided steel and exhibited better wear performance than bronze under the tested conditions. Due to their low specific density, light weight, high thermal and electrical conductivity, and wide availability, aluminium and its alloys are among the most commonly used metals for reducing overall system weight and the energy required for production and operation [19]. However, their relatively low hardness and greater sensitivity to abrasive wear often limit their use to applications with low friction and lower loads. To improve surface hardness, treatments such as conventional or hard anodising, plasma electrolytic oxidation, and heat hardening are applied [20–23]. Hard anodising can produce oxide layers 25–150 μm thick, increasing surface hardness by

5–8 times [24]. The anodising process also affects corrosion resistance. Although aluminium alloys are widely used, their tribological behaviour is typically studied under dry or oil-lubricated conditions.

In this study, the aluminium alloy (EN AW 6082) was tested before and after both conventional and hard anodising in a reciprocal sliding test against 3D-printed maraging steel (MS1) in demineralised water at two different sliding speeds. For comparison, the same tests were also conducted in conventional hydraulic oil ISO VG 46, conventional HFC 46 lubricant, and with conventional nitrided AISI 440C steel. The results of the tribological tests were used to analyse the influence of the surface treatment of the aluminium alloy and the effect of sliding speed when sliding against MS1, to assess potential material pairs for efficient proportional valves designed for water hydraulics.

Common tribological test methods designed to reproduce simplified contact geometries and loading conditions in sliding interfaces include ball-on-disc (BoD), pin-on-disc (PoD), and cylinder-on-disc (CoD) configurations, often implemented in reciprocating sliding tests to simulate short-stroke motions. Standardized procedures such as ASTM G99 for pin-on-disc and ASTM G133 for linearly reciprocating tests (ISO 7148, ISO 6601-6605) are widely employed in both academic and industrial studies for comparative friction and wear analyses [25–27]. The ball-on-disc configuration provides a well-defined Hertzian point contact at relatively high local pressures, making it suitable for fundamental contact stress and lubricant behavior characterization, whereas the pin-on-disc test offers a more stable contact area and medium–low nominal pressures relevant to a broader range of engineering materials. Although a cylinder-on-disc or line contact test would more closely approximate the geometry, contact mechanics, and lubrication regime of a hydraulic valve spool within a valve body, there is no single fully standardized tribological test that exactly reproduces the quasi-line contact of spool/body interfaces. For this reason, we selected an established pin-on-disc setup, which ensures methodological rigor and comparability through well-defined standards, and provides effective material screening and ranking; this choice is complemented by consideration of reciprocating cylinder-on-disc testing when higher representativeness is required. A reciprocating pin-/cylinder-on-disc configuration therefore represents a practical compromise between the representativeness of the real valve contact and the experimental control necessary for reproducible and comparable tribological assessment of material pairs in proportional hydraulic valve spool–body contacts.

The simplified tribological test configuration provides an effective platform for preliminary material screening and is routinely employed within our research group for assessing novel materials, surface coatings, reinforcements, and finishing processes for hydraulic components [28, 29], as well as for evaluating new lubricant formulations [30]. Comparable methodologies have recently been applied to investigate the performance of various ceramic materials intended for water-lubricated axial piston pump friction pairs using ball-on-disc tests [31], and to examine the tribological behaviour of valve plate pairs through ring-on-block experiments [32]. Nevertheless, such laboratory-scale tests should be complemented by full-component experiments or simulations incorporating the selected material pairs to ensure realistic performance assessment. Extensive research on axial piston pumps has demonstrated the importance of validating numerical simulations with experimental data, particularly when integrating wear prediction, running-in behaviour, and

friction analyses following innovative component redesign [33, 34]. This combined approach enables a comprehensive evaluation of overall component efficiency [35].

2. METHODOLOGY

2.1. Sample preparation

An EOS M290 3D metal printer using the SLM (selective laser melting) method was employed for sample preparation, with a laser beam speed of 1010 mm/s, power of 305 W, and a powder layer thickness of 50 µm. The material was in powder form with spherical grains ranging from 15 to 45 µm. As the material is melted in specific areas, the high temperature of the laser beam also locally hardens it. The final product exhibits anisotropic properties due to the layer-by-layer manufacturing process, which are eliminated by subsequent heat treatment. For the tribological tests (Figure 1.) 3D printed discs (25 x 25 x 6 mm), made of maraging steel (MS1/1.2709 also known as 18Ni-300) were used, representing a hydraulic valve body. The disc samples were polished in several steps using a RotoForce-3 automatic sample polishing and preparation device, achieving a final roughness of 0.1 µm.

The pin samples (Φ 10 × 20 mm, with a 12.5 mm radius) were prepared from two different materials: aluminium alloy (EN AW 6082/3.2315) and commercially available standard hardened stainless steel (AISI 440-C/1.4125). The pin samples were polished using the same procedure as the MS1 discs. After polishing, the aluminium samples were anodised using the standard procedure and underwent hard anodization. The pin samples represented hydraulic valve spools.

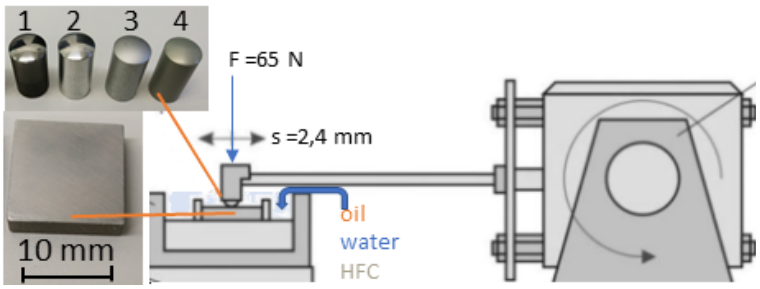


Figure 1. Experimental procedure: pins: 1- AISI 440-C; 2-Al 6082; 3-Al 6082 anodised; 4-Al 6082 hard anodised and disc - 3D printed MS1 at tribological tests set-up;

Materials microhardness was measured using a standard Vickers hardness tester, Leitz - Miniload 2, with an indentation force of 1.9 N (according to DIN 50133). However, with the measurement method employed, we were unable to measure the hardness of anodised aluminium samples successfully. Therefore, we assumed that the anodised samples had the hardness specified by the manufacturer. For soft anodised samples, this ranged from 200 to 250 HV. For hard anodised samples, it ranged from 500 to 600 HV. Before each test, the samples, clamps and pins were cleaned with ethanol and dried in air.

Table1. Mechanical properties of prepared samples

Material	Density [kg/m^3]	Young modulus [GPa]	Ultimate tensile strength [MPa]	Hardness [HV]
pin				
AISI 440-C	7.6	200	760-1960	~ 1330
Al 6082	2.7	70	260	~ 132
disc				
MS1	8-8.1	180	2080	~ 460

The first lubricant used was commonly available ISO VG46 hydraulic oil. The second lubricant was demineralised water (W). Additionally, commercially available HIDROL HFC 46 lubricant was tested. The selected HFC 46 is a water-glycol mixture (35–40% glycol content) with special additives that provide excellent corrosion protection. This lubricant contains no mineral oils, nitrites, PCBs, or phenols. The properties of the lubricants used were measured using an automatic viscometer, Anton Paar SVM 3001, and are presented in Table 2.

Table 2. Chemical properties od used lubricants

Lubricant	Density at 25°C [g/cm^3]	Density at 45°C [g/cm^3]	Kin. viscosity at 25°C [mm^2/s]	Kin. viscosity at 40°C [mm^2/s]
ISO VG46	0.86	0.84	100	41.4
Demin.water	0.997	0.99	0.89	0.66
Hidrol HFC 46	1.075	1.06	90	43

2.2. Experimental procedure

Tribological tests were conducted on a Cameron-Plint TE 77 high-frequency tribometer (Figure 1). During the test, the pin slides in a reciprocal mode on the 3D-printed disk-like plate. The average sliding velocity was set to two values: 0.05 m/s (corresponding to a frequency of 10 Hz) and 0.2 m/s (corresponding to a frequency of 40 Hz). The applied frequencies correspond to the dynamic opening and closing of the proportional spool valve. The maximum excitation frequency achievable with the tribometer was 40 Hz. The stroke length was 2.4 mm, representing the valve spool stroke in a CETOP 3 size 4/3-way slide valve. This value also represents the minimum stroke length available with the tribometer. The normal load was set to 65 N. Before each test, the pins and plate samples were fully immersed in the selected lubricant. During the tests, the current value of the friction coefficient was automatically recorded. The tests were conducted for 60 minutes, based on preliminary tests that showed a stable coefficient of friction and measurable wear. Each test (material and lubricant combination) was repeated three times, and the average values of the steady-state coefficient of friction were used for the comparative analysis.

2.3. Wear evaluation

After testing, the wear on the pins and wear tracks was analysed using a 3D digital microscope with a nano-point profilometer Hirox HRX-01 & NPS. The wear volume of the pins was calculated from the dimensions of the wear tracks. The 3D profile and dimensions of the wear calotte were determined with Mountain Map software [25]. The specific wear

rate was determined by dividing the measured wear volume by the product of the normal load and sliding distance. The surface films on the printed plates were also characterised for shape and size.

3. RESULTS AND DISCUSSION

3.1. Evolution of coefficient of friction

Figure 2 shows the evolution of measured coefficient of friction during individual tests of pin samples againts MS1 printed disks, in water and hydraulic oil at two different sliding frequences.

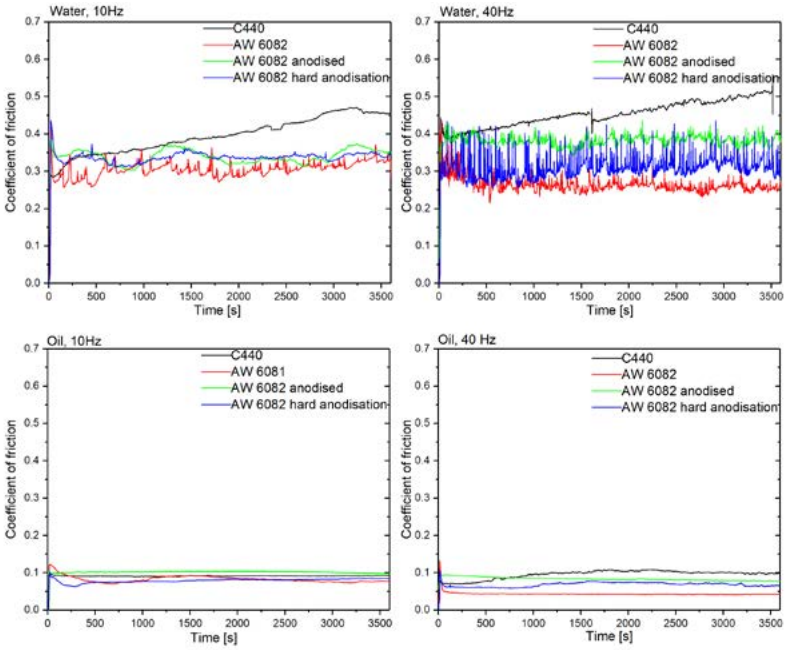


Figure 2. Evolution of the coefficient of friction during the reciprocal tests of of aluminium alloys pins against MS1 in water and hydraulic oil, at two different sliding speeds

The coefficient of friction measured in the water doesn't show typical running in, it rather has unusable value for all aluminium alloys. The fluctuations observed in case of water lubricated conditions indicate, that water incorporates some instabilities into the sliding contact. At lower frequency these instabilities are the most prominent for nature aluminium alloy, and tend to become smoothed after anodisation. At higher frequency these instabilities are simillar for nature alloy and standard anodisation, but more intense for hard anodized sample. In case of standard AISI 440-C these instabilities are not evident, although coefficient of friction is linearly increasing during the observed period of time at both frequences (up to 34 %) and has higher value compared to tested all aluminium alloys. If we compare these results to the measurements in ISO VG46 hydraulic oil, the coefficient of friction reaches a stable value almost immediately for all aluminium alloys similarly to AISI 440-C, at both frequencies (with in the first 2500 cycles).

3.2. Coefficient of friction

Figure 3 represents average values of measured steady coefficient of friction for all tested samples in both water and hydraulic oil, at both frequences.

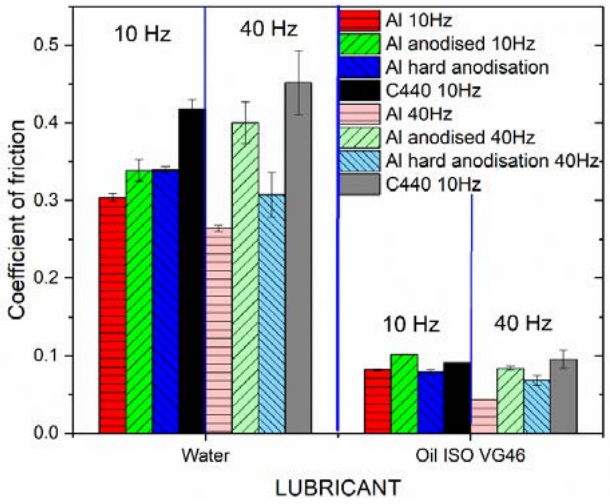


Figure 3. Average coefficient of friction during the reciprocal tests of aluminium alloys pins against MS1 in hydraulic oil and water at two different sliding speeds

When water was used as lubricant, at lower frequency, comperable values of coefficient of friction were measured for natrure aluminium alloy (around ~ 0.3), followed by anodised and hard anodised aluminium (~ 0.33), and considerably higher value was measured for AISI 440-C (~ 0.39). At higher frequency, nature and hard anodized aluminium alloy coefficient of friction even decresed (value around ~ 0.28 and 0.3 respectively). In water all aluminium alloys give good response compared to standard stainless steel, and from that point of view could be potentially good materials for water hydraulics. When we account for fluctuations in measured values seen in aluminium alloys, anodisation it is expected to enable smoother working conditions even in water.

Lower values of coefficient of friction were observed in hydraulic oil, for all tested samples compared to water lubricated conditions. The coefficient of friction decreased in oil around 4.5 times for all aluminium alloys and AISI 440-C, at both frequences. However even in oil, nature aluminium alloy showed lowest measured coefficient of fiction with values comparable to measured values for both anodised aluminium and AISI 440-C, all around ~ 0.09. At higher frequency, slight decrese was observed for nature aluminium alloy.

3.3. Specific wear

Based on repeated measurements in water and hydraulic oil at two sliding frequences, Figure 4 represents average values of calculated specific wear from analyzed pictures of wear tracs, of all tested pin samples for each material.

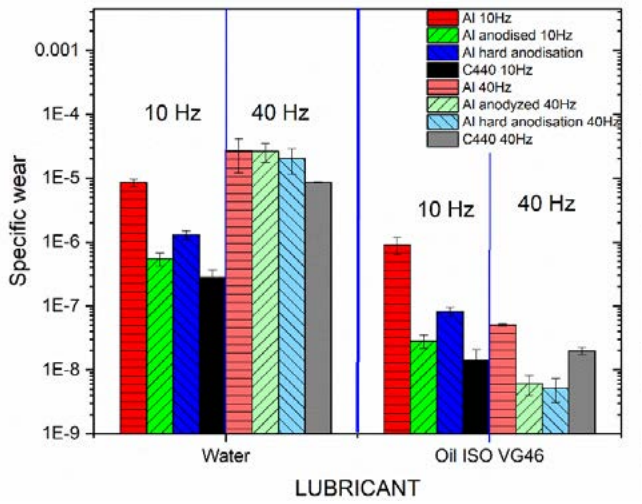


Figure 4. Average specific wear of aluminium alloys pins after sliding against MS1 in water and hydraulic oil, at two different sliding speeds

In water, the lowest measured specific wear was for AISI 440-C (~2.8 × 10^{-7} mm^3/Nm), fairly followed by anodised aluminium (~5.4×10^{-7} mm^3/Nm). Evidently anodization improved wear resistance of nature aluminium alloy (~ 8.6 × 10^{-8} mm^3/Nm), even comparted to hard anodization (~1.3 × 10^{-6} mm^3/Nm). However, at higher frequency difference between tested materials is lower. From all aluminium alloys (~2.3 × 10^{-5} mm^3/Nm) to AISI 440-C (~8.6 × 10^{-6} mm^3/Nm).

In hydraulic oil, we can see that both all aluminium alloys and AISI 440-C showed lower specific wear, compared to water lubricated conditions at both frequences. Smaller decrease was found at lower frequency (from 2 times for anodized aluminium up to 157 times for hard anodized aluminium alloy) compared to higher frequency (around 500 times for nature aluminium and AISI 440-C, up to around 4000 times for both anodized alloys). Again anodized aluminium alloy (~8.2 × 10^{-8} mm^3/Nm) fairly followed AISI 440-C (~1.4 × 10^{-8} mm^3/Nm) at lower frequency, with even better results at higher frequency (~6.1 × 10^{-9} mm^3/Nm and ~2 × 10^{-8} mm^3/Nm respectively).

Based on the measurements, we can conclude that the best results were accomplished by AISI 440-C followed by anodised aluminium, under harsh water lubricated conditions. In hydraulic oil, anodization improved aluminium properties to put it even in front of the AISI 440-C (for tested frequences and loads). Both anodization and hard anodization process improved tribological properties of nature aluminium alloy, however hard anodization did not significantly change properties compared to standard anodization, especially at lower frequences.

3.4. Comparison with HFC 46 as lubricant

The results of the same tribological tests in hidrol HFC 46, when nature aluminium and anodized aluminium alloy pins slide against MS1 at two different frequencies, are presented of Figure 5 and 6.

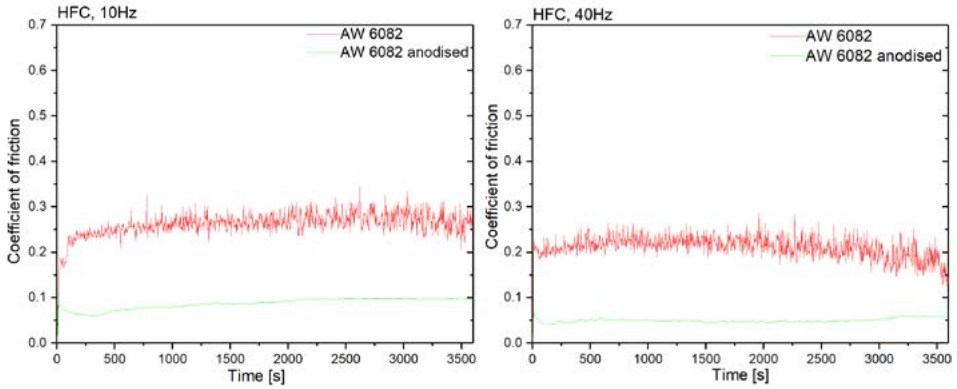


Figure 5. Evolution of the coefficient of friction during the reciprocal tests of aluminium alloys pins against MS1 at two different sliding speeds in HFC 46

From the results it is evident that anodization improves tribological properties of nature aluminium, and sliding frequency does not impact significantly measured properties (at sampled range 10 – 40 Hz). Measured coefficient of friction is smoother and more stable after anodization, at both frequencies (Figure 5), while average value of the coefficient of friction decreases 3.25 times at 10 Hz, up to 4.2 times at 40 Hz after anodization (Figure 6a). Specific wear also decreases after anodization, around 800 times at 10 Hz, up to 3000 times at higher frequency (Figure 6b). Specific wear of nature aluminium (~3.4 × 10^{-4} mm^3/Nm) is lower but not significantly affected by higher frequency (~2.4 × 10^{-4} mm^3/Nm).

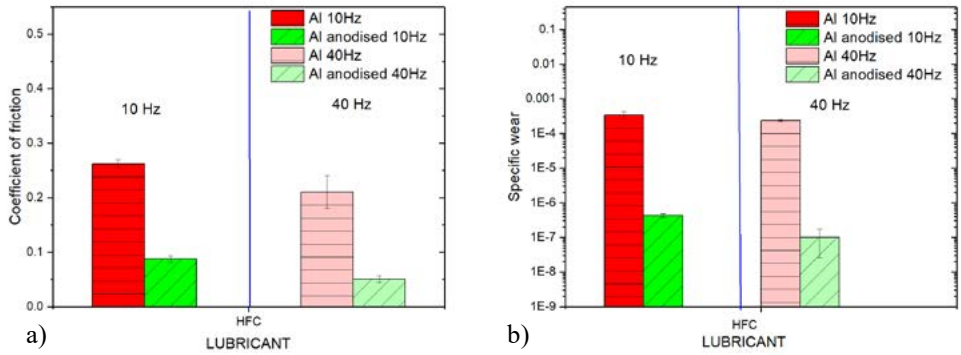


Figure 6. Average a) coefficient of friction b) specific wear of aluminium alloys pins after sliding against MS1 at two different sliding speeds in HFC 46

If we compare these results to the measurements in water and hydraulic oil, nature aluminium alloy responds similarly in water and HFC 46, with comparable values of coefficient of friction at both frequencies, and specific wear lower but comparable to the values in water. But after anodisation, measured coefficient of friction becomes comarable to the values measured in hydraulic oil, at both frequences. Also specific wear of anodised aluminum is one order (at 10 Hz) up to three orderes (at 40 Hz) lower compared to eater lubricated conditions, beeing cloaser to the values measured in hydraulic oil ISO VG46.

3.5. Surface analyses

After the tests, wear tracks of both pin and disc in sliding contact, were analyzed with digital microscope (Figure 7 and Figure 8). From the pictures it is clear, that the most significant wear and material removal area is evident in water at higher sliding speed, for all materials. Also, for the nature aluminium pins the stretches in the sliding direction are the most intensive at both sliding speeds, thick and prominent, suggesting abrasion as main wear mechanism.

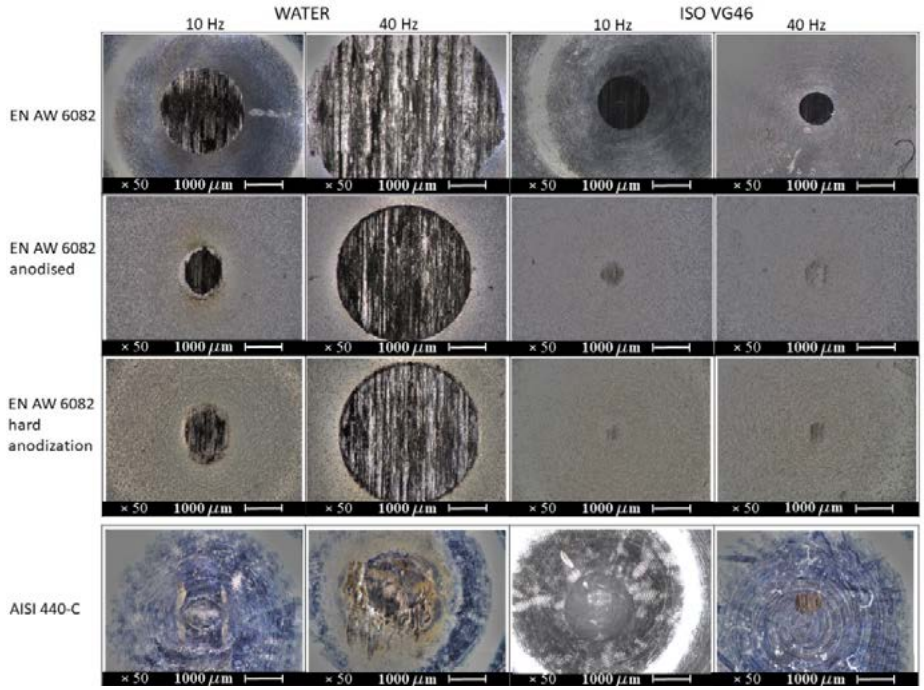


Figure 7. Digital micrographs of aluminium alloys pins after sliding against MS1 in water and hydraulic oil, at two different sliding speeds (magnification, ×50)

Aluminium alloy samples after both anodization and hard anodization suggest relatively smaller wear area, with area comparable to AISI 440-C, but different wear mechanism present. There were irregularities, that probably present peeled anodised protective layer in water, thus suggesting adhesive wear followed by two body abrasion. These adhesive debris and intensive delamination unevenly distributed over abrasive grooves were even more prominent at the four times higher sliding speed (increasing from 0.05 to 0.2 m/s). In case of AISI 440-C only some slight, rare unevenly distributed abrasive scratches were present.

The difference between nature and anodised aluminum is even more significant in hydraulic oil. Nature aluminum alloy shows stratches in sliding direction, less intense and thinner than in water, but comparable at both sliding speeds. In case of anodized and hard anodized aluminium in oil, only some irregular mocroploughings and pores are presents, even with smaller coverage area at higher sliding speed. These microplughings and microcracings are even less intense and with smaller coverage area in both anodized aluminium alloys comaperd to AISI 440-C.

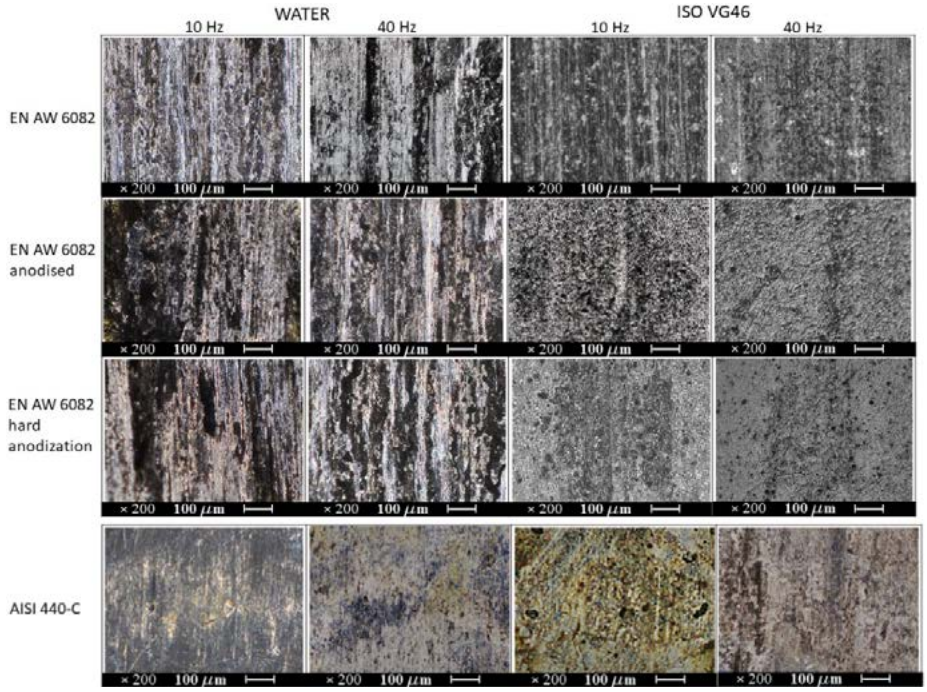


Figure 8. Digital micrographs of aluminium alloys pins after sliding against MS1 in water and hydraulic oil, at two different sliding speeds (magnification ×200)

The difference between nature and anodised aluminum is even more significant in HFC 46 (Figure 9), where the wear tracks of nature aluminum are comparable to the tracks seen in water at higher frequency. However, after anodisation wear track are comparable to the measurements in hydraulic oil, at both frequences. In case of nature aluminium alloy beside prominent intensive abrasive stratecs a lot of irregular embaded adhesive debris is pressent at both sliding speeds, with large wear area even compared to water lubricated conditions.

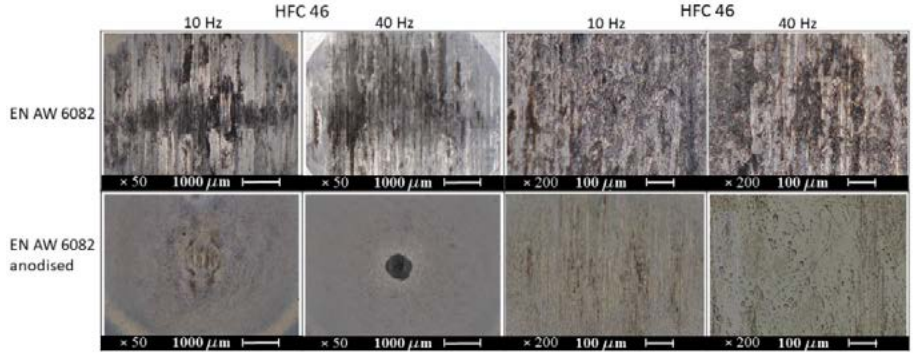


Figure 9. Digital micrographs of aluminium alloys pins after sliding against MS1 in HFC 46, at two different sliding speeds (magnification ×50 and ×200)

On contarry, in case of anodized aluminium alloy in HFC 46 relativelly small wear area was present at both slidin speeds, with some slight scratches at lower sliding speed and more

microplugings and pores at higher sliding speed. However all observed material changes are comparable to the wear tracks seen in case of hydraulic oil (Figure 7 - 8).

After the tribological tests the surfaces of MS1 discs were also analysed (Figure 10 - 12). The anodization (both standard and hard) improved tribological properties, and decreased contact area compared to nature aluminium, the effect is even greater in oil at both sliding frequencies (Figure 10).

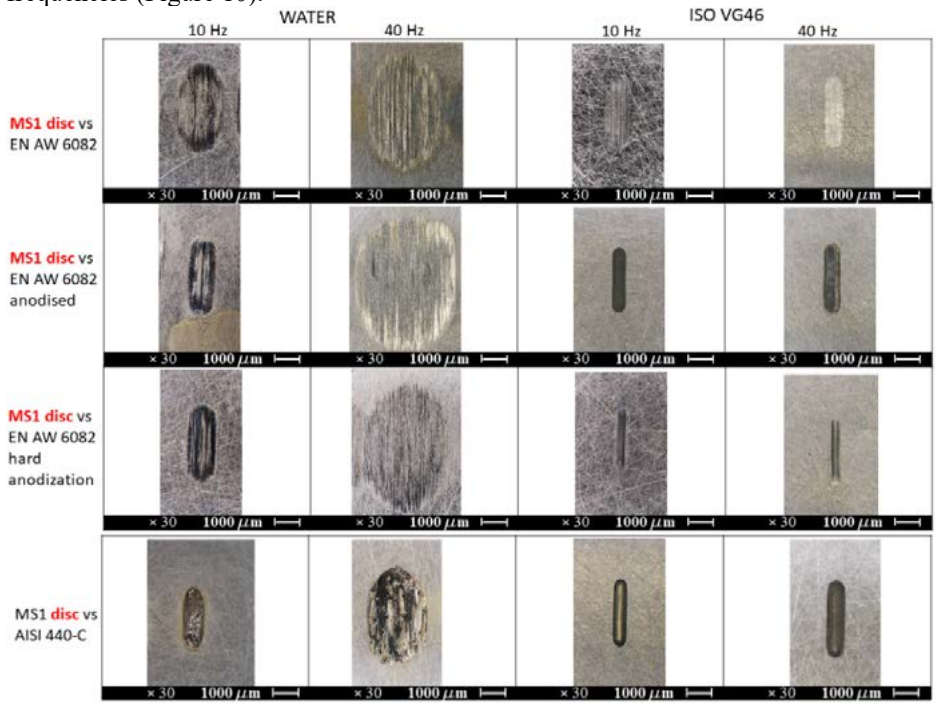


Figure 10. Digital micrographs of MS1 discs after sliding against aluminium alloys pins in water and hydraulic oil, at two different sliding speeds (magnification ×30)

The wear track widths at the lower sliding speed were generally comparable across most materials (~1070–1400 µm), with the exception of natural aluminium, with significantly wider track (~2300 µm) However, at higher sliding speed in water wear tracks become more prominent and intense for all tested materials *(~3500 µm)*. Under oil lubrication, the narrowest wear tracks were produced by hard-anodized aluminium pins at both sliding speeds (~350–450 µm). These were followed by C440 and conventionally anodized aluminium pins, which generated wear tracks of roughly 600–800 µm. In contrast, natural aluminium pins produced wider tracks, reaching ~1300 µm. The impact of higher sliding speed on MS1, is not significant in oil.

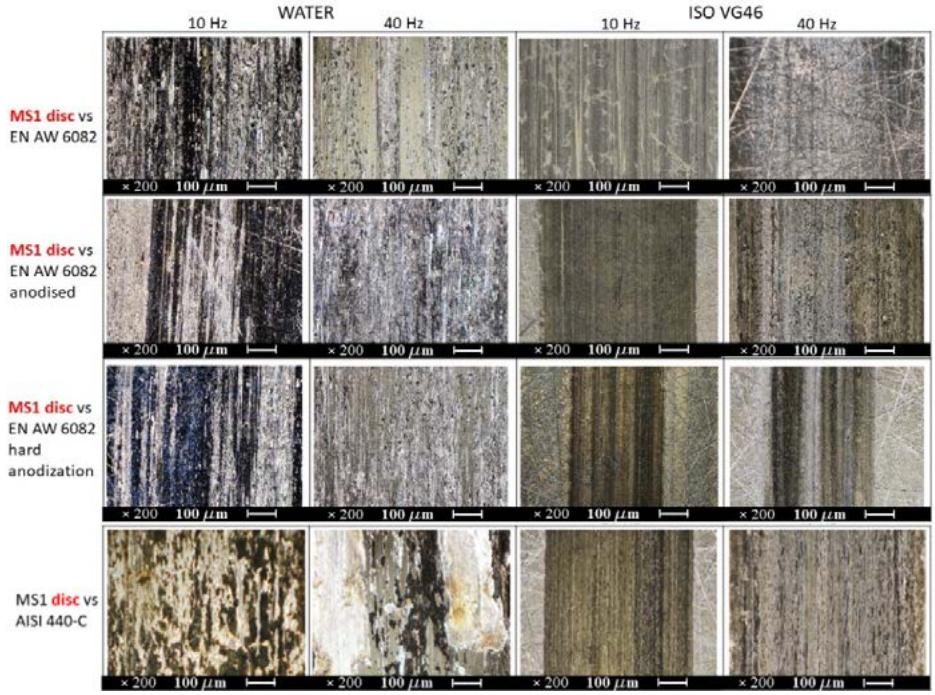


Figure 11. Digital micrographs of MS1 discs after sliding against aluminium alloys pins in water and hydraulic oil, at two different sliding speeds (magnification ×200)

At higher magnification (Figure 11) scratches and micro ploughing in sliding direction can be observed in all aluminium alloys in both water and hydraulic oil. Only the density and track width is much lower after anodization. In case of AISI 440C steel, intense delamination and adhesion is present especially at higher sliding speed. In case of HFC 46 (Figure 12), scratches are present in sliding direction, however the most significant difference is in the coverage area of the wear track between nature and anodized aluminium alloy.

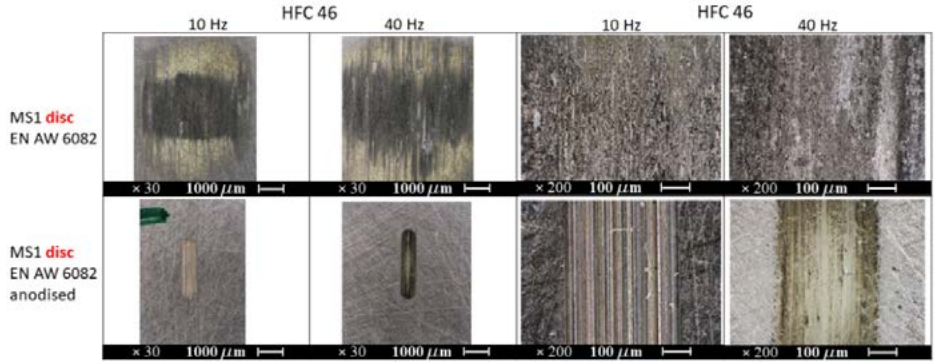


Figure 12. Digital micrographs of MS1 discs after sliding against aluminium alloys pins in HFC, at two different sliding speeds (magnification ×30, ×200)

4. CONCLUSIONS

In this study, the tribological properties of 3D-printed maraging steel were tested and analysed after sliding against three different versions of aluminium alloy AW 6082 at two different sliding speeds. The tests were conducted in water and compared to standard testing conditions in hydraulic oil ISO VG 46, as well as to sliding against the typical material used in hydraulic valves, nitride AISI 440C. A partial comparison was also made to the commercial water based lubricantHFC 46. The results of the tribological tests were used to analyse the influence of the surface treatment of the aluminium alloy and the effect of sliding speed when sliding against MS1, for potential material pairs in water hydraulics.

Sliding speed had no significant effect on the tribological performance of the tested materials in oil. Conventional anodization notably enhanced the surface properties of the aluminium alloy, achieving values comparable to those of AISI 440C steel. In contrast, under water lubrication, sliding speed emerged as a critical factor. At lower speeds, anodization substantially improved the specific wear resistance of the aluminium alloy, yielding results similar to those observed in steel. However, at higher sliding speeds, severe wear occurred on all aluminium specimens, although anodized samples still exhibited better performance than untreated ones.

The coefficient of friction in water was generally lower for the aluminium alloys compared to AISI 440C steel. While anodization did not significantly reduce the friction coefficient, it effectively stabilized the frictional response, producing smoother curves and more consistent sliding behaviour at both speeds.

Overall, the findings suggest that anodized aluminium alloys can be effectively employed in sliding contact with maraging steel, even under pure water lubrication, provided that the operational sliding speed is carefully controlled. Preliminary tests conducted in HFC 46 fluid further indicate that anodized aluminium in contact with MS1 steel demonstrates tribological performance comparable to that in hydraulic oil. This material pairing therefore holds promise for applications in standard hydraulic systems, offering a potential lightweight alternative to conventional steel components.

Funding

The presented research results were funded by Slovenian Research and Innovation agency (ARIS), grant No. P2-0231 and No. L2-50083

REFERENCES

1. Ralls AM, John M, Noud J, et al (2022) Tribological, Corrosion, and Mechanical Properties of Selective Laser Melted Steel. Metals (Basel) 12:1732. https://doi.org/10.3390/met12101732
2. Orgeldinger C, Seynstahl A, Rosnitschek T, Tremmel S (2023) Surface Properties and Tribological Behavior of Additively Manufactured Components: A Systematic Review. Lubricants 11:257. https://doi.org/10.3390/lubricants11060257

3. Guo L, Zhang L, Andersson J, Ojo O (2022) Additive manufacturing of 18% nickel maraging steels: Defect, structure and mechanical properties: A review. J Mater Sci Technol 120:227–252. https://doi.org/10.1016/J.JMST.2021.10.056
4. Bai Y, Wang D, Yang Y, Wang H (2019) Effect of heat treatment on the microstructure and mechanical properties of maraging steel by selective laser melting. Materials Science and Engineering: A 760:105–117. https://doi.org/10.1016/J.MSEA.2019.05.115
5. Klimek A, Kluczyński J, Łuszczek J (2023) Performance Analysis of Additively Manufactured Hydraulic Check Valves with Different Postprocessing. Materials 16:7302. https://doi.org/10.3390/ma16237302
6. Peng T, Wang Y, Zhu Y, et al (2020) Life cycle assessment of selective-laser-melting-produced hydraulic valve body with integrated design and manufacturing optimization: A cradle-to-gate study. Addit Manuf 36:101530. https://doi.org/10.1016/j.addma.2020.101530
7. Li D, Dai N, Wang H, Zhang F (2023) Mathematical Modeling Study of Pressure Loss in the Flow Channels of Additive Manufacturing Aviation Hydraulic Valves. Energies (Basel) 16:1788. https://doi.org/10.3390/en16041788
8. Hofmann U, Fankhauser M, Willen S, et al (2023) Design of an additively manufactured hydraulic directional spool valve: an industrial case study. Virtual Phys Prototyp 18:. https://doi.org/10.1080/17452759.2022.2129699
9. Guenther E, Kahlert M, Vollmer M, et al (2021) Tribological Performance of Additively Manufactured AISI H13 Steel in Different Surface Conditions. Materials 14:928. https://doi.org/10.3390/ma14040928
10. Zhu Y, Zou J, Chen X, Yang H (2016) Tribology of selective laser melting processed parts: Stainless steel 316 L under lubricated conditions. Wear 350–351:46–55. https://doi.org/10.1016/j.wear.2016.01.004
11. Guse F, Voshage M, Schleifenbaum JH, Schmitz K (2023) Cavitation Erosion Resistance of Additively Manufactured Materials. Chem Eng Technol 46:37–44. https://doi.org/10.1002/ceat.202200340
12. Casati R, Lemke JN, Tuissi A, Vedani M (2016) Aging behaviour and mechanical performance of 18-Ni 300 steel processed by selective laser melting. Metals (Basel) 6:. https://doi.org/10.3390/MET6090218
13. Bai Y, Yang Y, Xiao Z, Wang D (2018) Selective laser melting of maraging steel: mechanical properties development and its application in mold. Rapid Prototyp J 24:623–629. https://doi.org/10.1108/RPJ-05-2017-0104
14. Özer G, Khan HM, Tarakçi G, et al (2022) Effect of heat treatments on the microstructure and wear behaviour of a selective laser melted maraging steel. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering 236:2526–2535. https://doi.org/10.1177/09544089221093994
15. Vayre B, Vignat F, Villeneuve F (2012) Metallic additive manufacturing: State-of-the-art review and prospects. Mechanics and Industry 13:89–96. https://doi.org/10.1051/MECA/2012003
16. Venkata Krishna D, Singh J, Kiran Kumar K, Ravi Sankar M (2025) Tribological performance of 18Ni-300 maraging steel produced by DMLS technique. Tribology - Materials, Surfaces & Interfaces 19:214–226. https://doi.org/10.1177/17515831251336967
17. Tonolini P, Marchini L, Montesano L, et al (2022) Wear and corrosion behavior of 18Ni-300 maraging steel produced by laser-based powder bed fusion and conventional route. Procedia Structural Integrity 42:821–829. https://doi.org/10.1016/j.prostr.2022.12.104
18. Trajkovski A, Bartolj J, Spačal Ž, Majdič F (2025) Sliding contacts of 3D-printed direction control valves - tribological performance of suitable material pairs. In: The 19th Scandinavian International Conference on Fluid Power, SICFP'25. River Publishers, Denmark, pp 456–465
19. Musza A, Windisch M, Török M, et al (2023) Tribological Examination of Anodized Al-356 for Automotive Use. Coatings 13:1642. https://doi.org/10.3390/coatings13091642

20. Rodriguez L, Vieu A, Balsarin M, et al (2023) Physico-chemical characterisation and tribological behaviour of ground micro-arc oxidation coating on aluminium alloy – Comparison with hard anodised oxidation. Wear 516–517:204591. https://doi.org/10.1016/j.wear.2022.204591
21. Soffritti C, Fortini A, Nastruzzi A, et al (2021) Dry Sliding Behavior of an Aluminum Alloy after Innovative Hard Anodizing Treatments. Materials 14:3281. https://doi.org/10.3390/ma14123281
22. Bolelli G, Lusvarghi L, Barletta M (2009) HVOF-sprayed WC–CoCr coatings on Al alloy: Effect of the coating thickness on the tribological properties. Wear 267:944–953. https://doi.org/10.1016/j.wear.2008.12.066
23. Miramontes JC, Gaona Tiburcio C, García Mata E, et al (2022) Corrosion Resistance of Aluminum Alloy AA2024 with Hard Anodizing in Sulfuric Acid-Free Solution. Materials 15:6401. https://doi.org/10.3390/ma15186401
24. Nakase T, Sakuragi K (2017) Surface Treatment of Aluminium Alloy for Tribological Applications. KYB Technical Review 55:18–24
25. Polajnar M, Čoga L, Sharma SK, Kalin M (2025) Laboratory tribological screening test methodologies to evaluate the performance of metalworking lubricants for punching stamping. Tribol Int 211:110815. https://doi.org/10.1016/j.triboint.2025.110815
26. Gangwani P, Emami N, Kalin M (2025) Tribological behaviour of nano-titanium dioxide filled UHMWPE composites with a variety of micro fillers based on carbon, boron nitride and silicon dioxide under water-lubricated condition. Tribol Int 204:110479. https://doi.org/10.1016/j.triboint.2024.110479
27. Kalin M, Polajnar M, Kus M, Majdič F (2019) Green Tribology for the Sustainable Engineering of the Future. Strojniški vestnik – Journal of Mechanical Engineering 65:709–727. https://doi.org/10.5545/sv-jme.2019.6406
28. Strmčnik E, Majdič F, Kalin M (2019) Water-lubricated behaviour of AISI 440C stainless steel and a DLC coating for an orbital hydraulic motor application. Tribol Int 131:128–136. https://doi.org/10.1016/j.triboint.2018.10.032
29. Trajkovski A, Novak N, Pustavrh J, et al (2023) Performance of Polymer Composites Lubricated with Glycerol and Water as Green Lubricants. Applied Sciences 13:7413. https://doi.org/10.3390/app13137413
30. Trajkovski A, Matkovič S, Novak N, et al (2024) Glycerol aqueous solutions for the enhanced tribological behaviour of polymer composites sliding against steel. Tribol Int 192:109173. https://doi.org/10.1016/j.triboint.2023.109173
31. Mehdizadeh M, Akbarzadeh S, Shams K, Khonsari MM (2015) Experimental Investigation on the Effect of Operating Conditions on the Running-in Behavior of Lubricated Elliptical Contacts. Tribol Lett 59:6. https://doi.org/10.1007/s11249-015-0538-x
32. Jiao W, Liu W, Zhu Z, et al (2024) Effect of Loading Pressure on Tribological Properties for Valve Plate Pair during the Running-In Process. Tribology Transactions 67:247–258. https://doi.org/10.1080/10402004.2024.2320693
33. Ivantysyn R, Shorbagy A, Weber J INVESTIGATION OF THE WEAR BEHAVIOR OF THE SLIPPER IN AN AXIAL PISTON PUMP BY MEANS OF SIMULATION AND MEASUREMENT
34. Horn S, Ivantysyn R, Weber J Tribo-optimized lubricating interfaces in hydrostatic pumps with surface shaped slippers
35. Shorbagy A, Ivantysyn R, Weber J (2023) Holistic Analysis of the Tribological Interfaces of an Axial Piston Pump – Focusing on the Pump Efficiency. Chem Eng Technol 46:5–13. https://doi.org/10.1002/ceat.202200450

Biographies



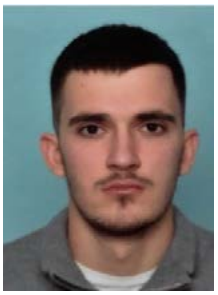
Ana Trajkovski received the PhD in Mechanical Engineering from University of Ljubljana in 2015. She is currently working as teaching assistant at the Department of Tribology and maintenance systems, Laboratory for Fluid Power and Controls, Faculty of Mechanical Engineering, University of Ljubljana. Her research areas include inovative materials and lubricants for sustainable hydraulics, smart control design and optimization of hydraulic components, water hydraulics. She has written many scientific papers and collaborated in many national and international conferences and research projects.



Jan Bartolj received the M.Sc. degree in Mechanical Engineering from University of Ljubljana in 2024. He is currently working as teaching assistant at the Department of Tribology and maintenance systems, Laboratory for Fluid Power and Controls, Faculty of Mechanical Engineering, University of Ljubljana. His research areas include optimisation of hydraulic components, additive manufacturing in hydraulics and water hydraulics.



Nejc Novak received the PhD in Mechanical Engineering from University of Ljubljana in 2025. He is currently working as researcher at the Department of Tribology and maintenance systems, Laboratory for Fluid Power and Controls, Faculty of Mechanical Engineering, University of Ljubljana. His research areas include analyses of hydraulic components wear mechanisms and lifetime, hydraulic fluid properties and filtration.



Jan Pustavrh received the M.Sc. in Mechanical Engineering from University of Ljubljana in 2022. He is currently working as researcher at the Department of Tribology and maintenance systems, Laboratory for Fluid Power and Controls, Faculty of Mechanical Engineering, University of Ljubljana. His research areas include design and control of multi-axis servo hydraulic systems, smart control and design of hydraulic components.



Mitjan Kalin is full professor and the head of the Department of Tribology and maintenance systems, at the Faculty of Mechanical Engineering, University of Ljubljana. He has written more than 200 international peer-review papers, gave more than 80 invited lectures at international conferences and is project leader of more than 50 major projects. His main

research areas are in wear and friction mechanisms of advanced materials, green lubrication technologies and nanoscale interface phenomena. In particular, he is recognised in academia and industry as a pioneer in boundary lubrication and tribochemistry of diamond-like carbon (DLC) coatings. Prof. Dr. Kalin was elected as a member of Slovenian Academy of Engineering, a Member of the International Academy of Engineering, and in 2024 to European academy of science and arts. He received a 2025 ASME Mayo D. Hersey Award, one of the highest academic recognitions in tribology.



Franc Majdič received the PhD in Mechanical Engineering from University of Ljubljana in 2010. Hi is currently working as assoc. professor at the Department of Tribology and maintenance systems, at the Faculty of Mechanical Engineering, University of Ljubljana and is the head of the Laboratory of Fluid power and controls. His research areas include water hydraulics, hydraulic components development and testing. He has written many scientific papers and collaborated in many international conferences. He is a project leader of more than 50 industrial and research projects and collaborated in several industrial patents.

Performance evaluation of a wear test rig for positive displacement pumps

Pascal Moor[1], Maximilian M. G. Kuhr[1], Christopher Albrecht[2], Peter F. Pelz[1]

[1] Chair of Fluid Systems – TU Darmstadt, pascal.moor@tu-darmstadt.de
[2] Leistritz Pumpen GmbH, 90459 Nürnberg

Abstract.

The operation with abrasive particles is not recommended for most types of rotating displacement pumps. For these pumps, the scientific knowledge of the wear behaviour and the influence of the many parameters on the wear behaviour is sparse. This is especially true for multi-screw pumps. Therefore, the chair of fluid systems has developed a test rig for scientific research of the wear in rotating displacement pumps with a focus on repeatability of the test parameters. The practical challenges in achieving such a setup and the results of the commissioning of the test rig are shared in this paper in order to aid likeminded researchers in their development of a suitable test setup.

Keywords. wear, abrasion, test rig, displacement pumps, screw pumps.

1. INTRODUCTION

In order to cater to the industrial markets and remain relevant within globally competitive markets, component and system manufacturers need to meet all three dimensions of quality to their customers satisfaction, namely effort, availability and acceptability [1]. If the manufacturers fail to do so, they lose the ability to develop new markets and even have to face being pushed out of existing markets. Many types of positive displacement pump have historically only been used for non-abrasive applications. With the expansion of markets however, the pump manufacturers increasingly face applications with abrasive particles suspended in the pumped fluids. These applications pose a significant challenge for the manufacturers and operators of the pumps. From a manufacturer's perspective, only strict adherence to the cleanliness requirements allows for optimal life expectancy, while from the operator's perspective, these levels of cleanliness are difficult or impossible to achieve in many applications. Especially for pumps in continuous production processes with particle-laden products e.g., in the chemical or food industries, the abrasiveness poses a constant challenge due to high wear rates in the pumps, thus limiting service life.

For such applications, pumps based on soft sealing partners, such as membrane pumps or lobe pumps are often chosen. These soft sealing mechanisms however limit the maximum efficiency that can be achieved per pumping stage even in optimal operation. Furthermore, the pressure levels that can be achieved with soft sealing mechanisms are limited [2]. For

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

diaphragm pumps, pump handbooks give a pressure limit of 17 bar for standard applications [3]. Hard sealing mechanisms as present in gear, vane, screw and lobe pumps allow for higher pressure levels, are however more susceptible to wear, when critical abrasive particles are present. The particles most critical are abrasive particles in a size close to the respective gap height, since these particles are likely to be entrained into the sealing gap by the relative motion of the sealing partners [4]. The sealing gaps are widened by wear, thus increasing internal leakage and reducing volumetric efficiency [5].

Wear to the sealing gaps can be described as a superposition of abrasive and erosive wear [4]. Abrasive wear, especially by means of three-body abrasion results from particles slightly larger than the gap height being drawn and rolled through the gaps when relative motion of the sealing partners is present. Erosive wear is caused by the flow through the sealing gaps, driven by the pressure difference. Abrasive particles smaller than the gap height are carried along by the leakage flow across the sealing gaps and cause erosion to the sealing partners. This mechanism does not require relative motion of the sealing partners [6].

The scientific literature does not offer adequate insight into the wear behaviour for many of the aforementioned pump types with hard sealing mechanisms when operated with abrasive particles. Wear in general cannot be characterised as a component's property, but is always dependent on the composition of the system that the components operate in. Schmidt [7] showed in her work on the wear behaviour in two-spindle screw pumps, that a correlation between the wear behaviour observed in standardised tribometer tests cannot be drawn to predict the wear behaviour in the field tests of the pumps. She performed a small series of experiments in an effort to analyse the wear behaviour. The tests were performed with water and 2 %-mass quartz-sand with a particle size of 40 – 150 µm (mean: 86 µm) at a pressure difference of 14 bar. The pressure was developed along the 3.3 chambers. With her work she showed the importance of performing wear experiments as close to the real application as possible, in order to not loose relevance to the real-world application.

Even when testing whole pumps, but removed from the real-world application, the pumps may show different wear behaviour when compared to real world applications, as was shown by the preliminary work for ISO/DIS 9632 [8] shows. Here pumps with similar real world wear behaviour showed different wear behaviour in the controlled test environment. This highlights the importance of well documented test rig setups for the repeatable experimental investigation of rotating displacement pumps.

The research paper that is presented here, covers the commissioning of a test rig, that has been developed in order to perform experimental research on the wear in pumps with hard sealing mechanism, as discussed above. The development of the test rig setup has been discussed in the scientific community at the screw machines conference in 2024 [9]. The presenting authors of this research paper aim at making their path to a newly developed test rig setup transparent. The test rig needs to offer a repeatable test environment. Due to logistical limitations and a lack in practical experience regarding the behaviour of such a test rig, strict control of all parameters for every experiment cannot be achieved. The repeatability with regards to the test results therefore relies on a combination of controlled parameters and monitoring of other parameters in a well-documented test environment, as will be elaborated on within this paper.

2. DESIGN AND COMMISSIONING OF THE TEST RIG

In order to achieve a test environment that allows for repeatable test parameters and offers the wide variability of adjustable parameters, the test rig has been designed with special focus on the conditioning of the suspension and on minimising dead-volumes in the pipelines. The test rig is comprised of two circuits, which are fed by a common suspension tank. The main circuit includes the measurement environment for the test pumps. The filtering circuit aids in the conditioning of the suspension. The conceptual design of the test rig is shown in figure 2.1. and explained below. Many of the design decisions could be evaluated during the commissioning phase.

2.1. Conceptional Design of the Test Rig

For the main circuit, the pipelines are kept short, while also considering the critical bend radius for the hoses and minimising the transitions in cross-sectional area. The suction side of the pump is connected to the tank, which is used for the preparation of the suspension and the temperature control. The mass flow as well as the density of the suspension is measured by a Coriolis-type flow meter located in the suction line. The difference in fluid temperature and in pressure are measured across the pump. Pump speed and torque are measured by a torque measuring shaft, which is coupled to a safety clutch that disconnects the pump from the motor in the event of a pump jamming. By means of these measurements, the operating point of the pumps can continually be monitored and changes to the pump behaviour caused by accumulative wear can be detected.

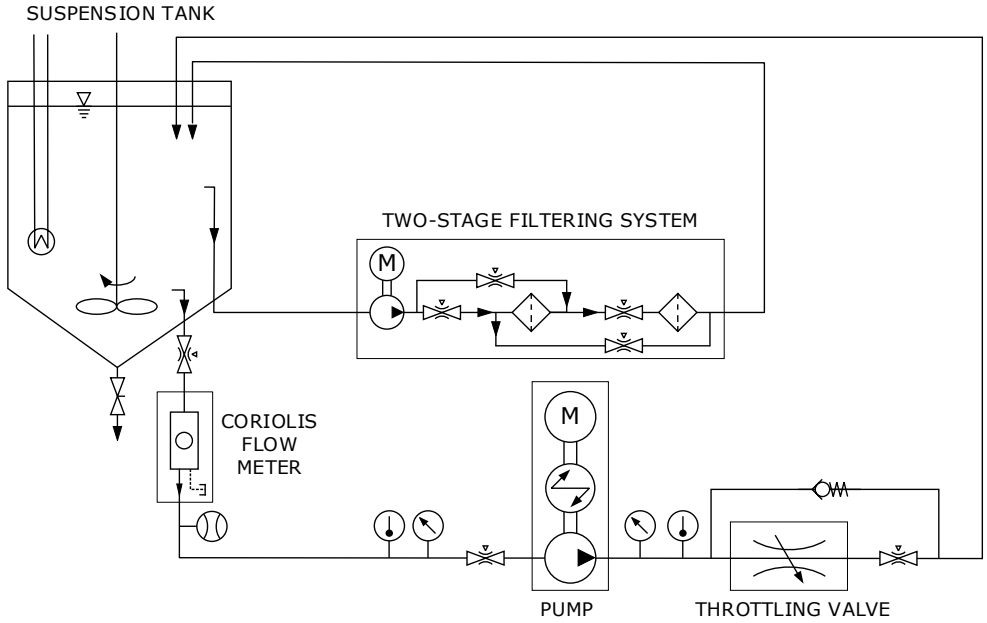


Figure 2.1. Hydraulic schematic of the test rig setup for wear analysis on pumps. [9]

In order to adjust the operating point of the pump, an adjustable pinch valve is used as a throttle. The valve, as shown in figure 2.2, is mechanically actuated and restricts the flow by pinching a flexible sleeve. After passing the throttle, the suspension can flow back to the

suspension tank. The outlet of the return line into the tank is oriented against the direction of flow in the tank caused by the mixer, in order to aid the mixing by disturbing the flow. The suction inlet is oriented in a way that the pump draws in the suspension against the direction of flow caused by the mixer.



Figure 2.2. Mechanically actuated pinch valve. [10]

The filtering section needs to allow for ease of assembly and maintenance. Therefore, a routing was chosen that allows for the sections that need to be drained, to remain small and with a low point at the filters. The filtering unit is comprised of a pump and two filters that are operated in series. The suspension is first pumped through a coarse filter in order to separate the large particles, that are still sharp and most relevant for the wear in the sealing gaps. For the fine filter, a 5 µm filter was chosen, instead of the 10 µm filter that was discussed in the previous paper in order to aid the oil cleanliness.

The experimental tests focus on the wear to the spindles and liner, thus for each test a new set of spindles and liner will be used. This pump assembly will then be operated at one operating point for each respective test. The operating point is defined by rotational speed, pressure, oil viscosity, particle concentration, particle size distribution and particle type. Each of these parameters can be set independently. The requirements and limitations that had to be faced in the design of the test rig, as well as the components that were chosen for the construction of the test rig are covered in depth by the previous research paper [9]. Based on the conceptual design, the test rig has been built up and first commissioning tests were able to be performed, offering first insights into the behaviour of the test rig and additional challenges that have to be faced.

2.2. Test Procedure

The test procedure consists of a fluid pre-conditioning cycle, the wear test, and a detailed cleaning cycle to provide a reproducible and uncontaminated environment for the next tests. Within the previous publication, this test procedure has not been shared yet and is therefore discussed below.

The pre-conditioning cycle of the fluid is split up in two phases. First the fluid is brought up to temperature, while the particles remain in a sedimented state. In order to aid the homogenous temperature distribution throughout the test rig, the pump is operated at slow speed to allow heating not only the fluid in the suspension tank but also in the pipeline assembly. Additionally, the mixer is operated at slow speed in order to avoid hot zones near the heating element. Care must be taken, to not bring the particles into suspension, that have

settled at the bottom of the tank since the previous test. When the desired test temperature of 40 °C is reached, the temperature is held at a constant level for the subsequent wear test. After the temperature pre-conditioning phase, the pump speed as well as the operating pressure are increased up to the operating point of the pump. Once the operating point has been reached, the mixer speed is increased in order to bring the particles into suspension, the wear test of the pump has started. The particle concentration in the suction line of the pump can continuously be monitored by means of the suspension density.

After the predefined test time the test pump is to be stopped, marking the end of the wear test. Now, the mixer and temperature control are turned off. Therefore, the pipelines are filled with particle-laden suspension. The particles remain in the pipelines and have a chance to settle at any given place in the test rig. This sediment needs to be removed from the pipelines in order to avoid introducing unsteady particle concentrations into the pumps during subsequent tests. Since disassembly of the test rig is not feasible in between all tests, the particles need to be flushed out of the pipeline, while in place. In-situ flushing of pipelines is common practice, especially in food industries or pharma applications, often called Cleaning-in-Place (CiP) [11]. By means of often turbulent flow of a cleaning aid, the pipes are flushed and sediments removed from the system [11]. Flushing the measuring section of the wear test rig in a similar way prohibits itself for two reasons. Firstly, the pipeline diameters of the test rig have specifically been chosen large in order to reduce erosive wear to the pipeline components by reduction of the flow velocity. Thus, turbulent flow can only be achieved with very high volume flows, which cannot be reasonably provided by the size of pump that would fit the available space or budget. Secondly, the particles shall remain in the suspension tank during the flushing cycle, in order to reduce effort in the test preparation and aftercare. Solutions with a suction side filter in the tank were not considered, since the size required to provide sufficient filter capacity would not fit the available space. Instead, the suspension is given time for sedimentation. Clean fluid for flushing is thus made available by allowing the particles to settle at the bottom of the tank. In order to not agitate the particles that have already settled at the bottom of the tank, the flushing cycle can only operate with low volume flow.

Instead of turbulent flushing, a concept based on mechanical agitation of the particles in the pipelines was chosen. A pipe cleaning spiral, fitted with a nylon brush, allows to agitate the particles, bringing them back into suspension. The suspended particles are then carried back to the tank by a low volume flow. In order to provide the necessary volume flow, a secondary flushing pump is used, that can be mounted in place of the test pump. With the flushing pump in operation, the cleaning spiral is inserted into the suction port in the tank and fed into the pipeline until the brush reaches the suction port of the pump. Thereby the sediments from the pipeline assembly can be flushed back to the tank while minimising disturbance to the sedimented particles in the tank.

This flushing concept is only possible thanks to the specific design of the pipeline assembly, which is based on the guidelines for food industries. Namely by avoiding dead spaces as well as sharp bends and aiming for smooth transition of diameters, it is possible for the brush and cleaning spiral to move freely through the pipes. Thanks to the design of the suction port of the secondary flushing pump, there is no risk of catching parts of the cleaning spiral or the brush in the moving parts of the pump. For changing the particle type, the return lines can be flushed accordingly by inserting the brush into the return line.

3. PERFORMANCE EVALUATION

The measures described above all focus on producing a repeatable test environment. The following paragraphs will show the performance of the systems and allow for an evaluation with regards to the repeatability of the test parameters.

3.1. Temperature Pre-Conditioning

The operating temperature for the test rig is 40 °C. The temperature is controlled to within ± 2 °C by a 9 kW heating element and two tube coil heat exchangers in the suspension tank. The temperature control allows for the test fluid to be preheated to 40 °C before starting the tests. For the preheating of the measuring section, the test pump is operated at slow speed, in order to not agitate the settled particles in the tank. The wear test is initiated by increasing pump speed and pressure to set the operating point and ramping up the mixer speed in order to bring the particles into suspension. Dependent on the environmental conditions at the time of the test and the operating point, the hydraulic losses may be great enough to keep the test rig at 40 °C or even heat up the test fluid. The temperature controller therefore adjusts the volume flow of coolant through the tube coil heat exchanger, thus allowing for continuous control of the temperature. Figure 3.1 shows the temperature in the tank and at the pump inlet during the pre-conditioning, confirming, that the temperature pre-conditioning cycle ensures repeatable test parameters, allowing for a consistent oil viscosity and density to be considered for the wear tests.

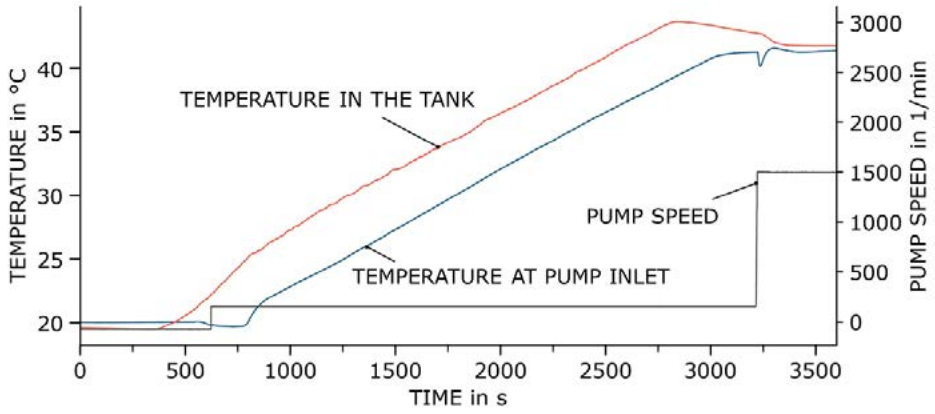


Figure 3.1. Fluid temperature pre-conditioning cycle and run-up of the pump.

3.2. Particle Concentration

The temperature pre-conditioning cycle discussed above must not introduce a critical amount of abrasive particles into the measuring pump. Additionally, the particle concentration introduced into the pump during the run-up of the pump and the mixer needs to be evaluated. The introductory paragraphs mention, that not all parameters can be controlled within the test rig. The particle concentration that is introduced into the pump is the most important parameter that can only indirectly be influenced by changes to the mass of particles in the tank and the run-up phase. Therefore, it is very important to monitor and document the particle concentration during all states and consider the findings when drawing conclusions from the measurement results.

Preliminary tests have shown, that with respect to the cleanliness achievable by the flushing cycles and the accuracy achievable in the measurement of the particle concentration, the temperature pre-conditioning cycle can be considered to not contribute to introducing particles into the measuring section. Fluid samples taken near the pump inlet in the tank after the pre-conditioning phase show, that no critical amount of particles was brought into suspension.

Figure 3.2 shows a pre-conditioning cycle after the oil has been heated to 40 °C with the pump and the mixer operated at slow speed, in order to provide uniform heating throughout the measuring section and the tank. Then, the pump is run-up to 1500 1/min in order to set the operating point, before the mixer speed is increased, in order to bring the particles into suspension. The density drops slightly below the expected reference value of 0.852 ± 0.002 kg/l during the temperature pre-conditioning cycle. The reference value was obtained using a Lauda TD3 tensiometer at a temperature of 40 °C. Once the pump has reached its operating speed of 1500 1/min, the density approaches the reference density, indicating that the pump does not stir up a critical amount of abrasive particles in the suspension tank.

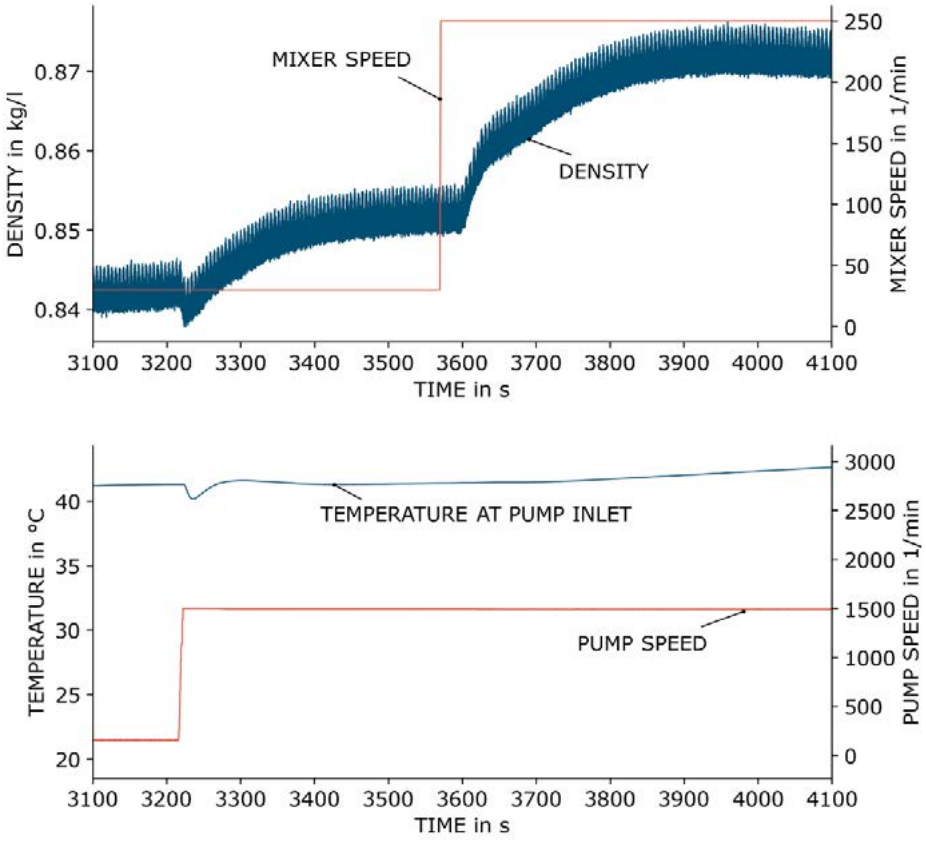


Figure 3.2. Suspension density and temperature at the pump inlet during pre-conditioning.

The density of clean oil at the same pump speed is shown in figure 3.3. Checking against the measurement in clean oil confirms, that for the run-up of the pump to 1500 1/min is safe to assume, that no critical amount of particles are stirred up from the bottom of the tank.

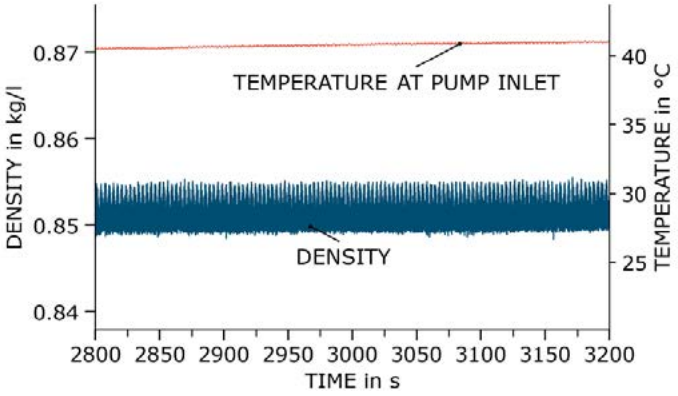


Figure 3.3. Clean oil density and temperature at the pump inlet.

Additionally, a fluid sample was taken from the tank, after the pump was run-up to 1500 1/min. After sufficient time for sedimentation of the sample, it could be confirmed, that no significant amount of particles was present in the oil. Thereby, both the temperature conditioning as well as the run-up cycle of the pump fulfil the requirements necessary for a repeatable test environment. This can also be confirmed by means of a sight glass in the suction line before the pump. Additionally, the fact that test pumps operated within these conditions did not show any signs of wear and upon disassembly did not show any signs of binding. Binding of the spindles within the liner due to particles within the small clearances has been observed with all pumps that were operated at test concentrations. The pre-conditioning phase can thus be considered to allow for sufficiently clean oil at the start of the tests, ensuring the repeatability of the tests.

After setting the operating point, the mixer speed is increased, in order to bring the particles into suspension. Figure 3.2 shows that the density increases. From the density, the particle concentration can be calculated. The supplier of the glass granule particles used for the commissioning tests, gives a range of 2.0 … 2.6 kg/l for the particle density. [12] Assuming a particle density of 2.2 ± 0.2 kg/l and an oil density of reference value of 0.852 ± 0.002 kg/l, a particle concentration of 1.53 % by mass can be calculated with a spread of +0.42 %, -0.31 %, matching the nominal particle concentration of 1.5 % by mass.

3.3. Flushing Cycle

Without clean pipelines at the start of a test, all other efforts to ensure a repeatable test environment are in vain. Unknown particle agglomerations, especially on the suction side may detach undetected and enter the pump. Thereby, the wear in the pump would no longer be able to be traced to known parameters. Since the density measurement only gives information about the density of the flow through the sensor, a manual inspection of the achievable cleanliness in the pipes has to be performed. The design of the test rig is such that critical dead volumes are minimised. Therefore, the remaining critical sections are the sharp transitions in diameter at the hose fittings and the step down in diameter at the pump

inlet. These critical sections were checked by disassembly and showed, that no particle agglomerations were to be found.

Additionally, the density measurements of clean oil in the test rig at an ambient temperature of 21 ± 1 °C were compared to the density after flushing at the same temperature. The density in clean oil was 0.868 ± 0.004 kg/l, while the density after flushing was 0.870 ± 0.004 kg/l. Thus, the flushing cycle offers sufficient cleanliness for the tests.

3.4. Cavitation

The commissioning test showed, that at high volume flows and high pressure drops ($\Delta p > 8$ bar), cavitation could be noticed. The cavitation presented as vibrations within the hydraulic hoses, that were transferred back to the test pump. The origin could be traced to the pinch valve. The high pressure drop at the pinch valve that is used to set the operating pressure of the pump, leads to cavitation within the pinch valve, which transmits the vibrations to the hose assembly of the test rig. These high system vibrations highlighted an issue in the mounting solution for the Coriolis-type flow meter, since vibrations from the return hose were able to be transmitted into the mount of the flow meter. The flow meter, which operates in a frequency range of $400 \ldots 600$ Hz thus gave inconsistent measurements for the density of the suspension. Therefore, the flow meter was mounted more rigidly and isolated from transmission of vibrations. Thereby the negative effect on the measurements was omitted. Further, in order to reduce vibrations within the test rig and reduce cavitation erosion to the pinch valve, concepts for further throttling were developed and will be added to the test section.

4. CONCLUSIONS AND OUTLOOK

Based on the results of the preliminary and commissioning tests, the test rig can be set up for the parameter studies to analyse the wear behaviour in screw pumps. Final adjustments to the throttling of the measuring section and to the filtering section can be performed. The performance of the test rig will continuously be monitored, and future changes documented in order to avoid the propagation of systemic changes to the experimental findings. The tests that will be performed, aim at the starting phase of wear in screw pumps in order to enhance the knowledge of critical parameters and better understand safe vs. critical operating points.

The experiences in the design and commissioning process of the test rig are further discussed below in order to aid future researchers by providing a condensed lessons learned. Future experiences in the operation and upkeep of the test-rig are intended to be shared as well.

Some practical findings to keep in mind during the design of a similar wear test rig, that became important during the commissioning tests are summarized below.

In order to reduce particle agglomerations within the pipeline system, smooth transitions are to be chosen, when variations to the inner diameter of pipes and connectors cannot be avoided. Adhering to that principle made the flushing concept described above possible. However, it can be noted that the step at the transition from the hydraulic hose to the barbed hose fitting leads to increased settling of particles at that transition. Therefore, for applications with particle suspensions it is recommendable to design the leading edge with sufficient taper.

Wear will not only affect the measuring section, but the filtering unit as well. Cumulative wear will lead to a decrease in volume flow, thus reducing the filtering rate. This reduction in the filtering rate can currently not be characterised since the filtering section is not equipped with a flow sensor. For ensuring consistent filtering, it is recommendable to add a flow sensor to the filtering section.

The suspension tank has been chosen with a focus on adaptability and dead volume free design. Therefore, a cylindrical tank from the food industries has been chosen that is equipped with a dished bottom and total drain. The tank has been equipped with two tube coil heat exchangers that are nested. These are oriented concentric to the central mixer and mounted to the wall of the tank. The mounting hardware serves double duty. They affix the coils in the tank and at the same time act as baffles preventing the formation of a vortex as a result of the centrally mounted mixer. Some of the mounts break through the fluid-air interface, thus allowing air entrapment into the flow around the mounts at high flow rates of the pump. The influence on the experimental results is expected to be minor, since the air bubbles remain close to the surface and are few in number. Nevertheless, it is advisable to avoid disturbing the fluid air interface to mitigate air entrapment.

Further focus needs to be put on the wear to the test rig itself. While it is designed with low wear rates to the components in mind, the test rig needs to be checked for wear periodically. Especially the throttling components, that are subject both to cavitation and highly erosive conditions need to be checked frequently. A standardised, sacrificial wear body may be added to the outlet of the pipeline assembly in order to pose as an indicator for cumulating wear in the test rig components.

Acknowledgement

We would like to thank Leistritz Pumpen GmbH for funding of and the substantial contributions to the project.

5. REFERENCES

[1] P. F. Pelz, P. Groche, M. E. Pfetsch, and M. Schaeffner, 'Mastering Uncertainty in Mechanical Engineering'. Cham: Springer International Publishing, 2021, ISBN: 978-3-030-78353-2.

[2] G. Vetter and W. Wirth, 'Understand Progressing Cavity Pumps Characteristics And Avoid Abrasive Wear', 1995, doi: 10.21423/R1F10N.

[3] I. J. Karassik, Ed., *Pump handbook,* 3rd ed. New York: McGraw-Hill, 2001, ISBN: 0070340323.

[4] G. Vetter, R. Kießling, and W. Wirth, 'Abrasive Wear In Pumps - A Tribometric Approach To Improve Pump Life', 1996, doi: 10.21423/R18H5T.

[5] G. Vetter, W. Wirth, H. Korner, and S. Pregler, 'Multiphase Pumping With Twin-Screw Pumps - Understand And Model Hydrodynamics And Hydroabrasive Wear', 2000, doi: 10.21423/R19X1R.

[6] K. Sommer, R. Heinz, and J. Schöfer, 'Verschleiß metallischer Werkstoffe'. Wiesbaden: Springer Fachmedien Wiesbaden, 2018, ISBN: 978-3-658-17850-5.

[7] S. Schmidt, 'Verschleiß von Schraubenspindelpumpen beim Betrieb mit abrasiven Fluiden'. Dissertation, Technische Fakultät, Universität Erlangen-Nürnberg, Erlangen, 1999.

[8] A. Crook, 'Contaminant Sensitivity of Hydraulic Pumps – Development of the ISO/DIS 9632 Test Method', *SAE Transactions*, vol. 100, pp. 50–53, 1991. [Online]. Available: http://www.jstor.org/stable/44471022

[9] P. Moor, M. Kuhr, and P. Pelz, 'Test rig setup for particle wear analysis in screw pumps', *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 1322, no. 1, p. 12013, 2024, doi: 10.1088/1757-899X/1322/1/012013.

[10] DUAL VALVES, 'Dual Pinch Valve Open Body Type'. [Online]. Available: https://dualvalves.com/wp-content/uploads/2025/04/DPVO.png (accessed: Jan. 6 2026).

[11] G. Sansebastiano, R. Zoni, and L. Bigliardi, 'Cleaning and Disinfection Procedures in the Food Industry General Aspects and Practical Applications' in *Food Safety*, A. McElhatton and R. J. Marshall, Eds., Boston, MA: Springer US, 2007, pp. 253–280, doi: 10.1007/978-0-387-33957-3_13.

[12] CL TRADE SERVICES, 'Strahlmittel - Glasgranulat für Strahlanlagen'. [Online]. Available: https://www.cltradeservices.de/strahlmittel/glasgranulat/ (accessed: Oct. 17 2025).

Biographies



Pascal Moor received the bachelor's degree in mechanical and process engineering from Technische Universität Darmstadt in 2019 and the master's degree in mechanical and process engineering from Technische Universität Darmstadt in 2022. He is currently working as a PhD researcher at the Chair of Fluid Systems, Faculty of Mechanical Engineering, Technische Universität Darmstadt. His research focuses on the experimental investigation of wear in rotating displacement pumps. He had a first opportunity to serve as a reviewer for a research journal in 2025.



Maximilian M. G. Kuhr received the bachelor's degree in Mechanical and Process Engineering from Technische Universität Darmstadt in 2013, the master's degree in mechanical and process engineering from Technische Universität Darmstadt in 2016, and the doctorate degree in Mechanical and Process Engineering from Technische Universität Darmstadt in 2022, respectively. He is currently working as Head of the Tribology and Cavitation Group at the Chair of Fluid Systems, Faculty of Mechanical Engineering, Technische Universität Darmstadt. His research areas include Rotor Dynamics, Cavitation, Wall Slip and Research Data Management. He has been serving as a reviewer for many highly-respected journals.

13



Peter F. Pelz received the degree of Dipl.-Ing. in mechanical engineering from Technische Universität Darmstadt in 1993 and the doctorate degree in mechanical engineering from Technische Universität Darmstadt in 2000. He is currently working as a Professor (W3) at the Chair of Fluid Systems, Faculty of Mechanical Engineering, Technische Universität Darmstadt, while also fulfilling his responsibilities as Vice President for Digitalisation, Infrastructure and Sustainability at Technische Universität Darmstadt. In addition, he is active in his role as a founding spokesperson for the research team Energy and Environment as well as his role as spokesperson for Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften (NFDI4Ing). His research areas include Sustainable Systems Design, Anticipation of Water Demand Scenarios, Multi-agent Systems & Mathematical Optimisation, Cavitation and Generic Flow in Turbomachinery and Systems, Availability of Fluid Power Systems, Influence of Molecular Structure, Additives and Surface Roughness on Wall Slip and Designing with Fluids. He has been serving as a reviewer for many highly-respected journals.

Electrification-Ready Tribology: Experimental and Simulation Insights into Wear and Performance of Surface Optimized Valve Plates

Roman Ivantysyn[1], Ahmed Shorbagy[1], Prof. Dr. Jürgen Weber[1], Raghavendra Reddy, Reddy Reddy[2], Mathias Rauschenberger[2], Torsten Schmidt[3]

1) *TU Dresden, Institute of Mechatronic Engineering, Helmholzstrasse 7a, 01069 Dresden, Germany roman.ivantysyn@tu-dresden.de*
2) *Inline Hydraulik GmbH, 12277 Berlin, Germany, R.Reddy@inlinehydraulik.com*
3) *Fraunhofer IWU, Reichenhainer Straße 88, 09126 Chemnitz, Germany, torsten.schmidt@iwu.fraunhofer.de*

Abstract.

Axial piston pumps in electrified, variable-speed duty cycles increasingly operate at low speed and start–stop conditions where hydrodynamic lubrication margins are reduced, making the cylinder block/valve plate interface especially wear-critical. In this study, a previously optimized bronze meso-structured (sinusoidal) valve plate is re-evaluated under thermal steady-state conditions on the TU Dresden test rig at 40 °C and 60 °C inlet temperature, including very low-speed operation. The bronze design retains its efficiency benefit across both temperatures and shows no measurable additional wear, with surface changes remaining below 1 µm. Motivated by this wear-free behavior, a lead-free steel/steel material pairing is pursued. A new steel wave pattern is obtained by repeating the THED simulation optimization for the updated material pairing and was consequently manufactured and tested. Experimental results show strong efficiency improvements compared to reference and bronze configurations, but also reveal that at elevated temperature short contact events can rapidly escalate into failure. The findings confirm that meso-structured valve plates can be designed predictively for a given material system, additional research is needed to combine them with steel/steel pairings.

Keywords. Surface-structure, lead-free material pairing, electrification, axial piston pumps, valve plate, THED gap simulation.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Positive displacement machines remain key components in fluid power systems due to their high power density and suitability for high-pressure operation. Axial piston pumps are widely used because of their high efficiency, variable displacement capability, and compactness as well as robustness in most heavy duty industrial applications. The ongoing shift toward electrified drivetrains has altered operating profiles. Frequent start-stop events, extended partial-load phases, and operation at low shaft speeds are increasingly required, while simultaneously high efficiency and reduced noise are expected. Under such conditions, the hydrodynamic lubrication margins of the major tribological interfaces are reduced and mixed lubrication or local contact is more likely to occur, particularly at the cylinder block/valve plate interface, which is spring pre-loaded for high speed performance, further increasing bearing loads at low-speed conditions.

In parallel, regulatory and sustainability-driven constraints have intensified the transition toward lead-free materials. In conventional pump designs, lead-containing alloys and coatings have been used to support run-in and to tolerate occasional metal-to-metal contact during early life and adverse operating points. It has been shown that surface shaping and meso-structuring can be applied to actively manipulate the lubricant film, increase hydrodynamic support, and reduce both friction and leakage losses, especially in low-speed regimes. Previous research has demonstrated efficiency gains from wave-shaped hydrostatic bearings as well as a significant reduction in wear [1], [2], [3], [4].

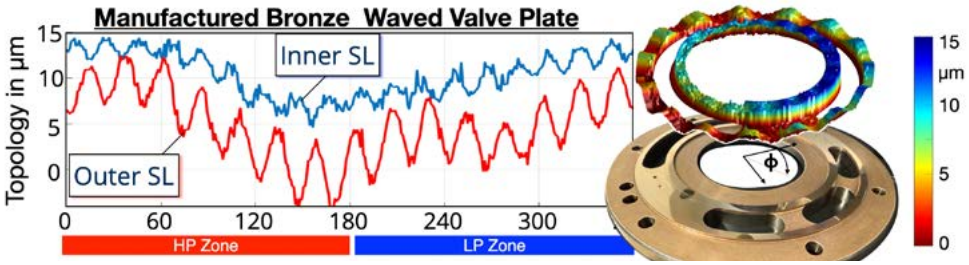


Figure 1.1. Manufactured and tested waved bronze valve plate [5].

In prior work, a meso-structured valve plate was designed using a thermo-elastohydrodynamic gap model in Caspar FSTI and was validated experimentally on an industrial test bench. A strong reduction in power losses (η_{total} +7%) at low speed was measured while high efficiency at nominal operating points was retained (η_{total} >90%). Importantly, the wave

structure remained intact after testing, and the run-in behavior typical of conventional designs was substantially reduced to remain below 1 µm total wear [5].

The present contribution extends the prior work in two directions. First, repeated measurements of the bronze meso-structured valve plate were performed under thermal steady-state conditions on the TU Dresden test rig at 40 °C and 60 °C inlet temperature, including very low-speed operation (<10 rpm), to confirm performance and wear behavior under controlled thermal boundary conditions. Second, the demonstrated low-wear behavior was used to motivate a next material step toward steel/steel pairings to avoid leaded materials and to reduce dependence on costly bronze/brass materials. Since the deformation behavior and lubrication conditions differ substantially with material pairing, a new micro-structure was designed by repeating the simulation optimization for steel/steel.

2. RESEARCH GOAL

The goal of this work is to validate and extend meso-structured valve plate technology for variable-speed axial piston pumps in a way that supports lead-free and cost-effective material strategies. Two specific objectives were pursued:

1. **Confirm improved thermal stability of the waved valve plate**: Simulation results indicated that the meso-structured valve plate experiences a superior thermal response when experiencing hotter fluid temperatures. In the industrial tests the fluid temperature was not maintained steady, not allowing for a definitive validation of this claim. Hence, these tests were repeated on a research test-rig at **40°C and 60°C inlet temperatures**, allowing for full thermal stability.
2. **Material transition and redesign for steel/steel**: Based on the wear-free behavior of the bronze design, a transition toward a **steel/steel** pairing was to be evaluated as a next step toward lead-free and cost-relevant pump development. Due to differences in material properties a new wave pattern had to be designed using the same simulation method as published previously [5], but with an updated material pairing. Prototype manufacturing and testing were to be performed to identify performance potential and limiting failure mechanisms.

3. TEST RIG AND MEASUREMENT METHODOLOGY

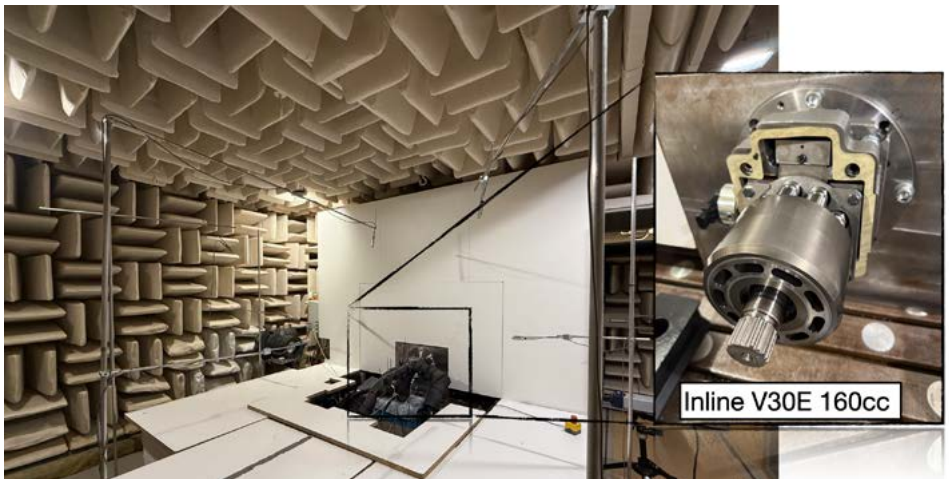


Figure 3.2. Test rig and 160cc test pump

The measurements for this work were performed in the pump test rig of the Institute of Fluid-Mechatronical Systems. This 250 kW test rig doubles as the anechoic sound measurement room, which is shown in Figure 3.2.

The pump used for this research is an V30E 160cc pump from the company Inline Hydraulik (also branded as HAWE), which is a high pressure industrial open circuit variable axial piston pump.

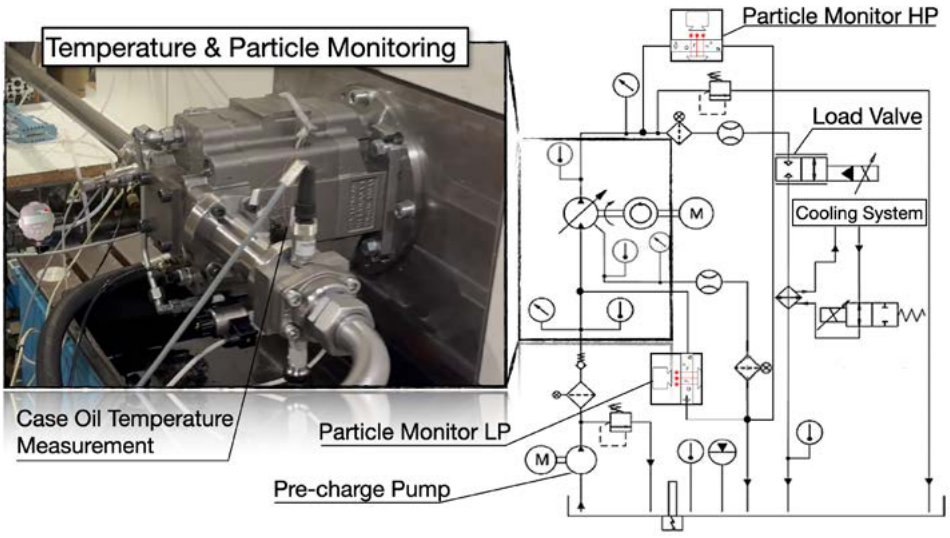


Figure 3.3. Schematic and sensor setup

The test rig features a 200 kW cooling system, capable of controlling the fluid temperature within ±1 K. In addition to the standard protocol for efficiency measurements such as pressure, temperature, flow and torque sensors, additional particle sensors as well as thermocouples were placed in strategic places (see Figure 3.3) to monitor pump performance and detect failures.

The measurement series were performed in the following order:

1. Reference measurements with a new pump (Ref1 - Kit1)
2. Repeat measurements of the used waved bronze valve plate (Bronze WVP- Kit1)
3. Unhardened waved steel valve plate (Steel WVP (S) - Kit1)
4. Hardened wave steel valve plate with a used rotating kit (Steel WVP (H) - Kit2)

The reason for the change in rotating kits, was that the unhardened steel wave had a significant failure contaminating the pump with metal particles. Hence, a rotating kit that was previously used for other measurements was used. Reference measurements were available for this replacement rotating kit, which will be compared to the Steel WVP (H) measurements only.

For each series the following measurement runs were performed:

Run1: Full operating span at 40˚C inlet
Run2: Full operating span at 60˚C inlet
Run3: Low Speed Tests
Run4: Repeat of full operating span at 40˚C inlet

Total test time: 25-30 hours.

4. PRIOR WORK AND STEADY MEASUREMENTS OF BRONZE VALVE PLATE

This paper is a continuation of the previously published meso-structured bronze valve plate [5]. This section will summarize the achievement of this paper and build upon them to lead up to the final results of the meso-structured bronze valve plate at steady state conditions.
Prior work on meso-/micro-structuring of lubricated interfaces in axial piston pumps has shown that tailored surface topographies can actively manipulate the fluid film to increase hydrodynamic load support and reduce both viscous

friction (hydromechanical) and leakage-related (volumetric) losses, especially under low-speed conditions where mixed lubrication is more likely.

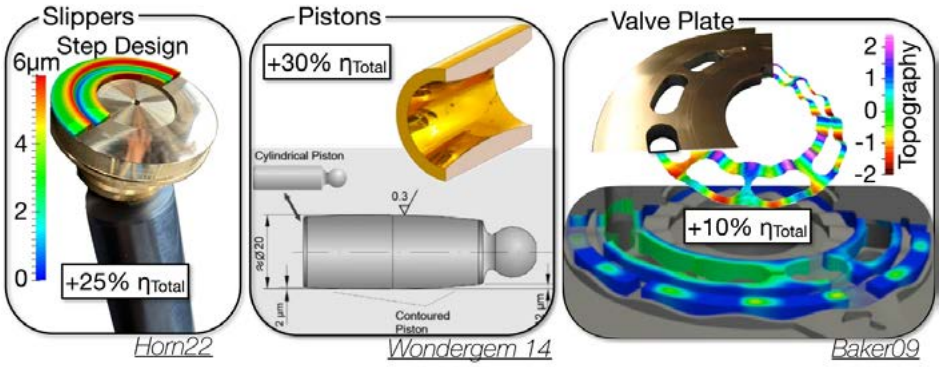


Figure 4.4. Previous published meso-structures in axial piston pumps

Across the three dominant loss-contributing interfaces—piston/cylinder, slipper/swash plate, and cylinder block/valve plate—different structure concepts and optimization targets have been published in the past and are visualized in Figure 4.4 [1], [2], [6]. The common outcome is improved low-speed efficiency with strongly reduced run-in wear when the structure is properly adapted to the interface physics. In particular, wave- or step-like meso-geometries were shown to provide measurable efficiency benefits at low speed while limiting wear to minor peak smoothing rather than bulk material removal.

Building on these findings, the present work focuses on the cylinder block/valve plate interface using the previously optimized waved bronze valve plate [5], and re-evaluates its performance under controlled *thermal steady-state* boundary conditions to remove the temperature transients inherent to the earlier industrial testing campaign.

4.1. Meso-Structured Waved Bronze Valve Plate at Steady State Conditions

Figure 4.5 summarizes the key efficiency benefit of the waved bronze valve plate at 25% displacement, showing the improvement in total efficiency (left) and the corresponding gain in hydromechanical efficiency (right) relative to the reference. The data indicate that most of the total-efficiency benefit originates from hydromechanical loss reduction, consistent with the intended mechanism of the wave structure in supporting the lubricating film. The four operating conditions used for the original design optimization are

superimposed and serve as fixed reference points for later power-loss breakdowns. The choice of 25% displacement is deliberate, since valve-plate-related losses contribute a larger fraction of total losses in this regime; therefore, a ~6 percentage-point total efficiency increase (in this case from 75% to 81%) is practically meaningful.

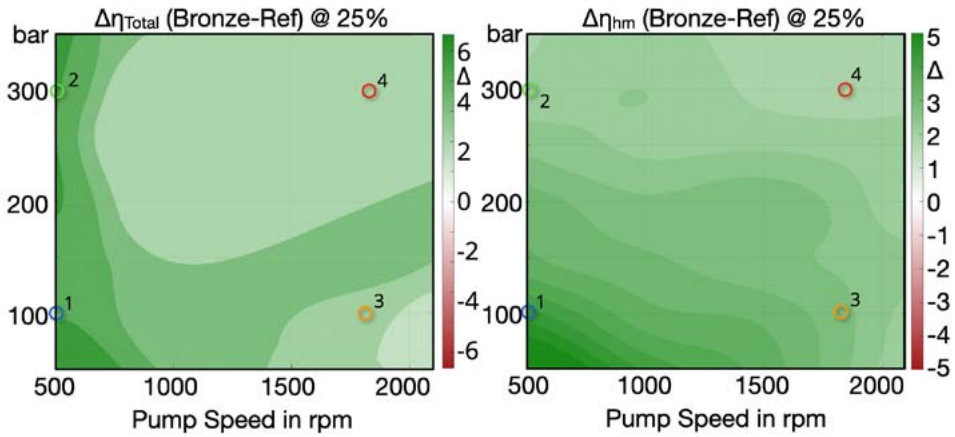


Figure 4.5. Improvement of the total and hm efficiency using a waved valve plate design in comparison to the reference at 25%

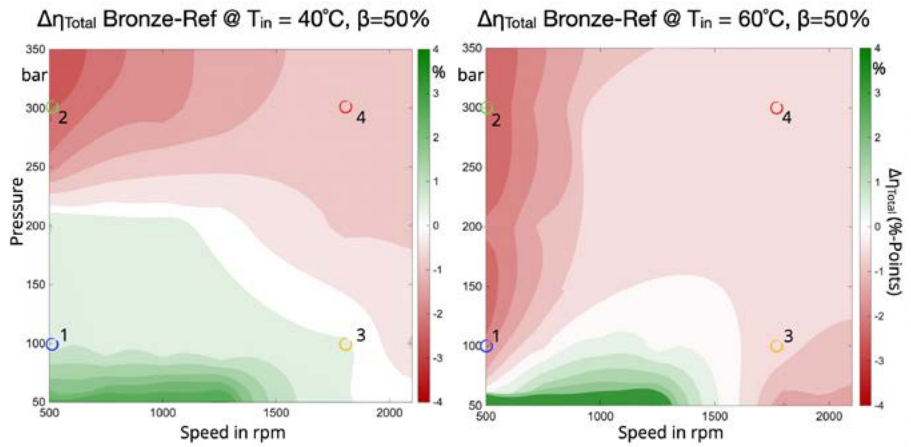


Figure 4.6. Difference between the measured total efficiency at 50% displacement for 40˚ and 60˚C inlet temperature between the meso-structured bronze VP and the reference.

Figure 4.6 provides the measured total-efficiency difference between the waved bronze and reference valve plates at 50% displacement for two inlet temperatures. The map indicates up to ~4 percentage-point improvements

primarily at low speed and low pressure, while the majority of the operating range remains close to unchanged (approximately ±1 percentage point). A localized efficiency reduction is visible in the low-speed/high-pressure region, and the overall efficiency level decreases with increased inlet temperature, most notably at higher pressures.
To connect the efficiency trends to energy terms, Figure 4.7 shows a stacked power-loss breakdown at the four design operating conditions for the reference and waved bronze valve plates at 40°C and 60°C inlet temperature. While total losses increase at 60°C for both configurations, the waved bronze valve plate shows a smaller loss increase than the reference (~15% lower increase), supporting the simulation-based expectation that the meso-structure provides improved thermal robustness of the valve-plate interface under hotter fluid conditions.

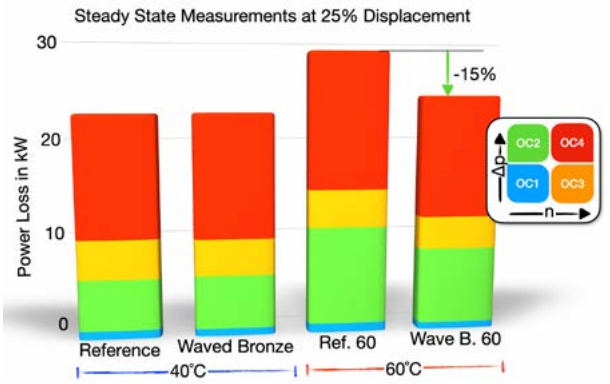


Figure 4.7. Measured efficiency improvement at 60˚C and 25% displacement across 4 main operating conditions.

This trend is further corroborated in Figure 4.8, where simulation and measurement are compared using a normalized representation: since the THED model predicts valve-plate interface losses rather than full-pump losses, values are normalized to the summed power loss at 40°C. The temperature-induced loss trend is captured well across most operating points, including a particularly strong agreement at low-speed conditions (OC1 and OC2). Deviations remain at higher-load points (notably OC3/OC4), but the agreement in both direction and magnitude is sufficient to validate the main thermal sensitivity trend relevant to variable-speed duty cycles.

In summary, the waved bronze valve plate demonstrates repeatable efficiency benefits under thermal steady-state conditions and shows a

reduced loss sensitivity to elevated inlet temperature; however, the most critical requirement for long-term applicability is that the small-amplitude structure remains intact, which is assessed next via detailed wear measurements.

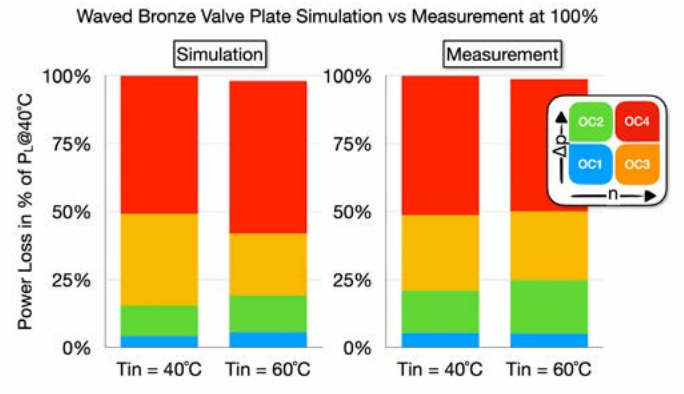


Figure 4.8. Comparison of predicted vs measured power losses at full displacement for 4 operating conditions at 40˚C and 60˚C normalized to the total power loss at 40˚C.

4.2. Evaluation of Surface Wear After Steady State Measurements

As the wave amplitudes are on the order of micrometers, even small amounts of wear can eliminate the functional hydrodynamic effect; therefore, *structure survivability* is a primary success criterion alongside efficiency. with no meaningful additional wear detected during the steady-state series.

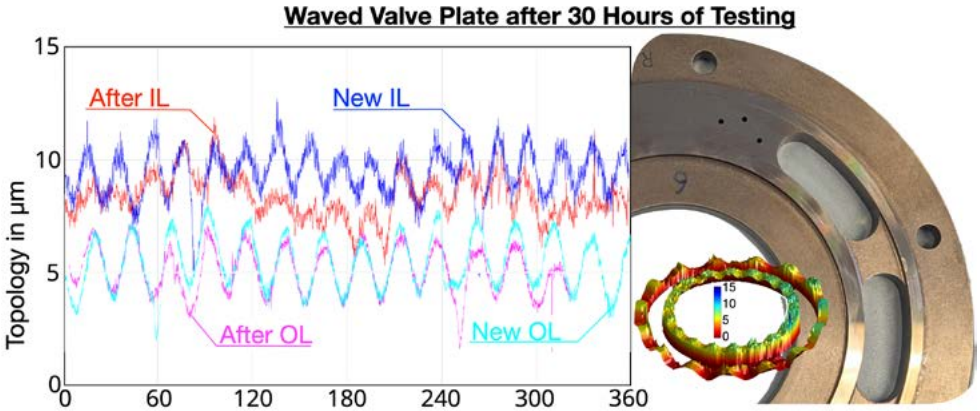


Figure 4.9. Meso-structured bronze waved valve plate new and after testing (IL = Inner sealing land; OL = Outer sealing land).

Figure 4.9 shows the waved bronze valve plate after the combined test exposure (industrial dynamic testing plus the additional steady-state campaign, totaling ~30 h). The 3D topography measurement confirms that the wave pattern remains clearly present across the sealing lands. The inner (IL) and outer (OL) sealing-land profiles, compared between the new condition and after testing, show only minor changes; while slight plastic deformation is visible—primarily at the IL—the wave form remains intact, and total material change stays below 1 µm. Importantly, the valve plate was measured after the industrial tests and again after the steady-state campaign,

The wear contrast to the reference valve plate is highlighted in Figure 4.10. Note that the color scales differ: the reference plate exhibits material loss up to approximately 5 µm at the inner sealing land and additional localized losses on the outer land, whereas the waved bronze valve plate primarily shows limited peak smoothing at the IL with the OL wave pattern largely preserved with material loss under 2 µm. Darker regions on the OL are predominantly attributable to plastic deformation rather than removal of the wave geometry, as can be seen in the surface profile in Figure 4.9.

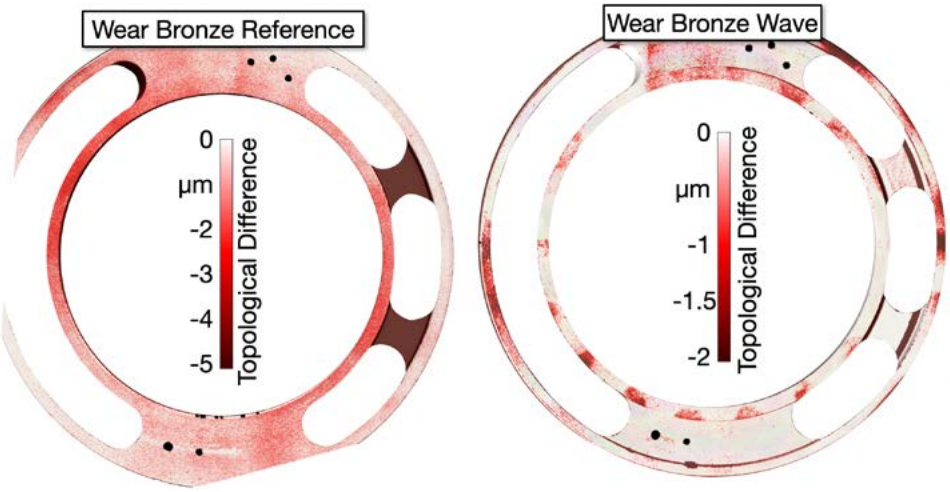


Figure 4.10. Material loss of both vale plates after industrial and steady state testing as compared to the new state.

In summary, the waved bronze concept not only improves partial load efficiency but also suppresses run-in wear to the micrometer level, motivating the hypothesis that a lead-free, cost-relevant steel/steel pairing could be feasible—provided the wave geometry is re-optimized for the

fundamentally different material and deformation behavior; the following section therefore describes the simulation-based redesign workflow for steel/steel.

5. STEEL/STEEL SIMULATION AND OPTIMIZATION

The valve plate design simulations were performed within the Caspar FSTI framework, which is a powerful simulation software tailored for gap design in axial piston pumps [7]. The used model of the pump was described in detail here [5], [8], [9], [10] and validated using temperature measurements of the cylinder block [8], [9], [11].

The meso-structured wave geometry discussed in Section 4 was originally designed for the bronze WVP configuration used in the previous study. Importantly, this component is not a monolithic bronze part, but a bi-metal configuration in which the valve plate body is predominantly steel with a bronze bearing layer (~0.5mm). This material pairing influences the thermo-elastic deformation response of the valve plate: the two materials exhibit different thermal expansion behavior and stiffness, such that the resulting thermal crowning and local gap shape can differ significantly from a full steel/steel pairing. Consequently, a wave pattern that is optimal for the bi-metal configuration cannot be assumed to remain optimal—or even stable—when transferred to a monolithic steel valve plate.

To evaluate whether the bronze-derived wave concept can be transferred with minimal redesign, two parametric studies were performed on a full steel valve plate by varying only the sealing-land radii (i.e., the hydrostatic bearing area), while keeping the kidney geometry unchanged. The inner sealing land diameter was varied from d_{GapIn} = 87.5 to 89 mm and the outer sealing land diameter from d_{GapOut} = 125.5 to 128 mm, and the four previously introduced design operating conditions (OC1–OC4) were simulated in a full DOE. In the first study, a *positive-wave* pattern (IL+1µm with +5µm offset; OL+3µm) was used, as this was the optimal design found for the bronze VP [5]. The positive wave is essentially the full wave without the negative amplitude. In a second study the full-wave pattern (IL±1 µm with +5µm offset; OL±3 µm, see Figure 1.1) was also analyzed on the steel VP. This full wave pattern reflects the realized waved bronze VP, which was chosen due to manufacturing limitations.

Figure 5.11 summarizes the results at T_{in}=35˚C as power-loss maps over the design-of-experiments domain for each operating condition. The color scale

indicates the magnitude of predicted power losses (note that the scale differs between OCs), while grey regions denote non-converged simulations. Non-convergence occurred primarily when contact was predicted or when the cylinder block lifted off, both indicating an unstable design for the tested geometry. The results demonstrate that the majority of radius combinations fail in at least one operating condition. Only a single candidate (Des20) converged in all four OCs; however, the lowest-loss region is shifted toward smaller values of d_{GapIn} and d_{GapOut} where designs such as Des72 approached favorable loss levels but still failed to converge in OC4. As a reference the bronze WVP had 88.5mm and 127.4 mm as d_{GapIn} and d_{GapOut} respectively. Nevertheless, both Des72 and Des20 showed problems at 60˚C inlet, a temperature when the full wave pattern was superior.

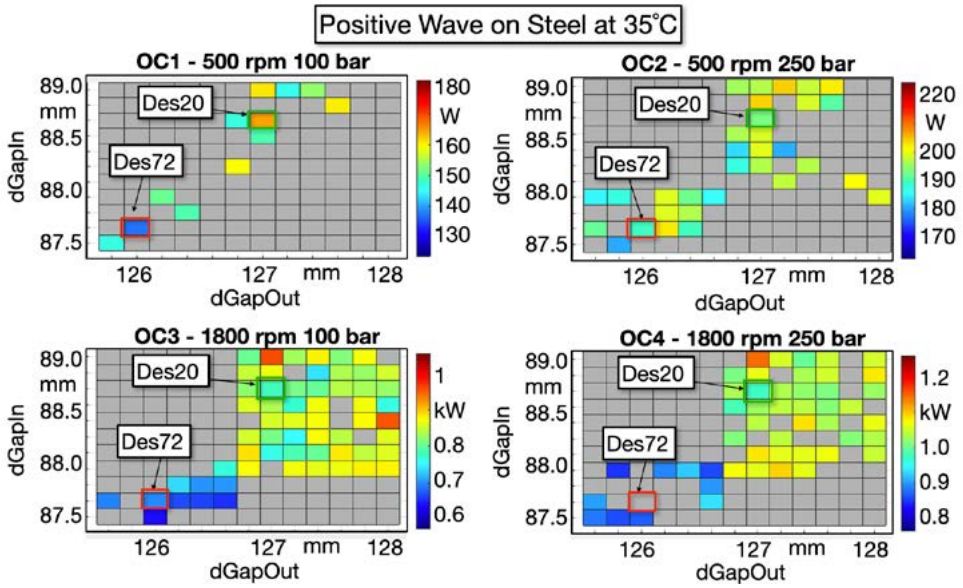


Figure 5.11. Powerlosses with variation of the sealing land width on a steel VP for a positive wave design adapted from the waved bronze VP

The results of the full wave pattern at T_{in}=35˚C are shown in Figure 5.12. While the loss levels were, on average, reduced compared with the positive-wave sweep, the same fundamental limitation remained: only one design converged across all four OCs, and the best-performing low-loss candidate (Des33) still failed in OC4. This indicated that the stability margin of the hydrostatic balance in steel/steel is highly sensitive and that “best loss” and “robust across OCs” do not coincide when only the sealing-land radii are varied. Des 33 converged in all OCs at 60˚C inlet temperature, making it the base for further amplitude studies.

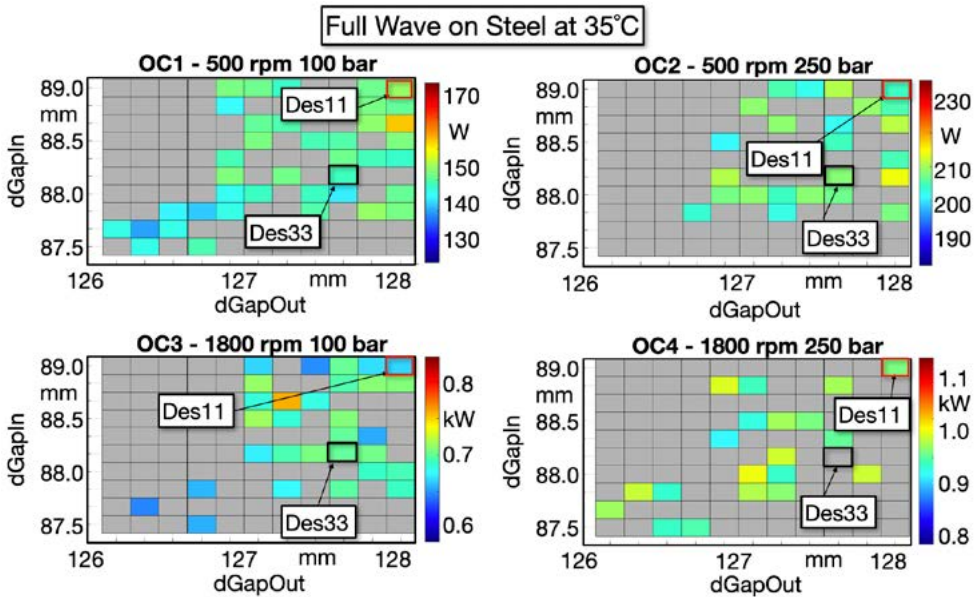


Figure 5.12. Powerlosses with variation of the sealing land width on a full steel VP for a full wave design adapted from the waved bronze VP

Since the bronze steady-state measurements in Section 4 revealed a strong temperature dependence, both geometrical studies were repeated at an elevated inlet temperature of $T_{in} = 60°C$. The outcome was a qualitative change in behavior: the formerly robust positive-wave candidate (Des20) failed at higher temperature (already in OC3), whereas Des33 became stable and exhibited very low predicted losses at elevated temperature.

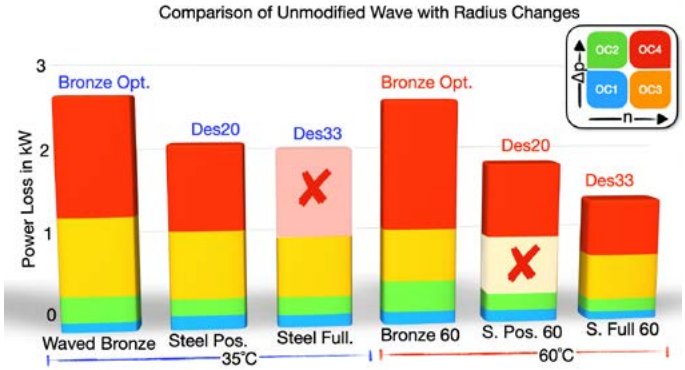


Figure 5.13. Direct adaption of full and positive wave from bronze design to a steel VP adjusting the sealing land width only.

Figure 5.13 condenses these findings by comparing stacked valve-plate power losses over the four operating conditions for essentially the same wave

pattern, only (differentiating between positive and full wave) for the bronze and steel VP at inlet temperatures of 35°C and 60°C. While the steel VP can, in principle, achieve lower valve-plate losses than the bronze design at both temperatures, none of the directly transferred configurations provided a single radius choice that remained stable and convergent across all four operating conditions and both inlet temperatures. In other words, the transfer produced attractive "point solutions," but no robust design envelope.

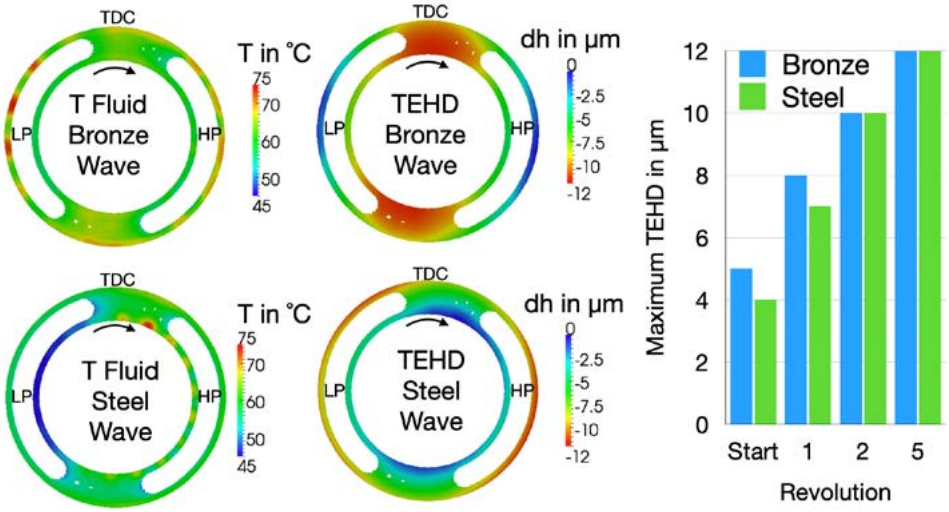


Figure 5.14. Thermal deformation effect on the fluid film between bronze and steel waved valve plate at OC4 (1800 rpm, 300bar, 100%)

The underlying cause of this temperature-dependent instability is the fundamentally different thermo-elastic deformation pattern of the full steel VP compared with the bi-metal bronze configuration. This is illustrated in Figure 5.14 for OC4, where the simulated fluid-film temperature distribution (left) and the gap-height change caused by thermo-hydrostatic-elastic deformation (THED contribution shown in the middle) are shown for both material systems. Although the *magnitude* of thermal deformation is comparable, the *spatial distribution* differs drastically: the bi-metal configuration produces a deformation pattern that creates a low spot in regions between the HP and LP kidney around TDC/BDC, whereas the full steel VP experiences the highest points in these critical areas. Because the wave pattern is designed to interact constructively with the thermo-elastic deformations, this shift explains why a geometry optimized for the bi-metal plate can become unstable when applied to steel/steel—particularly at OC4, where thermal gradients and pressure loading are most severe.

On the right side of the figure is the maximum THED of the valve plate with the simulation revolution. The simulation framework calculates the thermal conditions after each completed revolution and updates the thermal boundary conditions; thus, each revolution is associated with a modified thermal deformation field. When a design completes the first 3 revolutions and then fails, it is usually due to thermal deformations. This will be relevant later in this section.

As the direct adaptation of the bronze wave pattern onto the steel VP was shown to be unwise, a new wave optimization was performed. Due to time constraints, the optimization scope was restricted to a full-wave concept with the same phase and frequency as the bronze design, while varying predominantly the wave amplitudes and offsets. This restriction was motivated by prior sensitivity results indicating that amplitude is typically the dominant parameter, while phase and frequency have secondary influence for the bronze configuration.

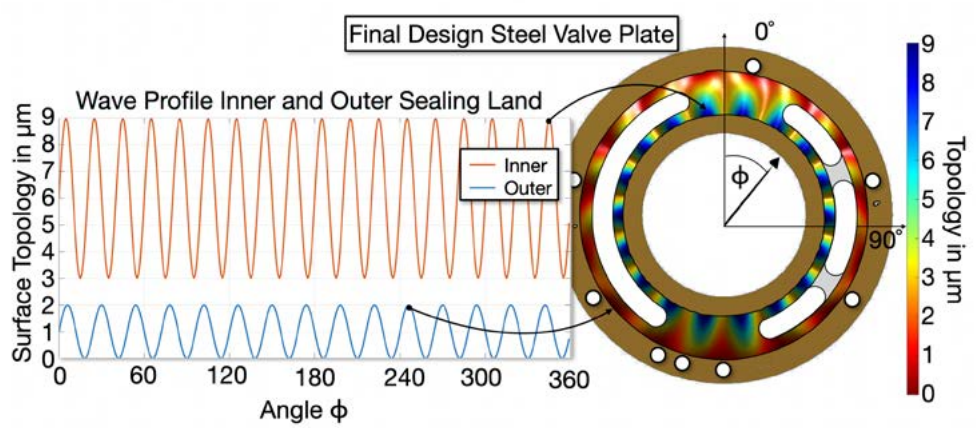


Figure 5.15. Final manufactured design for the steel valve plate with a full wave design and adapted sealing land width.

The resulting optimized steel design is shown in Figure 5.15 and differs significantly from the optimized bronze wave plate: whereas the bi-metal bronze plate used a larger outer-land amplitude (±3 µm) and a smaller inner-land amplitude (±1 µm) with a 7 µm offset, the optimized steel design essentially inverts this distribution by placing the dominant waviness on the inner sealing land (±3 with a 6 µm offset) and reducing the outer-land amplitude to ± 1µm. In addition, the steel design extends structuring across the full sealing-land circumference, including transition regions around TDC and BDC.

The performance predicted performance of this design is shown in Figure 5.16 in form of stacked power loss bars. The found design is predicted to substantially reduce power losses by more than 30% at 35°C inlet and 35% at 60°C inlet, when compared to the already well performing bronze wave design. This demonstrates the strong potential of steel/steel when the wave is designed for the correct thermo-elastic response. However, even with the redesigned wave pattern, a fully robust solution for OC4 at elevated temperature was not achieved within the limited optimization scope.

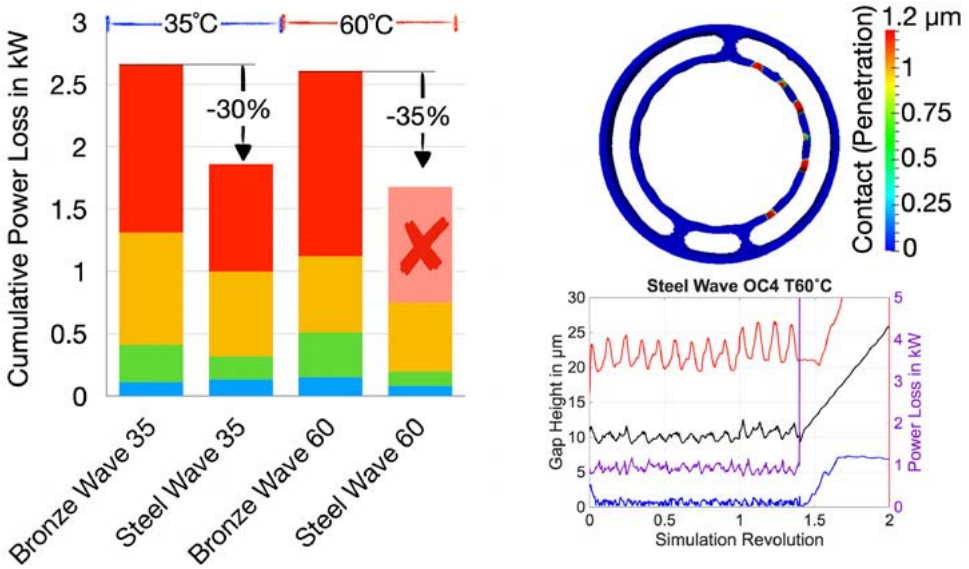


Figure 5.16. Final design steel valve plate after re-running full optimization (left). Right: Simulation result for OC4 Steel Wave 60, with predicted wear at the inner sealing land.

As shown on the right side of Figure 5.16 the steel design shows contact on the inner sealing land at OC4 at 60˚C inlet temperature. The graph on the bottom right of the figure confirms, that this is primarily due to a thermal deformation problem, as the first revolution delivered very good power losses of under 1 kW for this interface. For comparison, the standard valve plate produces about 2.4 kW power losses at this condition. However, after the simulation entered the second revolution the simulation predicted contact and crashed at the end of the high-pressure stroke. For this reason, the OC4 "result" should be interpreted as an initial-revolution indicator rather than a validated steady-state prediction, indicating that this design might experience problems at high speed, high pressure and high temperature.

Despite this remaining limitation, the significantly improved power losses and the stable behavior at 35°C across the majority of the operating domain motivated fabrication of the optimized steel valve plate to experimentally validate the low-temperature performance and to identify the limiting failure mechanisms in a controlled measurement campaign. A complete optimization over phase, frequency, amplitude distribution, offsets, and sealing-land geometry is expected to enable convergence across all operating conditions and both inlet temperatures; however, this extended design space was outside the available project timeframe. The following section therefore describes the manufacturing approach and the experimental validation of the optimized steel valve plate concept.

6. MANUFACTURING AND SURFACE MEASUREMENTS

Manufacturing of the steel meso-structured valve plates was carried out at Fraunhofer IWU (Chemnitz). IWU previously produced the step-structured slippers and the waved bronze valve plate discussed in Section 4 using precision turning, where the micro-topography is generated via a piezo-actuated tool path. For the present steel design, a laser-based structuring route was selected instead. Steel is well suited for ultrashort-pulse ablation, and the laser process offers (i) high geometric fidelity for shallow µm-scale features, (ii) flexible patterning without dedicated tooling, and (iii) robust repeatability for complex wave layouts. The applied laser setup with the resulting wave like structure is shown in Figure 6.17. It uses an femtosecond ultrashort-pulse TruMicro 2030 laser (average power: 20 W, wave length: 1030 nm, pulse rate: 300 fs–20 ps).

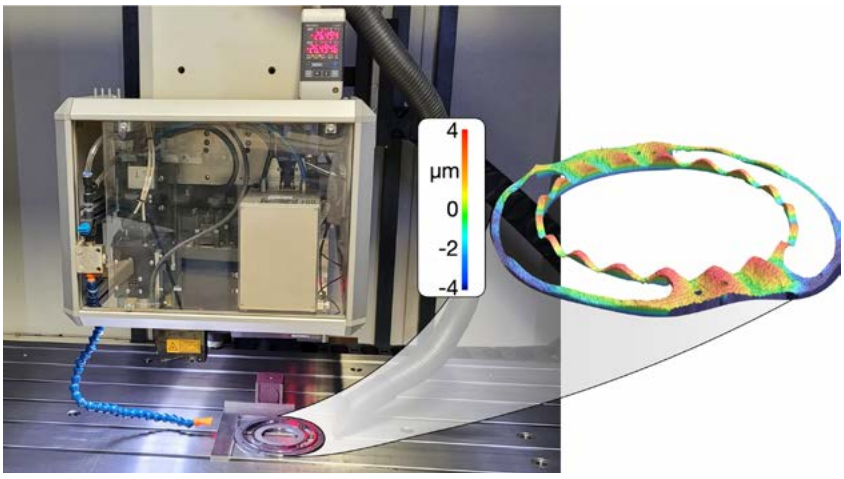


Figure 6.17. Laser structuring at the Fraunhofer IWU

A total of four steel valve plates were manufactured: two in an unhardened steel condition and two as hardened parts. The unhardened variant was intended to provide a more forgiving material response against the hardened cylinder block surface during possible mixed-lubrication events occurring at start-up or brief moments of dynamic changes. However, this assumption did not hold across the full operating envelope. In subsequent testing, the unhardened valve plates exhibited limitations at high-pressure conditions, whereas the hardened valve plates demonstrated clearly improved robustness at elevated pressures. The premature failure at high pressures motivated the focus on hardened steel valve plates for the high-pressure validation steps.
The manufactured valve plate and a superimposed 3D surface measurement are shown in Figure 6.18. The measured surface profiles on the inner land (IL) and outer land (OL) indicate that the realized wave amplitudes deviate moderately from the nominal targets: the OL is approximately ±1.5 µm (target ±1 µm), while the IL is approximately ±2.5 µm (target ±3 µm with offset). This deviation is likely influenced by the pre-existing flatness error of the valve plate prior to structuring (≈3.4 µm), which cannot be fully compensated by the laser structuring process. Nevertheless, the achieved geometry was considered acceptable, since the amplitudes remain within the tolerance band for which the simulations indicated robust performance.
Despite this deviation, the overall structure—frequency, phase relation, and relative amplitude concept—was realized consistently. The wave topographies are not visually apparent to the naked eye; instead, the structured region primarily becomes noticeable through characteristic color effects from light diffraction/reflection on the textured surface as captured in the photograph on the right.

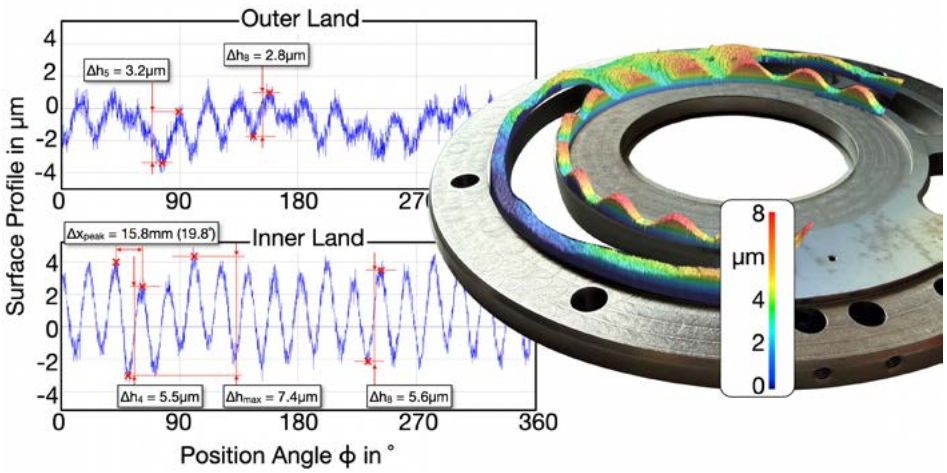


Figure 6.18. Manufactured waved full steel valve plate with surface scan.

With the steel valve plates manufactured and geometrically verified, the next section presents the experimental results for the steel/steel pairing, including performance mapping, temperature/particle monitoring, and the observed failure mechanisms under high-pressure and elevated-temperature conditions.

7. EXPERIMENTAL RESULTS: STEEL/STEEL VALVE PLATES (UNHARDENED THEN HARDENED)

Both variants of the meso-structured full-steel valve plate described in Section 6 (unhardened and hardened) were tested in the same measurement procedure described in section 3, using identical instrumentation for pressure, flow, torque, temperature and particle monitoring. The results are presented in two parts.

In Section 7.1, the unhardened steel waved valve plate (Steel WVP (S)) is directly compared against the waved bronze valve plate and the reference valve plate using **Rotating Kit 1**, for which complete reference measurements (Ref1) were available at both 40˚C and 60˚C inlet temperatures and including the low-speed measurements. Unfortunately, the unhardened steel plate failed prematurely, severely damaging the cylinder block and contaminating the pump with metallic debris, which necessitated a hardware change. Consequently, Section 7.2 focuses on the hardened steel valve plate measurements performed with a **replacement rotating kit (Kit 2)**. For Kit 2, a reference data set was available from other projects [12], but not for 60 °C and not for the ultra-low-speed conditions; furthermore, Kit 2 had already seen significant prior operation and therefore exhibits lower baseline efficiency than the new Kit 1. For this reason, the hardened steel results are evaluated primarily relative to Reference Kit 2 (Ref2) in terms of efficiency.

7.1. Rotating Kit 1 – Unhardened Steel Valve Plate

The unhardened steel wave valve plate was tested over the operating range 250–2100 rpm and 50–250 bar at 40˚C inlet temperature, before the test campaign was interrupted by a premature failure that occurred at 1800 rpm and 250 bar. Up to the onset of failure, Steel WVP (S) showed a remarkably strong efficiency potential and, in large parts of the map, outperformed both the new reference valve plate and the waved bronze valve plate—showing significant improvements even at full displacement.

Figure 7.19 summarizes this efficiency by showing the total efficiency difference of the Steel WVP (S) relative to Ref1 (left) and relative to the waved bronze valve plate (Bronze WVP) (right). The steel WVP yields

pronounced gains at low speed and low-to-medium pressures, reaching +10 percentage points over the reference and up to +16 percentage points over the bronze configuration at 250 rpm and high pressure (250 bar).

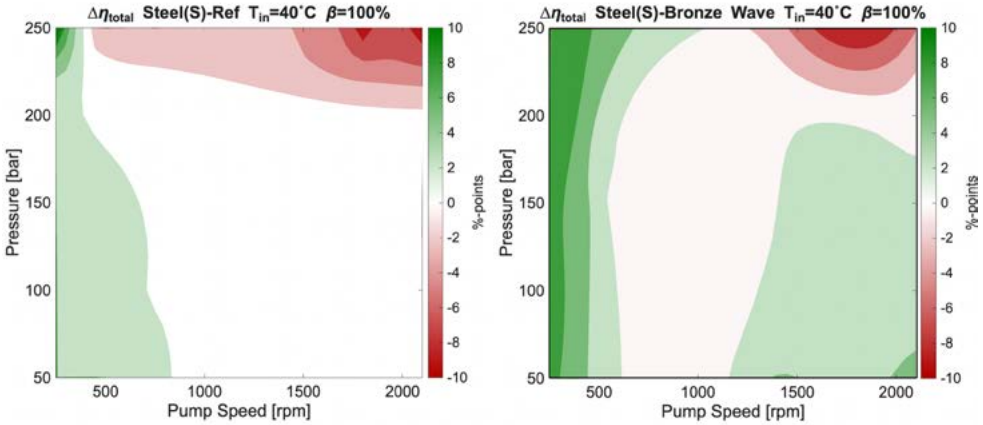


Figure 7.19. Efficiency difference between the unhardened steel WVP, the reference Kit1 and the bronze WVP at 40˚C inlet and 100% displacement

In the low-pressure region the advantage persists up to roughly 750 rpm, after which the performance approaches the reference level. At higher pressures, the efficiency initially remains improved but then degrades rapidly as the operating point approaches the failure region; immediately before failure, a strong drop is observed at high speed and 250 bar. Despite the failure, the magnitude of the low-speed improvement is significant. Increasing the total efficiency from 61% to 76% at 250 rpm, 250 bar and full displacement represents a substantial step toward electrification-relevant operating conditions, and the improvement over the already optimized waved bronze valve plate at low speed and high pressure indicates that the steel-specific wave interaction and thermal-mechanical behavior can produce different—and in parts more favorable—trends than the bronze-based design.

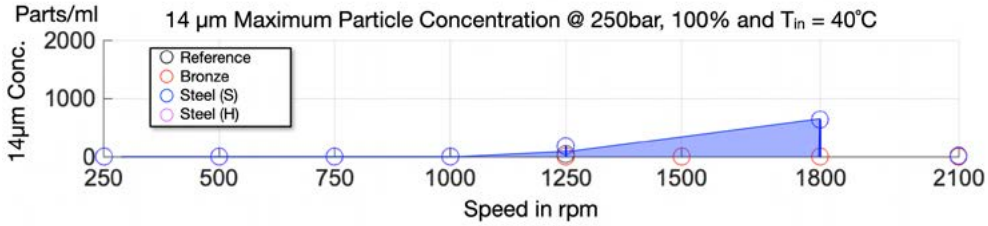


Figure 7.20. Maximum 14µm particle production at 250bar and 40˚C

Figure 7.20 shows the maximum concentration of 14 µm particles in the HP line at 250 bar and T_{in} = 40 °C for all tested valve plates. While Ref1 and the bronze WVP remain at low particle levels, Steel WVP (S) begins producing large particles already at 1250 rpm, and the particle concentration increases dramatically until the full failure at 1800 rpm. Notably the Steel WVP (H) did not produce any particles at this temperature and pressure. The resulting damage state is documented in Figure 7.21, where both the valve plate and the cylinder block exhibit severe wear, with grooves exceeding 0.5 mm in depth.



Figure 7.21. Failure of the unhardened steel valve plate after reaching 1800rpm and 250 bar at 40˚C.

This outcome highlights an important characteristic of steel/steel pairings: they are highly unforgiving once mixed lubrication and local contact occur. The meso-structure operates at the micrometer scale and is therefore sensitive to early damage. Moreover, the steel valve plate geometry was intentionally adapted (reduced sealing land width) to compensate for the additional hydrodynamic lift of the wave. Once wear occurs it removes or damages the meso-wave structure. After losing the additional hydrodynamic lift and the interface became critically underbalanced, accelerating contact, heat generation, and material removal. Post-test inspection further indicated plastic deformation of the unhardened plate, indicating that the high pressure loading deformed the VP, which is an effect that the simulation model cannot capture as only elastic deformation is considered.
These findings motivated the second steel configuration: a hardened wave valve plate intended to provide higher resistance against pressure-driven deformation and to improve robustness under the high-pressure operating

points. The following Section 7.2 therefore evaluates the hardened steel wave valve plate using Rotating Kit 2 and compares its efficiency performance to the corresponding reference for that kit.

7.2. Rotating Kit2 – Hardened Steel Valve Plate

The hardened steel waved valve plate (Steel WVP (H)) was tested with rotating kit 2 and is therefore evaluated exclusively against the corresponding reference measurement of kit 2 (Ref2). Figure 7.22 summarizes the resulting efficiency differences at 100% and 50% displacement. The largest gains occur at low pressure and in the operating regions that are most relevant for variable-speed duty cycles: at 100% displacement the total efficiency increases by more than 7 percentage points around 1000 rpm (from 80% to 87%), while at 50% displacement improvements of roughly 5 percentage points are observed at high speed (at 1800–2100 rpm from 70% to 75%). Importantly, these benefits are achieved with only minor efficiency drawbacks at high pressure and low-speed, where the efficiency decreases by approximately one percentage point.

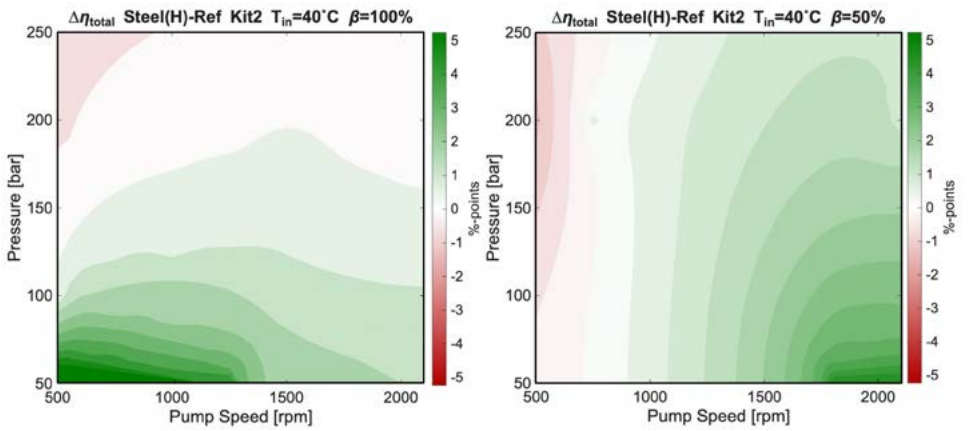


Figure 7.22. Efficiency difference between the hardened steel WVP and the reference for rotating kit 2 at 100% and 50 % displacement at 40˚C inlet.

The simulation-based performance trend is thereby validated and is further reflected in the operating-condition power-loss evaluation shown in Figure 7.23. In the design phase (Figure 5.16), the thermo-elastohydrodynamic optimization predicted a substantial reduction of valve-plate interface losses, primarily at OC3 and OC4 at 40 °C inlet temperature; the predicted relative reduction of interface losses corresponds to an absolute reduction on the order of 0.7 kW. Experimentally, a reduction of 1.1 kW

(~4%) is obtained when summing OC1–OC4 at 100% displacement. At 50% displacement the reduction increases to 3.6 kW (~17%). These results confirm that meso-structures can deliver a measurable reduction in losses even in a steel/steel pairing and across a wide speed range—an essential prerequisite for electrification-ready pump concepts where the machine is frequently operated in various speed domains.

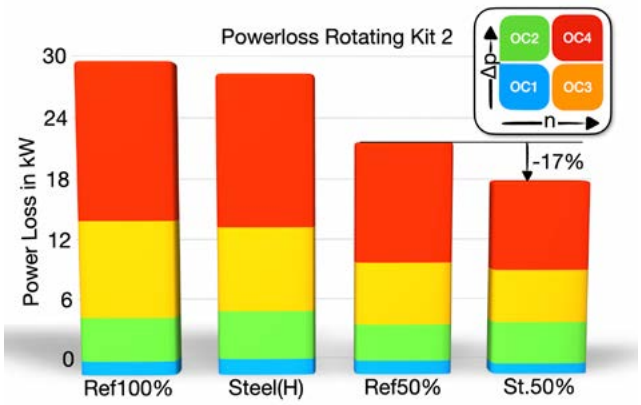


Figure 7.23. Efficiency improvements of the hardened steel VP at 100 and 50% at 40˚C inlet temperature for OC1-4.

Besides efficiency, the key system-level metric for electrified drives is torque demand, because required motor size, peak current capability, and thermal load scale directly with torque.

Figure 7.24 therefore compares torque losses at 50 bar and 200 bar. Due to the premature failure of the first steel valve plate and the lack of low-speed reference measurements for Kit 2, the comparison necessarily combines data from both rotating kits. As expected, Kit 1 exhibits lower torque losses overall, consistent with its higher efficiency baseline. Nevertheless, Steel WVP (H) (Kit2) shows very pronounced torque-loss reductions: at 50 bar and 20 rpm, the torque loss is reduced by approximately 66% (more than 10 Nm). By extrapolating the trend onto Kit1 (red dashed line), a further reduction is likely, amounting to 80% torque reduction when comparing the same rotating kit with each other. At 200 bar, the torque-loss reduction amounts to more than 30% at 250 rpm and remains around 20% over a broad speed range, which directly translates into a substantially reduced e-motor torque requirement and therefore offers a clear pathway to smaller and lower-cost electric drive systems.

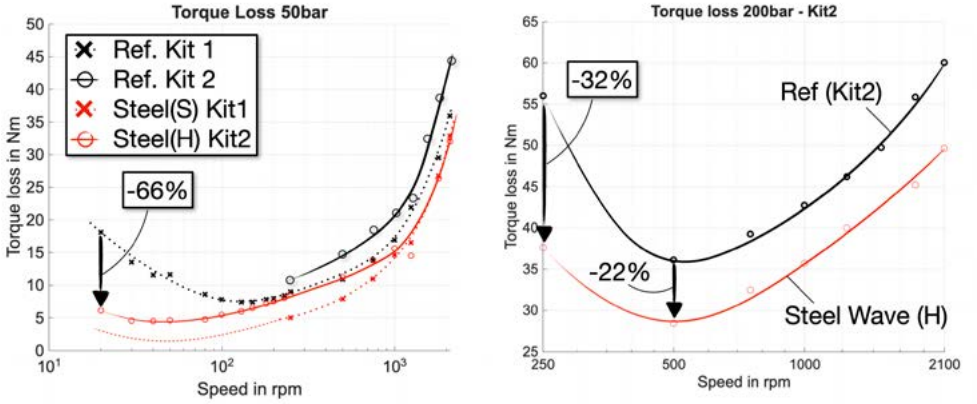


Figure 7.24. Improved low speed torque loss at 50 bar and 200bar for both rotating kits.

Up to this point, the hardened steel WVP confirms the simulation predictions and demonstrates strong potential. However, the results also confirm the key limitation of steel/steel pairings: once contact occurs under high temperature and elevated speed, wear becomes rapidly self-accelerating.

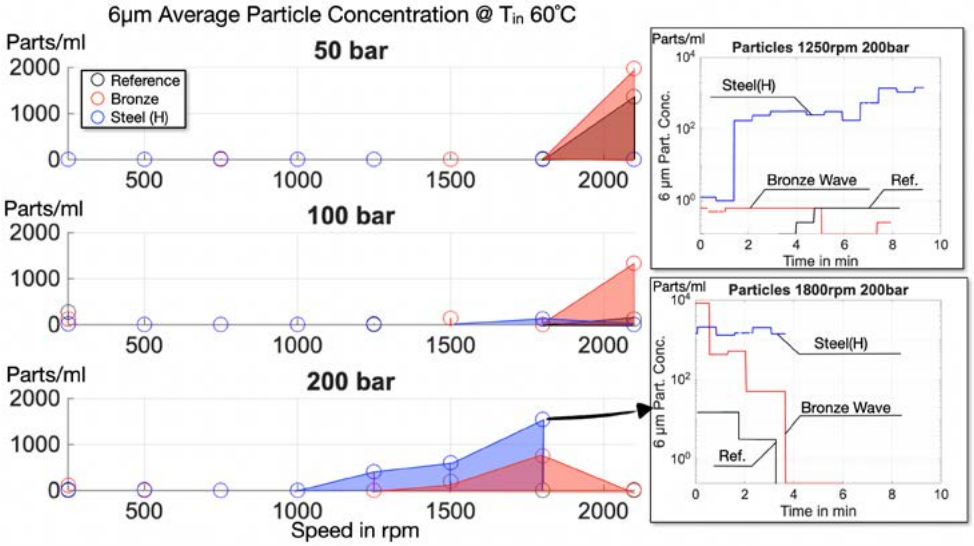


Figure 7.25. Average particle generation at 60˚C inlet temperature with zoom-ins at 1250 rpm and 1800 rpm at 200 bar.

Figure 7.25 shows the average 6 µm particle concentration in the HP-line for 60°C inlet temperature over the tested speed and pressure range. The 6 µm

class is chosen because it provides two practically relevant insights. First, it acts as an early warning indicator for impending failure in steel/steel interfaces: at 1250 rpm and 200 bar the particle concentration begins to rise and remains elevated over time even though the pump is still operating with acceptable efficiency. Continuing operation leads to further escalation, and at the final operating point (1800 rpm, 200 bar) the concentration increases strongly (note the logarithmic y-axis scaling), accompanied by the emergence of larger particles, which triggered the shutdown event. Second, the same measurement highlights the robustness of bronze-based pairings during short contact events: both the reference and the bronze valve plate can show elevated particle production at high speed and high temperature (e.g., the bronze waved valve plate at 1800 rpm and 200 bar), but the particle generation typically decays back toward zero, consistent with a gentle run-in mechanism. In contrast, steel/steel contact tends to remove micrometer-scale material and generate larger fragments, leaving grooves that promote further contact and accelerate damage. Post-test inspection of Steel WVP (H) showed severe grooving and wear depths on the order of several hundred micrometers, similar to the unhardened case, indicating that once the surface structure is damaged the interface can no longer act as a controlled hydrostatic/hydrodynamic bearing. If the VP would have had a similarly forgiving tribological material pairing, then the VP would have likely survived.

8. DISCUSSION

Across both rotating kits, the experiments demonstrate that meso-structured valve plates can be designed to significantly improve efficiency and reduce torque losses in operating regions that dominate variable-speed duty cycles. The steel designs, in particular, show that a hard/hard pairing does not inherently prevent performance benefits. On the contrary, the steel material allowed for a design of an even superior wave pattern, improving efficiency by more than 10% and torque loss by 60-80%, which directly supports electrification-driven requirements for compact and efficient motor-pump units.

At the same time, the measurements expose the central robustness challenge. While bronze-based systems can tolerate brief metal-to-metal contact through a forgiving run-in mechanism with limited wear volume, steel/steel contact can rapidly transition into severe, self-amplifying wear. Particle monitoring—especially in the 6 µm class—proved to be a sensitive early indicator of such failure onset, but preventing the underlying contact remains the primary requirement for a viable steel/steel implementation.

The simulation model predicted wear at high inlet temperature and elevated pressure, which the measurement confirmed. Previous designs optimized with the Caspar FSTI suite also indicated brief wear periods, something that a soft/hard material pairing usually accepted. However, the study has shown when designing a hard/hard pairing with meso-structures, contact needs to be avoided in all operating conditions, even when just brief. With a more time and exploring a larger design space, the author is confident that a suitable design can be found.

9. CONCLUSIONS AND OUTLOOK

This study confirms that meso-structured valve plates can improve the performance of axial piston pumps operated under electrification-relevant, variable-speed conditions. The structured bronze valve plate maintains the previously reported benefits and exhibits negligible wear, even when operated at steady-state thermal boundary conditions.

Extending the same design to steel/steel interfaces was not feasible, as the bi-metal bronze valve plate experiences different thermal expansion as a full steel one. Hence, a complete redesign had to be performed, yielding very different sealing land structures. The steel wave valve plate demonstrated even larger improvements in efficiency and torque demand, indicating that meso-structures can support lead-free and cost-relevant material strategies.

However, the steel/steel results also highlight that robustness, not peak efficiency, is the limiting factor. Once contact occurs at elevated temperature and speed, steel/steel interfaces can rapidly generate large wear volumes, destroy the micro-structure, and lose the hydrodynamic and hydrostatic balancing on which the design is based. Future work should therefore focus on mechanisms that allow short contact events without catastrophic damage, while simultaneously exploring a larger design space to avoid contact even at the most extreme conditions.

When combining these measured findings of over 80% torque reduction and 10-30% efficiency improvements, with the similarly impressive meso-structured slippers published in previous research [4], it becomes clear, that pump designs can be significantly improved when combining surface structures to all hydrostatic bearings. With these additions, meso-structuring offers a high-potential pathway to electrification-ready, lead-free axial piston pump designs that combine efficiency gains with reliable long-term durability.

ACKNOWLEDGEMENTS

The authors would like to thank Inline Hydraulik GmbH and the Fraunhofer IWU for supporting this research with donations and services. Without their kind support this research would have not been possible.

10. REFERENCES

[1] M. Ivantysynova and J. Baker, "Power loss in the lubricating gap between cylinder block and valve plate of swash plate type axial piston machines," *International Journal of Fluid Power*, vol. 10, no. 2, pp. 29–43, 2009, doi: 10.1080/14399776.2009.10780976.

[2] A. M. Wondergem, "Piston/Cylinder Interface of Axial Piston Machines - Effect of Piston-Micro Surface Shaping," 2014. doi: 1585442.

[3] A. A. Darbani, L. Shang, J. R. Beale, and M. Ivantysynova, "Slipper surface geometry optimization of the slipper/swashplate interface of swashplate-type axial piston machines," *International Journal of Fluid Power*, vol. 20, no. 2, pp. 245–270, 2019, doi: 10.13052/ijfp1439-9776.2025.

[4] R. Ivantysyn, S. Horn, and J. Weber, "Design of a Lead-Free Slipper Bearing for Low Speed Axial Piston Pump Applications," *International Journal of Fluid Power*, vol. 25, no. 2, pp. 183–202, 2024, doi: 10.13052/ijfp1439-9776.2524.

[5] R. Reddy, R. Ivantysyn, and P. J. Weber, "Advanded micro-surfacing: Tribological optimization for variable speed hydraulic drives," in *ASME/BATH 2025 Symposium on Fluid Power and Motion Control, FPMC 2025*, 2025, pp. 1–15.

[6] S. Horn, R. Ivantysyn, and J. Weber, "Tribo-optimized lubricating interfaces in hydrostatic pumps with surface shaped slippers," in *IFK2022*, Aachen, 2022.

[7] SHS GmbH, "Caspar FSTI The most powerful axial piston pump model." [Online]. Available: https://www.smart-hydraulic-solutions.com/casparfsti

[8] R. Ivantysyn and J. Weber, "Thermal Insights and Predictive Modeling of Axial Piston Pumps : A Journey to a High-Fidelity Digital Twin," in *Maha Conference 2024*, 2024, pp. 1–19.

[9] R. Ivantysyn, A. Shorbagy, A. Vedpathak, and J. Weber, “Investigation of the Heat Conduction in Axial Piston Pumps by Measurement and Simulation,” in *IEE GFPS 2024*, Hurtigsval, Sweden, 2024.
[10] R. Ivantysyn, J. Weber, A. Kunze, and W. Uffrecht, “Thermal Analysis of the Cylinder Block of an Axial Piston Pump – The Key to Monitoring Efficiency,” in *IFK2024*, 2024, pp. 1–21.
[11] R. Ivantysyn, W. Uffrecht, A. Kuntze, and J. Weber, “Efficiency Forecasting in Axial Piston Pumps : Leveraging High Speed Fluid Temperature Correlations for Condition Monitoring,” in *JFPS*, Hiroshima, 2024, pp. 1–15.
[12] R. Ivantysyn and J. Weber, “Advancing Thermal Monitoring in Axial Piston Pumps : Simulation , Measurement , and Boundary Condition Analysis for Efficiency Enhancement,” *International Journal of Fluid Power*, vol. 25, pp. 1–44, 2024, doi: 10.13052/ijfp1439-9776.2546.

Biographies



Roman Ivantysyn received the B.S. and M.S. degrees in Mechanical Engineering from Purdue University, West Lafayette, IN, USA. Since 2011, he has been a research assistant and Ph.D. candidate at the Chair of Fluid-Mechatronic Systems (LFD), TU Dresden, Germany, where his doctoral research addresses the thermal behavior of lubricating gaps in axial piston pumps with respect to efficiency and wear. His current research interests include large-scale mobile machinery and open-circuit displacement control, condition monitoring and lifetime studies of axial piston pumps, thermal analysis of pump systems, and surface shaping of hydrostatic bearings. Since 2024, he has also been CEO of Smart Hydraulic Solutions GmbH.

Analysis of Structure-Borne Noise Emissions of Axial Piston Pumps Using Blocked Forces and Virtual Point Transformation

Matthias Vogt[1,2,*] and Marcus Geimer[1]

[1]*Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*
[2]*Development of Axial Piston Units, Bosch Rexroth AG, Horb am Neckar, Germany*
E-mail: matthias.vogt@partner.kit.edu; marcus.geimer@kit.edu
[*]*Corresponding Author*

Abstract

With the shift towards electrification and the advancement of quieter internal combustion engines, the noise produced by axial piston pumps has gained increased attention. While fluid-borne noise emissions resulting from pulsating flow and airborne noise emissions through the housing of hydraulic pumps are well researched, studies on structure-borne noise emissions–transmitted to adjacent structures via the pump's mechanical mounting–are rare. One method to describe these emissions is through blocked forces, which allow for a receiver-independent characterization of structure-borne noise sources. In this publication, the excitation characteristics of an axial piston pump are analyzed based on blocked forces, which can be projected onto one or several points at the interface using the so-called virtual point transformation. The extent to which the pump's description differs when using one or two virtual points is investigated. The structure-borne noise excitation caused by the pump is examined during a speed run-up, allowing for the identification of dominant excitation degrees of freedom. It is indicated that for the axial piston pump under investigation, a source description utilizing a single virtual point is sufficient.

Keywords. Axial Piston Pumps, Structure Borne Noise, Blocked Forces, Virtual Point Transformation

1. INTRODUCTION

In addition to the airborne noise (ABN) directly emitted from the pump housing, hydraulic pumps also emit fluid-borne (FBN) and structure-borne noise (SBN) through the connection points into the overall system. Previous studies [1], [2], [3], [4], [5] indicate that these noise contributions have a significant impact on the overall sound power level of hydraulic systems. Fiebig [1] even states that hydraulic systems typically exhibit sound power levels that are 4 to 12 dB(A) higher than those of the pumps themselves, concluding that it is

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

unlikely that the ABN from the pumps housing constitutes the main sound power sources in such hydraulic systems.
While the excitation due to pressure pulsation is well-researched [6], [7] and established methods and standards [8], [9] for its determination exist, the SBN transmitted through the pump mounting points into the system remains largely unexplored. Fiebig's investigation [1] shows that this type of excitation can be a major cause of noise in hydraulic systems. In his study the dynamic forces generated by the pump are transmitted as SBN to the electric motor via the pump mounts. The motor and other large components act as noise radiators, converting this SBN energy into ABN. Therefore, knowing the pump emitted SBN at the interface also opens up various optimization potentials, such as optimizing the matching of excitation characteristics and system behavior through finite element simulation or designing decoupling elements between the pump and motor, as well as between the motor and the motor base.
A promising method to experimentally determine the SBN excitation at the pump mounting points independently of the receiver system is the "In situ Blocked Force Method" described in the ISO20270 standard [10], where the so-called blocked forces are used to characterize SBN sources. In [11], it has been demonstrated that this method can be successfully applied to an axial piston pump, thereby enabling the assessment of the SBN emission of hydraulic pumps.
In the following study, the source characterization based on experimentally determined blocked forces is further investigated. Therefore, the source description of an axial piston pump using one or two virtual contact points is compared. Furthermore, it is examined which excitation degree of freedom (DoF) are the most dominant for the investigated pump.

The publication is structured as follows: First, the theoretical foundations of the “Virtual Point Transformation” and “Component-based Transfer Path Analysis” are introduced. This is followed by a method for aggregating virtual point forces and torques into a single force and torque quantity. These concepts are then applied to an axial piston pump, and the results of the different methods are presented. Finally, the key insights are summarized, and an outlook on future work is provided.

2. METHODS

A fundamental prerequisite for analyzing SBN excitation from a hydraulic pump is the determination of quantities that describe the pump-induced SBN at the interface to a receiving system. Consequently, the following section introduces various concepts of transfer path analysis and source characterization.

2.1. Virtual Point Transformation

A critical aspect in the description of SBN sources is the definition of the interface between the source and the receiver system. Particularly when the excitation of a source is to be combined with the behavior of a passive receiver, the coupling points must be clearly defined. However, this is often not the case for quantities determined through measurement techniques, as source excitation and receiver behavior are measured in different ways or using different measurement setups.
One way to overcome this is the so-called Virtual Point Transformation (VPT). The concept was first introduced by van der Seijs et al. [12] and aims to project interface quantities (interface DoF) onto a set of interface displacement modes (IDMs), thereby enabling a clear

definition of the interface. When IDMs correspond to the six DoF (3 translational and 3 rotational) of a virtual point (VP), they are also known as rigid interface modes or virtual point DoF. A detailed description of the mathematical framework of the VPT is beyond the scope of this publication and has already been thoroughly described in various studies [12], [13], [14].

As the designation of the VP DoF as rigid interface modes implies, the classical VPT is based on the rigid body assumption. This means that the motions or forces projected onto a VP via VPT are assumed to behave as those of a rigid body. This condition can only be satisfied within a limited frequency range, depending on the structural dynamic behavior of the interface. In particular, the geometry and material properties of the interface significantly influence this range. If the rigid body condition is not fulfilled, the interface can no longer be fully described by the VP.

If a single VP is insufficient to describe the interface, as it is often the case for systems with multiple mounting points or flange-like contacts [15], [16], multiple VPs can be employed to represent the interface. The rigid body condition must then be satisfied only for each individual VP, thereby typically enabling the description of the source SBN excitation up to higher frequencies.

For the axial piston pump investigated, featuring a two-bolt mounting flange according to DIN ISO 3019-1 [17], two VPT setups were examined: one with a single VP (1VP setup) located at the center of the interface, and another with two VPs positioned at the mounting points (2VP setup). Figure 1 schematically illustrates both setups along with their respective interface DoF. While the 1VP setup requires only six DoF to describe the interface, the 2VP setup necessitates twelve DoF.

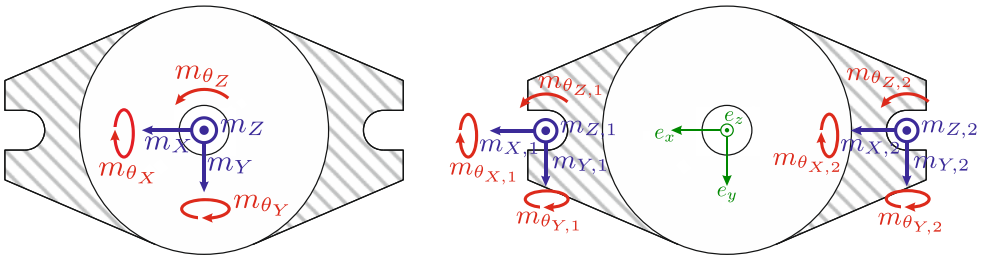


Figure 1: Schematic illustration of VPT setups for the axial piston pump. Left: 1VP setup with six DoF. Right: 2VP setup with twelve DoF.

2.2. Component-based Transfer Path Analysis

The primary objective of Transfer Path Analysis (TPA) is to identify the contributions of various noise sources and transmission paths to the overall noise level. A key step in TPA is source characterization, which aims to define a set of interface quantities that accurately describe the source's excitation at the interface with the passive receiving structure. Various interface quantities can be utilized for this purpose; however, for analyzing the source excitation itself, receiver-independent interface quantities are particularly valuable. When such quantities are used to describe the source, the TPA approach is referred to as component-based TPA (cTPA).[18]

The three most common methods of cTPA are illustrated in Figure 2. The “Free Velocity” method represents the extreme case where the source operates under freely vibrating conditions. In this scenario, the velocities measured at the interface $\mathbf{u}_2^{\text{free}}$ are independent of

the receiver and thus represent a pure characteristic of the source. However, this method is unsuitable for hydraulic pumps, as a motor is required to drive the pump. The other extreme case is the "Direct Blocked Force" method, which assumes that the receiver connected to the source is infinitely stiff. Under this assumption, the forces measured at the interface, known as blocked forces $\mathbf{f}_2^{\mathrm{eq}}$, are independent of the receiver. Nonetheless, this method presents practical challenges, as detailed in [11].

A method that avoids such extreme boundary conditions is the "In Situ Blocked Force" method, which was first introduced by Moorhouse et al. [19]. Moorhouse et al. demonstrated that the blocked forces can also be determined in systems that are not infinitely stiff by applying a matrix inversion procedure. By means of this two-stage approach, the influence of the receiving system can be removed, enabling the determination of the receiver-independent blocked forces. This approach allows for the determination of blocked forces *in situ*, meaning within the typical application and is adaptable to various SBN sources. Consequently, it has gained prominence in recent years, particularly in the automotive industry, and has been standardized in ISO 20270 [10]. A detailed description of its application to an axial piston pump is provided in [11], [13].

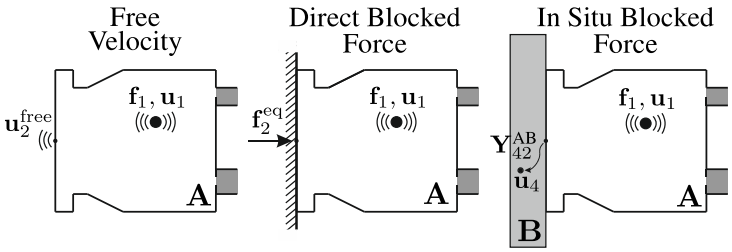


Figure 2: Common methods of source characterization in cTPA

The in situ blocked force method is based on the formula:

$$\mathbf{f}_2^{\mathrm{eq}} = (\mathbf{Y}_{42}^{\mathrm{AB}})^{+}\mathbf{u}_4 \tag{1}$$

which states that the blocked forces $\mathbf{f}_2^{\mathrm{eq}}$ can be determined by multiplying the vibration responses at the so-called indicator sensors $\mathbf{u}_4$ with the inverted transfer behavior $\mathbf{Y}_{42}^{\mathrm{AB}}$ from the interface to the indicators in the coupled state (denoted by $(\cdot)^{\mathrm{AB}}$). The symbol $(\cdot)^{+}$ denotes the pseudoinverse, as the transfer matrix $\mathbf{Y}_{42}^{\mathrm{AB}}$ is typically not a square matrix. From Equation 1, it follows that the method consists of two main steps:

Step 1**:** Determination of the transfer behavior $\mathbf{Y}_{42}^{\mathrm{AB}}$ from the interface to the indicators. This is typically achieved by exciting close to the interface using an impulse hammer or a shaker and measuring the vibration response at the indicators, usually with accelerometers.

Step 2: The source is in operation and the vibrations at the indicators $\mathbf{u}_4$ are measured. Using Equation 1, the blocked forces can then be calculated.

It is crucial to note that the blocked forces are determined exactly at the locations where the interface was excited with the impact hammer in Step 1. Because, in the in-situ blocked force method, the source and receiver are mechanically coupled, these excitation points are not located directly on the interface but rather in its vicinity, typically on the receiver structure. Consequently, the spatial positions of the blocked forces depend on the receiver

geometry in the region around the interface, and the resulting blocked forces are therefore not completely receiver independent.
This receiver-dependency can be eliminated by transforming the blocked forces into VP coordinates, which are uniquely defined by the source itself. However, instead of transforming the blocked forces $\mathbf{f}_2^{\mathrm{eq}}$ into VP blocked forces and torques $\mathbf{m}_2^{\mathrm{eq}}$ after performing the in-situ blocked force TPA, it is also possible to transform the transfer matrix $\mathbf{Y}_{42}^{\mathrm{AB}}$ beforehand. The measured matrix is usually denoted $\mathbf{Y}_{42,\mathrm{uf}}^{\mathrm{AB}}$, where the subscripts $(\cdot)_{\mathrm{u}}$ and $(\cdot)_{\mathrm{f}}$ are added to indicate that this corresponds to the original form of the matrix. The transformed matrix is denoted $\mathbf{Y}_{42,\mathrm{um}}^{\mathrm{AB}}$, as only the forces are transformed into VP coordinates, whereas the measured motions at the indicator sensors are not transformed. This has the benefit that by transforming only the forces into VP coordinates, the number of force DoF is reduced, while the number of motion DoF remains constant. This results in a higher overdetermination of $\mathbf{Y}_{42,\mathrm{um}}^{\mathrm{AB}}$ compared to $\mathbf{Y}_{42,\mathrm{uf}}^{\mathrm{AB}}$, leading to a more stable matrix inversion. In the ISO 20270 standard an overdetermination by a factor between 2 and 3 is suggested. By applying the VPT exclusively to the forces, the recommended overdetermination factor can be achieved with smaller number of sensors to capture the vibrations $\mathbf{u}_4$.
It is important to note, that both matrices, $\mathbf{Y}_{42,\mathrm{uf}}^{\mathrm{AB}}$ and $\mathbf{Y}_{42,\mathrm{um}}^{\mathrm{AB}}$, represent the transfer behavior $\mathbf{Y}_{42}^{\mathrm{AB}}$ schematically shown in Figure 2. However, $\mathbf{Y}_{42,\mathrm{uf}}^{\mathrm{AB}}$ uses the original physical excitation points of the impact hammer or shaker, whereas $\mathbf{Y}_{42,\mathrm{um}}^{\mathrm{AB}}$ uses the VP DoF.
Using this, the relation for the in situ blocked force TPA becomes

$$\mathbf{m}_2^{\mathrm{eq}} = (\mathbf{Y}_{42,\mathrm{um}}^{\mathrm{AB}})^{+}\mathbf{u}_4, \tag{2}$$

which allows for the determination of the blocked forces directly in the VP coordinates.
An additional benefit of the in situ blocked force method is that it allows for direct validation of the source description through the blocked forces. This validation step is referred to as On-Board Validation (OBV). For this purpose, additional validation sensors $\mathbf{u}_3$ are located on the receiver, which are not used for determining the blocked forces. The transfer behavior $\mathbf{Y}_{32,\mathrm{uf}}^{\mathrm{AB}}$ can be directly determined during the measurement of $\mathbf{Y}_{42,\mathrm{uf}}^{\mathrm{AB}}$, resulting in minimal additional effort. After the blocked forces have been determined using Equation 2, the movement at the validation points $\mathbf{u}_3$ can be predicted using the blocked forces through the relationship

$$\mathbf{u}_3^{pred.} = \mathbf{Y}_{32,\mathrm{um}}^{\mathrm{AB}}\mathbf{m}_2^{\mathrm{eq}}. \tag{3}$$

These predicted motions can then be compared with the directly measured signals $\mathbf{u}_3^{meas.}$, providing an initial validation of the blocked forces. If the blocked forces are unable to describe the excitation by the source, or if numerical errors occur during the inversion of the matrix $\mathbf{Y}_{42,\mathrm{um}}^{\mathrm{AB}}$, this will be evident in the OBV.

2.3. Aggregation of Blocked Forces and Blocked Torques Using Vector Norms

As already explained in [13], blocked forces and torques can be summarized using the vector norm, which is particularly useful when the goal is a straightforward evaluation of a SBN source. Here, the aggregation of blocked forces and torques is specifically used to analyze whether the source description with two VPs is advantageous or if a single VP is sufficient. For the single VP case, the procedure is identical to that in [13], and the force or torque vector norm is determined as follows:

$$\|\mathbf{F}_{1VP}\| = \sqrt{{F_x}^2 + {F_y}^2 + {F_z}^2} \tag{4}$$

$$\|\mathbf{M}_{1VP}\| = \sqrt{{M_x}^2 + {M_y}^2 + {M_z}^2} \tag{5}$$

In contrast, the 2VP setup introduces additional complexity because not all six force and torque components can be directly combined using vector norms without distinguishing between the two VPs. Moreover, the coupling between the two VPs must be taken into account: unlike in the single-VP case—where only one VP emits vibration power—the two-VP setup also involves power exchange between the VPs.

In general, for SBN sources described by multiple contact points or VPs, two special cases may occur. The first case arises when the VPs are not coupled, i.e., they behave completely independently. This condition is analogous to neglecting transfer-mobilities between coupling points in a power-based approaches. Previous investigations have shown that this simplification can be valid, for example, in systems with multiple point contacts that are connected to the receiver via resilient mounts [20]. In such cases, it is reasonable to treat the two VPs as separate sources. However, if the primary goal is to minimize the number of evaluation quantities, energetic aggregation of all force and torque components of the two VPs can also be considered. However, for the investigated axial piston pump, this first case is likely less relevant. Instead, it can be assumed that the pump's steel flange behaves like a rigid body over a broad frequency range, leading to strong coupling between the two VPs in the 2VP setup.

The second special case, which is therefore more plausible for the investigated axial piston pump, is that the VPs of a SBN source behave like points of a rigid body. In this case, the forces and torques of the two VPs can be geometrically combined into a total force and a total torque as follows:

$$\mathbf{F}_{2VP} = \mathbf{F}_{VP1} + \mathbf{F}_{VP2} \tag{6}$$

$$\mathbf{M}_{2VP} = \mathbf{M}_{VP1} + \mathbf{M}_{VP2} + \mathbf{r}_{VP1} \times \mathbf{F}_{VP1} + \mathbf{r}_{VP2} \times \mathbf{F}_{VP2} \tag{7}$$

Assuming rigid-body behavior is fully satisfied, these quantities are coherently combined in phase. Subsequently, the vector norms of the resulting force vector $\|\mathbf{F}_{2VP}\|$ and the resulting torque vector $\|\mathbf{M}_{2VP}\|$ can be calculated similarly to the 1VP setup. A key point is that the total torque in Equation 7 introduces terms of the form $\mathbf{r} \times \mathbf{F}$, which represent torques arising from the rigid coupling of the VPs [21]. The vectors $\mathbf{r}$ denote the relative vectors from a reference point to each respective VP. For later comparison with the single VP setup, the position of the sole VP in the 1VP setup can be chosen as the reference point. Provided that the interface behaves as a rigid body, the norms $\|\mathbf{F}_{1VP}\|$, $\|\mathbf{M}_{1VP}\|$, and $\|\mathbf{F}_{2VP}\|$, $\|\mathbf{M}_{2VP}\|$, should then be identical. However, it is evident that this method of combining forces and torques of the 2VP setup negates the primary advantage of describing the source with two VPs, which is to require the rigid body condition to be fulfilled for only one VP.

When neither of the two special cases—independence between the individual virtual points or fulfilment of the rigid-body condition—is met, a straightforward aggregation of the SBN excitation on a force or torque level is not possible. Instead, extended approaches, such as power-based methods, must be applied [22].

For the following investigation of the axial piston pump, the aggregation of the 2VP setup

under the assumption of rigid-body conditions is used to assess whether a description with two VPs is necessary, or if a single VP is sufficient.

3. RESULTS

In the following section, the results of the concepts introduced earlier will be presented, with a particular focus on comparing the 1VP and 2VP setups. It is important to note that all transfer functions, also called Frequency Response Functions (FRFs), were obtained through excitation with a shaker. This excitation method allowed for a sufficient signal-to-noise ratio to be achieved starting from 140 Hz. Consequently, all subsequent frequency-dependent quantities will be presented only for frequencies greater than 140 Hz. This lower frequency limit is sufficient to characterize the SBN of 9-piston axial piston pumps at rotational speeds of approximately 1000 rpm and higher, as the first harmonic at 1000 rpm occurs at 150 Hz. The upper considered frequency is 5 kHz, as this frequency is cited in the literature as the upper threshold for acoustic analysis of axial piston pumps [23].
All the results shown in the following were obtained with the support of the open-source Python package pyFBS [24], [25].

3.1. Virtual Point Transformation Quality

A straightforward method for analyzing the quality of the VPT is through the "Consistency" criterion. This metric assesses how accurately the measured forces or motions can be represented by the corresponding VP forces and motions. To determine the consistency, the measured transfer matrix is first transformed into VP coordinates and then mapped back to the original coordinates using the inverse VPT. If the rigid-body condition is fully satisfied, the reconstructed transfer matrix will match the originally measured matrix. However, if the rigid-body condition is violated, the reconstructed matrix will only contain the rigid-body motion, while the flexible dynamics are filtered out. Comparing the original and reconstructed matrices therefore serves as a quality indicator for how well the rigid-body assumption is fulfilled.
A detailed description of the consistency calculation is omitted here, as it would require a full introduction to the VPT, which is beyond the scope of this publication. The procedure is explained in [15], where it is also shown how, based on the original and reconstructed matrices, a single scalar value can be derived for impacts and sensors, using the vector norm. These values are referred to as overall sensor consistency and overall impact consistency, respectively.
In the following, the focus is on impact consistency, as the VPT is applied solely to the forces, as previously mentioned. The overall impact consistency for the 1VP and 2VP setups is illustrated in Figure 3. As demonstrated, the impact consistency for both setups remains approximately 100 % up to around 4.3 kHz. Beyond this frequency, the impact consistency of the 1VP setup declines sharply, whereas that of the 2VP setup remains high. According to these results, it can be concluded that the rigid body condition is satisfied for the 1VP setup only up to 4.3 kHz. For frequencies exceeding 4.3 kHz, the utilization of two VPs in the interface description provides a distinct advantage in impact consistency.
However, the determination of consistency is based solely on the measured transfer matrices and the geometric properties of the VPT. It does not allow for an assessment of whether the rigid body condition is also met during operation.

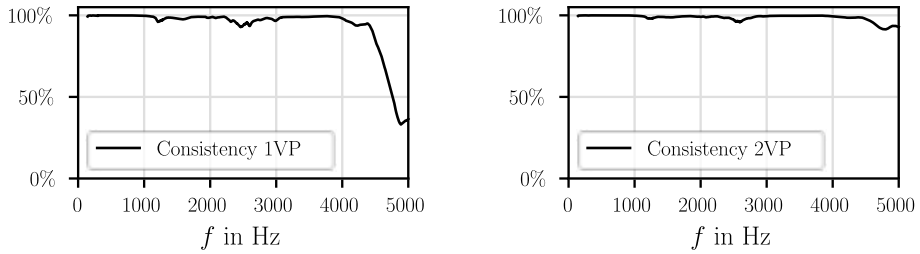


Figure 3: Overall impact consistency. Left: 1VP setup. Right: 2VP setup.

3.2. Results Stationary Operating Point

In the next step, the SBN excitation of an axial piston pump at a stationary operating point will be examined. Specifically, the operating point at $n = 2000$ rpm, $p = 200$ bar, and the maximum displacement volume $V_d = V_{d,max}$ will be considered. The pump investigated in the following study is suitable for use in a 6-ton mini excavator.

Figure 4 presents the results of the OBV for the 1VP and 2VP setups. In each case, the velocity measured directly at the validation sensors $\mathbf{u}_3^{meas.}$ is compared with the velocities predicted based on the blocked forces $\mathbf{u}_3^{pred.}$. The displayed velocity represents the vector norm of the three sensor channels of a triaxial accelerometer. A reference value of $v_{ref} = 10^{-6}$ m/s is used to calculate the dB value. Circles and crosses at the peaks of the harmonics were used to highlight the agreement at those frequencies.

It can be observed that both setups yield very similar results up to the 8th harmonic, i.e., 2.4 kHz, and that the direct measurements agree closely with the predictions within this frequency range. For the 1VP setup, the predictions become less accurate at certain higher harmonics, whereas for the 2VP setup, the predictions and measurements show good agreement, except at the 11th harmonic. Overall, it can be concluded that for both setups, the blocked forces and torques are capable of describing the SBN excitation of the axial piston pump, with the 2VP setup providing slightly better results for frequencies above 2.4 kHz.

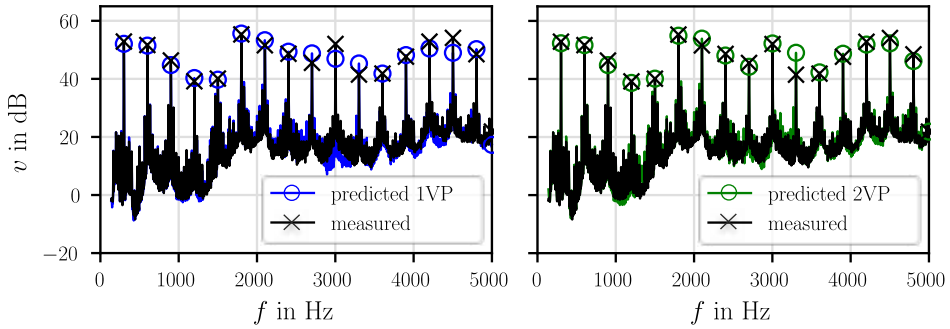


Figure 4: OBV for a stationary operating point. Left: 1VP setup. Right: 2VP setup.

The OBV method enables evaluation of whether interface quantities can describe the excitation caused by a SBN source. The primary objective of this study, however, is to quantitatively investigate the SBN excitation caused by the axial piston pump. Therefore, in the following the blocked forces and torques will be examined. Figure 5 presents the

aggregated blocked forces $\|\mathbf{F}\|$ and blocked torques $\|\mathbf{M}\|$ at the same operating point as the OBV, for both VP setups. For the calculation of the dB values a reference value of $F_{ref} = 10^{-6}$ N is used for forces and $M_{ref} = 10^{-6}$ Nm for Torques.
Before comparing the curves, it is important to first comprehend the differences in the vector norms between the two setups. In the 1VP setup, rigid-body behavior of the entire interface is assumed from the outset, and blocked forces and torques are determined under this assumption. The vector norms are then simply computed using Equations 4 and 5. In contrast, for the 2VP setup, blocked forces and torques are calculated under the condition that rigid-body behavior is satisfied for each VP individually, but not necessarily for the entire interface. The rigid-body assumption between the two VPs—and thus for the full interface—is only imposed later, when the forces and torques from the two VPs are aggregated using Equations 6 and 7.
Consequently, if $\|\mathbf{F}\|$ and $\|\mathbf{M}\|$ of the 1VP and 2VP setups agree, the rigid-body assumption must hold true during the aggregation of the 2VP setup via the vector norms, and thus also for the entire interface. Alternatively, if rigid-body assumption does not hold true, applying Equations (6) and (7) would be expected to produce discrepancies between $\|\mathbf{F}_{1VP}\|$ and $\|\mathbf{F}_{2VP}\|$ as well as $\|\mathbf{M}_{1VP}\|$ and $\|\mathbf{M}_{2VP}\|$. Therefore, the comparison of the vector norms between the two setups serves as a qualitative indicator of whether the rigid-body assumption is satisfied during operation.

Figure 5 shows that both $\|\mathbf{F}\|$ and $\|\mathbf{M}\|$ agree almost perfectly between the setups, with only minor deviations. This confirms what was already suggested by the evaluation of the impact consistency, namely that the interface behaves like a rigid body over a wide frequency range. However, it is interesting to note that while the consistency for the 1VP setup declines beyond 4.3 kHz, the determined vector norms still show good agreement at higher frequencies. This is likely because the consistency criterion is significantly more sensitive to violations of the rigid-body condition than the qualitative comparison of the norms. Yet, for the analysis of SBN excitation, the blocked forces and torques are of primary relevance, and these agree well for both setups. As a result, the considerably more complex interface description using two virtual points is not necessary, particularly as the dominant harmonics lie in the low-frequency range, where agreement is strongest.

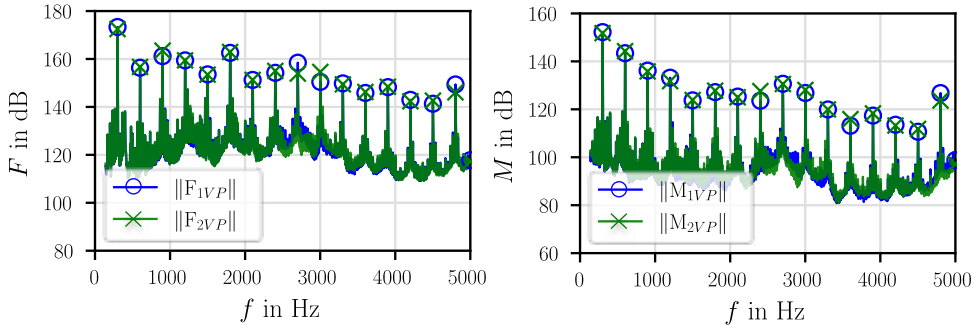


Figure 5: Comparison of SBN excitation for the 1VP and 2VP setup for a stationary operating point. Left: Norm blocked forces $\|\mathbf{F}\|$. Right: Norm blocked torques $\|\mathbf{M}\|$

3.3. Results Run-Up

As the final part of the analysis, the excitation of the SBN caused by the axial pistons during a speed run-up at a pressure of $p = 200\,\text{bar}$ and maximum displacement volume $V_d = V_{d,max}$ is examined. For each discrete speed step, a distinct excitation spectrum is obtained. Since analyzing each individual spectrum in detail would not be efficient, the spectra are consolidated by calculating the root-mean-square (RMS) value of each spectrum. The entire frequency range from 140 to 5000 Hz is considered in this calculation. However, as shown in Figure 5, it is expected that the resulting RMS value is primarily determined by the dominant low-frequency harmonics.

The corresponding blocked forces and blocked torques in each spatial direction are presented in Figure 6. Specifically, both the forces and torques of the 1VP setup $F_{i,1VP}$, $M_{i,1VP}$ and the total forces and torques per coordinate direction for the 2VP setup $\sum F_{i,2VP}$ $\sum M_{i,2VP}$ are shown. These results match almost perfectly, except for minor deviations in the torque around the z-axis, confirming the findings of the previous analysis.

For torques, the sum of the 2VP torques is additionally plotted with dotted lines, excluding the $\mathbf{r} \times \mathbf{F}$ terms of Equation 7. As expected, this omission has no effect on the torque in the x-direction $\sum M_{X,2VP}$, because the two VPs in the 2VP setup are displaced only along the x-direction (see Figure 1). However, for the other two directions, the torque is significantly underestimated when the $\mathbf{r} \times \mathbf{F}$ terms are omitted. This highlights the strong coupling between the two VPs and emphasizes that these terms must generally be taken into account when aggregating the torques of SBN sources described with multiple connection points.

Each plot also shows the vector norm of the 1VP forces $\|\mathbf{F}_{1VP}\|$ and torques $\|\mathbf{M}_{1VP}\|$. The vector norms of the 2VP setup are not shown, as they are nearly identical. This representation makes the dominant force or torque contribution at each rotational speed immediately visible, aiding in root-cause analysis and the development of countermeasures at critical operating points. If a critical operating point is identified, a detailed spectral analysis can help to determine harmonics responsible for strong SBN excitation.

For the axial piston pump under investigation, a key observation is that the force excitation is primarily dominated by the force in the y-direction, while the force in the z-direction contributes minimally to the overall excitation. Analysis of torque excitation revealed that all three rotational DoF significantly contribute to the overall excitation, depending on the rotational speed. For example, at 1500 rpm, the excitation is predominantly governed by torque around the z-axis, corresponding to axial torque excitation.

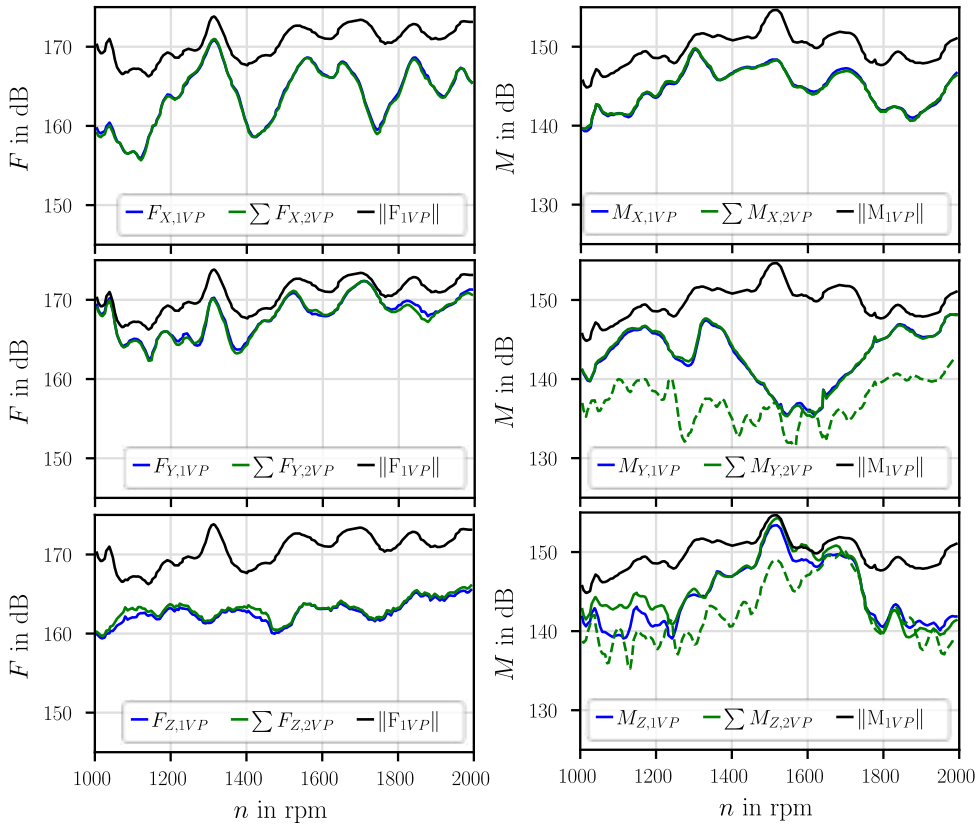


Figure 6: SBN excitation of an axial piston pump for a run-up. Left: Blocked forces. Right: Blocked torques (—— with $\mathbf{r} \times \mathbf{F}$ terms, - - - without $\mathbf{r} \times \mathbf{F}$ terms)

4. CONCLUSION AND OUTLOOK

To analyze the SBN excitation generated by an axial piston pump, the concepts of VPT and the in-situ blocked force TPA were introduced. For the investigated pump, a key objective was to compare the interface description of the pump using one versus two VPs.

Theoretical considerations indicate that that aggregating the interface forces and torques of multiple VPs using the vector norm is reasonable only when the VPs are largely independent or when the rigid-body assumption holds. In the latter case, however, it may be more appropriate to model the interface with a single VP from the outset.

This was also the case for the examined axial piston pump. When investigating the difference between the 1VP and 2VP setups, it was found that the coupling between the two VPs in the 2VP setup was very strong. Within the investigated frequency range, their interaction could be approximated as rigid body coupling. Consequently, representing the source with two VPs is unnecessary, and a single-VP representation in the center of the interface is more appropriate. However, it is important to note that this conclusion applies specifically to the pump under investigation and to the analysis of SBN excitation. As the OBV results show, describing an SBN source with multiple VPs can be advantageous—or even necessary—especially at higher frequencies.

Based on the blocked forces determined for a single VP, the SBN excitation of the pump can be analyzed for the individual DoF of the VP, enabling the identification and investigation of critical operating points. For the examined pump force excitation is predominantly in the y-direction, with minimal excitation in the z-direction. In contrast, for torque excitation, all directions contribute to the overall excitation, depending on the rotational speed of the pump.
A limitation of the applied approach is that forces and torques are evaluated separately and cannot be combined directly due to their differing physical units. A solution to this is provided by power-based approaches. By combining forces and torques with the corresponding velocities and angular velocities, power quantities for each DoF can be derived. This would allow for direct comparison between translational and rotational excitation.
Another promising direction is to study excitation in conjunction with a specific target system. Such an analysis would make it possible to evaluate which excitation DoF has the greatest influence on a specific target, like a microphone at the diver's ear, providing a more application-oriented understanding of the SBN.

5. ACKNOWLEDGEMENTS

Acknowledgements are given to Bosch Rexroth AG for the financial support and the provision of the acoustic test bench, as well as to the Bosch Sound Propagation expert team for their technical and professional support.

6. REFERENCES

[1] W. Fiebig, "Geraeuschminderung in hydraulischen Systemen," VDI Verlag, Düsseldorf, Germany, 337, 2008.

[2] R. Klop, "Investigation of Hydraulic Transmission Noise Sources," Purdue University, West Lafayette, IN, USA, 2010.

[3] R. Klop and M. Ivantysynova, "Investigation of Noise Sources on a Series Hybrid Transmission," *Int. J. Fluid Power*, vol. 12, no. 3, pp. 17–30, Jan. 2011, doi: 10.1080/14399776.2011.10781034.

[4] A. Waitschat, L. Nordmann, F. Thielecke, and V. Pommier-Budinger, "Enhanced Toolbox for the Combined Analysis of Fluid- and Structure-Borne Noise of Hydraulic Systems," in *BATH/ASME 2016 Symposium on Fluid Power and Motion Control*, Bath, UK: American Society of Mechanical Engineers, Sep. 2016, p. V001T01A001. doi: 10.1115/FPMC2016-1701.

[5] T. Opperwall and A. Vacca, "A Transfer Path Approach for Experimentally Determining the Noise Impact of Hydraulic Components," presented at the SAE 2015 Commercial Vehicle Engineering Congress, 2015, pp. 2015-01–2854. doi: 10.4271/2015-01-2854.

[6] T. Nafz, "Aktive Ventilumsteuerung von Axialkolbenpumpen zur Geräuschreduktion hydraulischer Anlagen," RWTH Aachen, Aachen, Germany, 2011.

[7] B. Müller, “Einsatz der Simulation zur Pulsations- und Geräuschminderung hydraulischer Anlagen,” RWTH Aachen, Aachen, Germany, 2002.

[8] International Organization for Standardization, *ISO 10767-1:2015 Hydraulic fluid power - Determination of pressure ripple levels generated in systems and components - Part 1: Method for determining source flow ripple and source impedance of pumps*, ISO 10767-1, 2015.

[9] International Organization for Standardization, *ISO 10767-1:1996 Hydraulic fluid power - Determination of pressure ripple levels generated in systems and components - Part 1: Precision method for pumps*, 1996.

[10] International Organization for Standardization, *ISO 20270-1:2019 Acoustics — Characterization of sources of structure-borne sound and vibration — Indirect measurement of blocked forces*, 2019.

[11] M. Vogt and M. Geimer, “An Approach for Evaluating Structure-Borne Noise Emission of Axial Piston Units Using Blocked Forces,” in *Proceedings of the 19th Scandinavian International Conference on Fluid Power (SICFP’25)*, Linköping, Sweden, 2025. doi: 10.13052/rp-9788743808251A35.

[12] M. van der Seijs, D. van den Bosch, D. Rixen, and D. de Klerk, “An improved methodology for the virtual point transformation of measured frequency response functions in dynamic substructuring,” in *Proceedings of COMPDYN 2013*, Kos Island, Greece, 2013, pp. 4334–4347. doi: 10.7712/120113.4816.C1539.

[13] M. Vogt and M. Geimer, “Comparison of different approaches for Evaluating Structure-Borne Noise Emission of Axial Piston Units,” in *Proceedings of DAGA 2025 (Deutsche Jahrestagung für Akustik)*, Copenhagen, Denmark, 2025.

[14] T. Bregar, A. El Mahmoudi, G. Čepon, D. J. Rixen, and M. Boltežar, “Performance of the Expanded Virtual Point Transformation on a Complex Test Structure,” *Exp. Tech.*, vol. 45, no. 1, pp. 83–93, Feb. 2021, doi: 10.1007/s40799-020-00398-1.

[15] M. V. van der Seijs, “Experimental Dynamic Substructuring: Analysis and Design Strategies for Vehicle Development,” Delft University of Technology, Delft, The Netherlands, 2016. doi: 10.4233/UUID:28B31294-8D53-49EB-B108-284B63EDF670.

[16] C. L. Balogh, T. Kimpián, and T. Ungvári, “An Examination of Component-Based TPA Performance and Feasibility for Steering Motors,” *J. Phys. Conf. Ser.*, vol. 2677, no. 1, p. 012011, Dec. 2023, doi: 10.1088/1742-6596/2677/1/012011.

[17] International Organization for Standardization, *ISO 3019--1:2001 Hydraulic fluid power - Dimensions and identification code for mounting flanges and shaft ends of displacement pumps and motors - Part 1: Inch series shown in metric units*, 2001.

[18] M. V. van der Seijs, D. de Klerk, and D. J. Rixen, “General framework for transfer path analysis: History, theory and classification of techniques,” *Mech. Syst. Signal Process.*, vol. 68–69, pp. 217–244, Feb. 2016, doi: 10.1016/j.ymssp.2015.08.004.

[19] A. T. Moorhouse, A. S. Elliott, and T. A. Evans, “In situ measurement of the blocked force of structure-borne sound sources,” *J. Sound Vib.*, vol. 325, no. 4–5, pp. 679–685, Sep. 2009, doi: 10.1016/j.jsv.2009.04.035.

[20] B. M. Gibbs and M. Villot, “Structure-borne sound in buildings: Advances in measurement and prediction methods,” *Noise Control Eng. J.*, vol. 68, no. 1, pp. 1–20, Jan. 2020, doi: 10.3397/1/37681.

[21] J. L. Meriam, J. N. Bolton, and J. L. /Kraige Meriam L. G., *Engineering Mechanics*, 9th ed. 2018.

[22] A. T. Moorhouse, “On the characteristic power of structure-borne sound sources,” *J. Sound Vib.*, vol. 248, no. 3, pp. 441–459, Nov. 2001, doi: 10.1006/jsvi.2001.3797.

[23] H. Münch, “Acoustically extended virtualization of industrial productsAkustisch erweiterte Virtualisierung von Industrieprodukten,” FAU Erlangen-Nürnberg, Erlangen, Germany, 2021. Accessed: Nov. 09, 2024. [Online]. Available: https://open.fau.de/handle/openfau/15906

[24] T. Bregar, A. El Mahmoudi, M. Kodrič, G. Čepon, M. Boltežar, and D. J. Rixen, “Introducing pyFBS: An Open-Source Python Package for Frequency Based Substructuring and Transfer Path Analysis,” in *Proceedings of the 39th International Modal Analysis Conference (IMAC XXXIX)*, M. S. Allen, W. D’Ambrogio, and D. Roettgen, Eds., 2021, pp. 81–89. doi: 10.1007/978-3-030-75910-0_10.

[25] T. Bregar *et al.*, “pyFBS: A Python package for Frequency Based Substructuring,” *J. Open Source Softw.*, vol. 7, no. 69, p. 3399, Jan. 2022, doi: 10.21105/joss.03399.

Biographies



Matthias Vogt received his B.Sc. (2019) and M.Sc. (2021) in Mechanical Engineering from the Karlsruhe Institute of Technology (KIT). From 2022 to 2023, he was a researcher at the Institute of Mobile Machines (Mobima) at KIT. Since 2023, he has been pursuing a Ph.D. at Mobima in cooperation with Bosch Rexroth AG. His research focuses on developing a method for evaluating the structure-borne noise emissions of axial piston pumps.



Marcus Geimer received his Diploma degree in mechanical engineering from the RWTH Aachen University, Aachen, Germany in 1990. In 1995, he received his Ph.D. from the Institute of Hydraulics and Pneumatics, today named Institute for Fluid Power Drives and Systems, RWTH Aachen University. He started his industrial career in 1995 in the field of construction at the company Krupp Berco Bautechnik GmbH, Essen (Germany). There, he was the leader of the research group for hydraulic breakers. In 2000, he changed to Bucher Hydraulics Gmbh, Klettgau (Germany), where he led the construction and customer development for mobile hydraulics. Since 2005, he is a full professor and director at the Institute of Mobile Machines (Mobima), at the Karlsruhe Institute of Technology KIT, Germany. Hydraulic and electric drives for the traction drive as well as for the working functions and controls for mobile machines are main research areas of the institute.

Review of Single-Sided and Double-Sided Axial Piston Machine Characteristics

Thomas Heeger, Liselott Ericson

Linköping University, Department of Management and Engineering, Division of Fluid and Mechatronic Systems, Linköping, Sweden, liselott.ericson@liu.se

Abstract.

Most conventional axial piston machines have a single rotating group. However, double-sided machines with another set of pistons can have the same displacement with different properties. Double-sided machines can counteract forces and moments between each side and therefore reduce structure-borne noise and bearing loads. Further, there are different setups of double-sided machines, with a main difference being if there is a phase shift between the sides. This paper qualitatively compares the properties of different machine setups.

Keywords. Axial piston pump, fluid power, pump design, double-sided machine.

1. INTRODUCTION

There are various types of positive displacement machines. One machine type which is very popular in mobile machinery is the axial piston machine type. For this machine type, there are different options available on the market. One significant difference can be how the rotating group is set up. Many machines on the market have one set of pistons (e.g., [1, 2]), which will be referred to as single-sided machines in this paper. However, several sets of rotating pistons can be driven on the same shaft and are available on the market, e.g., double row machines with two sets of pistons on different diameters (e.g., [3]), machines using modular back to back designs (e.g., [4]) where each set of pistons has its own housing, or tightly integrated machines where a mirrored design of two sets of pistons shares the same housing and the same inlet and outlet [5].

Some advantages of using two sets of pistons have been presented in the literature. For example, Innas BV presented their floating cup machine with two sets of pistons (e.g., [6]), and mentions the benefit of increasing the number of pistons, and the compensation of axial forces due to the mirrored design. This results in reduced bearing loads and is expected to reduce structure-borne noise. The high number of pistons also reduces both kinematic and compressible flow pulsations and thus reduces fluid-borne noise. Similarly, the Vickers MPEV3-040-2 aerospace motorpump included two inline machines driven in opposite directions by individual electric motors, which cancels both radial and axial piston forces. This results in reduced yoke and bearing loads. Further, the two inline machines shared one displacement controller [7]. Mehta [8] used a tandem pump with an angular between the mirrored sets of pistons to reduce the torque ripple by 75%. Danes and Vacca [9] aimed for destructive interference of the flow of a tandem pump in order to reduce fluid-borne noise. For low harmonics, measurements

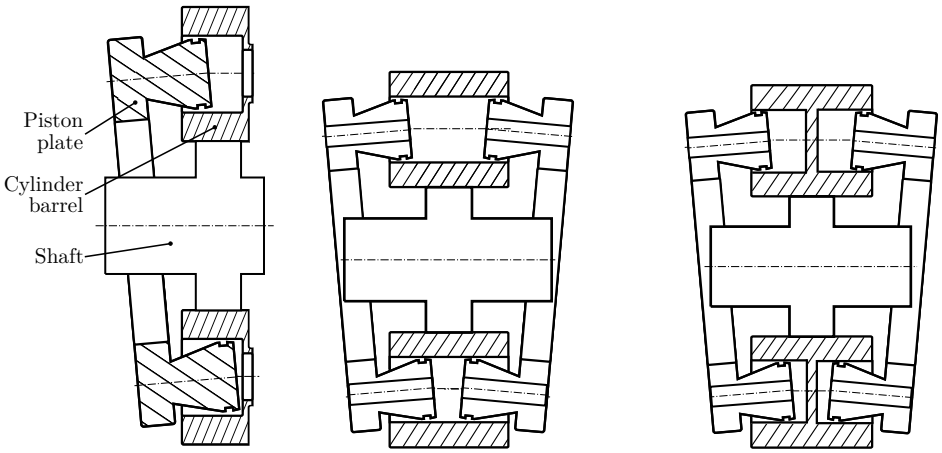
All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

showed significant reductions of sound pressure levels, whereas, contrary to the expectation from simulation, this was not the case for higher harmonics. Ericson and Heeger discussed using mirrored pump designs to achieve discretely variable by idling one pump half [10] or continuously variable displacement by rotating the valve plates [11].

1.1. *Machine Configurations*

Figure 1.1 sketches different axial piston machine configurations on the example of floating piston machines. Figure 1.1a shows a single-sided set up, whereas Figures 1.1b and 1.1c show double-sided machines with mirrored rotating groups. In Figure 1.1b, pistons from each side share a displacement chamber, whereas in Figure 1.1c, the cylinder barrel provides a separate displacement chamber for each piston. Note that further configurations are feasible, e.g., by using multiple cylinder barrels, but are not in scope of this paper.

Figure 1.1 shows floating piston machines, but the same machine configurations are applicable also for floating cup machines or conventional slipper type machines as sketched in Figure 1.2. As bent-axis machines do not provide a through-drive possibility [12], double-sided machines are not applicable for them.

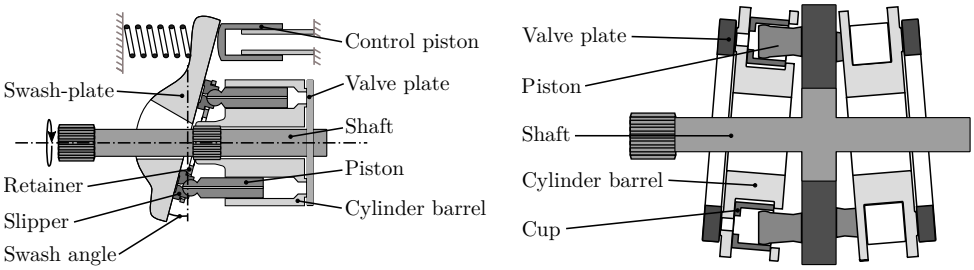


(a) Single-sided machine.

(b) Double-sided machine with connected sides.

(c) Double-sided machine with separated sides.

Figure 1.1: Schematic of different floating piston machine versions [10, 13].



(a) Sketch of main components of a slipper-type machine.

(b) Sketch of main components of a floating cup machine of mirrored design. Note the angular offset between the machine halves.

Figure 1.2: Schematic of axial piston machines [13].

1.2. *Research Gap*

Each of the aforementioned references was focusing on some specific design aspects. This paper aims to provide a wider comparison of pumps with one set of pistons with the two types of mirrored pump designs. The discussed aspects include flow, force, moment, and torque ripples; bearing loads; barrel tipping; aspect ratios; piston count.

1.3. *Limitations*

The simulations in Section 9 assume a inlet pressure of 3 bar and therefore cannot indicate differences in self-suction speed of the different machines. The pressure relief grooves in the study were manually tuned and therefore a comparison of commutation behaviour is not fully objective.

1.4. *Paper Outline*

Section 2 summarises background on flows, forces, and moments in axial piston machines. Based on this, Sections 3 to 8 qualitatively discuss the properties of the machine setups shown in Figure 1.1. Section 9 presents a lumped-parameter simulation study for specific designs. Finally, Section 10 summarises the properties for the different setups and shows conclusions.

2. Background on Flows, Forces, and Moments in Axial Piston Machines

The flow, force, and moment contributors in axial piston machines are crucial to understand their noise behaviour and bearing loads. Manring [14] and Ivantysyn and Ivantysynova [15] provided detailed information on this topic and Heeger et al. [16] summarised and visualised the noise contributors. The following section repeats this background information. Figure 2.1 sketches an exemplary sketch of nine pistons and valve plate kidneys.

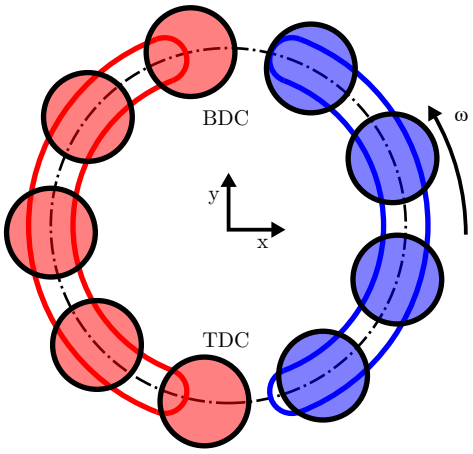


Figure 2.1: Sketch of nine pistons and valve plate kidneys. The sketch includes the location of bottom dead centre (BDC), top dead centre (TDC), as well as the coordinate system and direction of rotation.

In stationary operation, the same events are in principle repeated for each piston. In a simulation environment, neglecting e.g., manufacturing tolerances, the events are even exactly repeated for each piston and therefore piston periods are used to visualise the machine's behaviour. A piston period represents the time it takes between two pistons passing the same

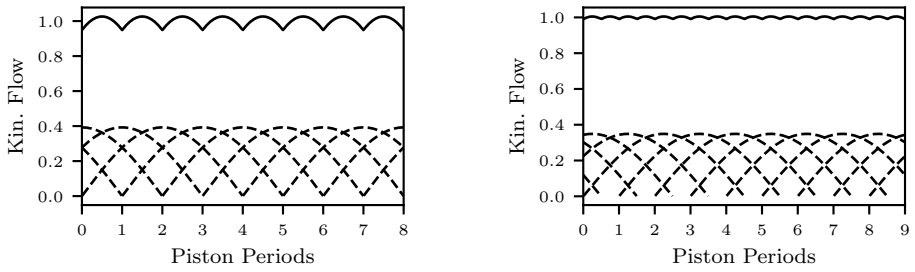
location, and the figures show the results for two consecutive piston periods. Note that the time of one piston period depends both on the machine's speed and piston count.

2.1. *Flow Ripples*

The overall flow ripple is the sum of the kinematic and the compressible flow ripples.

2.1.1. *Kinematic Flow Ripples*

In axial piston machines, the piston velocity and therefore the flow of one displacement chamber follow a sine shape. As the number of displacement chambers is finite, the total delivered flow is not constant and kinematic flow ripples as shown in Figure 2.2 occur.



(a) Machine with 8 pistons.

(b) Machine with 9 pistons.

Figure 2.2: Exemplary kinematic flow ripples. Dashed: contributions of the individual pistons, solid: total kinematic flow. The figures are normalised with the average total kinematic flow.

The kinematic flow ripple behaves differently for machines with odd and even piston counts z. For even piston counts, the chambers connect and disconnect from the ports in pairs, whereas for odd piston counts, there is half a piston period phase shift between the events of connecting and disconnecting. Consequently, odd piston counts yield lower peak-to-peak kinematic flow ripples but higher frequencies [15]. Conventionally, 7 or 9 pistons are used [14].

2.1.2. *Compressible Flow Ripples*

Pressure equalisation (and the fluid's compressibility) cause compressible flow during commutation. In mobile working machines, the compressible flow ripples are usually larger than the kinematic ones [17]. Without any commutation features, using so-called zero-lapped valve plates, very large compressible flow ripples as shown in Figure 2.3 occur. Various commutation features can reduce the compressible flow ripple amplitude, and reviews are provided e.g., by Johansson [17] and Heeger et al. [18]. The frequency of the compressible flow is not affected by whether the piston count is odd or even.

Increasing piston counts can decrease the compressible flow ripple, as the stroke and dead volume of each displacement chamber are reduced, thus reducing the compressible flow required for pressure equalisation.

2.2. *Force, Moment and Torque Ripples*

2.2.1. *Axial Force Ripples*

The axial force of a piston depends on its pressure level. A varying count of pressurised pistons therefore causes axial force ripples. When neglecting commutation effects, even piston counts do not have an axial force ripple, as the chambers connect and disconnect in pairs. Odd piston counts however do have force ripples in the magnitude of the force of one piston. Figure 2.4

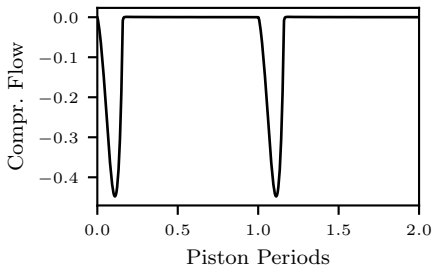
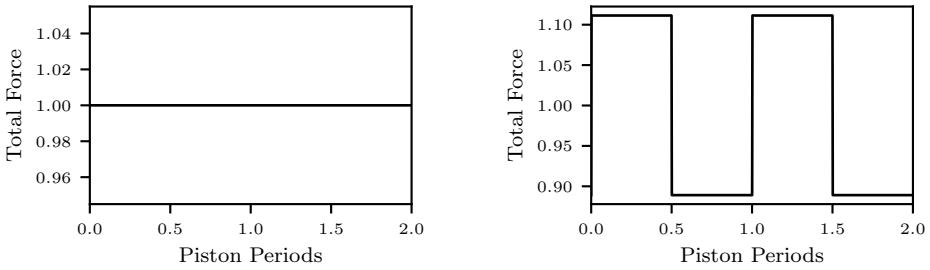


Figure 2.3: Compressible flow ripple for a zero-lapped valve plate, normalised with the average total kinematic flow.

shows the piston force ripple, neglecting commutation effects. Johansson [17] considers the axial force as the main source of structure-borne noise.



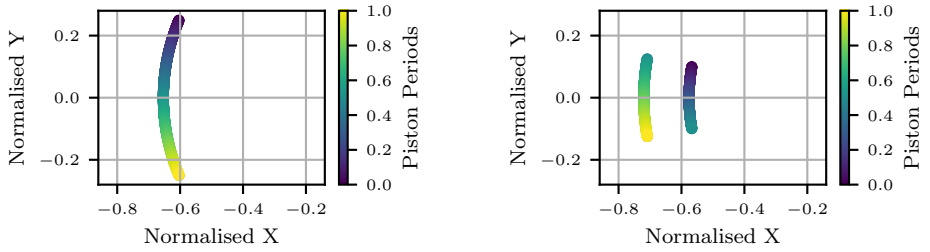
(a) Machine with 8 pistons.

(b) Machine with 9 pistons.

Figure 2.4: Exemplary normalised piston force ripples.

2.2.2. *Bending Moment and Drive Torque Ripples*

A varying count of pressurised pistons and the varying piston positions cause bending moment ripples, which also contribute to structure-borne noise. Figure 2.5 shows the position of the force resultants for one piston period 2.5.

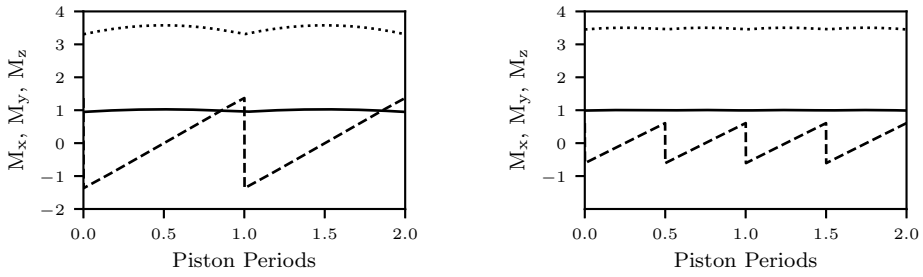


(a) Machine with 8 pistons.

(b) Machine with 9 pistons.

Figure 2.5: Exemplary location of resultant of piston forces, neglecting commutation effect. Normalised with the radius of cylinder barrel bore location.

Bending moments occur around the x- and the y-axis (refer to Figure 2.5 for the axis definition) and cause loads on the bearings. The moment around the z-axis is the drive torque. Figure 2.6 shows bending moment ripples and drive torque ripples.



(a) Machine with 8 pistons.

(b) Machine with 9 pistons.

Figure 2.6: Exemplary bending moment ripples (M_x - dashed, M_y - dotted) and drive torque ripples (M_z - solid line), normalised in relation to the average drive torque.

The bending moment ripple is increased for even piston counts, as the location of the resultant piston forces varies more. The ripple around the x-axis is larger than around the y-axis, as the normalised x-position is larger than the normalised y-position when the pistons enter / leave a port. Further, the x-moment has a varying sign and can therefore release geometric plays.

3. Bearing Loads and Barrel Tipping

In single-sided machines, the axial piston forces load a bearing. For double-sided machines, the axial forces from each side counteract each other. In case of the double-sided machine with connected sides, the forces and moments of each side are identical, leading to low bearing loads. For double-sided machines with a phase shift between the sides, the axial piston force is of alternating sign and of the magnitude of the force of one piston. Figure 3.1 visualises the resulting forces and moments for the three considered machine setups and a machine with 11 pistons. The origin for the moment calculation is in the intersection of the shaft axis and the mirroring plane to create the mirrored pumps. Moments are normalised by the average drive torque required to drive a single-sided pump [10].

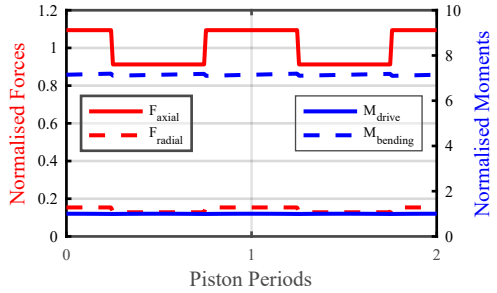
Mirroring a rotating group obviously doubles the displacement, the radial forces, and the drive torque, but reduces axial forces and bending moments. Note that for comparing single-sided and double-sided machines of the same displacement, the dimensions can be scaled according to the scaling laws presented by Achten [19].

3.1. *Bearing Loads*

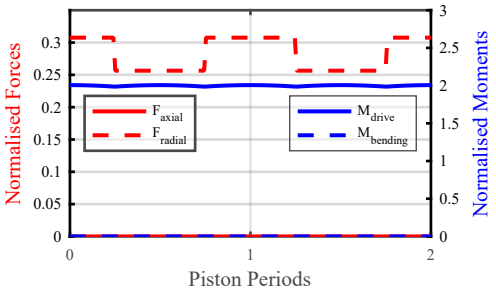
In case of a single-sided machine, the bearings are loaded with axial and radial forces and bending moments as shown in Figure 3.1a. As Figure 3.1b shows, the axial forces and bending moments from each machine side counteract each other in double-sided machines with connected sides. Therefore, the bearing are in principle only loaded with radial forces. If a double-sided machine with an angular offset between the two sides is used, the bearings are loaded with alternating forces with the magnitude equivalent to one pressurised piston as shown in Figure 3.1c.

3.2. *Barrel Tipping*

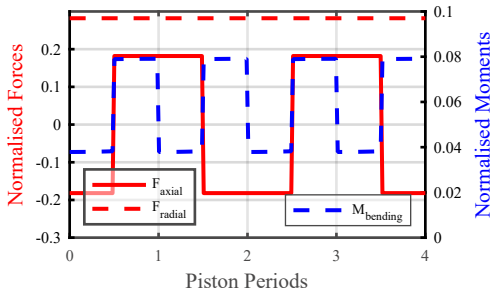
Achten et al. [20] investigate barrel tipping in single-sided floating cup and slipper-type machines, and show that hydrostatic, centrifugal, friction, spring, and reaction forces affect the cylinder barrel and therefore its tipping moment. In double-sided machines without an angular offset between the machine sides, the tipping moments from both machine sides counteract



(a) Normalised forces and moments for a single-sided machine.



(b) Normalised forces and moments for a mirrored machine with connected machine sides.



(c) Normalised forces and moments for a mirrored machine with independent machine sides and an offset of half a period between both sides. Due to the offset between the machine sides, there are twice as many piston periods per revolution.

Figure 3.1: Forces and bending loads on the bearings for different machine configurations. The double-sided machine in this figure have twice the displacement of the single-sided machine [10].

each other, thus providing higher robustness against barrel tipping. If there is an angular offset between the machine sides, the moments from each side are not fully compensated, but the resulting oscillating moments on the barrel are of lower magnitude than in single-sided machines.

4. Flow Ripple

4.1. *Interference of Flow from Double-sided Machines*

Danes and Vacca [9] aimed for destructive interference of the flow of a tandem pump in order to reduce fluid-borne noise. The considered tandem pump consisted of two pumps with nine pistons each. Simulation results showed that a phase shift of half a period eliminates the odd pressure ripple harmonics, whereas a phase shift of a quarter period eliminates harmonics of the order $2 + 4n$. Further, the pressure ripples for the tandem machine were lower than for a simulated single-sided machine of the same displacement, independent of the phase shift applied to the tandem machine halves. For the first and third harmonic, measurements showed significant reductions of sound pressure levels when using a phase shift of half a period between the pump halves, whereas, contrary to the expectation from simulation, this was not the case for higher harmonics. Danes and Vacca give insufficient symmetry in their experiment as a potential cause for this.

4.2. *Dead Volume*

Larger dead volumes cause increased compressible flow pulsations. Double-sided machines that require flow through the pistons (see, e.g., Figure 1.1) possess increased dead volume ratios.

5. TORQUE RIPPLE

Mehta [8] considered the torque ripple to be the main source of noise in hydraulic machines. He presented a slipper type machine with a central manifold and mirrored rotating groups on each side of the manifold. Mehta used this design to optimise the angular offset between the rotating groups to minimise the torque ripple. For even piston counts on each side, the ideal offset is half a period, whereas for odd piston counts, the ideal offset is a quarter of a period. When neglecting commutation and compressibility effects, the best offset angle reduces the peak-to-peak torque pulsations by 75% in comparison to the worst offset angle.

6. ASPECT RATIOS

Heeger et al. [21] expanded Achten's [19] method for the scaling of rotating groups of axial piston machines for floating piston machines. Figure 6.1 shows the rotating group's diameter, axial length, and volume for single-sided and double-sided floating piston machines of the same displacement. Obviously, the double-sided machine has smaller outer diameters and larger axial lengths. Also, its volume of the rotating parts is slightly lower and the total piston count is larger for the most compact machine.

In the context of the integration with electric machines, Heeger et al. [22] stated that single-sided machines are more advantageous for a compact integration via axial stacking, whereas double-sided machines are more advantageous for a compact radial integration within the core of an electric machine.

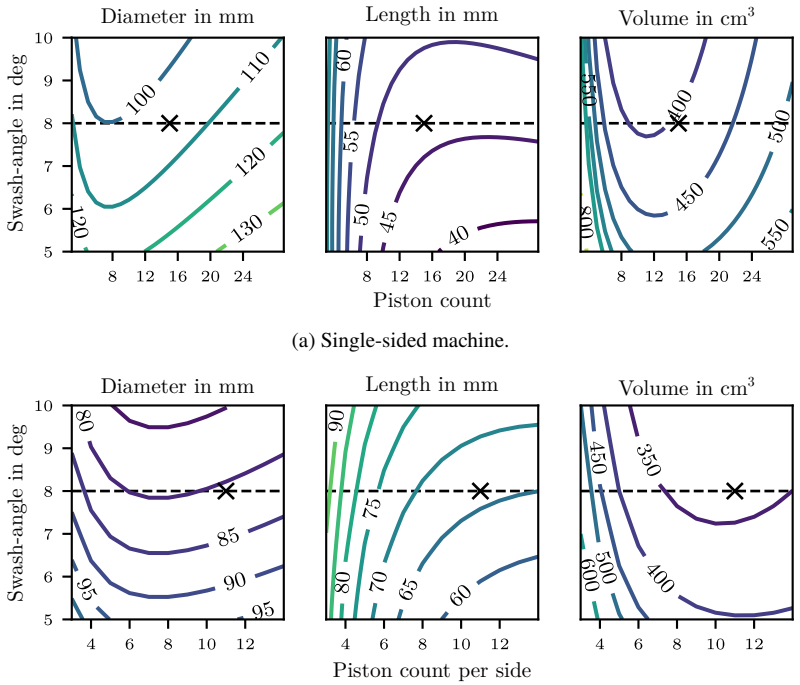
7. VARIABLE DISPLACEMENT

Variable single-sided machines vary their displacement by varying their swash-angle. Using double-sided machines opens opportunities for unconventional methods to vary displacement.

Valve plate rotation has been considered to alter the effective pre-compression for small rotation angles [23], and to provide variable displacement for large valve plate rotation angles [24]. Variable displacement via valve plate rotation however requires commutation to take place away from the dead centres and at significant axial piston speeds. This results in challenges such as cavitation, high-pressure peaks, and pulsations, and has therefore hindered a widespread application of valve plate rotation [25]. Note however that the Innas Hydraulic Transformer relies on valve plate rotation to alter the flow through each of its three ports [25].

In double-sided machines with connected sides as shown in Figure 1.1b each valve plate affects the effective pre- and de-compression angles. Therefore, the displacement and commutation behaviour can be altered at the same time. However, the high piston velocities at low setting ratios provide a significant throttling effect, and Heeger et al. [11] concluded that this method is mainly applicable to vary displacement at relatively large setting ratios but unsolved challenges remain for low setting ratios.

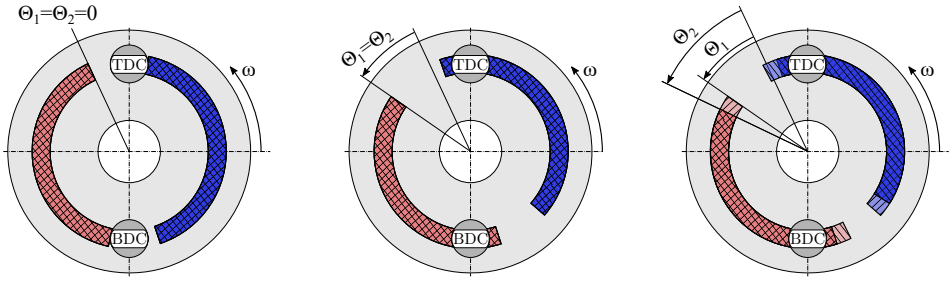
Alternatively, double-sided machines with separated sides can provide discretely variable displacement by circulating the flow of one machine side, e.g., using the setup sketched in Figure 7.2. This functionality is in principle available on the market where dual-flow machines can be equipped with shut-off valves (e.g., [27, 28]). This method can also be applied on double-sided machines with separated sides. However, when circulating oil at tank pressure level, the axial force is no longer balanced between the machine sides. To provide a better



(a) Single-sided machine.

(b) Double-sided machine with mirrored rotating group.

Figure 6.1: Rotating group outer diameter, axial length, and volume for floating piston machines with a displacement of 35 cc/rev. The crosses mark the designs of reference machines [21].



(a) Both valve plate rotation angles (Θ_1, Θ_2) are zero, and the valve plates are in the same position at maximum displacement.

(b) Joint valve plate rotation: both valve plates have been rotated by the same angle.

(c) Joint and relative valve plate rotation: valve plate 2 has been rotated by an additional angle.

Figure 7.1: Concept of joined and relative valve plate rotation [26].

balancing of the axial forces, an elevated low-pressure level can be achieved by a setup as sketched in Figure 7.2 [10].

8. Channel Design

The channel design depends on the machine type and on whether one or two cylinder barrels are used. When using just one cylinder barrel and axial ports, double-sided designs require flow through the pistons. This is a significant restriction, increasing the risk for cavitation and reducing self-suction speeds [21]. Mehta [8] avoided this for a slipper-type machine by using a

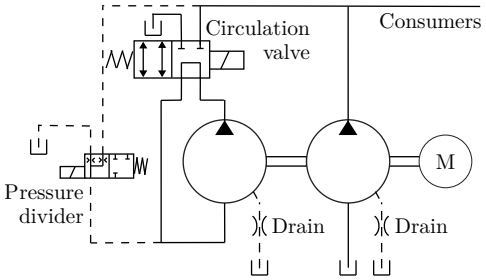


Figure 7.2: A possible setup for circulating flow from one pump half at an elevated pressure level [10].

central manifold and two cylinder barrels. A central manifold is however inconvenient for a radial integration of the hydraulic machine inside an electric machine.

The floating cup machines also uses two cylinder barrels and the flow goes through the cups, which can provide larger flow channels than pistons. The floating cup machine does not have a central manifold, instead it has two inlets and outlets, which can be combined through channels in the housing.

9. EXEMPLARY ROTATING GROUP SIMULATION STUDY

9.1. *Setup*

An exemplary simulation study aims to compare machines that are similar to traditional machines setups and to alternative machine setups. The study uses a lumped parameter model as sketched in Figure 9.1 using the software Hopsan [29]. Each machine has a displacement of 35 cc/rev. Table 9.1 shows the model inputs. The models inputs represent typical values for single-sided machines of bent-axis (A), inline (B), and floating piston (C) type, as well as for double-sided machines of floating cup (D) and floating piston (E) type. The inlet pressure is set to 3 bar. As commutation feature, pressure relief grooves as sketched in Figure 9.2 are applied. The pressure relief groove shape is set to triangular. The pressure relief grooves are manually tuned to reduce flow pulsations for operating with up to 3000 rpm and 350 bar. Only one quadrant is considered.

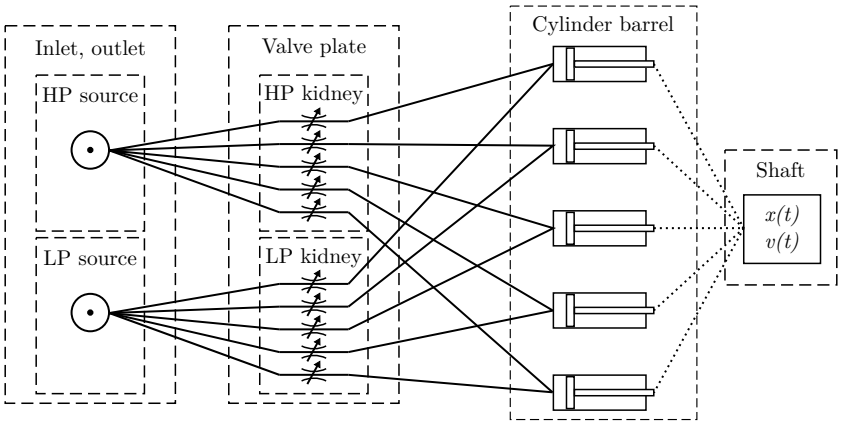


Figure 9.1: Simulation model for axial piston machines, modified from [11]. t - time, v - axial piston velocity, x - axial piston position.

Table 9.1: Simulated setups.

Setup	Machine type	Piston count	Swash angle in degrees	Dead volume ratio	Piston pitch parameter
A	Single-sided	7	40	0.2	0.13
B	Single-sided	9	20	0.8	0.33
C	Single-sided	15	8	0.2	0.13
D	Double-sided (separated)	2x12	8	0.2	0.27
E	Double-sided (connected)	2x11	8	0.7	0.13

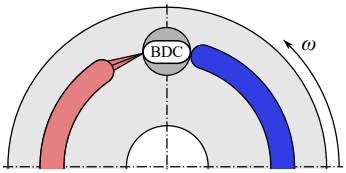
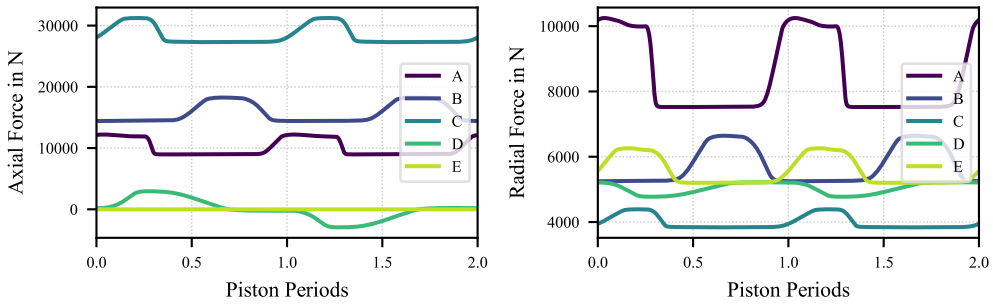


Figure 9.2: Pressure relief groove [18].

9.2. *Results*

Figures 9.3 to 9.6 show simulation results for the setups given in Table 9.1.

Figure 9.3 shows the axial and radial forces for the different setups. As expected, single-sided machines possess larger axial forces. The low displacement angle of the single-sided machine setup C causes high axial forces, and the bent-axis type machine setup A possesses large radial forces. The double-sided setups D and E provide low axial forces, but do not reduce the radial forces.



(a) Axial forces on bearings.

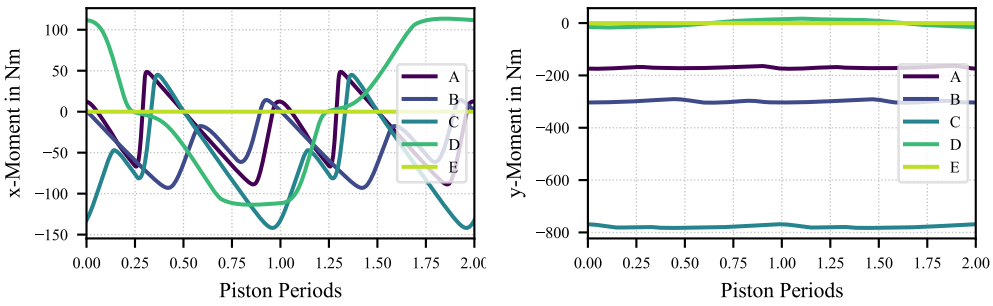
(b) Radial forces on bearings.

Figure 9.3: Comparison of bearing forces. Machine types see Table 9.1.

Figure 9.4 shows the bending moments for the different setups. For single-sided designs, increased displacement angles reduce the bending moment magnitudes. The double-sided setup E fully cancels the bending moments. The double-sided setup D which includes a phase shift between the sides has low moments around the y-axis but high moments around the x-axis.

Figure 9.5 shows the drive torque ripple. A higher piston count reduces the torque ripple. The double-sided design E without phase shift has significantly higher torque ripple than the double-sided design with phase shift D.

Figure 9.6 shows the flow ripples for the different setups. Similar to the drive torque ripple, a higher piston count reduces the flow ripple and a phase shift is advantageous for double-sided designs.



(a) Bending moments around the x axis.

(b) Bending moments around the y axis.

Figure 9.4: Comparison of bending moments on the cylinder barrel. Machine types see Table 9.1.

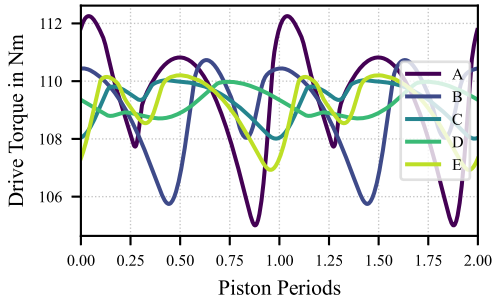
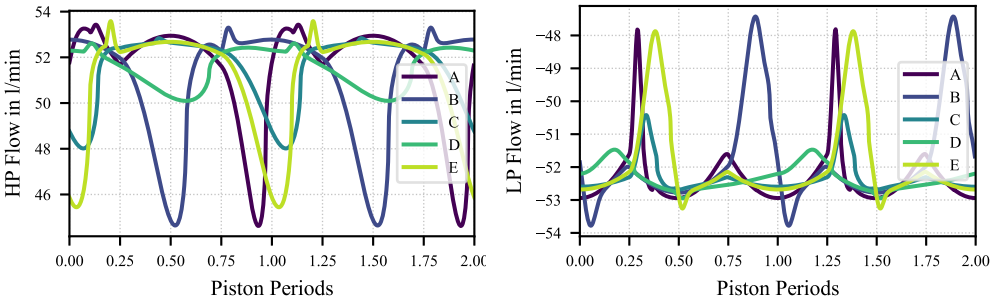


Figure 9.5: Comparison of drive torques. Machine types see Table 9.1.



(a) Delivered flow.

(b) Suction flow.

Figure 9.6: Comparison of flows. Machine types see Table 9.1.

10. SUMMARY AND CONCLUSION

Single-sided and double-sided machine can have the same displacement with different properties. Table 10.1 gives a simplified, qualitative summary of the properties of the different setups.

Table 10.1: Qualitative summary of properties, with "-" being the worst relative rating and "++" being the best relative rating.

Property	Single-sided	Double-sided (connected)	Double-sided (separated w phase shift)
Axial force ripple	o	++	o
Side force ripple	o	o	+
Moment ripple (x-axis)	o	++	-
Moment ripple (y-axis)	-	++	+
Drive torque ripple	o	o	+
Bearing load (average)	-	++	+
Flow ripple	o	o	+
Aspect ratio	baseline	longer and thinner	longer and thinner
Variable displacement	well-proven for slipper-type	option for valve plate rotation (not well proven)	option to circulate flow (proven)
Channel design	+	depends on machine type	depends on machine type

Double-sided machines can reduce noise contributors and bearing loads by cancelling of forces and moments between the machine sides.

Double-sided machines without a phase shift can in principle fully cancel axial forces and bending moments.

Double-sided machines with a phase shift increase the effective piston count, thereby reducing the amplitudes of most noise contributors. Axial forces and bending moments around the y-axis are largely compensated.

11. REFERENCES

[1] Bosch Rexroth AG, 'Axial piston variable pump A4VG Series 32, RE-E 92003/2023-10-30', 2023.

[2] Parker Hannifin Corporation, 'Axial Piston Pumps Series P2/P3 - Catalogue MSG30-2800/UK', 2025.

[3] Sunfab Hydraulics AB. "SCPD 56/26 DIN (3113EN2110 1.3 - 1)," Accessed: 2025-04-25. [Online]. Available: https://www.sunfab.com/media/vrmp155x/3113en2110scpd-56-26-din.pdf.

[4] Bosch Rexroth AG, 'Axial piston variable double pump A20VO, A20VNO, A20VLO series 10, RE 93100/2024-10-31', 2024.

[5] Bucher Hydraulics, 'Fixed Displacement Piston Pumps and Motors: AX Series, 200-FL-991635-EN-03 /10.2023', 2023.

[6] P. Achten et al., 'Design and Testing of an Axial Piston Pump Based on the Floating Cup Principle', in The Eighth Scandinavian International Conference on Fluid Power, SICFP'03, May 7-9 2003, Tampere, Finland, 2003.

[7] Swisher, Philip M., 'Product News PN 630-1, 5/91: New Integrated Motorpump, MPEV3-040'.

[8] V. Mehta, 'Torque Ripple Attenuation for an Axial Piston Swash Plate Type', PhD Thesis, University of Missouri - Columbia, 2006.

[9] L. Danes and A. Vacca, 'A tandem axial-piston unit based strategy for the reduction of noise sources in hydraulic systems', Energies, vol. 13, p. 5377, 2020, ISSN: 19961073. DOI: 10.3390/en13205377.

[10] L. Ericson and T. Heeger, 'Mirrored Dual-Displacement Axial Piston Pump with Pressure-Compensated Axial Force', in ASME 2025 International Desing Engineering Technical Conferences and Computers and Information in Engineering Conferences IDETC-CIE2025, August 17-20, 2025, Anaheim, CA, USA, 2025. DOI: https://doi.org/10.1115/DETC2025-164758.

[11] T. Heeger and L. Ericson, 'A New Degree of Freedom for Variable Axial Piston Pumps with Valve Plate Rotation', in Proceedings of the 17:th Scandinavian International Conference on Fluid Power, SICFP'21, May 31-June 2, 2021, Linköping, Sweden, 2021. DOI: https://doi.org/10.3384/ecp182.

[12] H. Murrenhoff, 'Grundlagen der Fluidtechnik, Teil 1: Hydraulik'. Aachen: Shaker Verlag GmbH, 2016, ISBN: 978-3-8440-4816-2.

[13] T. Heeger, 'Electro-Hydraulic Energy Converters: Development of Axial Piston Machines and Their Integration with Electric Machines', PhD Thesis, Linköping University, 2025, ISBN: 9789181182736. DOI: 10.3384/9789181182743.

[14] N. D. Manring, 'Fluid Power Pumps and Motors: Analysis, Design and Control'. McGraw Hill Book CO, 2013, ISBN: 978-0071812207.

[15] J. Ivantysyn and M. Ivantysynova, 'Hydrostatic pumps and motors: principles, design, performance, modelling, analysis, control, and testing'. New Delhi: Akademia Books International, 2001, ISBN: 9788185522166.

[16] T. Heeger, M. West, and L. Ericson, 'Harmonic Characterisation of Electrically Driven Pumps', in Advancements in Fluid Power Technology: Sustainability, Electrification, and Digitalization. Proceedings of the Global Fluid Power Society PhD Symposium 2024, 17-20 June, 2024, Hudiksvall, Sweden, L. Ericson and P. Krus, Eds., Hudiksvall, Sweden: Springer, 2025, pp. 357–376, ISBN: 978-3-031-84505-5. DOI: 10.1007/978-3-031-84505-5.

[17] A. Johansson, 'Design Principles for Noise Reduction in Hydraulic Piston Pumps', PhD Thesis, Linköping University, 2005, ISBN: 91-85457-15-9.

[18] T. Heeger, S. Kärnell, and L. Ericson, 'Challenges for multi-quadrant hydraulic piston machines', Energy Conversion and Management: X, vol. 22, no. March, p. 100 578, 2024, ISSN: 25901745. DOI: 10.1016/j.ecmx.2024.100578.

[19] P. A. J. Achten, 'Power Density of the Floating Cup Axial Piston Principle', in Proceedings of the ASME 2004 International Mechanical Engineering Congress and Expo (IMECE2004), November 13-19, 2004, Anaheim, California, USA, ASME, 2004, IMECE2004–59 006. DOI: 10.1115/imece2004-59006.

[20] P. Achten and S. Eggenkamp, 'Barrel tipping in axial piston pumps and motors', in Proceedings of 15:th Scandinavian International Conference on Fluid Power, SICFP'17, Linköping, Sweden, 7-9 June, Linköping, Sweden: Linköping University Electronic Press, 2017, pp. 381–391. DOI: 10.3384/ecp17144381.

[21] T. Heeger, M. West, K. Heybroek, and L. Ericson, 'Methodology for Dimensioning of Integrated Electro-Hydraulic Machines', in Proceedings of the 18:th Scandinavian International Conference on Fluid Power, SICFP23, 30 May - 1 June, 2023, Tampere, Finland, Tampere, Finland, 2023.

[22] T. Heeger, M. West, and L. Ericson, 'Comparative Analysis of Direct-Drive and Gearbox-Coupled Electro-Hydraulic Energy Converters', Energy Conversion and Management: X, no. in press, p. 101 463, 2025. DOI: 10.1016/j.ecmx.2025.101463.

[23] T. Grahl, 'Geräuschminderung an Axialkolbenpumpen durch variable Umsteuersysteme', O+P Ölhydraulik und Pneumatik, vol. 33, no. 5, pp. 437–443, 1989.

[24] Société Anonyme André Citroën, "Perfectionnements aux pompes à barillet," Patent BE578560A, 1959.

[25] P. A. J. Achten and Z. Fu, 'Valving Land Phenomena of the Innas Hydraulic Transformer', International Journal of Fluid Power, vol. 1, no. 1, pp. 39–47, 2000. DOI: 10.1080/14399776.2000.10781081.

[26] T. Heeger, S. Wegner, and L. Ericson, 'Analysis of Compensation Ratios and Control Torques of an Axial Piston Pump with Rotated Valve Plates', in 2022 Global Fluid Power Society Ph.D. Symposium, October 12-14, 2022, Naples, Italy, A. Senatore and E. Frosina, Eds., River Publishers, 2024, pp. 1–27, ISBN: 9788770047975. DOI: 10.13052/rp-9788770047975.001.

[27] Sunfab. "The sunfab dual flow pump," Accessed: 2025-08-14. [Online]. Available: https://www.sunfab.com/products/dual-flow-pumps/.

[28] P. Hannifin. "Parker F2 Twin-Flow hydraulic truck pump series," Accessed: 2025-08-14. [Online]. Available: https://ph.parker.com/se/en/product-list/axial-piston-fixed-pumps-series-f2.

[29] Linköping University (Division of Fluid and Mechatronic Systems). "Hopsan," Accessed: 2025-08-12. [Online]. Available: https://liu.se/en/research/hopsan.

Biographies



Thomas Heeger. He received his PhD in Fluid and Mechatronic Systems at Linköping University, Sweden, in 2025. His research interests include hydraulic pumps and their electrification.



Liselott Ericson. She received her Ph.D. in Hydraulics at Linköping University (LiU), Sweden, in 2012. She currently works as a professor at Fluid and Mechatronic Systems at LiU. Her research interests include pump and motor design, electro-hydraulic systems, and modeling and simulation.

Optimization of the Cylinder Block-Valve Plate Interface Considering Coupled Interaction from Piston Dynamics and Shaft Deformation

Ishan Suvarna[1], Lizhi Shang[1,*], Raul Rodriguez[2], Jeemon Kurian[2], Enrique Busquets[2]

[1]*Maha Fluid Power Research Center, Purdue University, West Lafayette, Indiana 47906, USA*

[2]*Bosch Rexroth, 100 Southchase Blvd, Fountain Inn, SC 29644, USA*

[*]*Corresponding author: shangl@purdue.edu*

Abstract

Cylinder Block/Valve Plate interface ideally operates with a thin fluid film separating the running surface, preventing contact and wear. At the same time, the film seals the interface, preventing leakage of high-pressure fluid in the displacement chamber. This thin film is maintained by balancing the hydrodynamic/hydrostatic pressure distribution with the external loads applied on the cylinder block. Extending the operating speeds and pressure of the pump can disrupt this balance, preventing the interface from sealing effectively, or generating wear that causes early failure of the machine. The factors limiting the operational range are therefore coupled physical phenomena, where the performance of the piston/cylinder block interface or the shaft-cylinder block spline connection may lead to the failure of the cylinder block/valve plate interface. Current state-of-the-art simulation tools focus on individual lubricating interfaces, which is not sufficient to accurately model physical behavior for finding the operational limits of the cylinder block/valve plate interface at extended operating conditions.

This paper proposes a coupled simulation method for the cylinder block/valve plate interface that includes consideration of the physical behavior at other lubricating interfaces. The paper presents simulation results when the cylinder block/valve plate interfaces is coupled with the piston/cylinder block interface and the shaft spline interface and demonstrates the significant difference in simulation results between the proposed coupled simulation method and the standalone simulations. Furthermore, the paper includes experimental test observations that validate the findings from the coupled model simulation. The coupled simulation is then used to find optimized designs for extended operating conditions, at higher pressures and higher speeds. The paper reports experimental results with the optimized design prototypes, which demonstrate performance improvement.

Keywords. Axial Piston Pump, Tribology, Lubricating Interface, Design Optimization

1. INTRODUCTION

Axial Piston Machines (APM) are among the most widely used hydrostatic pumps and motors. They are known for their ability to provide varying displacements, as well as high efficiency even at high operating speeds and pressures. Current trends in applications of hydraulic pumps and motors show continued demand for an increase in performance capabilities as well as longer operational lifespans. Therefore, it is important to understand

the behavior of APM when pushed to the limits of the operational window, to find the physical factors restricting performance, and to develop design solutions for further improvement of the machine's operation.

The cylinder block/valve plate interface is critical to the performance of an axial piston machine. The valve plate is responsible for regulating flow through the machine, and the design allows control over several key behaviors such as the flow ripple and noise generated during operation [1]. The lubricating film at this interface also regulates the leakage from the pressure ports, as well as providing hydrodynamic balancing against the loads acting on the cylinder block. Studies like Manring [2] have shown that improving the design of this interface could therefore address several issues affecting machine behavior at extremes of the operational range and improve overall performance. However, the far-reaching impact of valve plate design also increases the complexity of the design process.

Simulation models enhance the process of designing axial piston machines, reducing the need for expensive and time-consuming experimental runs, which can now be restricted to validating the final designs. Thanks to the simulation models, various design features can be parameterized and used in conjunction with optimization algorithms to generate several iterations of highly tuned efficient designs.

The lubricating interface performance of axial piston machines depends on interactive physical phenomena such as hydrodynamics, thermodynamics, heat transfer, elastic deformation due to pressure and thermal load, the micro motion of the moving parts, and fluid and solid properties. Various simulation methods have been developed targeting different aspects of this complex system.

Given the complex nature of flow within an axial piston machine, a full-scale CFD model would be particularly useful in understanding the flow behavior. Frosina et al. [3] demonstrated the use of commercial CFD models to predict flow behavior in axial piston pumps with complex design features to reduce flow ripple. Boland and Shang [4] used an axial piston pump model implemented using the CFD tool Simerics MP+ to study cavitation behavior and its scalability. CFD models can provide very detailed fluid behavior and can be very useful for studying the fluid dynamics within the pump. However, using a full-scale CFD model is computationally expensive. Ignoring or simplifying the fluid flow behavior at the lubricating interfaces can speed up the simulation but can ignore crucial flow behavior at these interfaces which can affect the accuracy of the results. Additionally, CFD approaches are restricted to fluid flow behavior, and often do not provide any further understanding on the dynamics of the solid components themselves.

Lumped parameter modelling approach is also widely used for positive displacement machines. This approach greatly reduces the computational cost of predicting flow behavior by considering the fluid volumes inside the machine as several lumped volumes, ignoring any spatial variation of fluid properties and any flow within the volume. Roccatello et al. [5] developed such a model for simulating axial piston machines and demonstrated its use in predicting pump characteristics like flow ripple and torque. Corvaglia et al [6] compared Lumped Parameter simulations to results provided by a full–scale 3D CFD model and found a good agreement between them. Lumped models are simple to develop and can provide quick results for a large variety of pump designs, compared to the speed of CFD models. They can also be used in conjunction with dynamic models to understand the motion of solid components within the machine. But the lumped volume assumption means there is no way to model detailed fluid behavior, particularly at the lubricating film interfaces.

Hasko et al [7] showed that film interfaces are a major contributor to power loss in an axial

piston machine. Therefore, a representative numerical model should be able to include the effects of lubricating films. A fluid structure interaction model for the cylinder block -valve plate interface was introduced by Zecchi and Ivantysynova [8], and for the piston-cylinder block interface by Pelosi and Ivantysynova [9]. These models involved coupling the lumped parameter model with a fluid simulation for the lubricating interfaces and demonstrated the improvement in modelling pump behavior without excessive computational cost. A tribological cylinder block/valve plate interface model was developed at IFAS and used to optimize the design geometry of the interface [10] to reduce wear due to fluid and solid friction. The benefits of detailed film modelling in developing new designs were demonstrated by Geffroy et al. [11] and the model was further validated experimentally in [12] A numerical lubricating film model was also used by Li et al. [13] to study wear characteristics at the cylinder block/valve plate interface under different operating conditions. More recently, Han et al. [14] used a cylinder block/valve plate lubricating film model to study the effects of hydrostatic pockets in reducing solid body contact in a floating piston pump.

It is clear that simulating the cylinder block/valve plate lubricating film interface would be useful in understanding how the interface performs in extreme operating conditions. However, isolated simulation models do not accurately capture the behavior of APMs. For example, the behavior at the cylinder block/valve plate interface would be affected by the forces created by the lubricating film along the piston interface. While analytical assumptions can be made for specific test cases, a general use simulation model should be able to model this coupled interaction accurately. Ransegnola developed a coupled model for an axial piston machine and used it to study the piston and slipper spin [15]. This model was able to extend methods previously used for singular lubricating interfaces, and couple them with a lumped parameter model for fluid volumes and a solid dynamics solver that resulted in a highly accurate simulation model of the entire axial piston machine assembly. Mukherjee et al. demonstrated the benefits of a coupled modelling approach by showing an improvement in estimation of power losses at lubricating interfaces with the new approach [16]. Gärtner et al. [17] developed and compared coupled models for the piston/cylinder interface, to better include the effects of the slipper/swashplate film and the slipper-piston ball joint dynamics. Coupling of multiple interfaces creates a more accurate model for the complex interactions in the APM and ensures better accuracy in studying the behavior of the cylinder block/valve plate interface.

The goal of this work is to extend the operational limits of axial piston machines through improved design of the cylinder block – distributor plate interface. The paper describes the physical behavior of the interface in extreme conditions such as high speeds and high pressures. A novel numerical modelling approach is proposed to simulate this behavior, addressing the limitations of previous numerical modelling methods for axial piston machines. Experimental validation of the improved model is done with two test cases, where the model is used to better understand the limiting factors of the interface design, and to test design changes addressing the limitations.

2. THE CYLINDER BLOCK/VALVE PLATE INTERFACE

The interface between the cylinder block and the valve plate serves three primary functions: to reduce friction between the stationary control plate and the rotating section of the APM, to for an effective seal preventing fluid from the high-pressure ports leaking out, and to provide hydrodynamic balancing for the forces and moments acting on the cylinder block

[1]. Since the film thickness distribution is primarily affected by the pressure load. the running surfaces of the cylinder block and the valve plate need to be carefully designed to ensure the lubricant film layer can sufficiently balance the load on the cylinder block. An ill-designed surface can disrupt the functioning of the film and affect its other functions.

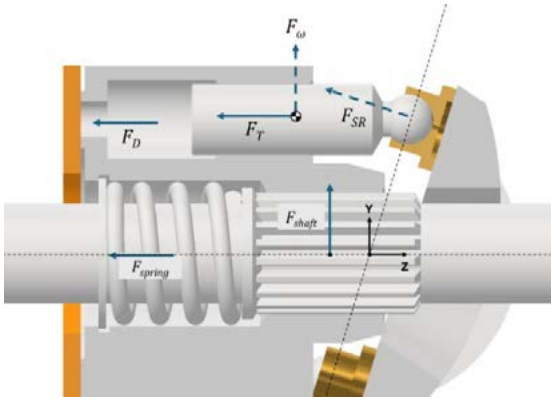


Figure 2.1. Overview of Cylinder Block Forces

Simulation models for use in understanding the behavior of the valve plate interface need to accurately include the effects of the forces and moments acting on the cylinder block. There are two main types of forces acting on the cylinder block [8,18]: forces that are directly applied on the cylinder block and reaction forces from other components transmitted through lubricating films. Figure 2.1 shows the locations where these forces are applied on the cylinder block. The direct forces are shown with dark arrows and the forces applied through film are shown in dashed arrows

Studies like Geffroy et al. [10] have shown that a numerical model of the forces on the cylinder block can be used to calculate the gap height distribution at the valve plate interface. While state of the art models can accurately include most forces, there are some factors which represent a challenge. The friction force acting on the cylinder block is a result of the lubricating interface between the block bore and moving piston surface. The friction includes both viscous losses generated by the fluid, and solid-solid friction when the surfaces are in contact with each other. Since these forces are primarily a result of lubricating film behavior, the best way to accurately calculate these forces is by simulating the films at each of the interfaces between the pistons and the cylinder block. Additionally, to properly simulate complex interactions, these film models would need to be coupled with the numerical model of the valve plate interface. This adds a high level of complexity for the model.

Another important factor governing the overall moment acting on the cylinder block is the shaft reaction force. The forces acting on the cylinder block perpendicular to the shaft axis are balanced by a reaction force provided by the shaft itself. This reaction force acts on the cylinder block on the region where it is in contact with the shaft, known as the crowning region. The location of the crowning region is important, as it determines the moment generated by the shaft reaction force on the cylinder block. The crowning region is defined by a characteristic curve added to the spline of the shaft, the peak of this curve considered as the center where the shaft would be in contact with the cylinder block, and the point where the transmission of forces and torque occurs.

The surfaces of the cylinder block and shaft in contact here can deform, as a result of the large forces acting on the cylinder block. This deformation increases if the axial piston machine is functioning at the limits of its operational range. The deformation can affect the nature of contact between the two surfaces. Figure 2.2 illustrates the effect that such deformation can have on the location of the crowning point.



Figure 2.2. Shaft Contact Behavior due to deformation

To understand better the exact behavior of contact between the shaft and the cylinder block considering deformation, a finite-element method simulation was conducted. A representative model of the Axial Piston Machine was constructed, and the simulation was conducted using Ansys Mechanical. Studying the contact region under the loads applied on the cylinder block, it is observed that contact is indeed not restricted to a single point but is instead a long region along the edge of the area of contact on the shaft spline, as shown in Figure 2.3.

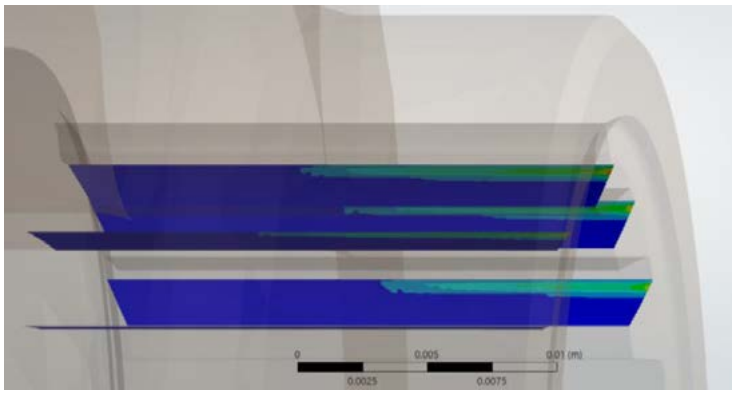


Figure 2.3. Shaft contact region from FEM Simulation

To calculate the contribution of shaft reaction moment on the block, and to accurately model the tilting of the cylinder block, it's important to define a single contact point. This point can be the center of the entire contact region, the ideal point at which the net contact force can be considered to be acting. To calculate this point, the net moment generated by the contact region on the cylinder block $\vec{M}$ is extracted from the Ansys model. Additionally, the contact pressure can be integrated with the contact area to calculate the net contact force $\vec{F_c}$. This contact force has two components, one contributing to the moment on the block, and the other countering the forces in the axial direction. To find the components of contact force normal to the moment $\vec{F}$, a simple vector calculation is performed, as described in equation 1.

$$\vec{F} = \vec{F_c} - \frac{\vec{F_c}.\vec{M}}{|\vec{M}|^2}\vec{M} \quad (1)$$

Using equations 2 and 3, the length of the normal moment arm $\vec{r}$ required by the contact force to generate the moment is calculated, which gives the location of the point at which the net contact force is acting. This is the ideal center of the contact region.

$$\vec{r} \times \vec{F} = \vec{M} \quad (2)$$

$$\vec{r}\,.\hat{F} = 0 \quad (3)$$

The inclusion of the effects of shaft deformation on the cylinder block moment, as well as the coupled modelling of piston films, represents the novelty of the modelling approach used in this study. As shown in the case studies in Section 4, the consideration of shaft deformation proves critical in the APM behavior at operational limits and ensure that design solutions proposed are indeed effective in improving performance.

3. SIMULATION MODEL

The primary goal of the simulation model for this study is to accurately capture the behavior of the cylinder block/valve plate interface. As described in section 2, the major factor affecting the behavior of this interface is the dynamics of the cylinder block, and the best way to accurately model this is using a coupled simulation model involving both the cylinder block/valve plate interface, and the piston/cylinder block interface. For this study, the lubricating interfaces are modeled using a modified version of the Reynolds equation [13], to include the effects of mixed lubrication and surface deformation, as shown in equation 4.

$$\nabla.\left(\phi_P \frac{Kh^3}{12\mu}\nabla\rho\right) = \nabla.\left(\rho v_m\left(\phi_R R_q + \phi_C h\right)\right) + \nabla.\left(\rho\frac{\phi_S}{2}R_q(v_t - v_b)\right) + \frac{\partial\rho\left(\phi_R R_q + \phi_C h\right)}{\partial t} \quad (4)$$

These modifications are necessary to account for the forces developed in the film due to varying flow regimes, and the effects of deformation on the gap height and the resulting balancing force produced by the film. To model the fluid chambers and hydraulic components like orifices, a lumped parameter fluid model is implemented to reduce computational cost. The kinematics of the various solid components in the APM are determined using a body dynamics solver.

To simulate the interaction between physical domains, the models are coupled with each other. This coupling is implemented through continuous information transfer between the models, simulating physical behavior in an APM. The lumped parameter and the solid body dynamics need to be solved with a much smaller timestep compared to the interface models, and forcing the interface model to be solved on this small timestep increases computational cost. So, the simulation is divided into independent timelines: one for each of the lubricating interfaces and one main timeline for the lumped parameter and body dynamics model. Information about the interface is interpolated between timesteps to be used by the lumped parameter and the body dynamics solver, ensuring continuous transfer of data between the models and improving the solution time.

The coupled model used in this study was built using Multics, a Multiphysics simulation tool developed by the Maha Fluid Power Research Center. Multics was first introduced by Ransegnola et al. [19] and used to perform a fully coupled simulation for external gear

machines. It included multiple models for various physical features such as tooth space volumes, floating bodies and lubricating interfaces. It also introduced the asymmetric timeline method used to improve simulation times with coupled models. The tool was further used to develop a coupled Swashplate APM model to study piston – slipper spin [15]. The flexible structure of Multics also allows for the addition of new models, which allows new physics to be introduced for various positive displacement machines. Mukherjee et al. further developed the Swashplate APM model to include the effects of thermal behavior across various physical domains in [20].

The simulation model proposed in this paper uses Multics to develop a fully coupled model consisting of the Cylinder Block/Valve plate and Piston/Cylinder Block interfaces, with mixed lubrication and deformation considerations at the interfaces, along with floating body models for moving components and lumped parameter modelling for the chambers. Additional considerations are also made to include the effects of shaft deformation in the dynamics of the cylinder block. Figure 3.1 shows an overview of the proposed model.

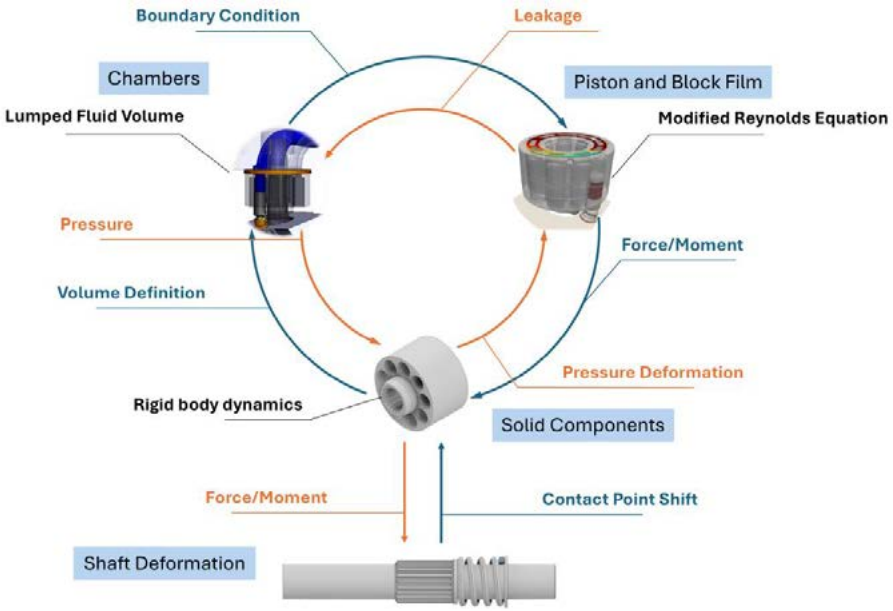


Figure 3.1. Overview of Swashplate APM Model

4. VALIDATION AND CASE STUDIES

The proposed coupled model is used in simulating the behavior of the cylinder block/valve plate interface of an axial piston machine near operational limits. Two test cases considering two different kinds of limiting behavior at the extremes of the operational range were considered:

- The first case studies the factors that define the maximum pressure capability of the pump, beyond which wear begins to appear on the valve plate running surface
- The second case studies the maximum power of an axial piston motor, beyond which the cylinder block begins to separate from the valve plate

For the two cases, simulation results for the baseline designs are validated with observations from experimental tests. Further, the simulation results are analyzed to determine the underlying factors, and to test design changes aimed at improving performance at higher pressure and operating speeds. Finally, the design solutions are validated experimentally. Additionally, the benefits of the novel considerations in the simulation model are highlighted for each of the test cases.

4.1. Study of Valve plate Wear at High Pressure

The cylinder block/valve plate interface is designed to hydrostatically balance the loads acting on the cylinder block. The lubricating film at this interface has a gap height in the order of microns, making it sensitive to the deformation of the running surfaces. When the machine is operated at high pressures, the surfaces can be deformed from the nominal plane, where 'bump' (local reduction of the film thickness) increases the change for contact and wear, and 'dent' (local increase of the film thickness) increases the leakage. This study will focus on the wear, which is a critical factor that determines the maximum operation pressure of an axial piston machine. Contact pressure is used as a parameter to measure and compare the degree of wear between the baseline design and test cases.

4.1.1. Baseline Design Simulations

The coupled numerical model is used to analyze the baseline design. The model simulates both the piston/cylinder interface and the cylinder block/valve plate interface simultaneously, to ensure the frictional forces and moments generated by the relative motion between the piston and the cylinder block are included when calculating the moments acting on the cylinder block. As shown in Figure 4.1, the simulation results from the proposed model indicate two regions of contact on the outer edge of the interface with a maximum magnitude of 1700 bar. The two contact regions, combined with the high sliding velocity between the running surfaces of the interface results in wear on the valve plate. To validate the simulation results, the location of the contact is compared to the observed location of wear in the experimental tests. As seen in Figure 4.2[A], the valve plate has deep wear marks along the outer edge of inlet and outlet ports, which are the boundaries of the interface with the rotating cylinder block.

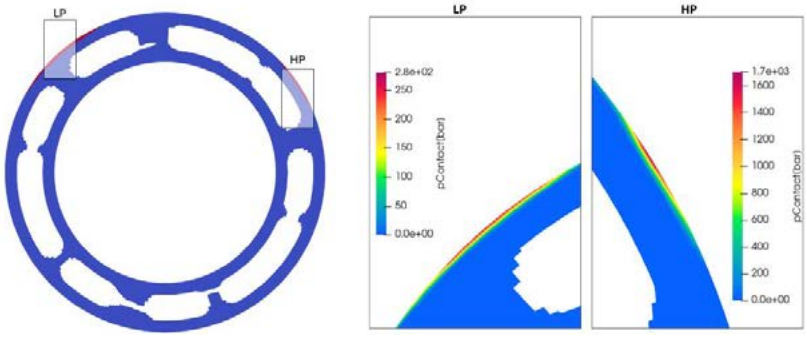


Figure 4.1. Contact Regions on Cylinder Block–Valve plate Interface.
HP and LP indicate areas of film connected to High-Pressure and Low-Pressure ports



Figure 4.2. Wear observed on valve plate after experimental tests. [A]: Baseline design, [B]: Improved Port Block, [C]: Increased Interface Area

Further investigation of the simulation results indicates that the contact pressure is a result of the combination of high tilting moments on the cylinder block under maximum pressure, the friction between the piston and bore, and the deformation of the solid bodies. Although the effect of the surface deformation of this interface was discussed in Chacon and Ivantysynova [21], the coupled simulation results indicated that the consideration of deformation alone is not sufficient to predict the contact pressure as well as the resulting wear. Figure 4.3 shows the results for the same design without considering the coupled effect from the piston–cylinder block interface. Therefore, the inclusion of the coupled model, despite the added complexity, ensures better representation of the pump's behavior at such extreme operation conditions. With the validation complete, further design solutions aimed at addressing the causes of wear are tested.

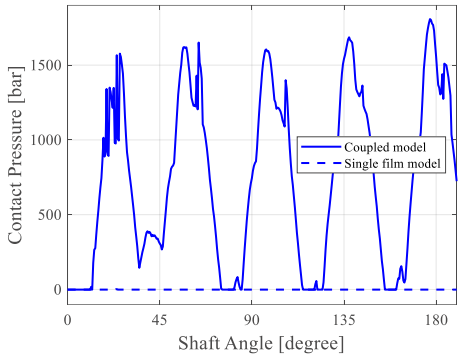


Figure 4.3. Comparison of Simulated Contact Pressure with Single Film Model

4.1.2. Design Improvements

The coupled simulation results indicated the causes of wear on the valve plate. Two design solutions were developed to reduce contact between the running surfaces.

A. Improved Port Block: The first design change was aimed at reducing the deformation of the valve plate surface by redesigning the port block. As shown in Figure 4.4, material added to the chamber walls increases support, which would reduce the deformation

caused by the high-pressure fluid in the chamber. This results in lower deformation of the valve plate, reducing contact with the moving cylinder block.

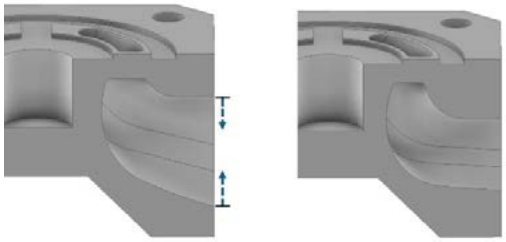


Figure 4.4. Illustration of Port Block design changes to reduce deformation

B. Increased Valve plate Interface Area: The second design change tested is increasing the sealing land area. The baseline design of the cylinder block/valve plate interface is insufficient to balance the large moments acting on the cylinder block at high operating pressures. Therefore, the interface area is increased by increasing the outer diameter and decreasing the inner diameter, as shown in Figure 4.5. This increased surface area can provide more moment balancing the cylinder block, therefore reducing the tilting and the resulting contact with the valve plate.

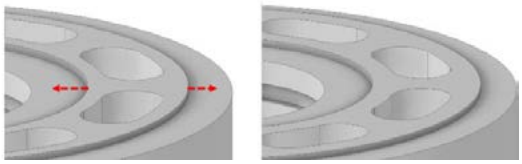


Figure 4.5. Illustration of valve plate interface area changes on Cylinder Block

4.1.3. Results

For the new port block design, simulation results indicate a reduction in maximum contact pressure, as seen in Figure 4.6.[A]. This result is validated by experimental tests for this design. As seen in Figure 4.2[B], the wear markings on the valve plate are still present, but the depth of wear is lower, showing a marginal improvement provided by the new port block.

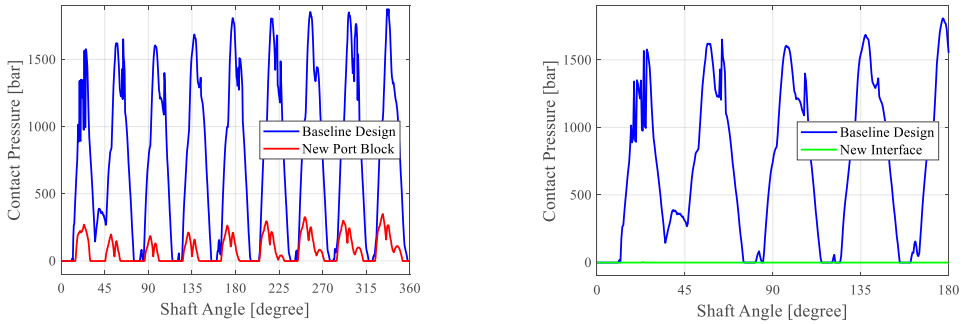


Figure 4.6. Comparison of Contact Pressure with new designs

Figure 4.6[B] shows the simulation results for the new valve plate interface design, with more film area. There is no contact observed, which shows that the cylinder block is sufficiently balanced by the new larger interface. Observing the results of the experimental tests with the new interface design in Figure 4.2[C], a significant reduction in wear is observed, matching the results of the coupled simulation model.

4.2. Study of Cylinder Block Separation at High Speeds

The hydrodynamic force in the lubricating film, the inertia force from the piston and the slipper, and the friction from the piston intensify with the increased piston machine speed. The combination of these effects tends to tilt the block, ultimately resulting in its separation from the valve plate. This failure mode can be observed as the sudden increase of drain flow rate, therefore the sudden increase of case pressure, as well as the sudden loss of high pressure when operating beyond the rated speed.

4.2.1. Baseline Design Simulations

Shaft with crowned spline is often used in piston machines to allow the block to tilt around a designed pivot point. As explained in Section 2, this pivot point may deviate from the shaft crowning point due to deformation. The proposed model considered the shift of the pivot location based on the FEM analysis. Coupled with the friction from the piston cylinder interface, this proposed model faithfully replicates the observation of the behavior of the cylinder block/valve plate interface in this case study. Those are:

- In motoring mode, the block separates from the valve plate when operating beyond the rated speed
- In pumping mode, no separation is observed, and the interface operates normally

The simulated film thickness in the cylinder block/valve plate interface is shown in Figure 4.7, where the block separation can be observed as the diverged film thickness. This figure also demonstrates that this behavior can only be replicated with the consideration of shaft deformation.

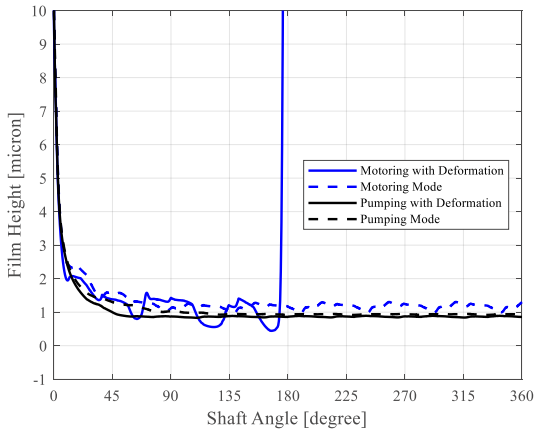


Figure 4.7. Effects of shaft spline deformation on block separation

4.2.2. Design Improvements

The coupled simulation results indicated the causes of wear on the valve plate. Two design solutions were developed to reduce contact between the running surfaces.

A. Hydrodynamic Balance Ring: Hydrodynamic balance ring is introduced to provide additional capability to counter the block tilting moments. Figure 4.8 illustrates the change of the geometry of the running surface of the cylinder block.

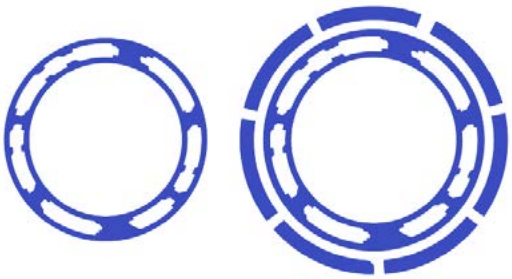


Figure 4.8. Modification of the running surface geometry of the cylinder block

B. Extended piston guide length: The length of the piston guide is defined by the distance between the beginning and end of the cylinder bore running surface. This typically is determined by the geometry of the bore undercut. It is widely understood that the increased length of the piston guide improves the sealing function of the piston cylinder interface and increases the friction due to larger area. However, the length contributes to another significant effect that determines the piston/cylinder performance – hydrodynamics. When operating in motoring mode, the piston moves outwards during the high-pressure stroke, where most of the friction happens. The high pressure tilts the piston in the bore, forming a region of wedge gap around the ends of the piston film.

In case of short bore, the film ends at the edge of the bore. The piston moves relative to the wedge gap, where the fluid is dragged out of the contact region, forcing the load to be carried by boundary lubrication. This is illustrated in Figure 4.9[A].

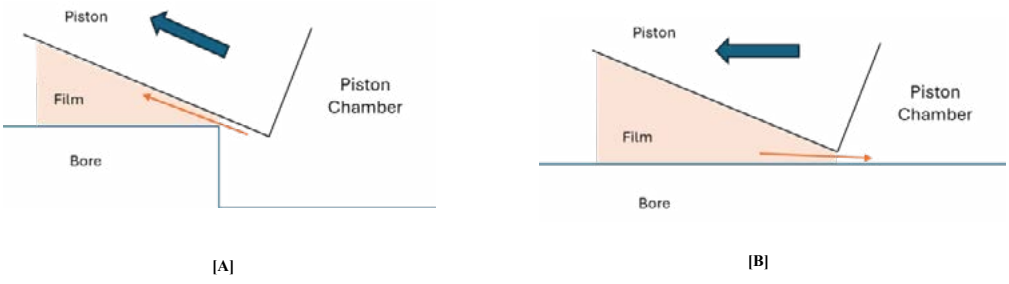


Figure 4.9. Illustration of lubrication behavior with different bore designs

In case of long bore, the film ends at the edge of the piston. The bore moves relative to the wedge gap, where the fluid is dragged into the contact, which helps carry the load hydrodynamically. This is illustrated in Figure 4.9[B]. Therefore, to improve cylinder

block/valve plate interface performance through reduction of friction from the piston in motoring more, an extended piston guide is designed as illustrated in Figure 4.10

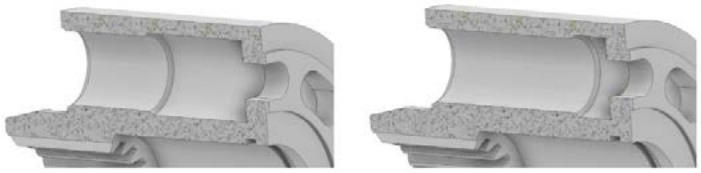


Figure 4.10. Illustration of bore design change

4.2.3. Results

Simulation results for the new interface with the hydrodynamic balance ring design are shown in Figure 4.11.[A]. The film gap heights now stay stable for the entire rotation of the cylinder block, as opposed to the baseline which is shown to separate. This indicates that the added hydrodynamic area reduces the tilting moment on the block and prevents separation.

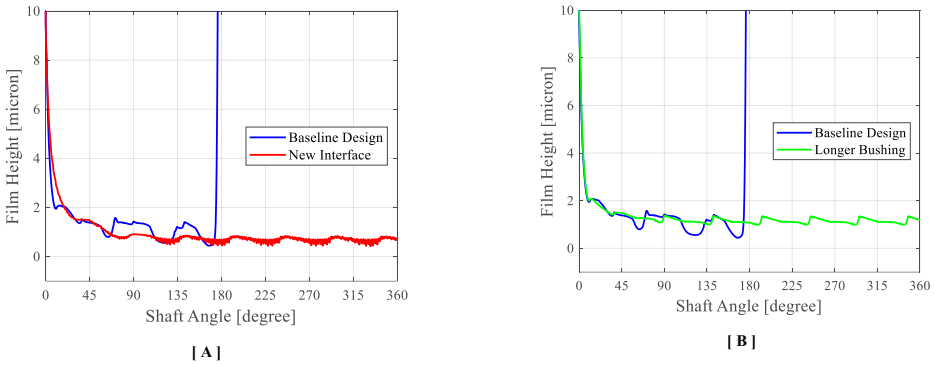


Figure 4.11. Comparison of Gap Height with new designs

Figure 4.11.[B] shows results for the new bore design with longer bushing length. The film gap height at the cylinder block/valveplate interface remains stable for the full rotation, showing that the tilting of the block has been reduced, and separation is prevented. This shows that improvement in performance of the cylinder block/valveplate interface can be achieved by changing the bore design at the piston/cylinder block interface. The coupled model used for this study enables the influence of design changes at other interfaces to be included when simulating the valve plate interface behavior, accurately showing the resulting improvements. Experimental tests show no separation with the new design, which validates the results of the coupled model.

5. CONCLUSIONS

The cylinder block/valve plate interface plays a major role in determining the operational limits of axial piston machines. The behavior of the lubricating film at this interface is dependent on the dynamics of the cylinder block, as well as the behavior of other lubricating

interfaces. Accurate numerical models need to include the effects from several interacting physical domains, in order to accurately simulate the behavior. While simplified models permit faster study of a specific parameter, several characteristics can only be understood with a full-scale model. A coupled model approach is proposed as a method to simulate such complex behavior. The proposed coupled model for the axial piston machine uses a fluid-structure interaction model to simulate the piston and valve plate lubricating interfaces, and the model is fully coupled with a body dynamics solver and lumped parameter models to simulate the solid components and the chamber volumes respectively. The effects of deformation of the cylinder block surface at the crowning region are included, to ensure that the effect of the shaft reaction moment is included.

The improvement of the simulation model is demonstrated in representing physical behavior of the axial piston machine at conditions outside the operational window. Two test cases are considered: valve plate wear at high pressures, and separation of the cylinder block at high speeds. For the two cases, the baseline designs were tested with the simulation model. The predicted behavior matched observations from experimental tests conducted for the same machine designs. Additionally, the results from the simulation model were analyzed to better understand the factors affecting performance, and design solutions were proposed addressing the causes. The improvements were first tested with the simulation model, then validated with experimental results. Table 5.1 summarizes the design changes and the resulting effects. It also highlights the improvement in simulation output with the proposed model's considerations, as opposed to the state of art models.

Designs	Proposed Model	State of Art Models	Experimental Observations
1. Valve plate Wear Study	**Operational Limit: Contact at High Pressure**		
Baseline	Contact	No Contact	Contact
New Port Block	Reduced Contact	-	Reduced Contact
Redesigned Interface	No Contact	-	No Contact
2. Cylinder Block Separation Study	**Operational Limit: Separation at High Speed**		
Baseline	Separation	No Separation	Separation
Redesigned Interface	No Separation	-	No Separation
Longer Piston Bore	No Separation	-	No Separation

Table 5.1. Summary of Test Cases and Design changes

The novel considerations of the simulation model, the coupled films and the shaft deformation, were found to be important in accurately capturing the behavior of the machine at such operating points, and ensuring the improved designs were functioning accurately. The consideration of the coupled film models, as opposed to an individual simulation of the cylinder block/valve plate interface, was found to be necessary for modelling the dynamics

of the cylinder block. Additionally, the shift in the crowning point due to deformation at the shaft ensures the simulation model captures the separation of the cylinder block from the valve plate at high speeds.

The simulation results show several cases where the cylinder block/valve plate interface is important in ensuring axial piston machines are operational at extreme operating conditions. Design changes at this interface solve issues like wear and leakage, while changes at other regions of the machine also help in improving the functioning of this interface. Therefore, the study effectively demonstrates that the operational limits of the cylinder block can be extended by optimizing the functioning of the cylinder block interface.

6. REFERENCES

[1] J. Ivantysyn, & M. Ivantysynova, Hydrostatic Pumps and Motors, Principles, Designs, Performance, Modelling, Analysis, Control and Testing. 2001.

[2] N. Manring, "Valve-plate design for an axial piston pump operating at low displacements." Journal of Mechanical Design 125, 2003. https://doi.org/10.1115/1.1541632

[3] E. Frosina, G. Marinaro, A. Senatore and M. Pavanetto, "Effects of PCFV and Pre-Compression Groove on the Flow Ripple Reduction in Axial Piston Pumps," 2018 Global Fluid Power Society PhD Symposium (GFPS), pp. 1-7, Samara, Russia, 2018. https://doi.org/10.1109/GFPS.2018.8472389

[4] H. Boland, & L. Shang, "An Analysis of Cavitation Phenomena Scalability in Axial Piston Machine." Proceedings of the ASME/BATH 2023 Symposium on Fluid Power and Motion Control. ASME/BATH 2023 Symposium on Fluid Power and Motion Control. Sarasota, Florida, USA. October 16–18, 2023. https://doi.org/10.1115/FPMC2023-113585

[5] A. Roccatello, S. Mancò, and N. Nervegna, "Modelling a Variable Displacement Axial Piston Pump in a Multibody Simulation Environment." ASME Journal of Dynamic Systems, Measurement, and Control, July 2007. https://doi.org/10.1115/1.2745851

[6] A. Corvaglia, M. Rundo, "Comparison of 0D and 3D Hydraulic Models for Axial Piston Pumps", Energy Procedia, pp. 114-121, Volume 148, 2018. https://doi.org/10.1016/j.egypro.2018.08.038

[7] D. Hasko, L. Shang, E. Noppe, E. Lefrançois, "Virtual Assessment and Experimental Validation of Power Loss Contributions in Swash Plate Type Axial Piston Pumps". Energies, 2019. https://doi.org/10.3390/en12163096

[8] M. Zecchi and M. Ivantysynova "Cylinder block/valve plate interface—A novel approach to predict thermal surface loads." Eighth International Fluid Power Conference, Dresden, Germany, Mar, pp. 26-28. 2012.

[9] M. Pelosi, and M. Ivantysynova. "A novel fluid-structure interaction model for lubricating gaps of piston machines." Proceedings of the 5th Fluid Structure Interaction Conference, pp. 13-24. 2009.

[10] S. Geffroy, N. Bauer, T. Mielke, S. Wegner, S. Gels, H. Murrenhoff, and K. Schmitz. "Optimization of the tribological contact of valve plate and cylinder block within axial piston machines." Eighth International Fluid Power Conference, Dresden, Germany, 2020. https://doi.org/10.25368/2020.46

[11] S. Geffroy, S. Wegner, S. Gels, H. Murrenhoff, and K. Schmitz. "New Design Concepts for the Tribological Contact of Cylinder Block and Valve Plate." Proceedings of the BATH/ASME 2020 Symposium on Fluid Power and

Motion Control. BATH/ASME 2020 Symposium on Fluid Power and Motion Control. September 9–11, 2020. https://doi.org/10.1115/FPMC2020-2725

[12] S. Geffroy, S. Wegner, S. Gels, and K. Schmitz. "Experimental investigation of new design concepts for the tribological contact between the valve plate and the cylinder block in axial piston machines." Scandinavian International Conference on Fluid Power, pp. 104-116. 2021. https://doi.org/10.3384/ecp182p104

[13] X. Li, J. Zhang, F. Lyu, and B. Xu. "The transition of wear and leakage characteristics of the cylinder block/valve plate interface in a wide range of operating conditions." Proceedings of the BATH/ASME 2023 Symposium on Fluid Power and Motion Control. BATH/ASME 2023 Symposium on Fluid Power and Motion Control., 2023. https://doi.org/10.1115/FPMC2023-111370

[14] H. Han, T. Heeger, L. Shang, & L. Ericson, "Simulation of Hydrostatic Pockets Between the Cylinder Block and Valve Plate of a Piston-type Pump". The 19th Scandinavian International Conference on Fluid Power, 2025. https://doi.org/10.13052/rp-9788743808251A12

[15] T. Ransegnola, L. Shang, and A. Vacca. "A study of piston and slipper spin in swashplate type axial piston machines." Tribology International 167, 2022. https://doi.org/10.1016/j.triboint.2021.107420

[16] S. Mukherjee, S. Sarode, C. Mujumdar, L. Shang and A. Vacca, “Effect of Dynamic Coupling on the Performance of Piston Pump Lubricating Interfaces.”, MM Science Journal, 2022. https://doi.org/10.17973/MMSJ.2022_10_2022075

[17] M. Gärtner, A. Holzer, F. Fischer, K. Schmitz, “EHL Simulation Model for an Abstracted Piston-Bushing Test Rig.”, Tribologie und Schmierungstechnik. 66., 2019. https://doi.org/ 10.30419/TuS-2019-0028

[18] S. Wegner, S. Gels, D. S. Jang, and H. Murrenhoff. "Experimental investigation of the cylinder block movement in an axial piston machine." Proceedings of the BATH/ASME 2015 Symposium on Fluid Power and Motion Control. BATH/ASME 2015 Symposium on Fluid Power and Motion Control., 2015. https://doi.org/10.1115/FPMC2015-9529

[19] T. Ransegnola, and A. Vacca. "Virtual design and analysis of the balancing element of an external gear machine considering cavitation and mixed lubrication effects." International Journal of Fluid Power 24, 2023. https://doi.org/10.13052/ijfp1439-9776.2414

[20] S. Mukherjee, L. Shang, and A. Vacca. "Numerical analysis and experimental validation of the coupled thermal effects in swashplate type axial piston machines." Mechanical Systems and Signal Processing 220, 2024. https://doi.org/10.1016/j.ymssp.2024.111673

[21] R. Chacon, and M. Ivantysynova. "An investigation of the impact of the elastic deformation of the end case/housing on axial piston machines cylinder block/valve plate lubricating interface.", 10th International Fluid Power Conference, Vol. 1, pp. 283-294, March 8 - 10, 2016.

Biographies



Ishan Suvarna received his Bachelor's degree in Mechanical Engineering from Birla Institute of Technology and Science, Pilani – Goa Campus in 2024, and is currently pursuing his Master's degree in Mechanical Engineering at Purdue University, while working as a graduate researcher at the Maha Fluid Power Research Center. His research interests include the development and application of numerical models to analyze performance of axial piston machines, particularly under extreme operating conditions.



Dr. Lizhi Shang is Assistant Professor of Mechanical Engineering and Agricultural & Biological Engineering at Purdue University. His research interests focus on designing and modeling hydrostatic pumps and motors, hydrodynamic pumps and turbines, fluid power systems, and advanced computational and experimental tribological analysis. His research aims to improve energy efficiency, reliability, and controllability of fluid power systems by conducting interdisciplinary research on both component and system level, and to explore and expand fluid power's use and fluid power technology in new applications. Dr. Shang joined Maha fluid power research center in 2020 as an assistant professor. He graduated as a Ph.D. student in 2018. His Ph.D. advisor is the late professor Dr. Monika Ivantysynova.



Raul Rodriguez received the bachelor's degree in Mechanical Engineering from Instituto Tecnológico de Estudios Superiores de Occidente in 2012. He is currently working as a Simulation and System Engineer at Bosch Rexroth, Fountain Inn, SC, USA. His work primarily involves the simulation and analysis of mechanical components and assemblies, with an emphasis on evaluating structural integrity and mechanical strength. His professional interests include mechanical system modeling, fatigue analysis, and the development of robust simulation methodologies for complex mechanical systems.



Jeemon Kurian brings over 20 years of industry experience as a simulation engineer, with a specialized focus on hydraulics and fluid system applications. Throughout his career, he has played a pivotal role in developing and validating innovative solutions for automotive and mobile machinery systems.
He holds a Bachelor’s degree from Visvesvaraya Technological University and a Master’s degree from Coventry University. Since joining Bosch Engineering in 2005, Jeemon has contributed significantly to technological advancements within the company. Previously, Jeemon Kurian was engineering manager responsible for powertrain development at Bosch Limited India until 2019. He currently leads the Simulation team in Engineering department at Bosch Rexroth in the United States. In this leadership role, he leverages his deep technical expertise to drive innovation and establish new benchmarks in simulation excellence. His work continues to strengthen Bosch Rexroth’s position as a global leader in industrial and mobile applications.



Dr. Enrique Busquets is currently responsible for the Bosch-Rexroth Mobile Solutions business in North America encompassing sales, engineering, product management and manufacturing. In addition to this current role, Enrique has been responsible for engineering with regional product and business development responsibility on hydraulic pumps, valves, electronics, software, telematics, and electrification in North America since April 2021 and the service and aftermarket business since 2024. Additionally, since 2018, Enrique has been responsible for the testing, validation, and vehicle integration infrastructure in North America. Previously, Enrique Busquets was the engineering manager responsible for systems and software development at Bosch Rexroth North America from 2018 to 2021. The technology focus was mobile applications with electronified hydraulics and electrified systems.
Enrique holds a bachelor's degree from the University of Texas at El Paso and a master's, and doctorate degree in mechanical engineering from Purdue University with emphasis on hydraulics and electronics controls. In addition to his professional activities, Dr. Enrique Busquets is the Bosch Rexroth industry sponsor and part of the industry advisory board at the Maha research center.

New Prototype of High-Speed External Gear Pump Filling Oil from Tooth Root

Rongfu Tang[1,2], Hong Ji[1,2], Bo Lin[3,4], Fei Sun[1,2]

[1]*Key Laboratory of Advanced Pumps, Valves and Fluid Control System of the Ministry of Education, Lanzhou University of Technology, Lanzhou 730050, China;* [2]*College of Energy and Power Engineering, Lanzhou University of Technology, Lanzhou 730050, China;* [3]*Guangxi Liugong Machinery Co., Ltd, Liuzhou 545001, China;* [4]*Guangxi Liugong Yuanxiang Technology Co., Ltd, Liuzhou 545001, China.*

tangrflut@163.com jihong@lut.edu.cn linbo@liugong.com 804216706@qq.com

Abstract.

To overcome the speed limitation of conventional external gear pumps (CEGPs) due to insufficient oil suction efficiency, a novel external gear pump (NEGP) based on tooth root-tooth tip distribution principle is proposed. A comprehensive flow-pressure mathematical model and flow field simulations are developed to analyze the tooth cavity suction process, revealing the key mechanisms responsible for inadequate oil filling in CEGPs. Additionally, the concepts of effective oil suction angle and efficient oil suction angle are introduced to facilitate a comprehensive quantitative analysis. Guided by these findings, an enhanced NEGP configuration integrating a "π"-shaped axial flow groove and a centrifugal flow baffle plate is designed to improve suction capability. This design significantly increases the effective suction area, suppresses centrifugal oil shedding, and enables dual-path oil intake via both tooth root and tooth tip, thereby extending the efficient suction angle. Both simulation and experimental results confirm the superior performance of the proposed NEGP, demonstrating a volumetric efficiency that is 12.1 percentage points higher than that of the CEGP at 3000 rpm.

Keywords. External gear pump, tooth root-tooth tip distribution principle, volumetric efficiency, effective oil suction angle, efficient oil suction angle

1. INTRODUCTION

External gear pumps are widely employed in construction machinery, automotive systems, aerospace, and other industrial applications due to their simple structure, low cost, and strong resistance to contamination. To align with the efficient operating range of electric motors, gear pumps are increasingly designed for higher speeds and miniaturization. However, the maximum rotational speed of conventional high-power external gear pumps (displacement

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

≥ 60 ml) is typically limited to 2500 rpm. Exceeding this threshold leads to a significant decline in volumetric efficiency, resulting in reduced energy utilization and difficulty in achieving high-speed design objectives.

The volumetric efficiency of hydraulic pumps is primarily influenced by two factors. The first is internal leakage, particularly that occurring at the gear end-face clearance, which accounts for approximately 70% of total leakage [1]. The second factor is suction efficiency, which becomes especially critical under high-speed operating conditions. Suction efficiency is governed by multiple parameters, including tooth number [2], tooth profile [3], size and orientation of the suction port [4,5], pipeline diameter and filter screen porosity [6,7], as well as pump speed [8], suction pressure [9], and the physical and chemical properties of the hydraulic fluid [10].

Various approaches to improving oil suction efficiency have been proposed in the existing literature. El-Hadj [11] suggested increasing the diameter of the suction port to reduce suction resistance. However, Corvaglia's subsequent research revealed that when the inlet diameter exceeds a certain threshold, internal recirculation occurs at the inlet, impeding oil entry into the tooth cavity and reducing its filling capacity [12]. Milani [13] experimentally compared the effects of single versus double suction ports on volumetric efficiency and found that the dual-port configuration significantly enhances pump performance. Bügener [14] applied biomimetic principles to optimize the suction channel structure, thereby improving flow streamline continuity and reducing local flow losses caused by vortices. These studies collectively provide valuable insights for addressing insufficient oil suction in gear pumps under high-speed operating conditions.

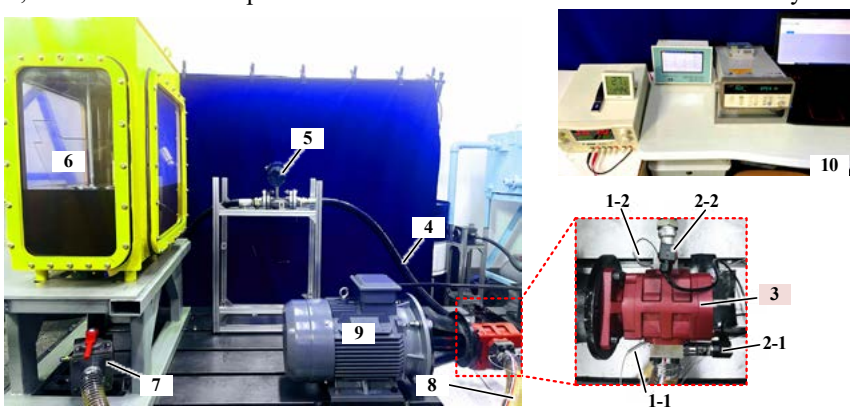
This study investigates the phenomenon of inadequate oil suction in CEGP during high-speed operation. A mathematical model of the flow-pressure dynamics during oil suction in the tooth cavity, along with a flow field simulation model, is established to systematically elucidate the underlying mechanisms contributing to reduced suction efficiency. Based on these findings, a novel external gear pump NEGP incorporating a tooth root-tooth tip distribution principle is proposed. Experimental results demonstrate that the NEGP effectively overcomes the speed limitation imposed by poor oil suction efficiency, thereby breaking through a key technical bottleneck. The findings of this research offer a theoretical foundation for the innovative design of high-speed external gear pumps.

2. ANALYSIS OF OIL SUCTION CHARACTERISTICS TO CEGP

2.1. CEGP Oil Absorption Efficiency Test

The experimental study focuses on a CEGP with a displacement of V_d=63cm^3/r. The tests were conducted under controlled conditions, with an oil temperature of 44.3°C and an ambient temperature of 27.3°C. The pump operated in an unloaded state to minimize the influence of internal leakage on volumetric efficiency, ensuring that the measured efficiency closely approximates the pump's suction efficiency. The test setup is illustrated in Figure 2.1, where HM46 anti-wear hydraulic oil is used in the reservoir. During testing, the motor drives the CEGP to draw oil from the reservoir through the suction pipe and discharge it back via the discharge pipe, forming a closed-loop circulation system. High-precision pressure sensors (accuracy: ±0.25%) and temperature sensors (accuracy: ±0.1°C) were installed at the pump inlet and outlet, respectively, to monitor oil pressure and temperature.

A turbine flowmeter with an accuracy of ±0.5% was mounted on the discharge line to measure the volumetric flow rate. After the pump reached steady-state operation, data were acquired using a data acquisition system at a sampling frequency of 10 Hz over a 20-second interval, and each test was repeated three times to ensure measurement reliability.



1-Temperature sensor 2-Pressure sensor 3-CEGP 4-Discharge pipe 5-Turbine flowmeter 6-Oil tank 7-Switch valve 8-Suction pipe 9-Motor

Figure 2.1. Experiment System for External Gear Pump

The time-averaged value was computed for each of the three data sets, and the arithmetic mean of these averages was calculated and denoted as Q_C. Subsequently, the volumetric efficiency η_C of the gear pump at various rotational speeds n was determined using Equation (2.1), with results presented in Figure 2.2. As shown by the experimental curve, η_C initially increases and then decreases with rising speed. Specifically, η_C gradually rises from 87.97% at 500 rpm, reaching a maximum of 94.2% at 2700 rpm. Beyond this point, η_C declines rapidly, dropping to 81.87% at 3000 rpm, representing a reduction of 12.33 percentage points from the peak efficiency. This trend clearly indicates a significant deterioration in oil suction performance at high operating speeds for the CEGP.

$$\eta_c = \frac{Q_c}{n \cdot V_d} \qquad (2.1)$$

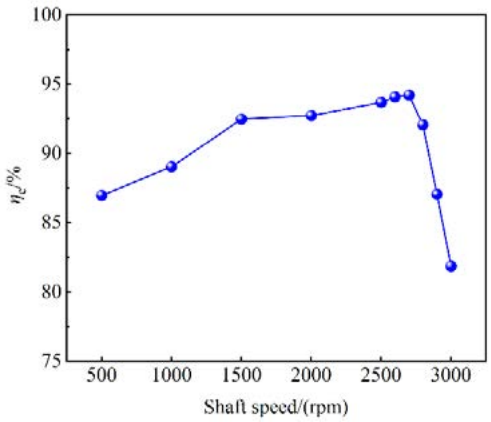


Figure 2.2. Volume efficiency of CEGP at different speeds

2.2. *CEGP Cavity Flow - Pressure Characteristics*

The CEGP employs a radial oil suction method from the tooth tip. The flow rate entering the tooth cavity is determined by the minimum flow area A formed during gear disengagement. To facilitate analysis, it is assumed that when the gears rotate to a specific angular position, the oil suction process can be approximated as quasi-static. Under this assumption, the minimum pressure difference Δp required for an instantaneous flow rate q to pass through the minimum flow area A is calculated. To clearly differentiate between the two suction configurations, the minimum suction area associated with radial suction from the tooth tip is denoted as A_r, and the corresponding minimum pressure difference is defined as Δp_r. For axial suction from the tooth root, the minimum suction area is denoted as A_a, and the minimum tooth cavity volume is represented as V. The oil flow paths for the two suction methods and the gear structural parameters are illustrated in Figure 2.3.

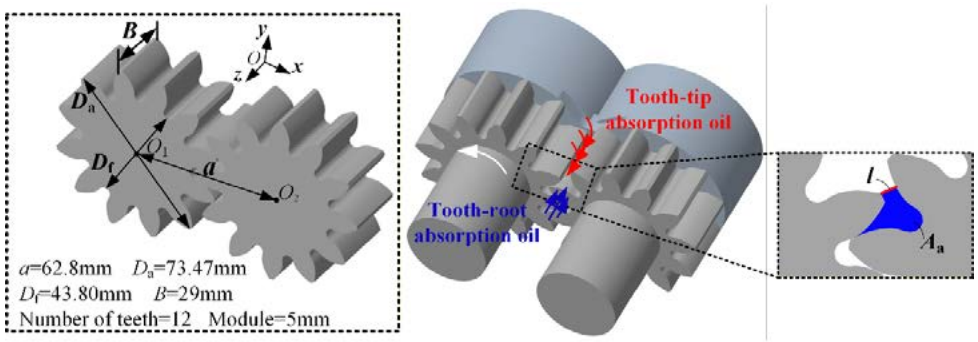


Figure 2.3. Diagram of different oil absorption directions and the gear structural parameters

By integrating 3D model analysis with data fitting, the variations of the shortest inter-tooth distance l, the radial oil suction area A_r, and the axial oil suction area A_a as functions of the rotation angle θ were systematically characterized in a quantitative manner. Consequently, the analytical expressions for A_r, A_a, and the minimum tooth cavity volume V were rigorously derived as follows:

$$l = \begin{cases} 0.063 \ (0° \le \lambda < 14°) \\ 45.83\theta^2 - 14.85\theta + 0.96 \ (14° \le \lambda \le 35°) \end{cases} \quad (2.2)$$

$$A_r = l \cdot B \quad (2.3)$$

$$A_a = 226.84\theta^2 - 1.9\theta + 10.74 \ (0° \le \lambda \le 35°) \quad (2.4)$$

$$V = A_a \cdot B \quad (2.5)$$

$$\begin{cases} \theta = \omega t \\ \lambda = \theta \cdot \dfrac{180}{\pi} \end{cases} \quad (2.6)$$

$$\omega = \frac{2\pi n}{60} \quad (2.7)$$

where ω represents the angular velocity of the gear.

As derived from reference [15], the relationship between flow rate and pressure can be expressed as:

$$q = C_{\mathrm{d}} A \sqrt{\frac{2\Delta p}{\rho}} \quad (2.8)$$

Equation (2.8) can be transformed into

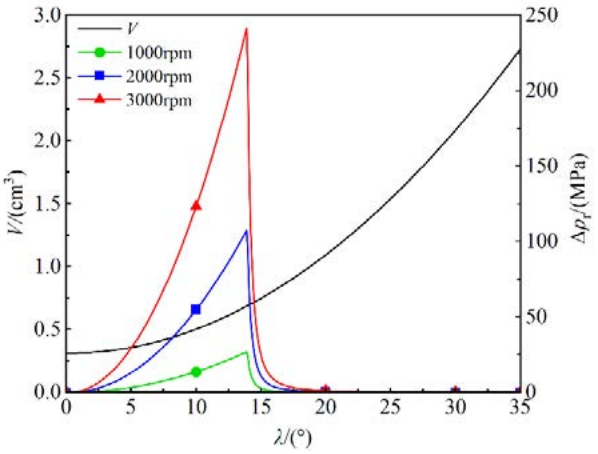
$$\Delta p = \frac{\rho}{2C_{\mathrm{d}}^2}\left(B\frac{\mathrm{d}A_{\mathrm{a}}}{\mathrm{d}t}\right)^2 \frac{1}{A^2} = \delta \cdot \frac{1}{A^2} \quad (2.9)$$

$$\delta = \frac{\rho}{2C_{\mathrm{d}}^2}\left(B\frac{\mathrm{d}A_{\mathrm{a}}}{\mathrm{d}t}\right)^2 = \frac{\rho}{2C_{\mathrm{d}}^2}\left[B\left(453.68\omega^2 t - 1.9\omega\right)\times 10^{-9}\right]^2 \quad (2.10)$$

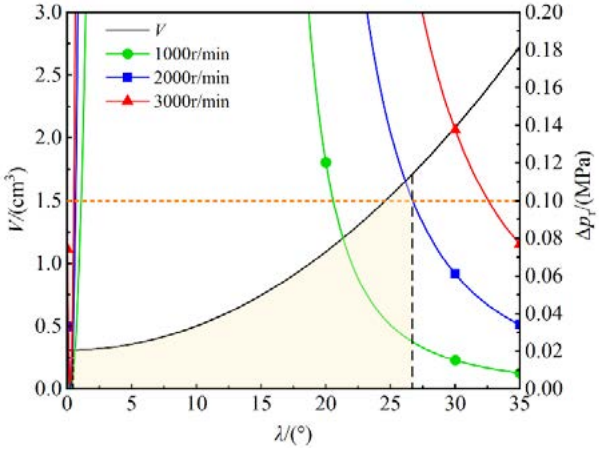
By substituting equation (2.3) into equation (2.8), the minimum pressure difference Δp_{r} required for radial oil suction can be determined. The resulting expression for Δp_{r} as a function of angular velocity ω is given by:

$$\Delta p_{\mathrm{r}} = \begin{cases} \delta \cdot \dfrac{1}{\left(0.063B\times 10^{-6}\right)^2} & (0° \le \lambda < 14°) \\ \delta \cdot \dfrac{1}{\left[B\left(45.83\omega^2 t^2 - 14.85\omega t + 10.74\right)\times 10^{-6}\right]^2} & (14° \le \lambda \le 35°) \end{cases} \quad (2.11)$$

Using the discharge coefficient $C_{\mathrm{d}} = 0.72$ [1] and fluid density $\rho = 860$ kg/m³, the variation curves of the tooth cavity volume V and the minimum pressure difference Δp_{r} with respect to the rotation angle θ are presented in Figure 2.4(a). The detailed profile of Δp_{r} within the range of 0 to 0.2 MPa is shown in Figure 2.4(b), where the yellow-shaded region indicates the minimum tooth cavity volume at a rotational speed of 2000 rpm under the condition that $\Delta p_{\mathrm{r}} \ge 0.1$ MPa.



(a)Δp_{r} global diagram



(b) Local graph of Δp_r from 0 to 0.2 MPa

Figure 2.4. V and Δp_r change curves with λ

As illustrated in Figure 2.4, under the tooth tip radial oil suction configuration, at rotational speeds of 1000, 2000, and 3000 rpm, the minimum proportions of the tooth cavity volume satisfying the condition $\Delta p_r \geq 0.1$ MPa account for 29.77%, 52.55%, and 83.07% of the total tooth cavity volume, respectively. This increasing trend demonstrates that higher rotational speeds require a larger angular range over which the gear pump must maintain sufficient oil suction pressure to ensure effective filling. However, the oil suction pressure is solely supplied by atmospheric pressure (approximately 0.1 MPa), which significantly reduces the available time for liquid filling. The underlying cause of this limitation is the inherently small radial oil suction area at the tooth tip, which severely constrains the flow runner capacity and thus impedes efficient oil delivery.

2.3. CEGP Internal Flow Field Characteristics

The theoretical analysis presented in Section 1.2 is confined to the flow–pressure difference relationship during the initial formation stage of the tooth cavity and does not yield an analytical solution for the internal flow dynamics after the cavity is fully developed. To address this limitation, the present study employs Pumplinx software to conduct numerical simulations. The RNG k-ε turbulence model is adopted to accurately capture turbulent flow characteristics. Boundary conditions are defined as follows: the inlet is assigned an absolute pressure of 0.1 MPa, and the outlet is set to 0.3 MPa. The oil properties are specified with a dynamic viscosity of 0.037 Pa·s and a density of 860 kg/m³, while the gear rotational speed is fixed at 2500 rpm. To ensure mesh independence, the outlet flow rate is monitored across different grid resolutions. The resulting simulation data are illustrated in Figure 2.5.

As illustrated in Figure 2.5, the time-averaged volumetric flow rate gradually converges with increasing cell density. When the cell count rises from 8.0×10^5 to 11.3×10^5, the flow rate variation is only 0.28%, which satisfies the criterion for mesh independence. Therefore, considering the balance between computational efficiency and numerical accuracy, a cell size of 8.0×10^5 is selected for all subsequent simulations.

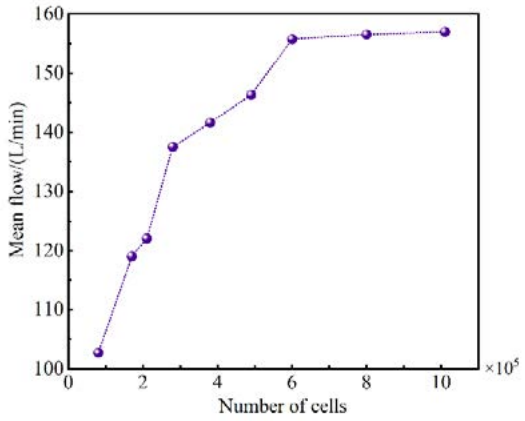


Figure 2.5. The output flow of the CEGP simulation model under different cell numbers

Furthermore, by monitoring the outlet flow rate in the CEGP simulation, the relationship between the volumetric efficiency η_{CS} and rotational speed is established, along with the difference between the simulated value η_{CS} and the experimental value η_C, as depicted in Figure 2.6. As illustrated in Figure 6, the trend of η_{CS} with respect to rotational speed is in good agreement with the experimental results. The maximum and average absolute differences $|\eta_{CS}-\eta_C|$ are 2.6% and 1.11%, respectively. The small discrepancy between simulation and test values indicates that the numerical model exhibits high reliability.

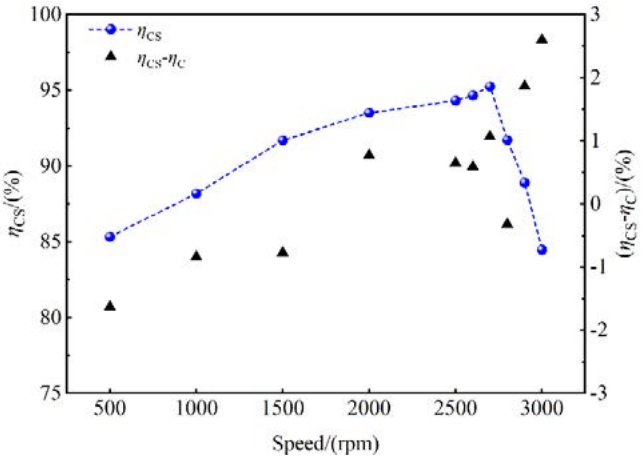
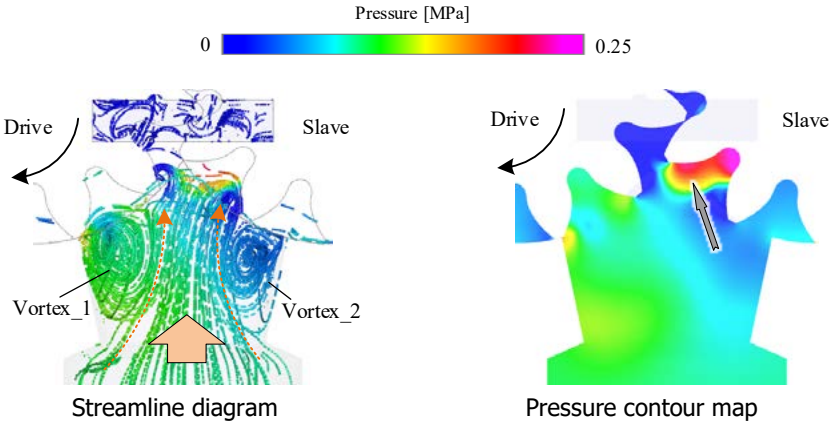


Figure 2.6. η_{CS} and(η_{CS}-η_C) of EGP at different rotational speeds

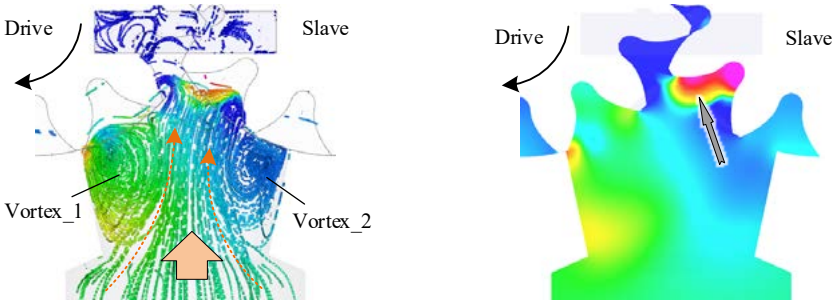
As shown in Figure 2.7, the streamline distribution within the CEGP at rotational speeds ranging from 2700 to 3000 rpm is presented. With increasing rotational speed, the number of streamlines entering the vortex region gradually increases, and the spatial extent of the vortex expands significantly. Concurrently, the effective flow area into the tooth cavity diminishes, and the density of streamlines decreases accordingly. This vortex structure impedes the normal inflow of oil into the tooth cavity. The underlying mechanism is attributed to the high centrifugal pressure generated within the fully formed tooth cavity due to the rotating gear's centrifugal force, which intensifies with rotational speed. Under this

pressure gradient, part of the oil is expelled radially outward and collides with incoming oil flowing in the opposite direction, thereby inducing localized vortex formation.



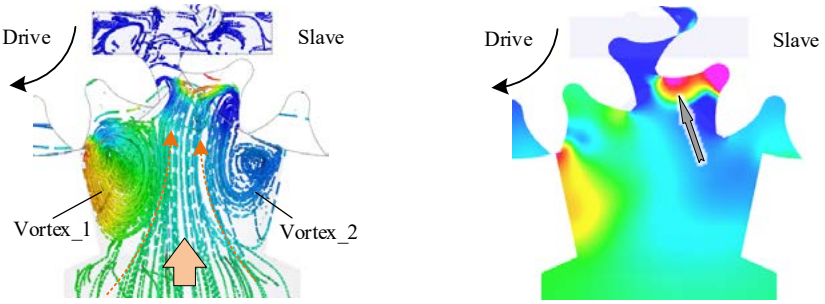
Streamline diagram Pressure contour map

(a) 2700rpm



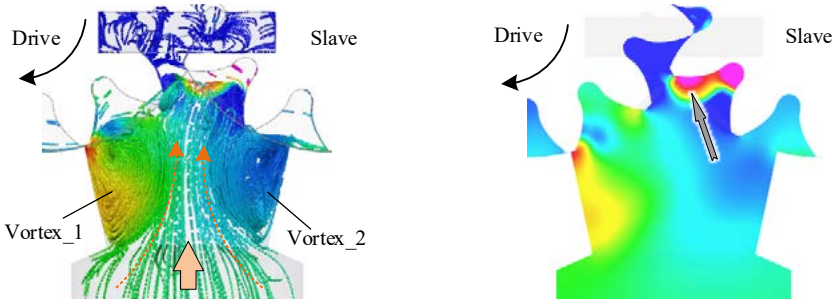
Streamline diagram Pressure contour map

(b) 2800rpm



Streamline diagram Pressure contour map

(c) 2900rpm



Streamline diagram　　Pressure contour map

(d) 3000rpm

Figure 2.7. Inlet flow diagram of CEGP at different speeds

Based on the above analysis, the high-speed oil suction deterioration process in CEGP can be characterized by three progressive stages. First, as the gear teeth begin to disengage and form a complete tooth cavity, the radial oil suction at the tooth tip is insufficient to meet the required flow rate, resulting in a low cavity filling efficiency. Second, under the influence of high rotational speed and centrifugal force, part of the oil within the fully formed cavity is expelled radially outward, further reducing the filling level. Finally, the expelled oil collides with the incoming radial flow, generating two prominent local vortices at the pump inlet that significantly impede the subsequent oil suction process.

2.4. Effective Oil Suction Angle and Efficient Oil Suction Angle

To further analyze the influence of flow characteristics at each stage on the tooth cavity's oil suction efficiency, two key angular parameters are defined: the effective oil suction angle and the efficient oil suction angle. The effective oil suction angle refers to the angular range over which the tooth cavity can freely absorption and discharge fluid, with its filling performance influenced by the radial oil suction area and the inlet vortex. The efficient oil suction angle denotes the portion within the effective oil suction angle where the cavity filling rate is not constrained by these two factors and can achieve an ideal, loss-free filling state.

The geometric definitions of these angles are illustrated in Figure 2.8 and described as follows:

Effective oil suction angle α: This denotes the angular displacement between the line c_1, which connects the edge position of the pump body's oil suction port to the gear center, and the $+x$ reference axis.

Radial oil suction area influence angle β: As introduced in Section 1.3, this represents the angular displacement during which $\Delta p_r \geq 0.1$ MPa, reflecting the extent to which the radial oil suction area limits oil inflow.

Inlet vortex influence angle γ: This is the angle between line c_1 and line c_3, where c_3 is drawn from the gear center tangent to the boundary of the inlet vortex influence zone. It quantifies the angular extent of obstruction caused by the inlet vortex.

Finally, the efficient oil suction angle φ is determined by the equation $\varphi=\alpha-\beta-\gamma$, representing the optimal angular range for oil suction that remains unaffected by hydraulic losses.

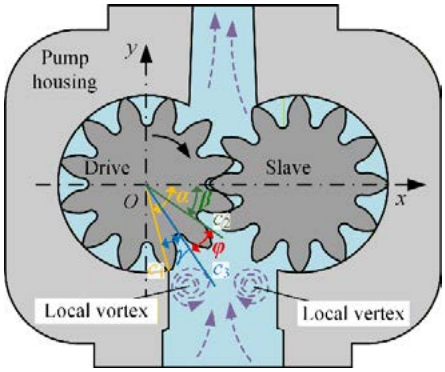


Figure 2.8. Diagram of α, β, γ, φ

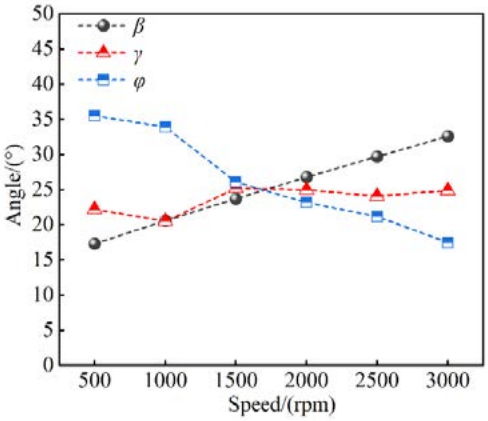


Figure 2.9. The change curve of the β, γ and φ with speed

The variation of each angular parameter with rotational speed is presented in Figure 2.9. As shown in the figure, the radial oil suction area influence angle β increases significantly with rising rotational speed, whereas the inlet vortex influence angle γ exhibits a moderate upward trend with a relatively limited increment. Under these conditions, the efficient oil suction angle φ, primarily governed by the rapid expansion of β, decreases sharply as rotational speed increases, leading to a substantial reduction in the effective liquid filling time of the tooth cavity.

3. Design Principles and Verification for NEGP

Based on the aforementioned analysis, the efficient oil suction angle is primarily constrained by two factors: the radial oil suction area and the inlet-localized vortex. Therefore, the NEGP structural optimization proposed in this study is conducted with respect to these two aspects.

3.1. Design of the New Distribution Structure for NEGP

(1) Centrifugal Flow Baffle Plate and Radial Flow Distribution Groove

According to the analysis in Section 2.3, the local vortex at the inlet is primarily caused by the centrifugal force generated during high-speed gear rotation, which drives the oil out of the tooth cavity. To mitigate this phenomenon, a centrifugal flow baffle plate is incorporated into the pump housing to prevent oil from being expelled under centrifugal force. Additionally, to maintain an adequate low-pressure oil suction passage in the tooth tip region, a radial distribution groove is machined at the center of the flow baffle plate, thereby balancing the dual requirements of flow containment and sufficient oil supply.

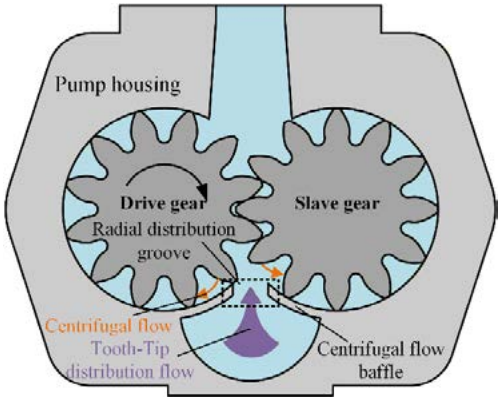


Figure 3.1. Diagram of Tooth-Top distribution flow

(2) "π" -shaped axial flow distribution groove

As described in Section 2.2, single radial oil suction at the tooth tip requires a relatively high suction pressure differential due to its limited flow area, making it prone to becoming a flow bottleneck under high-speed operating conditions. According to Equation (3.1), the centrifugal force acting on a fluid element is directly proportional to the radius of gyration and the square of the rotational speed. Consequently, the centrifugal pressure in the tooth root region is lower than that in the tooth tip region.

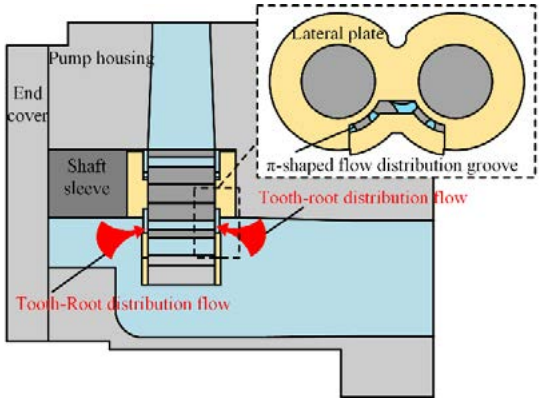


Figure 3.2. Diagram of Tooth-Root distribution flow

To address this imbalance, a π-shaped axial flow distribution groove is machined into the floating lateral plate adjacent to the gear tooth root and connected to the axial suction runner. This feature establishes an auxiliary suction path characterized by lower centrifugal

pressure, thereby compensating for insufficient oil supply at the tooth tip by enhancing suction capacity at the tooth root. As this design primarily leverages and enhances the inherent suction potential in the tooth root region, it is termed tooth-root flow distribution. The detailed configuration is illustrated in Figure 3.2.

$$p = \rho\omega^2 \frac{r^2}{2} + p_0 \qquad (3.1)$$

where r is the radial radius.

The radial distribution groove and the π-shaped axial distribution groove collectively constitute the root-to-tip flow distribution system. When flow allocation is based on suction area, the radial suction flow rate $q_{r,c}$ can be calculated using Equation (3.2). Likewise, the volume delivered to the tip tooth cavity $V_{r,c}$ is derived from Equation (3.3).

$$q_{r,c} = q \cdot \frac{A_r}{A_r + 2A_a} \qquad (3.2)$$

$$V_{r,c} = V \cdot \frac{A_r}{A_r + 2A_a} \qquad (3.3)$$

Substituting equation (3.2) into equation (2.9) yields

$$\Delta p_{r,c} = \frac{\rho}{2C_d^2} \cdot q_{r,c}^{\ 2} \frac{1}{A_r^2} \qquad (3.4)$$

The results calculated from Equations (3.3) and (3.4) are presented in Figure 3.3.

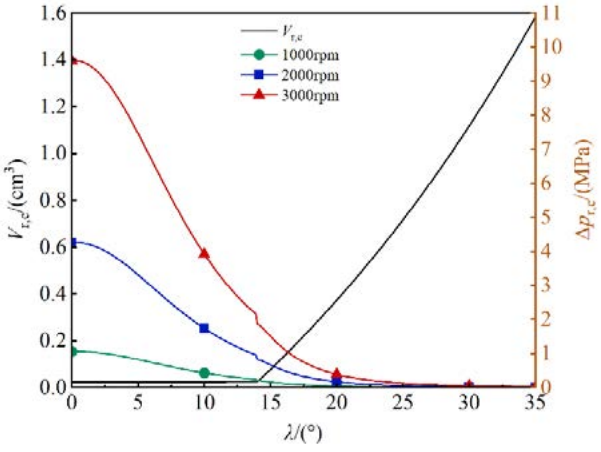


Figure 3.3. $V_{r,c}$ and $\Delta p_{r,c}$ change curves with λ

It can be observed that the pressure difference $\Delta p_{r,c}$ required for parallel oil suction is significantly lower than that of Δp_r under single radial oil suction. At rotational speeds of 1000, 2000, and 3000 rpm, the proportions of the total tooth cavity volume experiencing a pressure difference exceeding 0.1 MPa under the parallel oil suction mode are 4.26%, 22.5%, and 40.45%, respectively, reflecting reductions of 25.51, 30.05, and 42.62 percentage points relative to the single radial oil suction mode. These findings provide theoretical validation that the parallel oil suction configuration at both the tooth root and

tooth tip effectively improves oil suction conditions at high rotational speeds and exhibits superior oil suction performance.

3.2. Internal Flow Field Characteristics of NEGP

Under the same simulation conditions as those applied to CEGP, the fluid domain model of NEGP was analyzed. The volumetric efficiency of NEGP at various rotational speeds is presented in Figure 3.4. The results show that the volumetric efficiency increases rapidly with rising rotational speed within the range of 500 to 2500 rpm, while the rate of increase diminishes in the high-speed range from 2600 to 3000 rpm. Compared with the simulation data for CEGP, NEGP exhibits a volumetric efficiency improvement of 2.18%, 5.0%, and 9.61% at 2800, 2900, and 3000 rpm, respectively—demonstrating that the new distribution structure offers significant performance advantages under high-speed operating conditions.

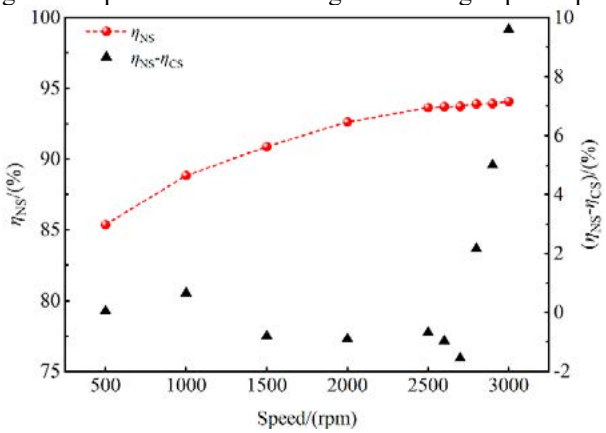


Figure 3.4. Volumetric efficiency of NEGP' simulation model

Figure 3.5 shows the internal streamline distribution of the NEGP at a specific angle under a rotational speed of 3000 rpm. As can be seen from the figure, no local vortices occur in the inlet area, and the oil is clearly divided into three independent and smooth streamlines entering the tooth cavity.

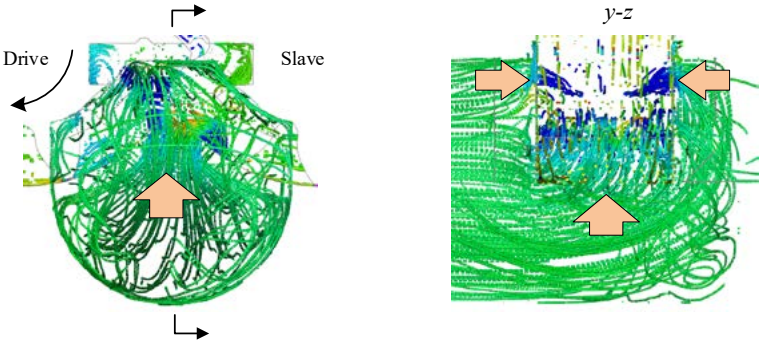


Figure 3.5. Inlet streamline figures of NEGP under 3000rpm

Figure 3.6 further illustrates the variation in angular parameters of NEGP. For NEGP, the vortex influence angle γ remains consistently zero, attributable to the optimized inlet flow field that effectively eliminates local vortices. Although the oil suction area influence angle

β increases rapidly with rotational speed, leading to a reduction in the effective oil suction angle φ, the value of φ at 3000 rpm is increased by 227.54% compared to CEGP. This significant expansion provides substantially improved oil suction duration for the gear pump under high-speed operating conditions, thereby providing a fundamental explanation for its enhanced performance.

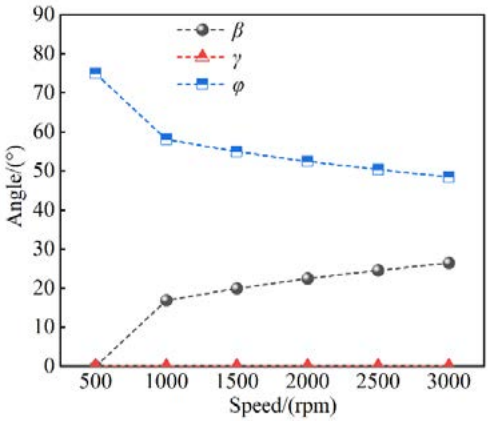
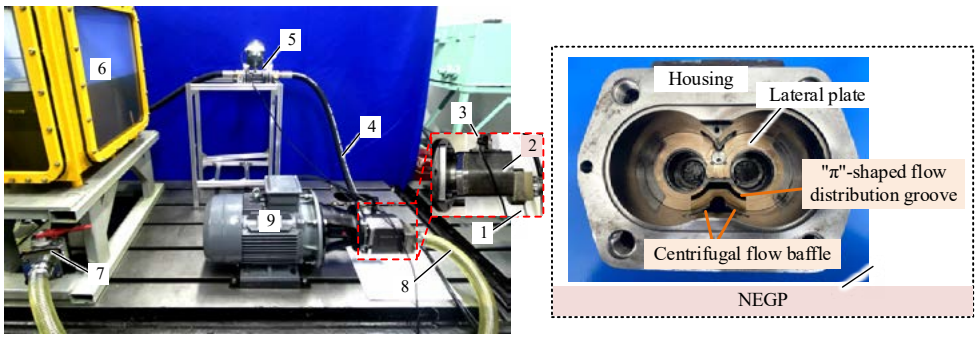


Figure 3.6. The change curve of the β, γ and φ with speed of NEGP

3.3. Experimental Verification to NEGP

To further validate the rationality of the root-to-tip parallel flow distribution principle employed in NEGP, a test bench as shown in Figure 3.7 was constructed, with operating conditions maintained consistently with those of the CEGP test.



1-Temperature sensor 2- NEGP 3- Pressure sensor 4-Discharge pipe 5-Turbine flowmeter 6-Oil tank 7-Switch valve 8-Suction pipe 9-Motor

Figure 3.7. Experimental bench of NEGP

The volumetric efficiency η_N of the NEGP varies with rotational speed, as shown in Figure 3.8. As can be observed from the figure, the volumetric efficiency of NEGP increases rapidly with rising speed within the range of 500 to 2500 rpm, while the rate of increase slows in the range of 2600 to 3000 rpm, reaching a peak value of 93.97% at 3000 rpm. When the rotational speed exceeds 3000 rpm, η_N begins to decrease, dropping to 79.51% at 4000 rpm. Compared with the efficiency η_C of CEGP, η_N exhibits only a small difference within the 500-2700 rpm range. However, at 2800, 2900, and 3000 rpm, η_N improves by 1.71, 6.92,

and 12.1 percentage points, respectively. Based on these results, 2700 rpm can be identified as the critical speed for significant improvement in NEGP's volumetric efficiency. The experimental results clearly demonstrate that NEGP effectively extends the speed range for efficient oil suction and offers distinct performance advantages under high-speed operating conditions.

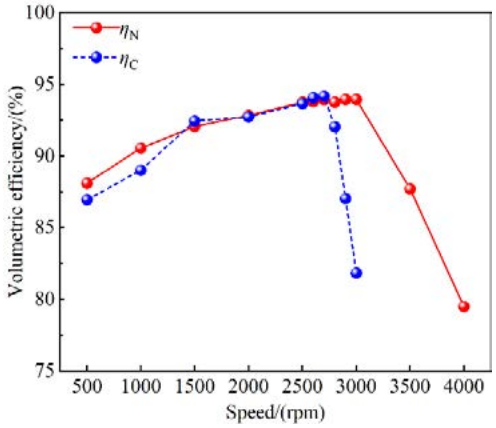


Figure 3.8. Volumetric efficiency of NEGP under different speeds

4. CONCLUSION

The intrinsic mechanism underlying the decline in oil suction efficiency of CEGP under high-speed operation was el1ucidated. This process can be divided into three stages: First, during the initial cavity formation stage, the narrow radial oil suction area at the tooth tip results in insufficient initial filling rate. Second, the centrifugal effect induced by high-speed rotation expels oil from the cavity, further reducing the filling rate. Third, the expelled oil collides with radially incoming oil at the pump inlet, generating two large local vortices that impede radial oil suction at the tooth tip. Collectively, these factors lead to a significant reduction in the effective oil suction angle of CEGP, inadequate effective filling duration, and ultimately result in poor high-speed oil suction performance and low volumetric efficiency.

A novel external gear pump based on the tooth root-tooth tip parallel distribution principle was proposed. Its innovative design incorporates a π-shaped axial distribution groove on the floating side plate to enhance oil suction capacity in the tooth root region. Additionally, a centrifugal flow baffle plate featuring radial distribution grooves is integrated into the pump housing to suppress centrifugal oil ejection while preserving volumetric oil suction at the tooth tip. The axial and radial distribution grooves serve as the primary pathways for centrifugal oil suction at the tooth root and volumetric oil suction at the tooth tip, respectively, forming a dual-channel oil supply system in NEGP.

Simulation and experimental results validate the superior performance of NEGP. Flow field analysis reveals the absence of local vortices at the NEGP inlet, with the effective oil suction angle being significantly larger than that of CEGP. Experimental data show that the volumetric efficiency of NEGP reaches a peak of 93.97% at 3000 rpm, and its efficient operating range is extended. Comparative performance analysis indicates that, with 2700

rpm identified as the critical speed, the volumetric efficiency of NEGP at 2800, 2900, and 3000 rpm exceeds that of CEGP by 1.71, 6.92, and 12.1 percentage points, respectively.

5. REFERENCES

[1] J. Lin, C. He, 'Hydraulic Components', Beijing: Machinery Industry Press, 1988.

[2] V. Ivanov, D. Karaivanov, S. Ivanova, et al., 'Gear mesh geometry effect on performance improvement for external gear pumps', MATEC Web Conf., EDP Sciences, 2019. https://doi.org/10.1051/matecconf/201928701007

[3] X. Zhao, A.Vacca, 'Analysis of continuous-contact helical gear pumps through numerical modeling and experimental validation', Mech. Syst. Signal. Pr., pp. 352-378, 2018. https://doi.org/10.1016/j.ymssp.2018.02.043

[4] G. Yang, W. Wang, F.Huang, et al., 'Cavitation characteristics of gear pump under different suction port sizes and speeds', Chin. Hydraul. Pneum., (01): 74-79, 2020. https://doi.org/10.11832 /j.issn.1000-4858.2020.01.012

[5] Bügener N, Klecker J, Weber J. Analysis and improvement of the suction performance of axial piston pumps in swash plate design[J]. Int. J. Fluid Power., 15(3): 153–167, 2014. https://doi.org/10.1080/14399776.2014.968436

[6] M. Suresh Kumar, K. Manonmani, 'Computational fluid dynamics integrated development of gerotor pump inlet components for engine lubrication' Proc. Inst. Mech. Eng., Part D: J. Automob. Eng., 224(12): 1555-1567, 2010. https://doi.org/ 10.1243/09544070JAUTO1594

[7] M. Suresh Kumar, K. Manonmani, 'Numerical and experimental investigation of lubricating oil flow in a gerotor pump', Int. J. Auto. Tech-Kor., 12(6): 903-911, 2011. https://doi.org/10.1007/s12239-011-0103-z

[8] D. del Campo Sud, 'Analysis of the suction chamber of external gear pumps and their influence on cavitation and volumetric efficiency', Univ. Politéc. Catal., Dep. Mec. Fluid., 2012.

[9] Y. Zhang, Z. Chen, J. Liao, 'Effects of pump parameters on the oil absorption characteristics of bidirectional involute internal gear pumps', Res. Eng., 20: 101405, 2023. https://doi.org/10.1016/j.rineng.2023.101405

[10] B. Wieczorek, Ł. Warguła, O. Zharkevich, et al., 'The study of the impact of oil type and throttling pressure on the vibrations of a gear pump', Adv. Sci. Technol. Res. J., 19(5), 142–155, 2025. https://doi.org/10.12913/22998624/201216

[11] A. A. El-Hadj, S. Z. B. Abd Rahim, 'Optimization of an external gear pump using response surface method', J. Mech., 36:567-575, 2020. https://doi.org/10.1017/ jmech.2020.7

[12] A. Corvaglia, M. Rundo, S. Bonati, et al., 'Simulation and experimental activity for the evaluation of the filling capability in external gear pumps', Fluids, 8(9): 251, 2023. https://doi.org/10.3390/fluids8090251

[13] M. Milani, L. Montorsi, F. Paltrinieri, 'Experimental investigation of the suction capabilities of an innovative high speed external gear pump for electro-hydraulic actuated automotive transmissions', Int. J. Fluid Power., pp. 243-272, 2024. https://doi.org/10.13052/ijfp1439-9776.2527

[14] N. Bügener, J. Klecker, J. Weber, 'Analysis and improvement of the suction performance of axial piston pumps in swash plate design', Int. J. Fluid Power., 15(3): 153-167, 2014. https://doi.org/10.1080/14399776.2014.968436

[15] A. Vacca, G. Franzoni, 'Hydraulic fluid power: fundamentals, applications, and circuit design' Hoboken: John Wiley & Sons, 2021.

Biographies



Rongfu Tang received the bachelor's degree in mechatronic engineering from Lanzhou University of Technology in 2019 and has been pursuing a Ph.D. in the same discipline since 2020. His primary research focuses on fundamental studies and applications of modern hydraulic components.



Hong Ji received a Bachelor of Engineering degree in Fluid Transmission and Control from Gansu University of Technology in June 1993, and subsequently earned a Doctor of Engineering degree from Zhejiang University in March 2005. He is currently a Professor and Doctoral Supervisor at the School of Energy and Power Engineering, Lanzhou University of Technology. His primary research focuses on fundamental studies and applications of modern hydraulic components, as well as hydraulic technology for construction machinery.



Bo Lin obtained Ph.D. in Analytical Chemistry from Wuhan University and is currently serving as a Senior Engineer at Guangxi Liugong Machinery Co., Ltd. Her primary research focuses include the development of lubricating materials for construction machinery, the application of non-metallic materials, and advancing material technologies for performance under extreme operating conditions.



Fei Sun received his Doctor of Engineering degree in Mechatronic Engineering from Lanzhou University of Technology in 2024. He is currently serving as a Lecturer in the School of Mechanical and Electrical Engineering at Lanzhou University of Technology, where his research primarily focuses on the integrated design of electro-hydraulic power systems.

Duty-Based Dumping Control for Underground Load-Haul-Dump Vehicle Based on Reinforcement Learning

LuLu Gao [1,*], Zihang Zhang[1], Johannes Sprink[2], Katharina Schmitz[2]

[1] *School of Mechanical Engineering, University of Science & Technology Beijing, Beijing, 100083, China*

[2] *ifas, RWTH Aachen University, Campus-Boulevard 30,52074 Aachen, Germany*

[*] *Corresponding author: Tel.: +86-15201463530; E-mail address: gaolulu@ustb.edu.cn*

Abstract.

Underground load-haul-dump (LHD) vehicles, with the excellent maneuverability and efficiency, are widely used to the materials removal in underground mines and constructions. Owing to the harsh operation environment, the autonomous functions of LHDs have been urgently needed for the consideration of safety, efficiency, and cost. Dumping is one of the core functions for underground load-haul-dump (LHD) vehicles. Typical underground environment including narrow space and varying dumping locations restricted the unconstrained autonomous dumping. This paper probes into the autonomous dumping control of LHDs under narrow and constrained environment based on the reinforcement learning (RL) frame. Firstly, the kinematic and dynamic models of a LHD are established considering the data-dependency of the RL method. The duty-based RL control method, is proposed to realize the coordinative control of the driving system and the working system in dumping duty, in which the safety in avoiding collision and working efficiency were considered. A test platform was established with the LHD to validate the methods. For the RL-based method, the performance of different algorithms was evaluated. A typical dump duty was performed to validate the method. The results shown that the method could finish the dumping operation autonomously under constrained environment. The duty-based RL control method shown great coordinative characteristics and environmental adaptability with the RL frame under constrained environment.

Keywords. Underground load-haul-dump (LHD), Dumping operation, reinforcement learning (RL), Duty-based, Coordinative

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Underground load-haul-dump (LHD) vehicles are essential equipment in underground mining and construction due to the excellent maneuverability and operational efficiency[1-2]. However, the harsh and confined working conditions pose significant challenges to automation, driving an urgent need for autonomous LHDs functionalities to enhance safety, productivity, and cost-effectiveness. The autonomous operations of LHDs evolve through five stages: driver-only, assisted, semi-autonomous, highly-autonomous, and fully autonomous, with current operations transitioning from assisted to semi-autonomous[3].

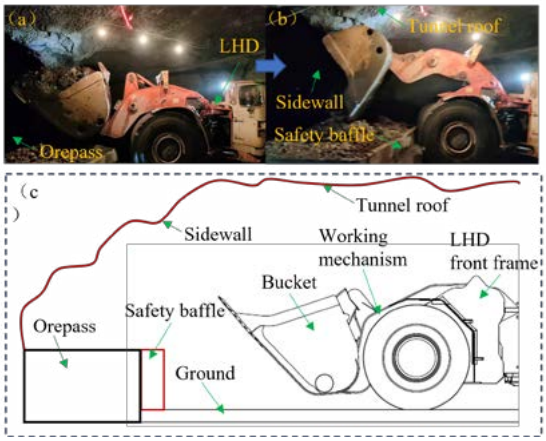


Figure 1: Typical dumping environment of underground LHD

The working environment of LHDs is narrow and constrained, as shown in Fig. 1, which functions include loading, hauling, and dumping materials. The autonomous loading function has been focused for many years considering the decisive effect on working efficiency. However, the interaction between materials and bucket is complex and the characteristics of materials varies significantly[4], which blocked the practical application and made the autonomous loading challenging and long term developing-needed. The autonomous hauling can be realized with the trajectory tracking of articulated vehicles, which has been applied in the practical working process. The dumping operation, one of the core function of LHDs, involved the control of the driving system and the working mechanism. Although the operation could be performed with the pre-set automation program, the driving system and working mechanism were not controlled coordinately as manual control. Current autonomous dumping would be inefficiency and unsafe in constrained environment. In addition, the variation of dump point and environment constraints could not be covered. Therefore, the improvement of autonomous dumping function would be practical and effective for the autonomous development transition[5-6].

The dumping function of LHDs could be realized with the pose (including trajectory and attitude angle) control of bucket. For the pose control, the traditional proportional-integral-derivative (PID) method would be the first choice considering the simple structure and stable performance. There are two problems to be addressed before the application of PID. First, the target/reference values of the controlled lift, dump cylinders, and the driving system should be provided via the inverse kinematics of the vehicle and the working mechanism.

The kinematics of the six link mechanism provided a feasible solution obtained from cumbersome calculation. Second, the parameters of PID must be tuned individually. Different algorithms had been used to tune the parameters, such as fuzzy logic, neural network (NN), and genetic algorithm (GA). Some modern control methods were also used to the trajectory tracking of bucket, including mode predictive control (MPC), optimal control, and RL-based methods[7-9]. However, most of these researches studied the trajectory control of bucket in open area and unconstrained environment, or only the working mechanism of LHDs[10] was mainly focused.

Inspired by the aforementioned researches, this paper explores the duty-based RL autonomous dumping control method. First, the models of the LHD and the corresponding test platform are established considering the data-dependency of the RL methods. Then, the duty-based RL control method (DRLC) is proposed with the RL frame, in which the completion rate, working efficiency, and environment constrains are covered. The performance of different algorithms and preview horizon in the methods, including deep deterministic policy gradient (DDPG), twin Delayed Deep Deterministic policy gradient algorithm(TD3), SAC, and distributional SAC (DSAC)[11], are evaluated under typical dumping operation. The main contributions of this paper include: (1) An environment-adapted control method for autonomous dumping of LHDs is established based on the RL frame; (2) The performance of different algorithms and preview horizon are uncovered under typical dumping operations; (3) The performance of RL-based methods is evaluated in terms of safety and efficiency.

2. KINEMATIC AND DYNAMIC MODEL

The working mechanism of the LHD is a typical reverse Z-shape six link mechanism, as shown in Fig. 2, the actuators of the whole system are the lift and dump cylinders. During the lifting process, if only the lift cylinder is controlled, the angle of bucket would vary, which would lead the scattering of the ore materials. Therefore, the whole autonomous dumping operation in constrained environment can only be realized with the coordinate control of the driving system and the working mechanism.

2.1. Model of LHD

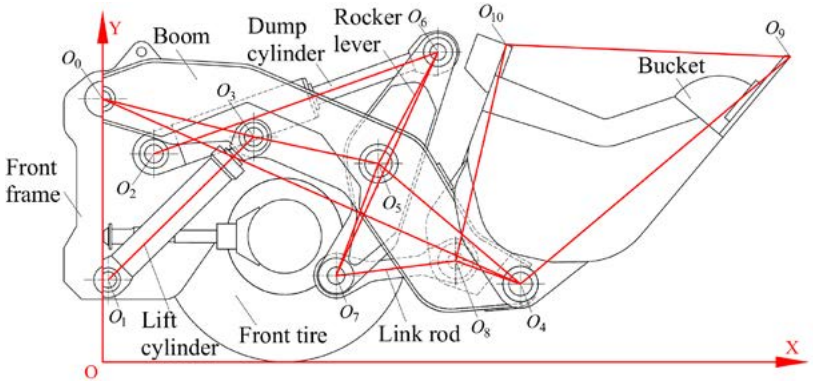


Figure 2: Working mechanism of LHD

The kinematic model of LHD and the working mechanism are involved during the dumping operation control. For working mechanism, as shown in Fig. 2, the boom is connected with the front frame (at O_0), rocker lever (at O_5), bucket (at O_4), and the lift cylinder (at O_3). The two connecting points of cylinders are O_1 and O_2, respectively. The motion of the working mechanism can be described in the coordinate frame (XOY), in which the vertical direction across the O_0 is defined as Y axis, and the horizontal tangent direction of front tire is defined as X axis.

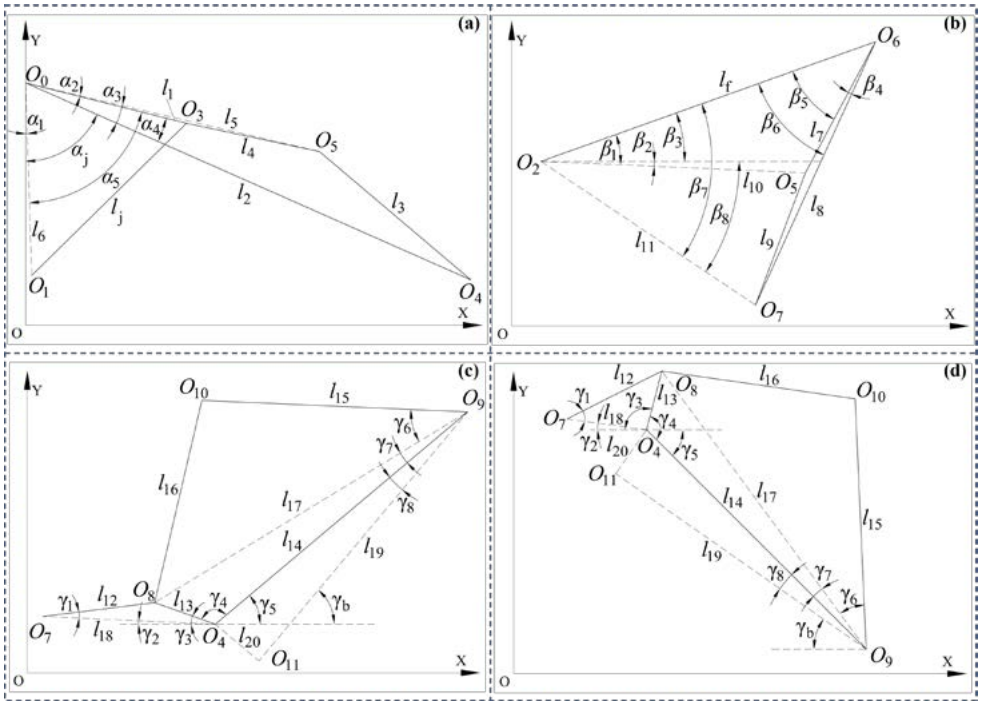


Figure 3: Geometric relationship of different states of working mechanism

The geometric relationship of bucket in dumping state is shown in Fig. 3 (d), and the coordinate of bucket tip O_9 can be expressed as:

$$\begin{cases} X_9 = X_4 + l_{14}\cos\gamma_5 \\ Y_9 = Y_4 + l_{14}\sin\gamma_5 \\ X_{10} = X_9 - l_{15}\cos(\gamma_6 + \gamma_7 - \gamma_5) \\ Y_{10} = Y_9 + l_{15}\sin(\gamma_6 + \gamma_7 - \gamma_5) \end{cases} \tag{1}$$

in which γ_b is the angle between bucket bottom and horizontal direction., the forward kinematics of the working mechanism can be obtained with the related geometry relationship in Figure 3.

Considering that the dumping operation is finished with driving system and working mechanism (no steering maneuver is involved), the longitudinal dynamic model of the LHD should be established. The longitudinal dynamic model of LHD in the X axis of global coordinate frame can be described as:

$$m\ddot{x} = F_{tx} - F_r \tag{2}$$

in which, m is the mass of LHD, F_{tx}, F_r are the longitudinal driving force, and total resistance force, respectively, x is the longitudinal distance of LHD in global coordinate frame.

The longitudinal force of tires can be given by Magic Formula:

$$F_{tx} = S_y + Dsin\left\{Carctan\left[\begin{matrix} B(X_m - S_x)(1-E) \\ +EarctanB(X_m - S_x) \end{matrix}\right]\right\} \tag{3}$$

where X_m denotes the slip ratio of the tire, which are defined as η; F_{tx} denotes the longitudinal force of tire; the parameters S_x, S_y, B, C, D, and E are constant, where in this study $S_x = 0$, $S_y = 0$, and other parameters are calculated by equations with vertical load. The slip ratio of different tire can be given by:

$$\eta = \begin{cases} (r_d\omega_t - v_x)/r_d\omega_t \times 100\% & Drive \\ (r_d\omega_t - v_x)/v_x \times 100\% & Brake \end{cases} \tag{4}$$

where r_d denotes the rolling radius of different tires, ω_t denotes the angular velocity of tires, and v_x denotes the longitudinal velocity of LHD. The final drive torque in the wheel overcomes the rolling resistance force and braking force to produce the longitudinal driving force, and the detailed model can be expressed as:

$$I_w\dot{\omega}_t = T_d - f_r - F_{tz}r_d - F_{tx}r_d \tag{5}$$

where I_w is the rotational inertia of the wheel, and T_d is the driving torque, F_{tz} is the vertical load of tires. f_r is the rolling resistance force, which can be expressed as:

$$\begin{cases} f_r = 0.00792 + 0.00025v_x & v_x \neq 0 \\ f_r = 0 & v_x = 0 \end{cases} \tag{6}$$

The longitudinal dynamic model of LHD can be obtained by combing equation (2) - (6). Therefore, the bucket pose in global frame could be given by combing the kinematic and dynamic models:

$$\begin{cases} X_{11} = x + X_4 + l_{14}\cos\gamma_5 \\ Y_{11} = y + Y_4 + l_{14}\cos\gamma_5 \end{cases} \tag{7}$$

where X_{11}, Y_{11} are the coordinates of bucket tip O_9 in global frame. y is the vertical distance of local and global frames.

2.2. Model validation

To validate the involved models, an LHD test system is established, as shown in Fig. 4. The control system is a two-layer structure system, in which an industrial computer is used in the upper layer for the functions of perception, navigation, and control decision. A 16 line LiDAR and inertial measurement unit (IMU) are used to obtain the vehicle state and the operation environment in the test site. The control target of the driving system and the working mechanism, including the velocity of LHD, angle of boom and displacement of dump cylinders, are given to the lower controller, including five EPEC controllers. The EPEC controllers are used to control the longitudinal velocity and the hydraulic related actuators, including steering, lift, and dump cylinders. Only the control of the driving system and the working mechanism during dumping operation are considered in this paper considering that the steering system is not involved during the late stage of dump operation. A displacement sensor is used to obtain the displacement of dump cylinder. For pressure sensors are used to record the pressure data of the two

cylinders during operations. Two angle sensors are used to measure the angle of boom and the spatial angle state of bucket. In addition, a revolution sensor is used to obtain the rotation speed of front tire, i.e., the longitudinal velocity of LHD. The controllers can communicate with each other via CAN bus. The details of the involved devices are shown in Tab. I.

Table 1: Detail parameters of the devices

Device	Parameters
Industrial computer	Nvidia jetson agx Xavier
IMU	XW-GI5610
LiDAR	LSLIDAR C16
Displacement sensor	BTL6, 0~10V, 0~1000mm
Angle sensor	SVCK15, ±180°
Revolution sensor	HC3806-1024, 12V
Pressure sensor	SUP-P300, 0-30MPa
EPEC controller	EPEC 3724

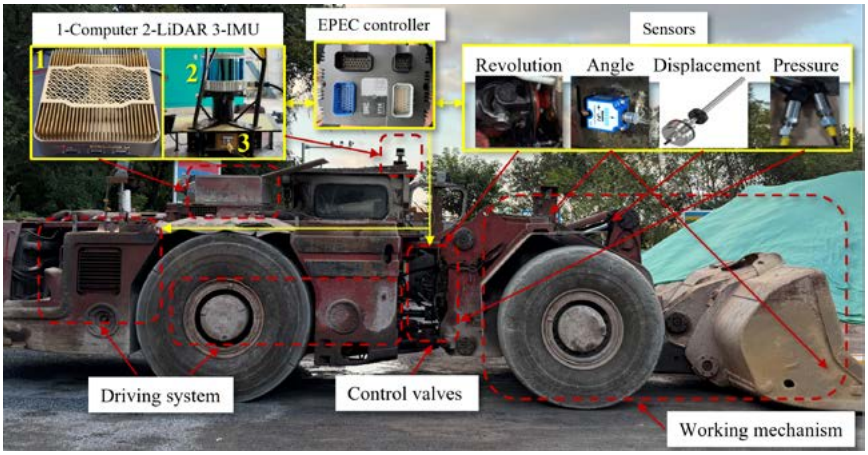


Figure 4: Schematic of test system

The dynamic model of LHD and kinematic model of the working mechanism were validated, separately. For the validation of longitudinal dynamic model, same control signal of driving system was used in the model simulation and field test, and the responses of longitudinal velocity and position were compared in Fig. 5. The longitudinal velocity of test and simulation results coincided well during the test interval, and only minor differences appeared during the free-running period, i.e., 39~41.8 s and 9.2~10.1 s. This error was brought on by the rolling resistance force model. The used resistance force model was established based on the paved road condition, while the test road was an unpaved road. The maximum velocity of the test was about 7.1 km/h, and the corresponding result was about 6.9 km/h. Fig. 5 (b) shown the position result of test and simulation. As it can be seen, the longitudinal position was about 26.4 m, and the next stop position was about 52.6 m. The simulation results given the 25.3 m and 52.1 m. It meant that the maximum position error was less than 4.2 %. These results indicated that the established dynamic model could reveal the dynamic characteristics of LHD in longitudinal direction.

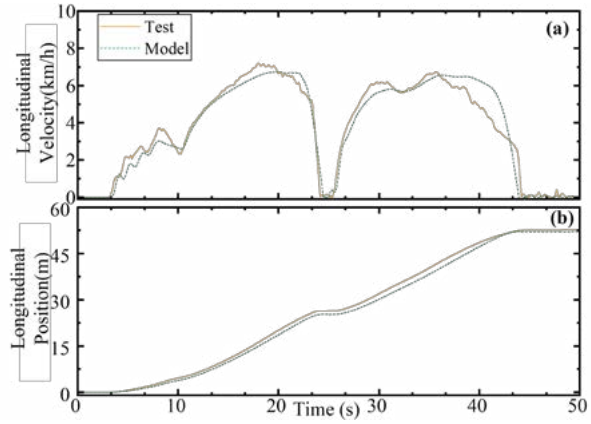


Figure 5: Comparing results of longitudinal dynamic model validation

For the validation of working mechanism, the displacement of cylinder and angle of boom and bucket were recorded. And the coordinate of bucket tip point O_9 was manually measured. Then the displacement results and angle results were input into the forward kinematic model, and the results of bucket angle and coordinate of bucket tip point were obtained continuously as the simulation results. According to the results in Fig. 6 (a)(b), the displacement of dump cylinder varied between 1256.8 – 1852.8 mm and the angle of boom varied between 65.3° - 109.4°. The comparing results of bucket tip position was shown in Fig. 6 (c), in which nine scatter coordinates are the actual measured values. The detailed values of bucket tip position were compared in Tab. II. The average error in Y-axis was larger than the error in X-axis, and the average error rate was about 0.2% and 0.5% in X and Y directions, respectively. The angle of bucket varied between -59.8° - 46.2° according to Fig. 6 (d). The maximum angle error was about 1.4°, at about 36 s. The error rate of angle was about 1.1%. These results indicated that the kinematic model was accurate enough for the research of dumping control.

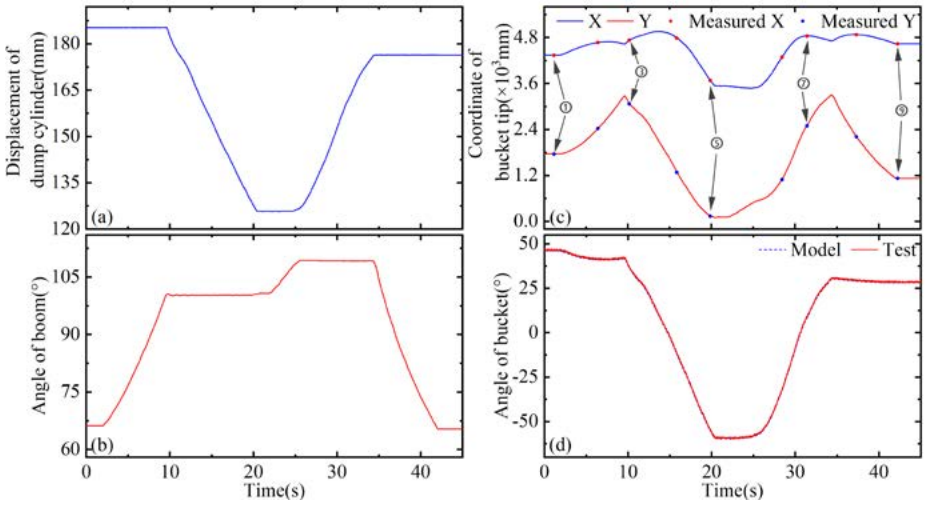


Figure 6: Comparing results of kinematic model validation

Table 2: Detail values of bucket tip position

	Model(mm)	Measured(mm)	Error(mm)
1	x=4343.6, y=1757.0	x=4336.9, y=1761.7	ex=6.7, ey=4.7
2	x=4669.3, y=2418.5	x=4664.7, y=2413.3	ex=4.6, ey=5.2
3	x=4732.0, y=3077.8	x=4728.2, y=3088.9	ex=3.8, ey=11.1
4	x=4788.9, y=1291.9	x=4782.7, y=1300.3	ex=6.2, ey=8.4
5	x=3686.0, y=150.4	x=3679.2, y=157.6	ex=6.8, ey=7.2
6	x=4291.3, y=1105.0	x=4283.7, y=1114.7	ex=7.6, ey=9.7
7	x=4841.0, y=2504.9	x=4832.2, y=2521.1	ex=8.8, ey=16.2
8	x=4876.7, y=2210.8	x=4870.4, y=2221.3	ex=6.3, ey=10.5
9	x=4632.9, y=1117.3	x=4628.7, y=1113.2	ex=4.2, ey=4.1

Based on the above analysis results, it can be concluded that the established dynamic and kinematic models could reveal the characteristic of driving system and working mechanism during dumping operation. And the models were accurate enough for the RL-related method.

3. AUTONOMOUS DUMPING CONTROL METHODS

The RL frame was used to realize the coordinate control of the driving and the working system in this section. During the manual operation, the operator can freely perform the control according to current states and target duty until the required duty is completed. Therefore, we designed the duty-based RL control method (DRLC), the duty-driven characteristics of manual operation is considered.

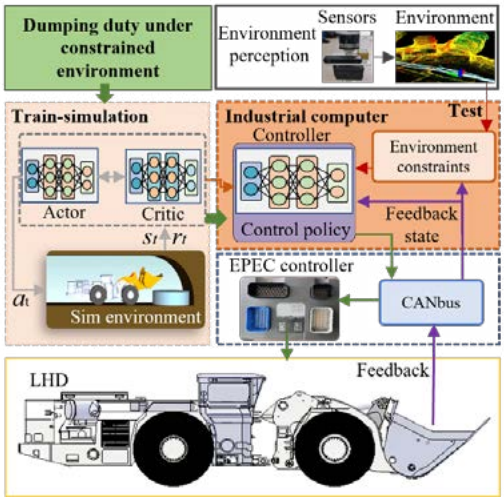


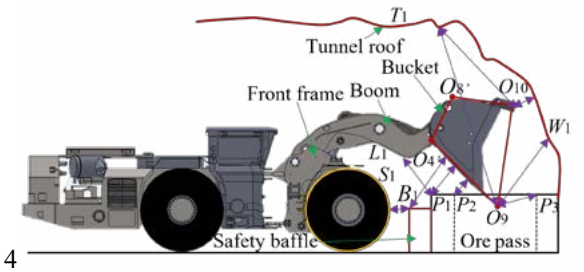
Figure 7: Schematic of RL-based dump control method

As stated before, the LHD is operated in narrow underground mine. The excellent efficiency of manual LHD originates from the preview characteristic of operators. It means that the operator could see the target ore pile, path, and ore pass in advance. Reasonable operation for driving system and working mechanism could be implemented according to the current states of LHD and the information of duty and

environment. Inspired by this characteristic, the DRLC is designed in this section. We designed an autonomous dump system for LHD based on the RL frame in this section to improve the efficiency and safety, as shown in Fig. 7.

For the whole system, the environment information, including the position of ore pass, tunnel, safety baffle, and LHD, were obtained with related sensors, as stated in test system. The environment constraints, including the tunnel roof, side wall, and safety baffle, are extracted. The trained control policy with RL frame in the controller gives the target trajectory of bucket point considering the collision risk, states and actuator constraints of LHD and further specifies the actuator targets of EPEC controllers, as shown in Fig. 7. The target trajectory contains the attitude angle of bucket, and X and Y location in global frame. Special attention is spent on the design of policy training based on RL.

The RL-based controller is a trained policy in essence based the RL train frame. The policy can give the control signals according to the feedback states, defined in s_t. The collision constraints during dumping operation is shown in Fig. 8. As it can be seen, the whole situation is simplified as the plane view. It means that the longitudinal symmetry line of LHD coincides with the symmetry line of the circular ore pass. The tunnel roof, sidewall, and front tire are denoted as T_1, W_1, and S_1, respectively. The outline of bucket is marked as O_4'$O_9O_{10}O_8$'. The outline of safety baffle is marked as B_1. For the ore pass, the out point is marked as P_1, and the inner points are marked as P_2 and P_3, respectively. The lower point of boom is marked as L_1.



4

Figure 8: Collision constraints during dumping operation.

For the safety baffle, the collision, with front tire S_1, lower point of boom L_1, and bucket O_4'$O_9O_{10}O_8$', might occur. For ore pass, the collision might occur between ore pass, low point of boom, and bucket. As it can be seen from Fig. 8, although the outline of bucket and ore pass is intersecting, no collision occur. It means that the middle channel of ore pass is free for bucket. For the main moving parts, the collision might occur between bucket and tunnel roof, bucket and sidewall of tunnel. All of the safety constraints is expressed as:

$$C = S(g(f_1,\ f_2)) \tag{8}$$

in which C is the collision results, it has two values, i.e., 1 and 0. The value 1 denotes the collision occurs, and the value 0 denotes the motion of LHD is free to collision. $g(\cdot)$ is the closest distance calculation function. f_1 contains O_4'$O_9O_{10}O_8$', S_1, and L_1, f_2 contains T_1, W_1, B_1, and P_1 - P_3. $S(\cdot)$ is specified as:

$$S(x) = \begin{cases} 1 & x \le x_{lim} \\ 0 & x > x_{lim} \end{cases} \tag{9}$$

in which x_{lim} is the threshold for collision detection. Then the collision reward could be given by:

$$r_c = -c_6 C \tag{10}$$

4. RESULTS AND DISCUSSION

4.1. Simulation validation of ADC

A dumping operation is performed firstly in simulation environment to validate DRLC. The control policy was obtained with DSAC algorithm. The dump operation scene is concretized, as shown in Fig. 9. As it can be seen, the ore pass is concretized as a hollow cylinder with a height of 0.8 m, an inner diameter of 2.8 m, and an outer diameter of 1 m. The safety baffle is a rectangular prism with a section height of 0.6 m and a width of 0.4 m. The distance between the center of the front wheel and the center of the ore pass is 22.0 m. The distance between the center of the front wheel and the left side of safety baffle is 20.0 m. The motion of bucket is constrained with 3.8 m in the top direction, which is the distance between the tunnel roof and front wheel center. The side wall around ore pass is simplified as a circular arc with 3.0 m radius.

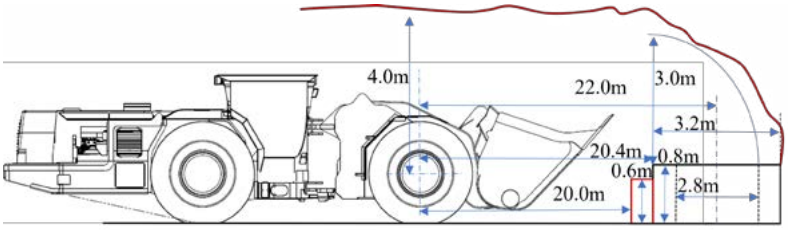


Figure 9: A typical dumping scene in constraint environment

The results of DRLC are shown in Fig. 10 - Fig. 11. The state results of bucket tip were shown in Fig. 10, in which the subfigures (a), (b), and (c) are the results of displacement in X direction, in Y direction, and attitude angle, respectively.

As it can be seen from Fig. 10 (a), the LHD was planned to run longitudinally from start point 0 m to final point 21.9 m. The control result coincided well reaching the target location. The front point of bucket was within the range of ore pass, which is very important for the dumping operation. As it can be seen from Fig. 10 (b), the vertical direction distance, i.e. the Y axis was planned to lift from 1.74 m to 3.19 m, and then fallen from 3.19 m to 0.72 m. The motion of simulation coincided well avoiding obstacles and reaching the target location. For the attitude angle of bucket, i.e., γ, shown in Fig. 10 (c). The angle varied slightly during the lifting stage. The variation was about 5.4°, during 40.8° and 46.2°, which made sure there was no scattering dumping the lifting stage. Then bucket was operated to tilt during 21 s - 29 s. The final tilt angle was about -54.2°.

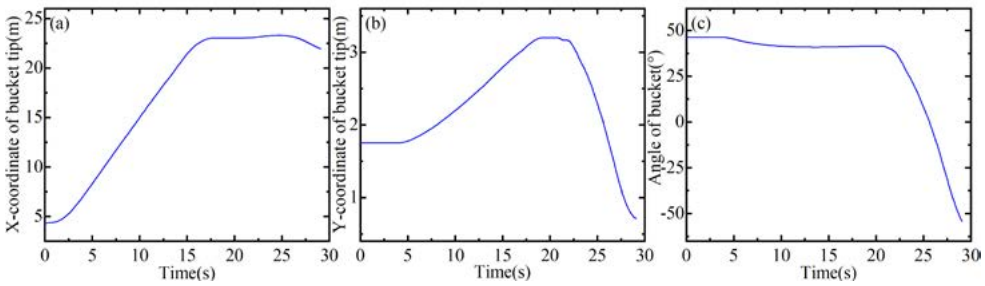


Figure 10: Simulation results of displacement and angle in dump operation.

In order to analyze the dumping operation, the trajectories of bucket tip and front tire were recorded in Fig. 11. The location of ore pass and safety baffle was also indicated in the figure. As it can be seen, the DRLC can control the trajectory of bucket front tip to finish dumping operation. The final operation was finished within the range of ore pass. In addition, the LHD stop in front of safety baffle with an appropriate distance according the results of tire trajectory. Based on the above analysis, it can be concluded that the DRLC can control LHD to finish dump operation autonomously and precisely.

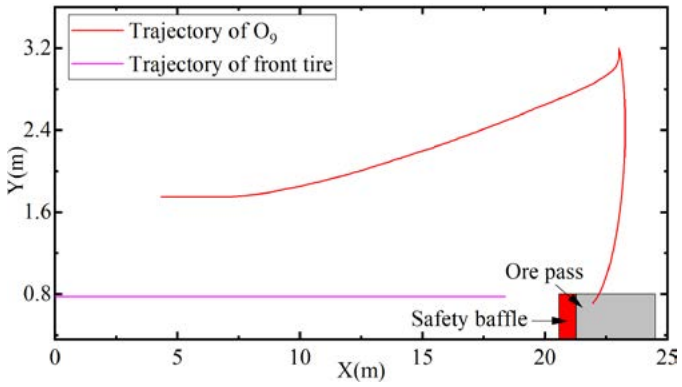


Figure 11: Bucket tip trajectory results of RLPC

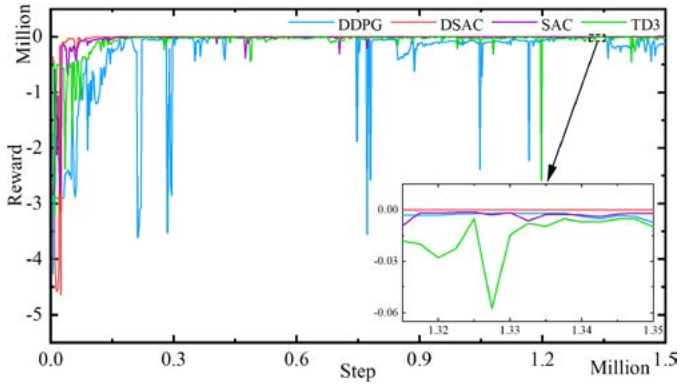


Figure 12: Reward results of different algorithm.

In order to explore the performances of different RL algorithms, trains of different algorithms, including DDPG, TD3, SAC, and DSAC, were performed based on the same dump operation and RLPC. For the train process, the same environment was used, and same values were used for the same parameters.

The reward of different algorithms was shown in Fig. 12. The maximum train step is 1.5 million. As it can be seen, DSAC and SAC show a faster increase trend at the start stage, and the TD3 is also fast than DDPG. The DSAC and SAC also show a more stable increase trend during the middle stage. In addition, the detailed figure also indicated that the DSAC shown higher reward during the whole process, which indicated that it given better comprehensive performance during the dumping operations.

4.2. Test results and discussion

In order to validate the proposed methods further, the test of dumping operation was performed with the aforementioned test platform. The system was implemented with the control policy trained with DSAC. The velocity and longitudinal location results were shown in Fig. 13 (a) and (b), in which the output of RL controller was marked as "Sim", and the actual results of LHD was marked as "Test".

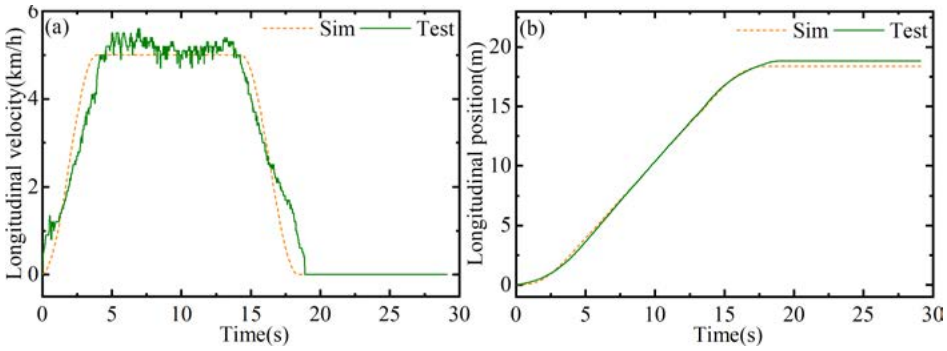


Figure 13: Velocity and longitudinal distance test results of RL-based method.

As it can be seen, the proposed method could control LHD run to target ore pass location. The max deviation during running process was about 0.2 m. The last stop point in the test was about 20.0 m. The error of longitudinal control was less than 1.2%. It was less than the error threshold for engineering application, i.e., 5%. The fluctuation of the velocity aroused during the acceleration and deceleration phases, as shown in Fig. 13. The actual velocity results indicated that the proposed method could guarantee LHD run as expected, although more fluctuation occurred in the actual velocity result. The maximum velocity error was about 0.33 m/s. The longitudinal result validated the drive system control performance of the proposed method.

The results of lift and dump cylinder displacement, and the attitude angle of bucket, γ, are shown in Fig. 14. As it can be seen, the dump cylinder kept static during the driving stage, 0 - 20 s. The displacement of dump cylinder at this stage was about 184.3 cm. The attitude angle of bucket varied slightly during 44.5° and 39.8° and the variation was about 4.7°. The lift operation started at 4.0 s, the LHD was still run with the 5.0 km/h velocity. The dump operation was performed at 21.4 s after the LHD stopped and

the lift cylinder arrived at the target displacement. The results of boom angle shown that the lift and dump cylinder were controlled coordinately between 21.4s to 29.1s. The range of boom angle varies from 100.0° to 111.3°, while the displacement of dump cylinder changed between 184.3 cm to 130.7 cm. The final attitude angle was about -52.1°, which guarantee the dump of ore materials. Therefore, it can be concluded that the proposed method, DRLC, could finish the dump operation autonomously and coordinately.

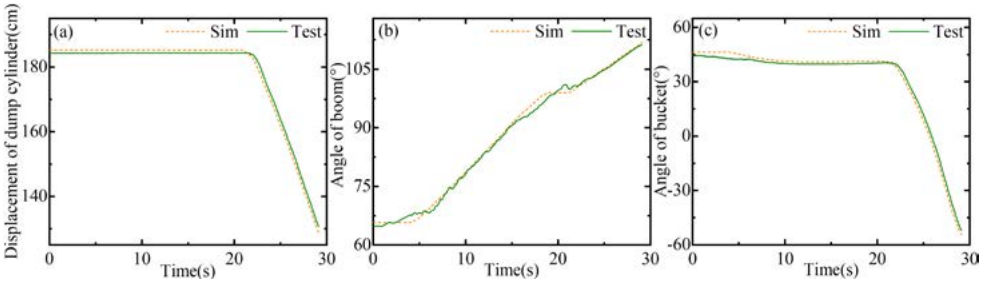


Figure 14: Cylinder displacement test results of RL-based method.

In order to further analyze the dump operation during test, the trajectory of bucket tip in the LHD frame is shown in Fig. 15. As it can be seen, the start and stop point of test and reference almost coincided with each other, although the difference occurred during the internal process. The final resulting bucket tip position was about 22.5 m in longitudinal direction, and 0.52 m in vertical direction. Therefore, it can be concluded that the proposed method could control the bucket to finish the dump operation safety. The whole dump operation was finished within 29.1 s according to the results.

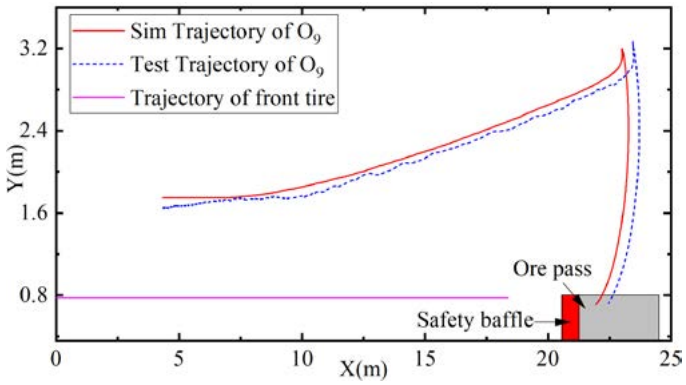


Figure 15: Trajectory of bucket tip in test.

Based on the above results, the proposed method in this paper could finish the dump operation with 29.1 s. The coordinate control in the proposed method guarantee the better state of bucket in lift process. The variation of attitude angle was about 5°, and meet the error threshold for engineering application. The less variation meant that less ore materials was scattered during lift process. The appropriate stop point result also indicates that the method could guarantee collision safety during the dump motion.

Therefore, it could be concluded that the method could finish dump operation with excellent safety, efficiency, and environmental adaptability.

5. CONCLUSION

This paper proposed an autonomous dump control method based RL frame for LHD under constrained environment. The involved dynamic and kinematic model during dump operation were established and validated with a LHD test platform. The kinematic model errors in terms of longitudinal, vertical position, and attitude angle were 0.2%, 0.5%, and 1.1%, respectively, and the longitudinal dynamic model error was about 4.2 %. The corresponding design details about the duty-based method (DRLC) was presented, in which the constraints of driving system and working mechanism actuators, and collision safety were considered, while also accounting for the completion status of the dumping operation. The effect of different RL algorithms was explored for dump occasion, in which the DSAC shown better performance. The field tests of autonomous dump operation with DRLC was performed. The results shown that the proposed method could finish the dump operation with 29.1 s under the premise of ensuring safety and efficiency. Therefore, the proposed method could finish dump operation with excellent safety, efficiency, and environmental adaptability.

6. REFERENCES

[1] D. Cardenas, P. Loncomilla, F. Inostroza, I. Parra-Tsunekawa, J. Ruiz-del-Solar, Autonomous detection and loading of ore piles with load–haul–dump machines in Room & Pillar mines, J. F. Robot. 40 (2023) 1424–1443.

[2] S. Dadhich, F. Sandin, U. Bodin, U. Andersson, T. Martinsson, Field test of neural-network based automatic bucket-filling algorithm for wheel-loaders, Autom. Constr. 97 (2019) 1–12.

[3] C. Borngrund, F. Sandin, U. Bodin, Deep-learning-based vision for earth-moving automation, Autom. Constr. 133 (2022) 104013.

[4] R. Hoeger, H. Zeng, A. Hoess, HAVEit: The future of driving. Final Report., (Deliverable D61.1) (2011) 358.

[5] A. Hemami, F. Hassani, An Overview of Autonomous Loading of Bulk Material An Overview of Autonomous Loading of Bulk Material, (2015).

[6] R. Filla, M. Obermayr, B. Frank, A study to compare trajectory generation algorithms for automatic bucket filling in wheel loaders, 3rd Commer. Veh. Technol. Symp. (2014) 588–605.

[7] Y. Meng, H. Fang, G. Liang, Q. Gu, L. Liu, Bucket trajectory optimization under the automatic scooping of LHD, Energies 12 (2019) 1–18.

[8] K.M. Elbayomy, Z. Jiao, H. Zhang, PID controller optimization by GA and its performances on the electro-hydraulic servo control system, Chinese J. Aeronaut. 21 (2008) 378–384.

[9] R. Song, Z. Ye, L. Wang, T. He, L. Zhang, Autonomous Wheel Loader Trajectory Tracking Control Using, 2022 Am. Control Conf. (2022) 2063–2069.

[10] D. Eriksson, R. Ghabcheloo, Comparison of machine learning methods for automatic bucket filling: An imitation learning approach, Autom. Constr. 150 (2023) 104843.

[11] J. Duan, Y. Guan, S.E. Li, Y. Ren, Q. Sun, B. Cheng, Distributional Soft Actor-Critic:

Off-Policy Reinforcement Learning for Addressing Value Estimation Errors, IEEE Trans. Neural Networks Learn. Syst. 33 (2022) 6584–6598.

Biographies



Lulu Gao received the B.Tech. and Ph.D. degree in school of mechanical engineering from University of Science and Technology Beijing, Beijing, China, in 2014 and 2021. From February 2021 to July 2023, he was a postdoctoral with the school of mechanical engineering from University of Science and Technology Beijing. His current research interests include motion control of hydraulic power system, reinforcement learning application, intelligent mining heavy-duty vehicles.



Zihang Zhang received the bachelor's degree in vehicle engineering, in 2022, from the University of Science and Technology Beijing, where he is currently working toward the master's degree in vehicle engineering. His research areas include control of underground Load-Haul-Dump vehicle working mechanisms, simulation of hydraulic power systems, and reinforcement learning.



Prof. **Katharina Schmitz** studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz’s awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Energy efficiency evaluation of hybrid distributed-centralized load sensing system for electrified excavator

Hongyun Mu[a], Min Cheng[a,*], Xiongfeng Tang[a]

a State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, China

Abstract. A hybrid distributed-centralized load sensing (HDC-LS) system is a potential solution for reducing energy consumption in electric construction machinery. It balances load differences by establishing a hybrid distributed-centralized actuator (HDCA) through the parallel connection of a distributed actuator alongside the original hydraulic cylinder. The characteristic of the excavator actuator, which is highly involved in compound motions, is exploited to equip the arm with the HDCA in this study. A parameter design approach is proposed that fully considers multiple HDCA factors, including load balance capacity, nominal power, and installation space. Compared with the conventional EHA, the integrated HDCA reduces the volume and mass by 43.9% and 27.3%, respectively, for a nominal power of 5 kW. Simulation results for a 6 T electrified excavator show that battery energy consumption is reduced by 2.2% – 15.6% for HDCA nominal powers of 3 – 7 kW, compared with the conventional load sensing (LS) system.

Keywords. Electric excavator, Load sensing, Load differences, Integrated, Installation space, Energy efficiency evaluation.

1. INTRODUCTION

With the international community increasingly focusing on policies of carbon neutrality and carbon peaking, the development of new, highly efficient construction machinery has become more urgent in recent years [1]. The load sensing (LS) system matchs pressure by controlling the pump outlet pressure to be one pressure margin higher than the load pressure. Offering advantages such as simple and reliable operation control, it is widely used in small and medium-sized excavators. However, the system exhibits substantial energy losses that originate from two dominant mechanisms.

(1) Control valve losses. These losses are termed meter-in (upstream) or meter-out (downstream) throttling losses, depending on their location relative to the actuator. The former can be partially reduced by variable pressure margin control, but this approach requires extra sensors and signal-conditioning electronics [2]. The latter can be addressed by potential energy recovery, but this requires additional energy-conversion components, the implementation of which is often constrained by space limitations. The independent metering control system is an effective solution for reducing these throttling losses [3]. However, the excessively low pressure differential across the valve orifice can easily trigger oscillations, resulting in poor system stability [4].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

(2) Load differences loss. Previous studies have tackled this problem through four mainstream approaches. 1) The hydraulic transformers. Due to their complex structure, production costs remain high, posing significant challenges to commercialization [5]. 2) The multi-pressure sources. These systems require multiple components (e.g., pumps, accumulators, switching valves), leading to complex structures, high costs, and large space requirements [6]. 3) The multi-chamber cylinder. When the actuator load changes, frequent mode switching by the control valve is required to select a new working chamber. This results in pressure shocks and poor system controllability [7]. 4) The distributed control systems. Purely distributed systems suffer from high installed power and large installation space.

To address the issue of low energy efficiency caused by load difference, a hybrid distributed-centralized load sensing (HDC-LS) system was proposed in our previous work [8]. The results obtained from an excavator's boom showed advantages of high energy efficiency and ease of implementation. In this study, the parameter design criterion for the HDCA is further discussed to systematically consider its key factors, including load balance capacity, nominal power, and installation space. A parameter design approach is proposed for the HDCA and the control valve to make tradeoffs among nominal power, installation space, and energy-saving capacity. Furthermore, an integrated HDCA with a four-chamber hydraulic cylinder is designed and validated on a 6 T electrified excavator under different nominal powers.

2. LAYOUT OF THE HDC-LS SYSTEM

As shown in Figure 2.1, the basic mechanism of the HDC-LS system is to establish an HDCA by connecting an electro-hydrostatic actuator (EHA) (distributed actuator) in parallel alongside the original hydraulic cylinder (centralized actuator). The four-quadrant operation of the EHA is exploited to balance the load difference. When the HDCA is the highest-load actuator, the EHA applies an auxiliary force to adjust the meter-in chamber pressure down to that of the second-highest actuator, thereby reducing LS pump energy consumption. Conversely, when the HDCA is not the highest-load actuator, the EHA applies a resistive force to achieve pressure matching, thereby recovering the lost energy of the pressure compensator. Furthermore, the EHA can also recover the potential energy under overrunning loads during load balancing. The principle of the EHA controller is shown in Figure 2.1 (d), with details referenced in our previous work [8]. In this study, the case of the HDCA installed on the excavator's arm is considered to discuss a design criterion. There are three reasons as follows.

(1) The nominal power of the HDCA should be high enough to achieve full load balance and energy recovery under different conditions. The arm experiences more complex four-quadrant loads than the boom. Therefore, accurate and uniform evaluation of the load difference under compound motion is necessary.

(2) The HDCA includes many components such as an electric motor, a valve block, and a hydraulic pump. The increased mass has a significant impact on the required force and, thereby, on the energy consumption.

(3) The HDCA is sensitive to size constraints when installed on the arm of the excavator or other machinery. The HDCA should be designed carefully within limited installation space to avoid structural interference.

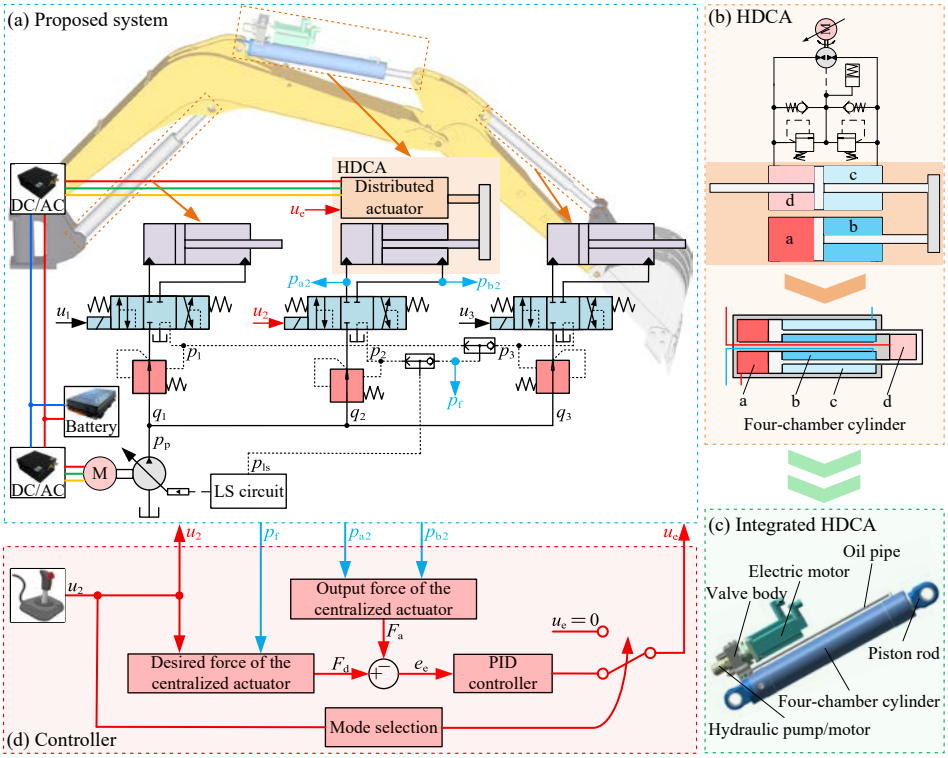


Figure 2.1. Working principle of the HDC-LS system

3. PARAMETER DESIGN APPROACH

The flow diagram of the parameter design is shown in Figure 3.1. For the distributed actuator, the key criterion for parameter design is to identify the relationship between EHA nominal power (HDCA nominal power) and energy-saving capability. The load forces of three actuators during the excavating cycle are obtained through load characteristic modelling. The EHA output force is calculated assuming fully balanced load differences, and its probability distribution is derived. A power coverage ratio is calculated by combining the probability distribution of EHA velocity, and this ratio is used as an index for evaluating energy-saving capability. For the centralized actuator, the meter-out orifice of the control valve is redesigned based on the selected HDCA nominal power to compensate for loads exceeding the nominal power.

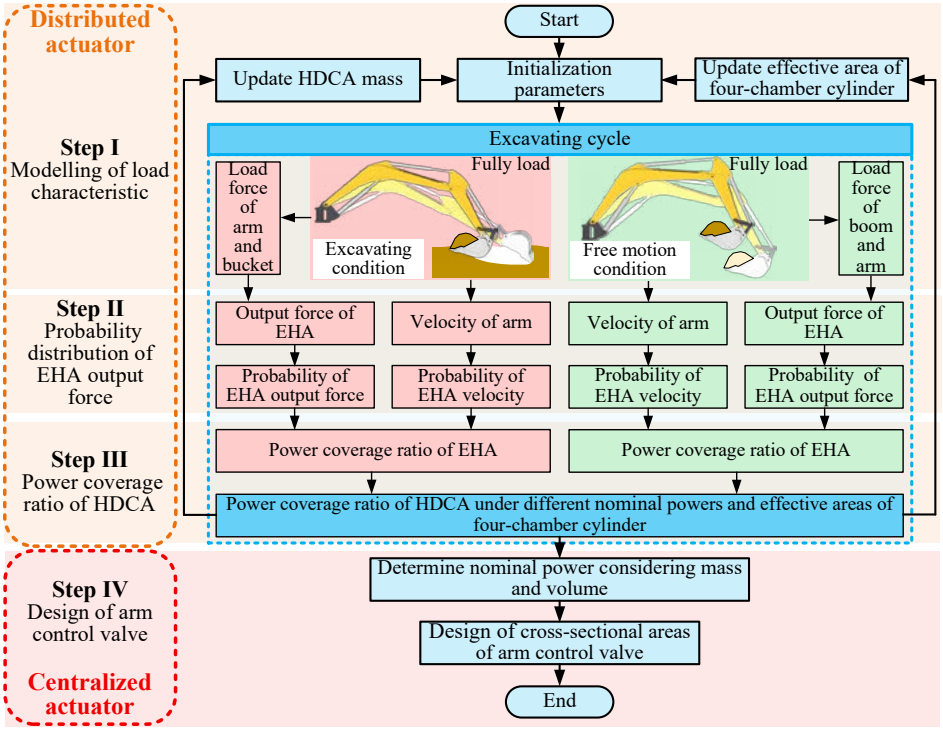


Figure 3.1. Flow diagram of the parameter design

3.1. Distributed actuator

3.1.1. Modelling of load characteristic

A typical excavating cycle is shown in Figure 3.2(a), which consists of four working steps: digging, lifting, dumping, and lowering. To evaluate the energy-saving capability of the HDCA, the excavating cycle is divided into two conditions: soil interaction condition (Figure 3.2(b)) and free motion condition (Figure 3.2(c)).

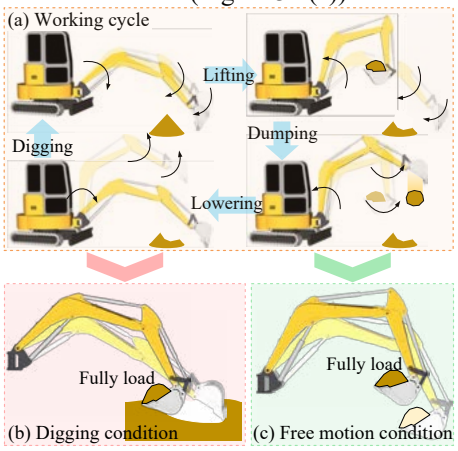
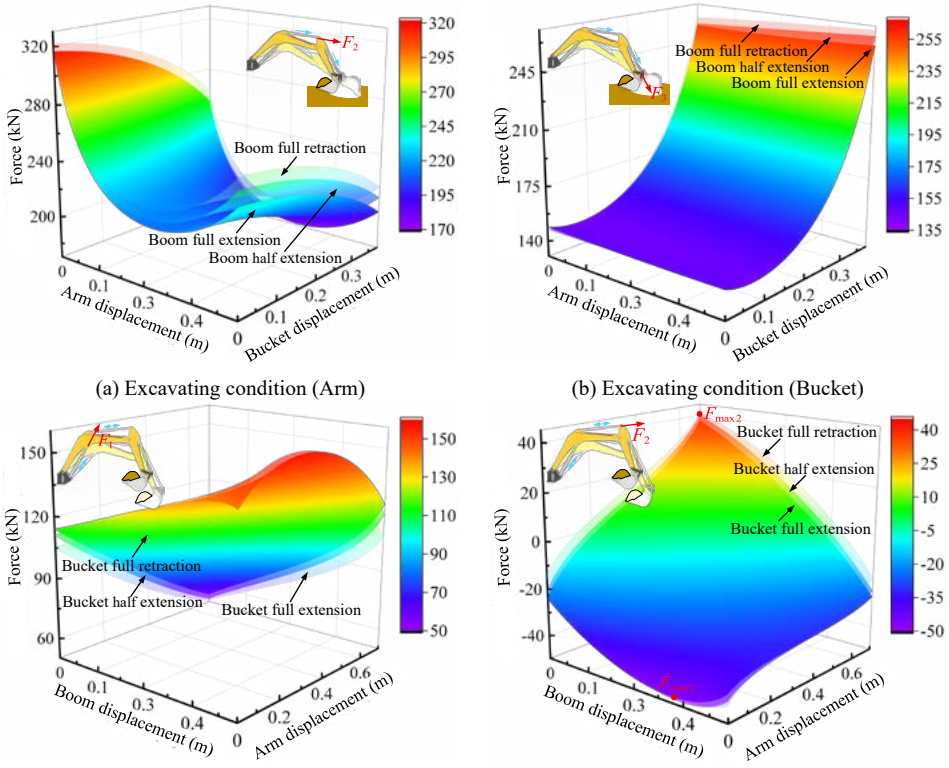


Figure 3.2. Typical excavating cycle and its two conditions

1) Full load excavating condition. The arm and bucket cylinders typically extend in a compound motion for soil excavation, while the boom has a rare or small movement. The digging force is applied to the bucket tip by the soil. Specifically, the mass of the HDCA impact on energy consumption is considered in the modelling process. Furthermore, to guarantee that the load forces represent the operating conditions of the arm and bucket cylinders across various postures, the modelling process considers different boom displacements. 2) Full load free motion condition. The boom, arm, and bucket cylinders perform a compound motion to complete the soil dumping. Similarly, the effects of different bucket displacements on the load forces are considered. During the excavating condition, static equilibrium balances the excavating force and manipulator gravity against the hydraulic cylinder output. During the free-motion condition, it balances the combined gravity of the soil and manipulator against the hydraulic cylinder output. The load force of the actuators is shown Figure 3.3.



(a) Excavating condition (Arm)

(b) Excavating condition (Bucket)

(c) Free motion condition (Boom)

(d) Free motion condition (Arm)

Figure 3.3. Load forces of the actuators

3.1.2. Probability distribution of EHA output force

The four-chamber cylinder structure is shown in Figure 3.4. To ensure the digging force of the arm, the effective area A_{a2} of chamber "a" in the four-chamber cylinder of the centralized actuator must not be less than that of the original cylinder. 1) Full load excavating condition. When the excavator is in excavating condition, both the arm and bucket actuators are typically extended. Ignoring friction, the steady state force equilibrium equation for the arm and bucket can be expressed as:

$$\begin{cases} F_2 = (p_{a2}A_{a2} - p_{b2}A_{b2}) + F_{Eha} \\ F_3 = p_{a3}A_{a3} - p_{b3}A_{b3} \end{cases} \tag{3.1}$$

where F_2 and F_3 represent the load forces on the arm and bucket, respectively (positive direction is defined as cylinder extension); F_{Eha} indicates the output force of the EHA; p_{a2} and p_{b2} denote the pressure in chambers "a" and "b"; A_{b2} is the effective area in chamber "b"; p_{a3} and p_{b3} are the head-side and rod-side pressures of the bucket cylinder; and A_{a3} and A_{b3} represent the corresponding effective areas. Here, p_{b2} with the meter-out orifice fully open for potential energy recovery. p_{b3} can be derived from the orifice equation, which is:

$$q = C_d A_d \sqrt{\frac{2\Delta p}{\rho}} \tag{3.2}$$

where q represents the flow rate; C_d indicates the flow coefficient; ρ denotes the oil density; A_d is the cross-sectional area; and Δp represents the differential pressure. Therefore, p_{b2} and p_{b3} can be expressed as:

$$\begin{cases} p_{b2} = 0, & \text{Arm extends} \\ p_{b3} = \dfrac{\rho q_{b3}^2}{2C_d^2 A_{db3}^2}, & \text{Bucket extends} \end{cases} \tag{3.3}$$

where q_{b3} and A_{db3} represent the flow rate and cross-sectional area of the bucket control valve, respectively. When the EHA adjusts the load pressure of the centralized actuator so that it is consistent with the load-pressure feedback from other actuators, the load differences are fully balanced, that is:

$$p_{a2} = p_f = p_{a3} \tag{3.4}$$

where p_f represents load pressure feedback from other actuators. From Eqs. (3.1) and (3.4), F_{Eha} can be inferred as:

$$F_{Eha} = F_2 - \frac{F_3 + p_{b3}A_{b3}}{A_{a3}} A_{a2} \tag{3.5}$$

From Eq. (3.5), F_{Eha} can be calculated for different displacements of the arm and bucket, as shown in Figure 3.5. 2) Full load free motion condition. Similarly, F_{Eha} can be obtained according to the direction of motion of the boom and arm, as shown in Figure 3.6.

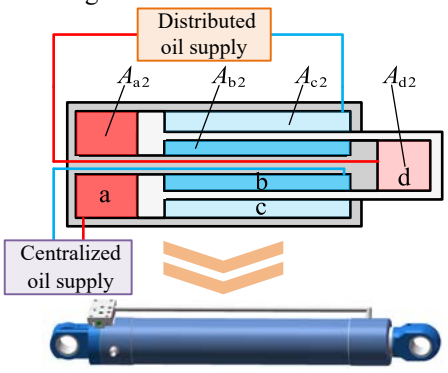


Figure 3.4. Structure of the four-chamber cylinder

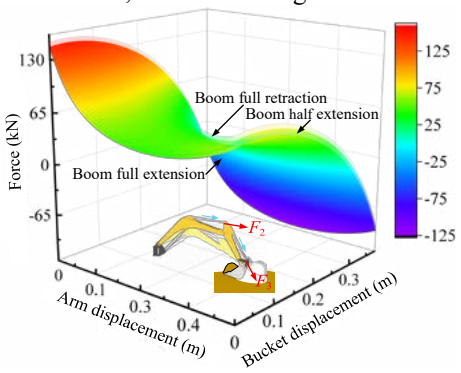
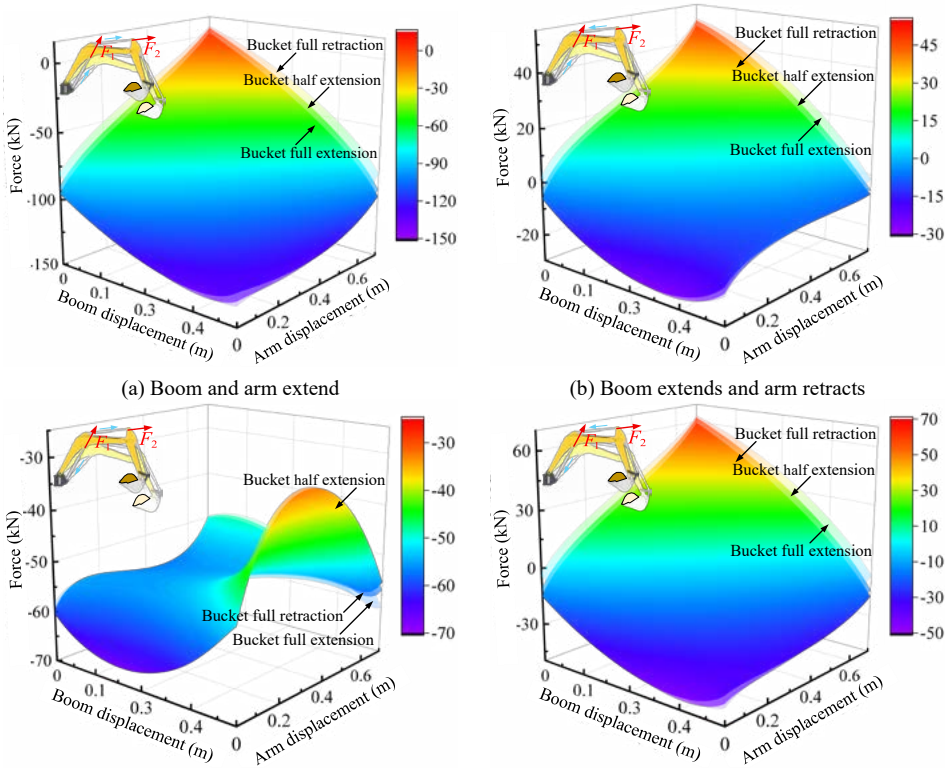


Figure 3.5. Output force of the EHA under excavating condition



(a) Boom and arm extend

(b) Boom extends and arm retracts

(c) Boom retracts and arm extends

(d) Boom and arm retract

Figure 3.6. Output force of the EHA under free motion condition

3.1.3. Power coverage ratio of HDCA

As shown in Figure 3.5 and Figure 3.6, F_{Eha} is stochastic under different motion conditions. To better describe this complex load characteristic, its probability distribution is calculated for the typical excavating cycle by uniformly segmenting F_{Eha} several zones. The probability distribution and segmenting can be expressed as:

$$\begin{cases} \begin{cases} P_{\mathrm{FDig}} = \dfrac{n_i}{n_{11}} \cdot 100\%, \\ F_{\mathrm{D1}} + \dfrac{F_{\mathrm{U1}} - F_{\mathrm{D1}}}{z_1}(i-1) \le F_{\mathrm{Eha}} < F_{\mathrm{D1}} + \dfrac{F_{\mathrm{U1}} - F_{\mathrm{D1}}}{z_1} i, \end{cases} & i = 1,2,3,\ldots,z_1 \quad \text{Excavating condition} \\ \begin{cases} P_{\mathrm{FFre}} = \dfrac{n_j}{n_{21}+n_{22}+n_{23}+n_{24}} \cdot 100\%, \\ F_{\mathrm{D2}} + \dfrac{F_{\mathrm{U2}} - F_{\mathrm{D2}}}{z_2}(j-1) \le F_{\mathrm{Eha}} < F_{\mathrm{D2}} + \dfrac{F_{\mathrm{U2}} - F_{\mathrm{D2}}}{z_2} j, \end{cases} & j = 1,2,3,\ldots,z_2 \quad \text{Free motion condition} \end{cases} \tag{3.6}$$

where P_{FDig} and P_{FFre} represent the probabilities within the output force of segment i and j, respectively; F_{U1}, F_{D1}, F_{U2} and F_{D2} indicate the upper and lower limits; z_1 and z_2 denote the number of segments; n_i and n_j are the number of sample points within each segment; and n_{11} and n_{21} represent the numbers of sample points for F_{Eha} Figure 3.5 and Figure 3.6 (each actuator displacement step corresponds to a sample point). During each motion phase, the velocities of the arm actuator typically tend toward a normal distribution. Therefore, the probability β_{v} at the velocity v_{Eha} can be expressed as:

$$\beta_{\mathrm{v}}=\frac{\mathrm{e}^{-\frac{\left(v_{\mathrm{Eha}}-\mu\right)^{2}}{2\sigma^{2}}}}{\sqrt{2\pi\sigma}}\cdot 100\% \tag{3.7}$$

where μ and σ represent the mean and standard deviation, respectively. The total number of the EHA power sample points is defined as w. In each segment, m output-force samples are calculated based on the probability distribution, and m velocity samples are generated randomly according to the velocity distribution. Considering the efficiencies η_{Pum} and η_{Cyl} of the hydraulic pump and cylinder, the power P_{Eha}^{k} corresponding to each sample point for the distributed side can be expressed as:

$$P_{\mathrm{Eha}}^{k}=\begin{cases}\dfrac{F_{\mathrm{Eha}}^{k}v_{\mathrm{Eha}}^{k}}{\eta_{\mathrm{Pum}}\eta_{\mathrm{Cyl}}}, & \left(F_{\mathrm{d}}^{k}v_{\mathrm{d}}^{k}\geq 0\right)\ \text{Under resistive condition}\\ F_{\mathrm{Eha}}^{k}v_{\mathrm{Eha}}^{k}\eta_{\mathrm{Pum}}\eta_{\mathrm{Cyl}}, & \left(F_{\mathrm{d}}^{k}v_{\mathrm{d}}^{k}<0\right)\ \text{Under overrunning condition}\end{cases},\ k=1,2,\cdots,w \tag{3.8}$$

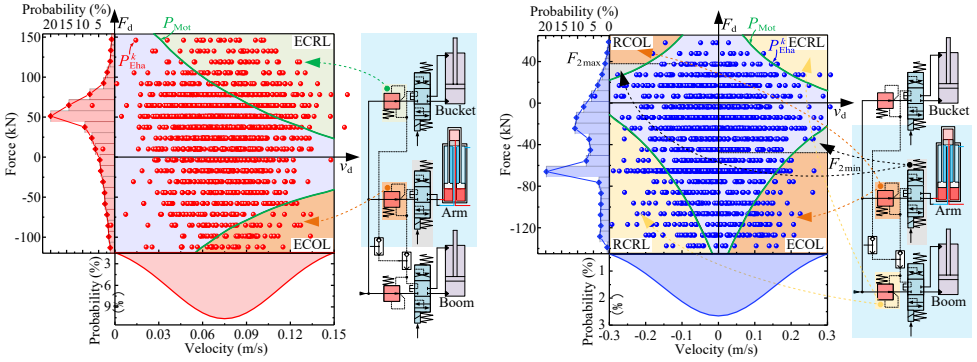
where F_{Eha}^{k} and v_{Eha}^{k} represent the output force and velocity corresponding to each sample point, respectively. To achieve better load balance, a higher nominal power of the EHA is preferable to cover all P_{Eha}^{k}, but is limited by the installation space of the arm. To evaluate the load balance capacity, the power coverage ratio of the HDCA is introduced as an evaluation index. The schematic of the power coverage ratio under different motion conditions is shown in Figure 3.7. Here, the red and blue points represent power sample points P_{Eha}^{k}, and the green curve is the nominal power P_{Mot} of the EHA electric motor. P_{Eha}^{k} are distributed according to the joint probability distribution of the velocity and the output force of the distributed side. The light blue area represents the sufficient nominal power ($P_{\mathrm{Eha}}^{k}\leq P_{\mathrm{Mot}}$), while other colored areas represent the insufficient power ($P_{\mathrm{Eha}}^{k}>P_{\mathrm{Mot}}$). The EHA operates in two-quadrant modes during the soil interaction condition: extension condition with resistive load (ECRL) and extension condition with overrunning load (ECOL). When $P_{\mathrm{Eha}}^{k}>P_{\mathrm{Mot}}$ (light green and orange regions in Figure 3.7(a)), the excess power is consumed by the pressure compensator (ECRL mode: bucket compensator; ECOL mode: arm compensator). During the free motion condition, the EHA operates in four-quadrant modes: ECRL, ECOL, retraction condition with overrunning load (RCOL), and retraction condition with resistive load (RCRL). Similarly, when $P_{\mathrm{Eha}}^{k}>P_{\mathrm{Mot}}$ (yellow, orange, and black regions in Figure 3.7(b)), the excess power is consumed by the pressure compensator or control valve. Under ECRL and RCRL modes, it is consumed by the boom compensator. Under RCOL and ECOL modes, the arm control valve takes priority and consumes the excess power, while the arm pressure compensator compensates for the insufficient power. Therefore, the power coverage ratio P_{Cov} of the EHA can be expressed as:

$$P_{\mathrm{Cov}}=\frac{w_{1}}{w}\cdot 100\% \tag{3.9}$$

where w_{1} presents the number of sample points ($P_{\mathrm{Eha}}^{k}\leq P_{\mathrm{Mot}}$). The distributions of the excavating condition and the free motion condition are determined to be 1/3 and 2/3, respectively, through statistical analysis of multiple typical excavating cycles. Therefore, the power coverage ratio P_{Cyc} of the HDCA under excavating cycle can be expressed as:

$$P_{\mathrm{Cyc}}=\frac{P_{\mathrm{CovDig}}}{3}+\frac{2P_{\mathrm{CovFre}}}{3} \tag{3.10}$$

where P_{CovDig} and P_{CovFre} represent the power coverage ratios of the EHA for the excavating condition and free motion condition, respectively.



(a) Excavating condition (b) Free motion condition

Figure 3.7. Schematic of the power coverage ratio under different motion conditions

The power coverage ratio of the HDCA needs to consider the effects of its mass and the effective area of the four-chamber cylinder:

(1) HDCA mass. The mass of the HDCA can be equivalently transferred to the boom when the arm is equipped with the HDCA, as shown in Figure 3.8. Thus, the equivalent mass m_1' of the boom can be expressed as:

$$m_1' = m_1 + k_{\mathrm{L}} m_{\mathrm{Hdca}} \tag{3.11}$$

where k_{L} represents the coefficient; m_1 indicates the mass of the boom; and m_{Hdca} denotes the additional mass of the HDCA.

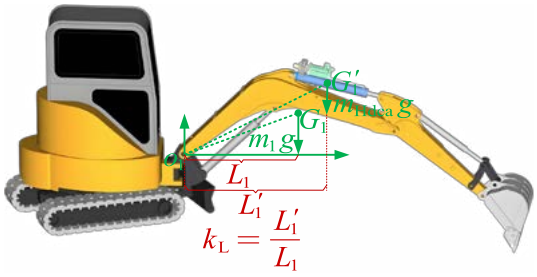


Figure 3.8. Equivalent mass of the boom

(2) Effective area of the four-chamber cylinder. To guarantee that the EHA can output maximum force within pressure limits in the EHA circuit, the effective areas A_{c2} and A_{d2} of the symmetrical chambers are selected to be half of A_{a2}.Therefore, the area coefficient k_{Are} can be described as:

$$k_{\mathrm{Are}} = \frac{2A_{\mathrm{b2}}'}{A_{\mathrm{a2}}}, \quad 0 < k_{\mathrm{Are}} < 1 \tag{3.12}$$

where A_{b2}' represents the new effective area of chamber "b" in the four-chamber cylinder.

The comparison of volume and mass between HDCA and conventional EHA (arm EHA with purely distributed drive in [9]) is shown in Figure 3.9. It can be observed that the percentage decrease in volume and mass of the HDCA gradually decreases with P_{Mot}. When P_{Mot} is 5 kW, the volume and mass are reduced by 43.9% and 27.3%, respectively. By combining Eqs. (3.1) – (3.12), P_{Cyc} for different values of P_{Mot} and k_{Are} can be calculated, as shown in Figure 3.10.

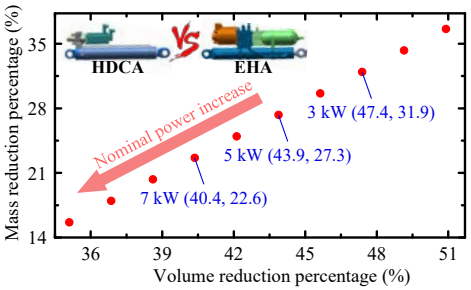


Figure 3.9. Comparison of volume and mass between HDCA and EHA

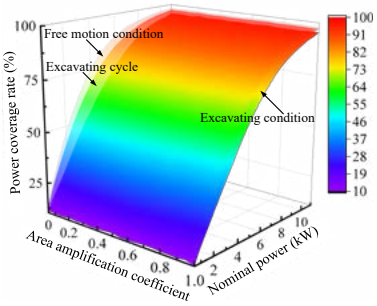


Figure 3.10. Power coverage ratio of the HDCA

3.2. Centralized actuator

When the arm is in overrunning mode, it is necessary to avoid the HDCA overspeed because of insufficient output force. Therefore, the cross-sectional area acquires to be redesigned. The energy-saving effects of the system are evaluated for $k_{\mathrm{Are}} = 0.43$ and P_{Mot} of 3 kW, 5 kW, and 7 kW. Correspondingly, P_{Cyc} is 56.8%, 76.2% and 87.3%, respectively, from Figure 3.10. Considering the limitations of the EHA components, the actual output force F_{EhaAct} can be expressed as:

$$F_{\mathrm{EhaAct}} = \min\left(\left|\frac{P_{\mathrm{Mot}}}{v_{\mathrm{Eha}}\eta_{\mathrm{Pum}}\eta_{\mathrm{Cyl}}}\right|, F_{\mathrm{Com}}\right), \quad \text{EHA under overrunning condition} \tag{3.13}$$

where F_{Com} represents the maximum output force limited by the EHA components. The back pressure p_{Bac1} and p_{Bac2} required at the control valve can be expressed as:

$$\begin{cases} p_{\mathrm{Bac1}} = \max\left(p_{\mathrm{s}}, \dfrac{F_{\mathrm{Min2}} - F_{\mathrm{EhaAct}}}{A_{\mathrm{b2}}}\right), & \text{Arm extends} \\ p_{\mathrm{Bac2}} = \max\left(p_{\mathrm{s}}, \dfrac{F_{\mathrm{Max2}} - F_{\mathrm{EhaAct}}}{A_{\mathrm{a2}}}\right), & \text{Arm retracts} \end{cases} \tag{3.14}$$

where F_{Min2} and F_{Max2} represent the maximum gravitational forces that the EHA needs to overcome in overrunning mode (shown in Figure 3.3(d)); and p_{s} indicates the minimum pressure to guarantee control performance. The flow rate q_{Bac1} and q_{Bac2} of the control valve can be expressed as:

$$\begin{cases} q_{\mathrm{Bac1}} = A_{\mathrm{b2}} v_2, & \text{Arm extends} \\ q_{\mathrm{Bac2}} = A_{\mathrm{a2}} v_2, & \text{Arm retracts} \end{cases} \tag{3.15}$$

Here, v_{Eha} can be expressed through the control valve opening k_{Ope} as follows:

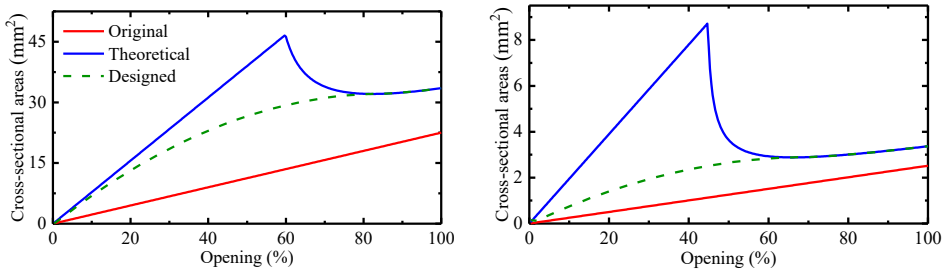
$$v_{\mathrm{Eha}} = k_{\mathrm{Ope}} v_{\mathrm{Max2}} \tag{3.16}$$

where v_{Max2} represents the maximum velocity of the arm. From Eq. (3.2), the designed cross-sectional area A_{d1} and A_{d2} can be calculated using the flow formula as follows:

$$\begin{cases} A_{\mathrm{d1}} = \dfrac{q_{\mathrm{Bac1}}}{C_d\sqrt{\frac{2(p_{\mathrm{Bac1}} - p_{\mathrm{T}})}{\rho}}}, & \text{Arm extends} \\ A_{\mathrm{d2}} = \dfrac{q_{\mathrm{Bac2}}}{C_d\sqrt{\frac{2(p_{\mathrm{Bac2}} - p_{\mathrm{T}})}{\rho}}}, & \text{Arm retracts} \end{cases} \tag{3.17}$$

where p_{T} represents the tank pressure ($p_{\mathrm{T}} = 0$). From Eqs. (3.13) – (3.17), A_{d1} and A_{d2} (blue curve) for different k_{Ope} values are obtained when P_{Mot} is 3 kW, 5 kW, and 7 kW, respectively, as shown in Figure 3.11. To reduce the difficulty of valve machining, A_{d1} and

A_{d2} are modified to obtain a smooth and monotonically increasing design curve (green curve).



(a) Arm retracts (b) Arm extends

Figure 3.11. Cross-sectional area of the arm control valve (5 kW)

4. SIMULATION STUDY

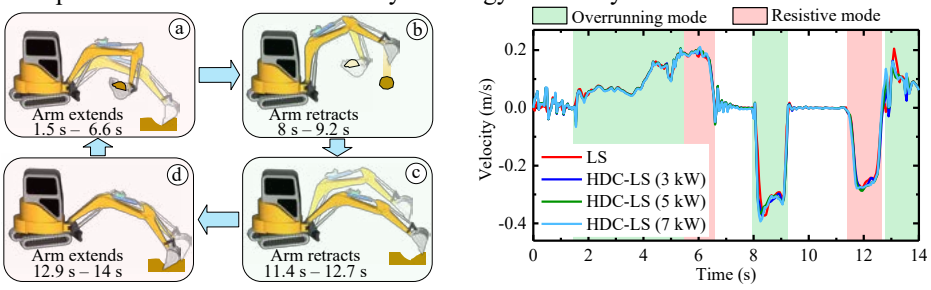
4.1. Simulation modelling

A 6 T electric excavator driven by a conventional LS system is taken as the reference prototype. A simulation model of this prototype is built in the AMESim environment. Based on the simulation model, the arm cylinder is replaced with an HDCA to establish the HDC-LS system. The models are simulated under the same excavation cycle to evaluate the energy-saving effects and the velocity performance at HDCA nominal powers of 3 kW, 5 kW, and 7 kW.

4.2. Comparative analysis

4.2.1. Motion performance

As shown in Figure 4.1(a), the HDCA enters the extend mode and the retract mode according to the arm control signal during a typical excavation cycle. As shown in Figure 4.1(b), during the entire motion process, the arm actuator maintains nearly identical velocities with smooth dynamic speed variations because the flow control method of the conventional LS system is retained. Therefore, the designed cross-sectional area of the control valve avoids overspeed dropping caused by insufficient HDCA nominal power. This guarantees the consistent motion profiles and evaluation accuracy of energy efficiency.

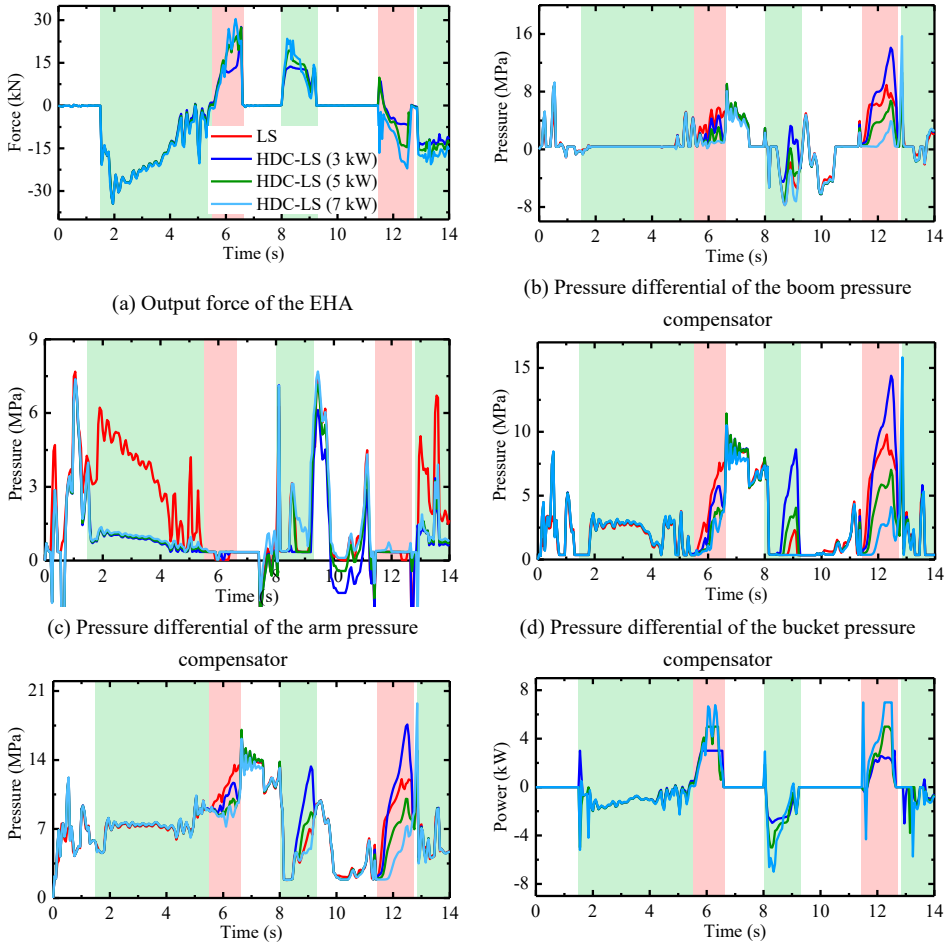


(a) Typical excavating cycle of the excavator (b) Arm

Figure 4.1. Excavating cycle and actuator velocities

4.2.2. *Energy consumption performance*

As shown in Figure 4.2, by controlling the EHA output force to adjust the pressures in chambers a and b of the four-chamber cylinder, the pressure differential and the energy losses across the pressure compensators are reduced. The output energy of the LS pump is reduced effectively since the EHA outputs an auxiliary force. When the EHA outputs a resistive force, it prioritizes recovering potential energy while balancing load difference.



(a) Output force of the EHA

(b) Pressure differential of the boom pressure compensator

(c) Pressure differential of the arm pressure compensator

(d) Pressure differential of the bucket pressure compensator

(e) Outlet pressure of the LS pump

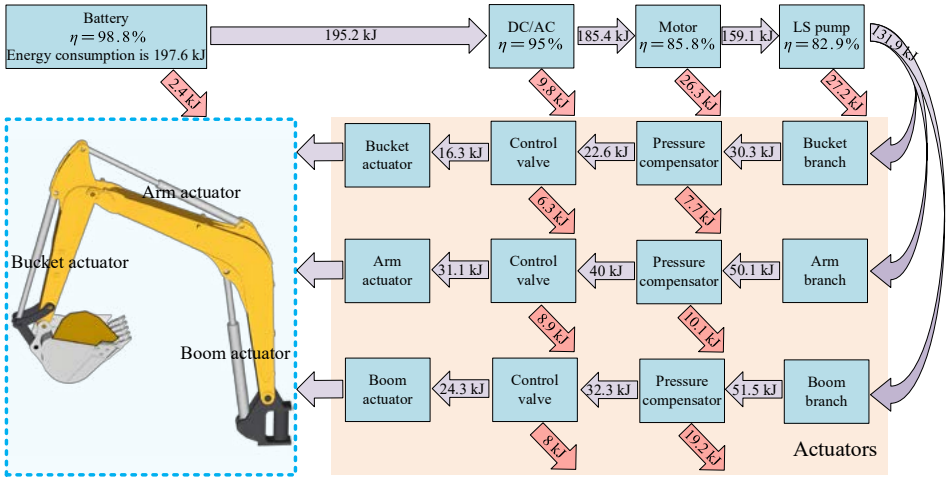
(f) Power of the EHA

Figure 4.2. Simulation results of the conventional LS and HDC-LS systems

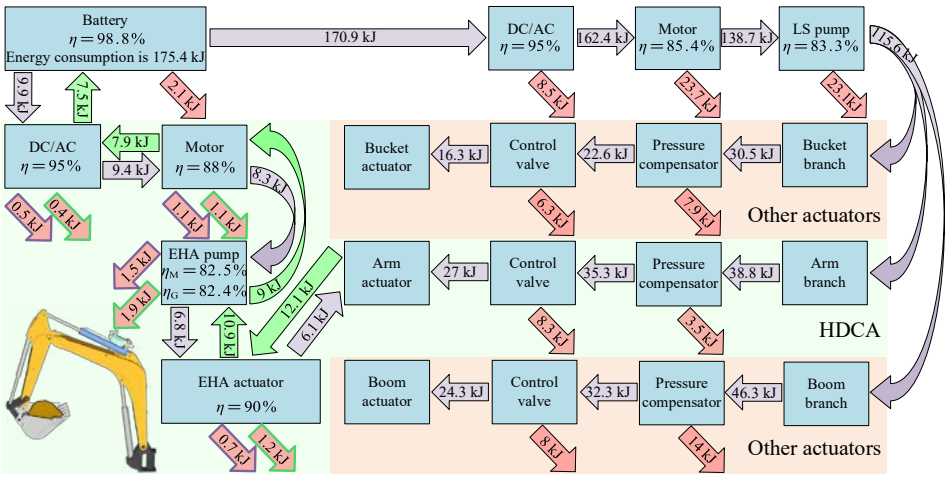
The system efficiency can be calculated as:

$$\begin{cases} \eta_{\mathrm{C}} = \dfrac{E_{\mathrm{Ct}} - E_{\mathrm{Cp}}}{E_{\mathrm{ct}}} \cdot 100\% \\ \eta_{\mathrm{P}} = \dfrac{E_{\mathrm{Pt}} - E_{\mathrm{Pp}}}{E_{\mathrm{Pt}}} \cdot 100\% \\ \eta_{\mathrm{S}} = \dfrac{E_{\mathrm{T}} - E_{\mathrm{P}}}{E_{\mathrm{T}}} \cdot 100\% \end{cases} \tag{4.1}$$

where η_C and η_P represent energy-saving efficiency of the pressure compensator and LS pump, respectively; η_S indicates the energy-saving efficiency; E_{Ct}, E_{Pt} and E_T denote the energy consumption of the pressure compensator, LS pump, and battery in the conventional LS system; and E_{Cp}, E_{Pp} and E_P are the energy consumption of the pressure compensator, LS pump, and battery in the HDC-LS system. Based on Eq. (4.1) and combined Figure 4.3, the system efficiencies for different HDCA nominal powers are calculated in Table 1. It can be observed that both the energy loss in the pressure compensator due to the load difference and the output energy by the LS pump are significantly reduced. As the nominal power of the HDCA increases, the load difference effect of the EHA becomes more pronounced, leading to increasing energy savings. At nominal powers of 3 kW, 5 kW, and 7 kW, the energy consumption of the battery is reduced by 2.2 %, 11.2 %, and 15.6 %, respectively.



(a) LS system



(b) HDC-LS system (5 kW)

Figure 4.3. Comparison of the energy flow for different layouts

Table 1 System efficiency for different HDCA nominal powers

Name	LS system	HDC-LS system (3 kW)	HDC-LS system (5 kW)	HDC-LS system (7 kW)
$E_{\mathrm{Ct}}/E_{\mathrm{Cp}}$	37 kJ (Boom 19.2 kJ) (Arm 10.1 kJ) (Bucket 7.7 kJ)	33.9 kJ (Boom 19.7 kJ) (Arm 3 kJ) (Bucket 11.2 kJ)	25.4 kJ (Boom 14 kJ) (Arm 3.5 kJ) (Bucket 7.9 kJ)	21.5 kJ (Boom 10.4 kJ) (Arm 4.1 kJ) (Bucket 7 kJ)
$E_{\mathrm{Pt}}/E_{\mathrm{Pp}}$	131.9 kJ	128.6 kJ	115.6 kJ	108.8 kJ
$E_{\mathrm{T}}/E_{\mathrm{P}}$	197.6 kJ	193.2 kJ	175.4 kJ	166.8 kJ
η_{C}	—	8.4%	31.4%	41.9%
η_{P}	—	2.5%	12.4%	17.5%
η_{S}	—	2.2%	11.2%	15.6%

5. CONCLUSION

In this study, a parameter design approach suitable for the HDCA when installed on the arm of electric excavator is proposed. Furthermore, the HDCA is designed to achieve an integrated drive. Compared with the conventional EHA, the integrated HDCA reduces the volume and mass by 43.9% and 27.3%, respectively, for a nominal power of 5 kW. A 6 T electric excavator simulation model of the HDC-LS system is established. The energy-saving effects of the HDC-LS system are evaluated under HDCA nominal powers of 3 kW, 5 kW, and 7 kW. The energy losses of the pressure compensator (boom + arm + bucket) are reduced by 8.4%, 31.4%, and 41.9%, respectively, through the EHA load balancing effect. This approach reduces battery energy consumption by 2.2%, 11.2%, and 15.6%. With the increase in HDCA nominal power, the energy-saving effect of the proposed system becomes increasingly significant.

6. REFERENCES

[1] T. Do, T. Dang, T. Dinh, K. Ahn, 'Developments in energy regeneration technologies for hydraulic excavators: a review', *Renew Sustain Energy Rev*, vol. 145, p. 111076, 2021.

[2] T. Lin, Y. Lin, H. Ren, H. Chen, Z. Li, Q. Chen, 'A double variable control load sensing system for electric hydraulic excavator', *Energy*, vol. 223, p. 119999, 2021.

[3] J. Chen, H. Jiang, X. Kong, C. Ai, 'Mode switching control of independent metering fluid power systems', *ISA Trans.*, vol. 158, pp. 735–748, 2025.

[4] L. Bachmann, J. Liu, J. Weber, "Investigation of the temperature influence on electrohydraulic valve control and presentation of a novel compensation approach for independent metering valve systems," in *Proc. ASME/BATH 2023 Symp. Fluid Power Motion Control*, Sarasota, FL, USA, p. V001T01A031, 2023.

[5] D. Fassbender, V. Zakharov, T. Minav, 'Utilization of electric prime movers in hydraulic heavy-duty mobile-machine implement systems', *Autom. Constr.*, vol. 132, p. 103964, 2021.

[6] M. Huova, M. Linjama, 'Energy efficient throttling control of a multi-pressure system using a genetic algorithm and model predictive control', *Proc. Inst. Mech. Eng. Part I: J. Syst. Control Eng.*, vol. 236, pp. 406–417, 2022.

[7] H. Bartlett, 'A symmetric multichamber hydraulic cylinder with variable piston area: An approach to compact and efficient electrohydrostatic actuation', *J. Mech. Des.*, vol. 143, p. 083501, 2021.

[8] H. Mu, M. Cheng, X. Tang, R. Ding, W. Ma, ‘A hybrid distributed-centralized load sensing system for efficiency improvement of electrified construction machinery’, *Energy*, vol. 314, p. 134123, 2025.
[9] R. Ding, Y. Jiang, F. Yang, L. Peng, Z. Liu, W. Ma, ‘Design and energy consumption analysis of a decentralized independent electro-hydraulic control system’, *Ain Shams Eng. J.*, vol. 16, p. 103759, 2025.

Biographies



Hongyun Mu was born in Zunyi, China, in 1996. He received the B.S. and M.S. degrees in mechanical engineering from Guizhou University, Guiyang, China, in 2019 and 2023, respectively. He has been working toward the Ph.D. degree in mechanical engineering with State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing, China. His research interests include electro-hydraulic control system and energy-saving technology of mobile machinery.



Min Cheng was born in Jingzhou, China, in 1987. He received the B.S. and Ph.D. degrees in mechatronics from Zhejiang University, Hangzhou, China, in 2009 and 2015, respectively. He is currently an Associate Professor with the State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, China. His research interests include electro-hydraulic control systems of mobile machinery, energy saving and motion control of hydraulic systems, and mechatronic systems design.



Xiongfeng Tang was born in Yiyang, China, in 2002. He received the B.S. degree in mechatronic engineering from Chongqing University, Chongqing, China, in 2023. He has been working toward the M.S. degree in mechanical engineering with State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing, China. His research interests include high-performance electro-hydraulic transmission and control.

Innovative fan drive solution for mobile machines

Paulo Teixeira[1], Pedro Condessa, Stefan Kock[2] and Martin Laube[3]

[1]Bosch Rexroth Ltda, Advanced Engineering External Gear Units, Pomerode, Brazil
[2]Bosch Rexroth AG, Advanced Engineering External Gear Units, Nürnberg, Germany
[3]Bosch Rexroth AG, Head of Product Area External Gear Units, Lohr a. Main, Germany
E-mail: paulo.teixeira3@boschrexroth.com.br, pedro.condessa1@gmail.com, stefan.kock@boschrexroth.de, martin.laube@boschrexroth.de

Abstract

Hydraulic fan drive systems offer advantages for thermal management in mobile machinery, particularly in construction and agricultural equipment applications. Compared to traditional mechanical or direct-drive electric systems, hydraulically actuated fans enable flexible installation, provide good control and reduced power consumption by decoupling fan speed from engine speed. In the design of hydraulic fan drive systems for mobile machinery, optimizing the design to be both compact and affordable is essential for meeting customer expectations. For the status quo, two independent valves are required to achieve both fan speed control and radiator cleaning functionality provided by fan reversal. Simplifying the hydraulic system by using one valve instead of two would enable for size and cost reduction. This paper presents a patented and innovative hydraulic fan drive solution incorporating a single integrated valve in an external gear motor to control both fan speed and reversing direction. The objective of this paper is to verify the feasibility of the one-valve solution by means of a 1D simulation model. Simulation results demonstrated that the one-valve solution, when integrated with its enhanced feed-forward control, delivers performance comparable to the current two-valve solution, particularly in dynamic response and maintaining consistent flow characteristics despite pump speed fluctuations. This provides compelling evidence that the one-valve solution offers a simplified, cost-effective, and dynamically robust alternative for future mobile hydraulic systems.

Keywords. Mobile hydraulics, Fan drive.

1. INTRODUCTION

Mobile machinery applications, such as construction and agricultural equipment, operate in competitive markets where cost efficiency and system compactness are critical factors in equipment design. These machines generate heat during operation, primarily from the engine and the hydraulic system itself. Thermal management is necessary for maintaining optimal operating temperatures and it is achieved by matching heat rejection with minimal energy consumption. The engine's heat is absorbed by coolant fluid, then carried to the radiator, where the heat is exchanged with ambient air [1]. For diesel engines of mobile

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

machinery, it is estimated that the power dissipated in the heat exchanger is 45% of the power output [2] and 30% of fuel energy [3]. To improve heat rejection, a cooling fan actively forces the air across the radiator. The faster it spins, the more air it pulls through the radiator, and the more engine power it consumes. Therefore, regulating the fan speed is required for efficient thermal management.

Fan drives can be mechanical, electric, or hydraulic. Mechanical systems, using pulleys and belts, often overcool due to fixed speed ratios. Electric drives suit low to medium power demands. Hydraulic fan drives are a common solution for heavy-duty off-road vehicles due to their power delivery, ability to control fan speed independent of engine speed and assembly flexibility [1]. This ability means that the fan is not wasting energy by rotating faster than it needs to, just because the engine is running at high speed. This work focuses on hydraulic fan drives.

Optimizing the design of hydraulic systems used in fan drive application to be compact, and affordable is essential for meeting customer expectations. It needs to be compact to fit in increasingly physically optimized machines and affordable to keep up to the market competitiveness [4]. However, the design should still meet functional and reliability performance. As one of the world's leading suppliers for drive and control technologies, Bosch Rexroth offers, among other things, hydraulics, electric drive and control technology, gear technology as well as linear motion and assembly technology. In the fan drive application, the company established a strong presence with its solutions featuring external gear motors and valves manifold. To maintain competitiveness, two strategic approaches were considered for product improvement.

As presented in Figure 1.1, the architecture of the hydraulic system for fan drive is defined by three subsystems: 1) hydraulic pump, usually powered by diesel engine, 2) hydraulic motor, which drives the fan, and 3) valves, when specific functions are necessary, which can be in an external manifold or integrated in the motor. Integrated options are increasingly used in the market because they optimize space, reduce the number of pipes and hoses, and decrease number of connections, which are potential leak points.

Different combinations of hydraulic pumps and motors are possible. In this paper, the focus is a system with external gear pump with external gear motor, both with fixed displacements. For this combination, several possible functions can be implemented by valves such as:

- fan drive speed control,
- radiator cleaning functionality provided by fan reversal,
- cavitation mitigation for the period when the pump stops, and the motor continues to rotate for a certain time due to the fan rotational inertia.

The need for each of the functions depends on the specification of the machine and its application.

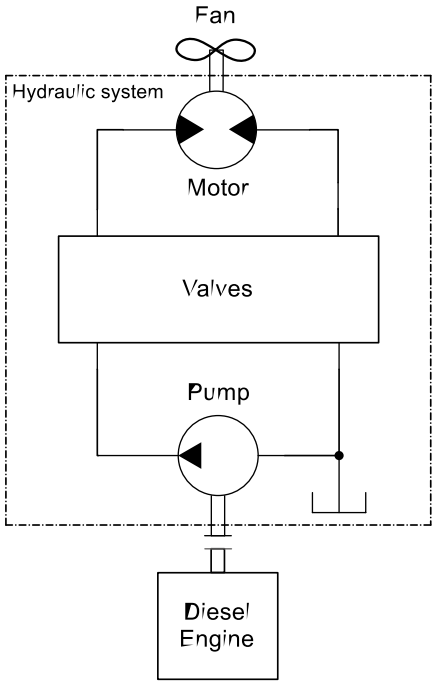


Figure 1.1 – Schematics of a generic hydraulic system for fan drive with fixed displacement pump and motor

In the status quo solution, available in the market, the speed control and the fan reversal is performed by two different hydraulic valves. This paper presents and patented and innovative solution that combines the two functions in one single valve proportional direction valve [5]. The novel concept is called then one-valve design and the status quo, two-valve design. Combining two functions in one valve allows for size and cost optimization.

The proportional directional valve is however sensible to input flow rate from the pump, which is a challenge. In mobile machinery the pumps are usually powered by diesel engines. If the diesel engine speed varies, the pump speed also changes and consequently the input flow rate. For the novel solution, this could cause then a variation in the fan speed, which is not desirable. To overcome this, a feedforward controller is proposed.

Although, the motivation behind the one-valve design is to reduce size and cost, in this paper the objective is to verify the feasibility, by means of 1D simulation, that the novel one-valve design with feedforward controller can have similar or better functional performance than two-valve design, which is the current status quo. Representative operation cycles derived from typical mobile applications will be defined and applied to the model to compare its performance to the two-valve solution.

In the following next section, it will be explained how the status quo design works. Later, a novel concept will be explored the by explaining the difference to status quo, the challenged raised by it, and how to overcome it. Then, the simulation model developed, and the

simulation results for the performance verification will be presented and discussed. Finally, the main findings and outline the next steps will be summarized.

2. STATUS-QUO SOLUTION

A hydraulic system for fan drive may be required to perform several distinct functions to meet the full range of operational demands. The specific system utilized for this investigation and presented in Figure 2.1 is designed with three primary capabilities delivered by valves: fan speed control for thermal management optimization, a reversing function for radiator cleaning, and an integrated anti-cavitation to protect the hydraulic motor.

In the status quo solution, the speed control is performed by proportional pressure relief valve (V1) and the fan reversal by on/off 4/2 directional valve (V2). The status quo solution using two independent valves (V1 and V2) inherently increases system size, and cost. Typically, an anti-cavitation valve is used for cavitation mitigation. In this paper no change is proposed for this function.

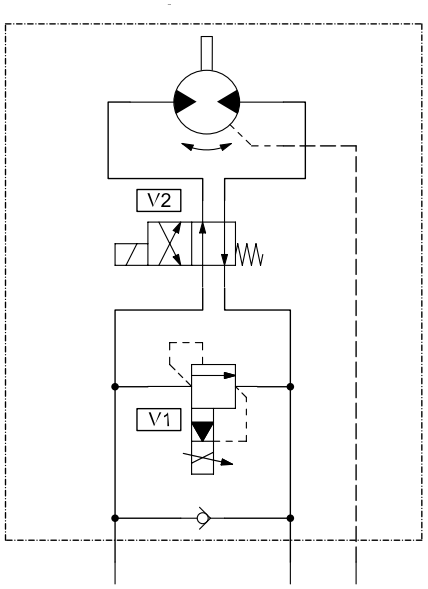


Figure 2.1. Hydraulic schematics of the status quo two-valve design (pump is omitted)

2.1. Speed control

Thermal management, which is crucial for maintaining optimal operating temperatures in mobile machinery, relies on cooling fans actively forcing air across the radiator. The rate of heat rejection is directly proportional to the volumetric airflow through the radiator, which in turn is directly proportional to the fan rotational speed. In pursuit of higher operational efficiency, the control system is designed to regulate the fan's rotational speed to the minimum level required by the current thermal conditions [1]. The logic of speed control using a proportional pressure relief valve is shown in Figure 2.2.

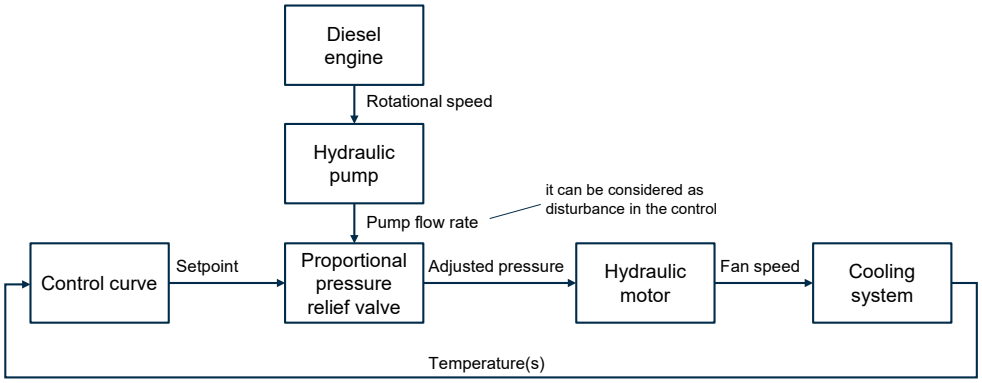


Figure 2.2. Speed control logic using proportional pressure relief valve

At lower temperatures, a reduced cooling airflow is needed, and therefore a lower fan speed is sufficient. The relationship between system temperature and required fan speed is typically defined by a control curve (often referred to as a "cooling curve") defined to meet machine's specific cooling requirements. Figure 2.3 illustrates an example of cooling curve. Fan can be set to standstill if the temperature is below a defined threshold [6].

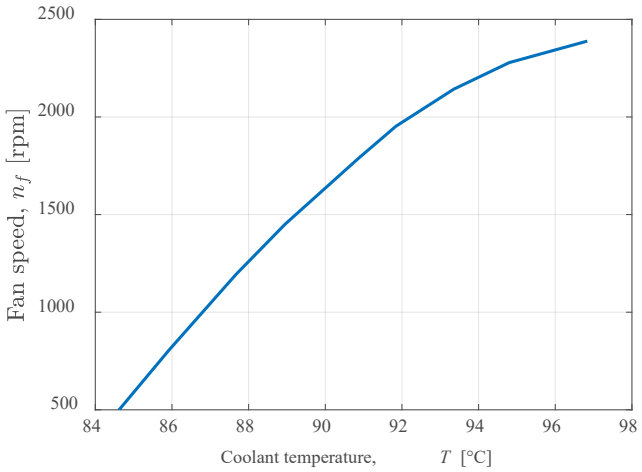


Figure 2.3. Typical fan speed control curve as a function of engine coolant temperature, adapted from [3]

The proportional pressure relief valve (V1 of Figure 2.1), which modulates system pressure in response to an electrical setpoint. Because the motor's torque is a direct function of the pressure differential across it, controlling the pressure effectively controls the torque. Since primary load on the motor is the fan's aerodynamic drag, it requires greater torque to achieve higher speeds. Consequently, a direct relationship is established where system pressure, as regulated by the valve, dictates the fan's rotational speed.

To ensure the engine does not overheat during an electrical failure, the system is designed to be fail-safe. If the control signal is lost, the fan automatically runs at maximum speed. This is achieved using negative control logic: as the electrical signal increases, the fan speed

decreases. Therefore, a zero signal commands maximum fan performance. As presented in Figure 2.4, the Bosch Rexroth KBVS.3B valve is a common example of a component that uses this principle in fan drive systems.

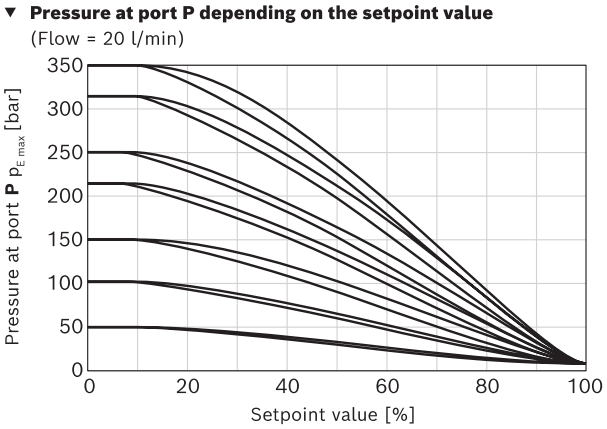


Figure 2.4. Pressure to setpoint relationship of KBVS.3B [7]

While a constant electrical setpoint to the proportional pressure relief valve should ideally maintain a constant system pressure and, therefore, a stable fan speed, the valve's inherent pressure-flow (P-Q) characteristic (Figure 2.5) causes a deviation from this ideal behaviour. As the pump's output flow increases – typically due to an acceleration of the diesel engine – a greater portion of this flow must be diverted through the valve. According to the P-Q curve, this results in in a change in system pressure even though the electrical setpoint remains fixed. This effect makes the fan speed to change under variations in pump speed. This inherent sensitivity of the status quo design will serve as the performance benchmark against the novel solution.

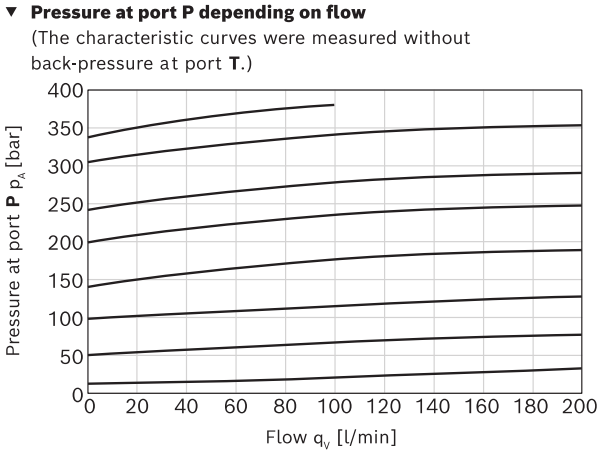


Figure 2.5. Pressure to flow relationship of KBVS.3B [7]

2.2. Reversal of speed

In many off-highway and agricultural applications, machinery operates in environments with high concentrations of airborne particulates, such as dust, chaff, and other debris. This debris can accumulate on the surface of the radiator, clogging the cooling and severely impeding airflow. A clogged radiator leads to a reduction in heat exchange efficiency, which can cause the engine and other hydraulic systems to overheat, reducing overall productivity and potentially causing damage to critical components [1].

To mitigate this issue, different radiator cleaning solutions have been developed, mainly by forcing air in the opposite direction to dislodge accumulated debris and carry the dirt away. Two usual methods are 1) changing the pitch of the fan blades [8] and 2) reversing the fan's direction of rotation [1]. This paper will focus on the latter approach, where the fan's rotational direction is reversed to generate a burst of air in the opposite direction of normal operation.

In the hydraulic system under investigation, this reversing function is accomplished by changing the fluid path to the hydraulic fan motor. As presented in Figure 2.6, this is typically achieved using a 4/2 on/off directional control valve that switches the pressure and return lines. During normal cooling operation, the pump's pressure port P is connected to working port A of the motor, while the port B is connected to the reservoir T. To initiate the cleaning cycle, the solenoid “a” is actuated, valve shifts, connecting the pressure port P to the working port B and working port A to reservoir T. This change in flow direction causes the motor and, consequently, the fan to rotate in the opposite direction.

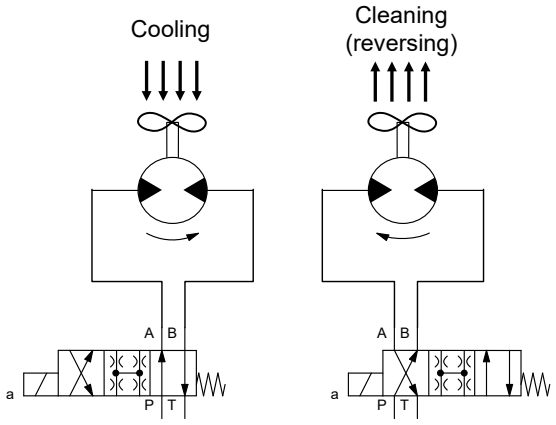


Figure 2.6. Directional positions for cooling and cleaning (reversing) modes

The activation of the reversing cycle is managed by the machine's controller. This can be initiated manually by the operator via a button in the cab or programmed to occur automatically at periodic intervals. The frequency of these automatic cleaning cycles can be tailored to the specific application and expected level of debris.

3. NOVEL SOLUTION

The idea behind the novel solution is to use a simplified hydraulic system combined with a feedforward control system to achieve the functionalities required for the fan drive

application with similar performance as status quo solution. As shown in the hydraulic schematics of Figure 3.1, in the novel concept, the speed control and fan reversal functions are combined in one single proportional directional valve (V3). Therefore, the novel concept is called one-valve design. The spool design is explained in section 3.1 and the feedforward controller in section 3.2.

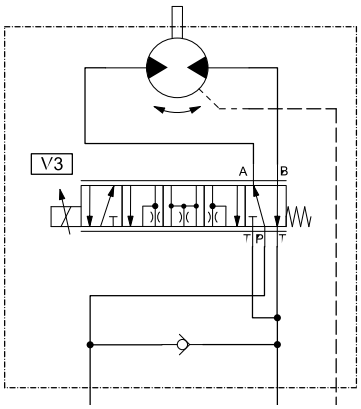


Figure 3.1. Hydraulic schematics of the novel concept one-valve design (pump is omitted)

3.1. Spool valve design

This novel solution employs a single proportional directional valve whose spool geometry is designed to achieve both fan speed control and direction change. This valve features a spool specifically designed for fan drive applications through control edges to achieve the performance required. The spool position can be continuously set from 0 to 100% by an electrical signal, which is referred as setpoint. Figure 3.2 illustrates a simplified representation of the spool valve geometry and its 5 functional stages.

Stage	Setpoint	T A P B T	Mode	Fan speed
1	0%		Cooling	Maximum
2	25%		Cooling	Intermediate
3	50%		Standstill	Stop
4	75%		Cleaning	Intermediate
5	100%		Cleaning	Maximum

Figure 3.2: Simplified representation of the spool in its five stages

Please find below an explanation of the five stages:

1. To meet the fail-safe requirement, at "zero" setpoint the maximum fan speed in the cooling mode is achieved by connecting P to A and B to T. The meter-in (P to A) edge is designed to achieve pressure drop of ~3 bar at 40 L/min to meets the market requirements, which is similar to other parts in body solutions of Bosch Rexroth such as SM12 open control block [10]
2. By increasing the setpoint, the P-A and B-T connections are maintained, but the A-T connection partially starts to open, causing some of the flow that would go to the motor to now divert to the reservoir. This causes the motor to decrease its speed, but still kept in the same direction and therefore cooling mode.
3. When the setpoint reachs 50%, then all four ports PABT are connected by specific designed metered edges. This allows for a smooth transition when reversing. Also, if the controller keeps the 50% setpoint, then the motor is set to standstill mode, i.e., it stops.
4. By this stage, the reversing starts connecting P to B and A to T, but not at full speed since there is a connection between B to T, as a mirror of stage 2. Therefore, this stage is in the cleaning mode with partial speed.
5. When the connection between B to T closes, then the fan operates at cleaning mode at full speed, as a mirror of stage 1.

To control the speed of the fan, the flow rate going to the motor (Q_m) is reduced by diverting part of the pump flow rate (Q_p) to reservoir, which is known as bleed-off flow rate (Q_b) as shown in Equation 3.1.

$$Q_m = Q_p - Q_b \qquad (3.1)$$

This technique allows for flow control without throttling, i.e., restricting, the path from pump to the load [11]. The challenge of the proposed flow control is the hight sensitivity of the controlled motor flow rate to the pump flow rate. To explain, if the speed of the diesel engine increases, the pump flow rate also increases, and if the bleed-off flow rate remains the same, the motor flow rate and consequently the fan speed will increase, which is not desired. As will explained in the next section, to overcome this sensitivity, a feedforward controller is developed where it takes account the diesel engine speed as input to keep a stable fan drive speed.

3.2. Control with feedforward

The objective of the control with feedforward is to keep the fan speed stable, even with fluctuations in pump speed. As shown in the Figure 3.3, it calculates the spool position setpoint (x') based on the fan speed setpoint (n_f'), determined from control curve according to the measured temperature (T), and adjusted to the measured pump speed (n_p), which is commonly available in the electronics of modern machines. It allows to the controller to automatically adjust the bleed-off flow rate to compensate for a possible fluctuation in pump flow rate (Q_m), which keeps a constant controlled flow rate to the motor (Q_m), therefore a stable fan speed (n_f).

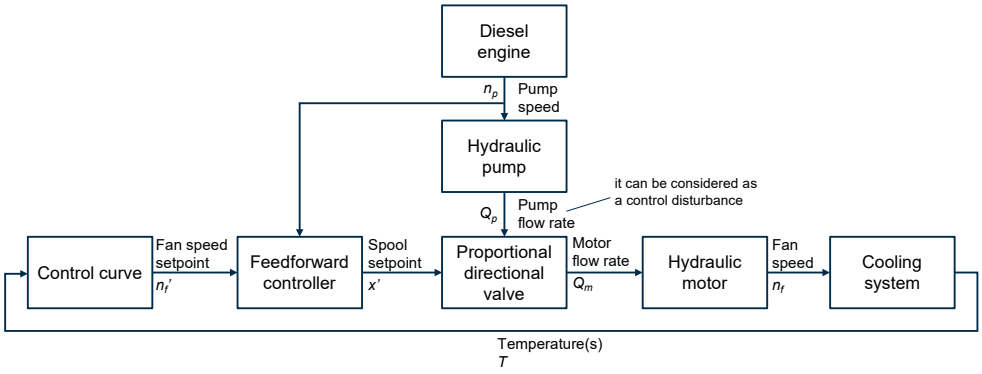


Figure 3.3. Speed control logic with proportional directional valve

To understand behind how the control calculates the spool setpoint based on the fan speed setpoint and the pump speed, the Equation 3.1 is reorganized to isolate the bleed-off flow rate (Q_b):

$$Q_b = Q_p - Q_m \tag{3.2}$$

Knowing the measured pump speed (n_p), the pump displacement (V_p), and the volumetric efficiency of the pump ($\eta_{p,v}$), it is possible to calculate the flow rate as follows:

$$Q_p = n_p V_p \eta_{p,v} \tag{3.3}$$

Similar way, it is possible to calculate the motor flow rate (Q_m) knowing the desired motor speed, which is the same as the fan speed setpoint (n_f'), the motor displacement (V_m), and the volumetric efficiency of the pump ($\eta_{m,v}$):

$$Q_m = n_f' V_m \eta_{m,v} \tag{3.4}$$

Consolidating Equations 3.3 and 3.4 in 3.2:

$$Q_b = n_p V_p \eta_{p,v} - n_f' V_m \eta_{m,v} \tag{3.5}$$

The bleed-off flow rate is controlled by the spool opening, which can be described as a turbulent orifice [12]:

$$Q_b = C_d A \sqrt{\frac{2}{\rho} \Delta p} \tag{3.6}$$

Where C_d is discharge coefficient, normally considered 0,7, A the orifice area of the bleed-off control edge, which varies according to spool position, and (Δp) the pressure differential across the bleed-off control edge.

The pressure differential across the motor can be considered the same as the differential across the bleed-off control edge, since its orifice area is significantly more restrictive than the meter-in and meter-out. As explain in the section 2.1 for the status quo design, the relationship between pressure differential across the motor (Δp) and fan speed setpoint (n_f') is known:

$$\Delta p = f(n_f') \tag{3.7}$$

The orifice area of the bleed-off control edge (A) depends on the spool setpoint (x') and its relationship is defined in the spool design:

$$x' = g(A) \qquad (3.8)$$

Combining Equations 3.5, 3.6, 3.7 and 3.8, then:

$$x' = g\left(\frac{n_p V_p \eta_{p,v} - n'_f V_m \eta_{m,v}}{C_d \sqrt{\frac{2}{\rho} f(n'_f)}}\right) \qquad (3.9)$$

Equation 3.9 therefore governs the behaviour of the feedforward controller by continuously calculating the spool setpoint (x') according to fan speed setpoint (n'_f), and the measured pump speed (n_p).

4. SIMULATION MODEL

Having clarified how the status quo solution works, which is the performance benchmark, and the challenges involved in the novel solution, a simulation model was created to support the product development. The use of a simulation model significantly accelerates the design cycle and reduces costs by enabling early performance validation and optimization of spool geometry reducing the design iterations. The objective of the model in this phase is to confirm the feasibility of the novel solution by performing a preliminary assessment of system performance in comparison to status quo.

The model was created using AMESim, a mechatronic systems simulation platform [13]. For this comparative analysis, models for both the existing and novel solution were created. The schematic diagrams for both are presented in Figure 4.1 side-by-side, with key components labelled for clarity. Parameters used in the model such as the size of the pump, motor and fan are not from any specific machine, but those fall within the range commonly found in fan drive application. If necessary, in the future it is possible to parameterize the model to validate a specific configuration.

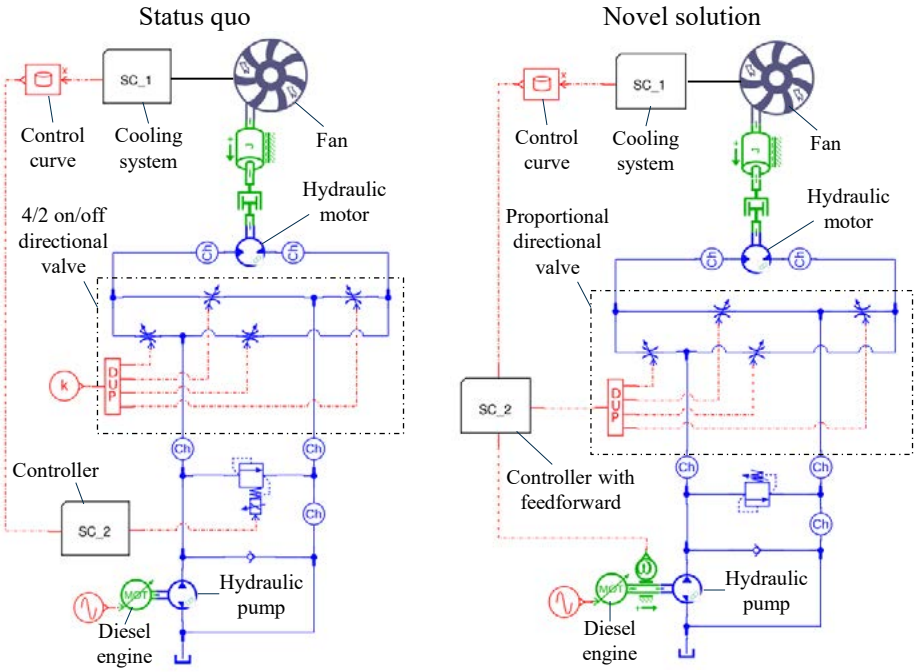


Figure 4.1. Schematics of simulation models

For the 4/2 on/off directional valve (status quo model) four orifices with turbulent conditions were used, one for each connection P-A, P-B, A-T and B-T. The orifice area (A) of each connection varies according to the spool setpoint (x'), where the relationship is defined by the spool geometry including the notches. The time response of the valve is defined by a first order response, where a 99% of the setpoint is achieve within 100 ms. Same approach is used for the proportional directional valve (novel solution model). It was considered a fan with a torque of 50 Nm at 2500 rpm. For the pump and motor, it was chosen Rexroth external gear units of 28 ccm and 19 ccm respectively, where volumetric efficiencies were taken from datasheet curves [14] [15]. The oil properties followed the ISO VG 46 classification. The influence of temperature variation on the parameters of the hydraulic components was not considered. Other models are standard from AMESim library. The main parameters of the model are available in Table 4.1.

Table 4.1. Main parameters of simulation model

Component	Description	Value	Unit
Motor	Displacement	19	ccm
	Volumetric efficiency	0.92	-
Pump	Displacement	28	ccm
	Volumetric efficiency	0.95	-
Fan Drive	Static friction	5	Nm
	Running friction	4	Nm
	Viscous friction coefficient	0.0184	Nm/rpm
	Moment of Inertia	0.07	kg.m²
Valves	First order time constant	0.0217	s
Proportional Pressure Relief Valve	Flow Rate Pressure Gradient	6.15	L/min/bar
Fluid	Fluid density	865	kg/m³

5. RESULTS

The performance of the novel solution (one-valve design) was evaluated under a significant disturbance by varying the pump speed and compared to the status quo (two-valve design), as illustrated in the Figure 5.1. To show that feedforward controller is indeed necessary, the novel solution behaviour was presented both with and without this controller. As presented in Graph b, a periodic disturbance was introduced by varying the pump speed sinusoidally between approximately 1,000 rpm and 2,400 rpm over a 20-second cycle. The fan speed of status quo, novel solution without feedforward, novel solution with feedforward is presented in Graph a. The objective was to maintain a constant fan speed of 1,400 rpm.

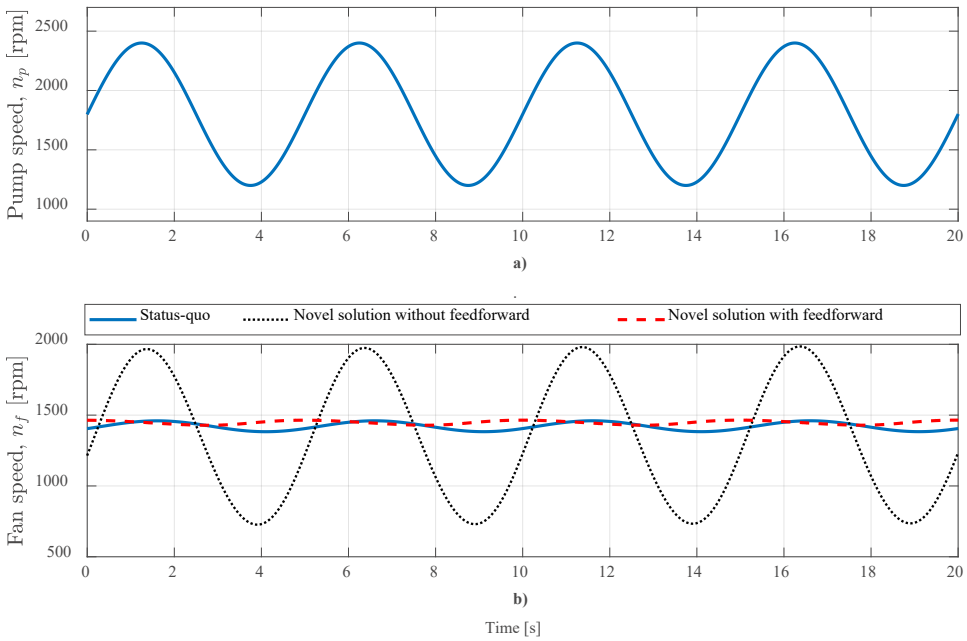


Figure 5.1. Fluctuation in fan drive speed due to variation in pump speed (simulation)

- **Status quo (benchmark):** The existing control system (solid blue line) serves as the reference for performance. It demonstrates partial disturbance rejection, with oscillations of approximately ±38 rpm amplitude.
- **Novel solution without feedforward:** When compared to the benchmark, this control strategy (dotted black line) was unable to compensate for the disturbance. The fan speed exhibited large, unstable oscillations with amplitude of approximately ±623 rpm, demonstrating a clear failure to regulate the system and a significant degradation in performance.
- **Novel solution with feedforward:** The proposed controller incorporating a feed-forward mechanism (dashed red line) demonstrated better performance than Status quo. It effectively attenuated the disturbance from the pump with oscillations of approximately ±18 rpm amplitude.

These results indicate that the inclusion of a feedforward control loop provides a substantial improvement in disturbance rejection and system stability over both the status-quo and non-feed-forward solutions. Other investigations were also carried out using the simulation model, such as the assessment of peak pressure and velocity during reversal, where the requirements were met However, these were not presented in detail since the focus of this paper is on the stability in speed control.

6. CONCLUSION

This paper successfully demonstrated by simulation, that the novel one-valve solution, when integrated with its enhanced feed-forward control, can overcome the challenge of fan speed instability caused by engine speed fluctuations when using proportional directional valves. The feedforward control strategy shows to be capable of providing robust disturbance rejection to pump speed variation, using the status quo solution, which is a pressure-controlled fan drive, as a performance benchmark.

Simulation results revealed that the status quo exhibited fan speed oscillations of approximately ±38 rpm under variable pump speeds. The proposed novel solution with feedforward controller effectively compensated for these disturbances by achieving ±18 rpm oscillation, while without controller it presented low performance with ±623 rpm oscillations.

The findings show that the use of a feedforward controller can enable the use of a cheaper and more compact hydraulic system, while maintaining low sensitivity to variations in engine speed. Furthermore, the development process underscored the value of simulation as a powerful tool for rapidly designing and validating control concepts prior to hardware implementation.

Future work should prioritize the experimental validation of these simulation results on a physical test bench to confirm the real-world performance of the hydraulic system and controller. Further investigations could also explore the system's robustness against other operational variables, such as changes in valve’s time response and geometry tolerances.

7. REFERENCES

[1] Bosch Rexroth AG, “Less Emission, Lower Cost. Rexroth Hydrostatic Fan Drives , RE 98065.” Sep. 01, 2010.

[2] A. Martyr, A. J. Martyr, and M. A. Plint, *Engine testing: theory and practice*, 3. ed. Oxford: Butterworth-Heinemann, 2007.

[3] J. Feuerstein, “Electronically Controlled Hydrostatic Fan Drive System for the Internal Combustion Engine,” presented at the International Off-Highway & Powerplant Congress & Exposition, Milwaukee, Wisconsin, United States, Sep. 1990, p. 901586. doi: 10.4271/901586.

[4] V. Jambukar, V. Aher, and V. Wakchaure, “Hydraulic System for Fan Drive Application to Cool Engine Effectively,” vol. 2, no. 7, 2014.

[5] P. Teixeira, “Fan drive device with proportional valve,” US12429070 B2, Sep. 30, 2025

[6] Bosch Rexroth AG, “Hydrostatic fan drive - Infinitely variable speed control and power on demand, RE98745.” Aug. 20, 2019.

[7] Bosch Rexroth AG, “Proportional pressure relief valve, pilot operated, decreasing characteristic curve KBVS.3B, RE18139-07.” Oct. 2022.

[8] A. Sinha, M. Jha, P. Vishwakarma, and U. G. Student, “Mechanically Operated Reversible Radiator Fan Blade,” *IJIRSET*, vol. 5, no. 9, 2007.

[9] HydraForce, “Solenoid Cartridge Valve Spool, 4 port, 2 position, SV10-40.” Sep. 25, 2024.

[10] Bosch Rexroth AG, “Open center control block in sandwich plate design SM12, RE 64122.” Jul. 2016.

[11] J. S. Cundiff, *Fluid power circuits and controls: fundamentals and applications*. Boca Raton: CRC Press, 2002.

[12] H. E. Merrit, *Hydraulic control systems*. New York: Wiley, 1967.

[13] “Simcenter Amesim | Siemens Software.” https://plm.sw.siemens.com/en-US/simcenter/systems-simulation/amesim/ (accessed Jan. 14, 2026).

[14] Bosch Rexroth AG, “External gear pump High Performance AZPF, RE 10089.” Nov. 14, 2024.

[15] Bosch Rexroth AG, “External gear motor High Performance AZMF, RE 14028.” Nov. 19, 2024.

Biographies



Paulo Teixeira received the bachelor's degree in Mechanical Engineering from Federal University of Santa Catarina (UFSC) in 2012, the master's degree in Mechanical Engineering with a concentration in Mechanical Systems Design, in Federal University of Santa Catarina (UFSC) in 2015. He is currently working as Product Engineer at the Advanced Engineering External Gear Units Department at Bosch Rexroth. His research areas include hydraulics, mobile application, fatigue assessment, finite element method (FEM), and reliability.



Pedro Condessa received his bachelor’s degree in Mechanical Engineering from the Federal University of Santa Catarina (UFSC) in 2025. He worked as an intern in Advanced Engineering External Gear Units Department at Bosch Rexroth, where he focused on thermal-hydraulic system simulation and analysis. His research interests include hydraulic systems, system modeling and simulation, and data-driven analysis.



Stefan Kock received his Mechanical Engineering degree from the Technical University of Dresden, Germany, and his Ph.D. from RWTH Aachen University, where his research focused on wind turbine drive trains and torque measurements. He is currently Team Leader for Advanced Engineering for External Gear Units at Bosch Rexroth. His research areas include drive train technology, mechatronic systems, fatigue assessment, finite element method (FEM), and fluid power drives.



Martin Laube received his Diploma degree as Dipl.-Ing. (univ.) in Mechanical Engineering from the Technical University of Munich (TUM) in 1996. He specialized during his studies in the field of Design and Development with a focus on machine elements, vehicle technology and fluid technology. He started his industrial career in 1996 as a development engineer for railway systems at Mannesmann Rexroth, Lohr a. Main. After holding various positions in sales and engineering for industrial applications at Mannesmann Rexroth and Bosch Rexroth (since 2001), he assumed the position of Engineering Director for Hydraulic Power Units, Manifolds and Systems at Bosch Rexroth AG in 2013. Since July 2024, he has been Vice President and Head of Product Area External Gear Units in the Business Unit Mobile Components at Bosch Rexroth AG.

An Electro Hydraulic Hybrid System to Downsize Motor Power of Electric Wheel Loader Working Function

Yihan Qiao, Jiaming Wu, Feng Wang* and Bing Xu

State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, Zhejiang, 310027, China

** Corresponding author, E-mail: dieter@zju.edu.cn*

Abstract.

The electro-hydrostatic variable speed drive system (VSD) is one of the effective ways to improve energy efficiency of electric wheel loader working function. However, the VSD system still has the limitations including large installed motor power, low-efficiency operating points of the electric motor-pump, and high dynamic response requirement of components. Therefore, an electro-hydraulic hybrid variable speed drive system (EH-VSD) is proposed to reduce the installed power with maintaining the energy efficiency. Unlike VSD system that relies on motor speed control, the proposed solution incorporates a high-pressure hydraulic accumulator in parallel configuration. The electric motor-pump supplies the average power demand, while the accumulator responds to peak load requirements. An energy management strategy based on the dynamic programming is developed to manage the power distribution of two energy sources. Simulation results show that the installed motor power of EH-VSD system reduces by 54.5% compared to VSD system (from 110 kW to 50 kW). The EH-VSD system achieves a significant reduction in installed power while maintaining a high energy efficiency comparable to that of the VSD system, effectively avoiding substantial throttling losses.

Keywords. Electro-hydraulic hybrid system, variable speed drive, dynamic programming, installed power, electric wheel loader.

1. INTRODUCTION

With the growing global demand for energy conservation and emission reduction in construction machinery, traditional machines powered by internal combustion engines are increasingly unsuitable due to their high energy consumption, substantial carbon emissions, and severe environmental impact [1]. Electrification technology, which replaces conventional engines with electric motors powered by batteries or cables, enables zero emissions at the point of use and significantly improves system energy efficiency. Consequently, it has become a prominent research focus in the field of construction machinery [2].

To reduce energy consumption in the working system of electric wheel loader, an electro-hydrostatic variable speed drive (VSD) system has been proposed for application in wheel loaders [3], as illustrated in Fig. 1.1. This system controls the velocity of the hydraulic cylinder by adjusting the motor speed, thereby avoiding throttling regulation and improving overall energy efficiency [4]. Qu et al. applied an open-circuit VSD working system with an

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

integrated motor-pump unit in a skid steer loader, achieving a steady-state energy efficiency of up to 59.9% [5]. Guo et al. proposed a VSD steering system for wheel loaders, which reduced energy consumption by 22.8% compared with conventional systems [6].

Although the VSD working system significantly improves energy efficiency, it still has certain inherent limitations due to its system configuration: 1) The high-magnitude and transient load demands characteristic of working hydraulic system requires substantial installed power when relying solely on motor-pump actuation. 2) The frequent load cycle variation compels the motor-pump unit to respond rapidly, often operating outside its optimal efficiency range. 3) The concurrent demands for high dynamic response, sustained energy efficiency, and high peak power result in increased system cost [7-10].

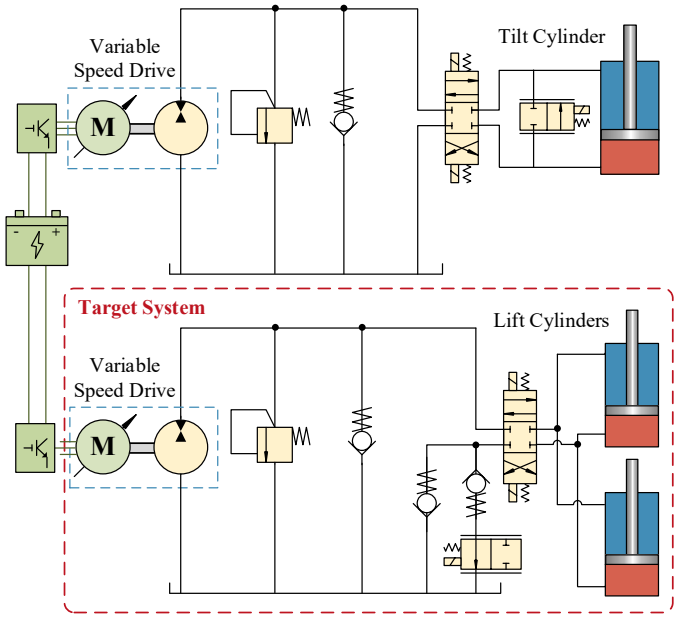


Fig. 1.1. The electro-hydrostatic variable speed drive (VSD) system (baseline system)

Compared with purely electric drive systems, electro-hydraulic hybrid drive systems integrate a secondary energy storage component, which expands the controllable dimensions and enables flexible energy distribution and conversion between different power sources, thereby facilitating improved overall performance [11-13]. Some recent studies have focused on electro-hydraulic hybrid control strategies for VSD working systems. Kang et al. proposed using an accumulator as a built-in regulator to enhance the dynamic performance of the VSD system, aiming to achieve faster response and broader frequency bandwidth [14]. Shang et al. introduced an energy recovery module into the variable-speed system, which disabled the generator mode of the motor under forward load conditions [15]. This configuration allowed the elimination of the system's power resistor and reduced electric motor heating temperature by 39%. Seo et al. employed a secondary control part in the VSD system, consisting of a high-pressure accumulator, a servo valve and mode-switching valves, to reconfigure the hydraulic circuit according to operational modes [16].

Therefore, an electro-hydraulic hybrid variable speed drive system (EH-VSD) is proposed for the working function of electric wheel loaders. Unlike VSD system that relies solely on motor speed regulation, the proposed solution incorporates a high-pressure accumulator in parallel configuration. The electro-hydraulic parallel design leverages the respective advantages of both components: the electric motor and pump unit (VSD unit) supplies the average power demand, while the accumulator responds to peak load requirements. This configuration effectively reduces the installed power, leading to a more compact and lightweight system, making it more suitable for powertrain layout design and contributing to lower electric drive costs. Meanwhile, the proposed system still utilizes the variable speed drive to maintain the throttle-free advantage.

The main contribution of this study is to propose an electro-hydraulic hybrid variable speed drive system (EH-VSD) to downsize the electric drive system. An energy management strategy based on the dynamic programming is developed to ensures the VSD unit operates within its high-efficiency region, which further improves the energy efficiency of the system. The comparison between proposed system and baseline system (VSD) is conducted in terms of system performance, installed power and energy efficiency. The rest of the paper is organized as follows: Section 2 presents system mathematical modelling and working modes, section 3 shows control strategy and optimization method, section 4 shows simulation study and comparison, and section 5 concludes the study.

2. SYSTEM CONFIGURATION

2.1. System overview

The schematic of the electro-hydraulic hybrid variable speed drive (EH-VSD) working system is illustrated in Fig. 2.1. In the working system of a wheel loader, the lift function accounts for over 80% of the total energy consumption and is characterized by high load power [17]. Therefore, this study focuses on the lift function as a representative case to investigate the performance of the proposed EH-VSD system.

The system's power sources consist of a VSD unit and a high-pressure hydraulic accumulator module. The VSD unit is composed of a lithium iron phosphate ($LiFePO_4$) battery pack, an inverter, a permanent magnet synchronous motor (PMSM), and a hydraulic pump. The PMSM is selected for its fast dynamic response, making it well-suited for variable-speed operation. A fixed-displacement, unidirectional hydraulic pump is employed in the system. The high-pressure accumulator module serves as a secondary power source, supplying additional energy to the system when required. The proportional valve 1 (PV1) is used to regulate the flow discharged from the accumulator and also allows the hydraulic pump to operate independently. Both the hydraulic pump and the accumulator adopted in the proposed system are commercially available hydraulic components.

The hydraulic system adopts an open-circuit configuration, in which the low-pressure pipeline and the pump inlet are directly connected to the hydraulic reservoir. The system switches between different operating modes using a directional valve (DV), with specific control strategies detailed in Section III. The proportional valve 2 (PV2) is employed to control the lowering speed of the hydraulic cylinder and remains fully open in the left position during other working modes. The check valve 1 (CV1) ensures unidirectional rotation of the hydraulic pump by preventing backflow from the accumulator, thereby

avoiding interference with pump operation. The check valves 2 (CV2) and 3 (CV3) are used to control the flow direction between the hydraulic cylinder and hydraulic reservoir, enabling appropriate flow paths for different working modes. A relief valve (RV) is included as a protective component to prevent system pressure from exceeding the designated working pressure.

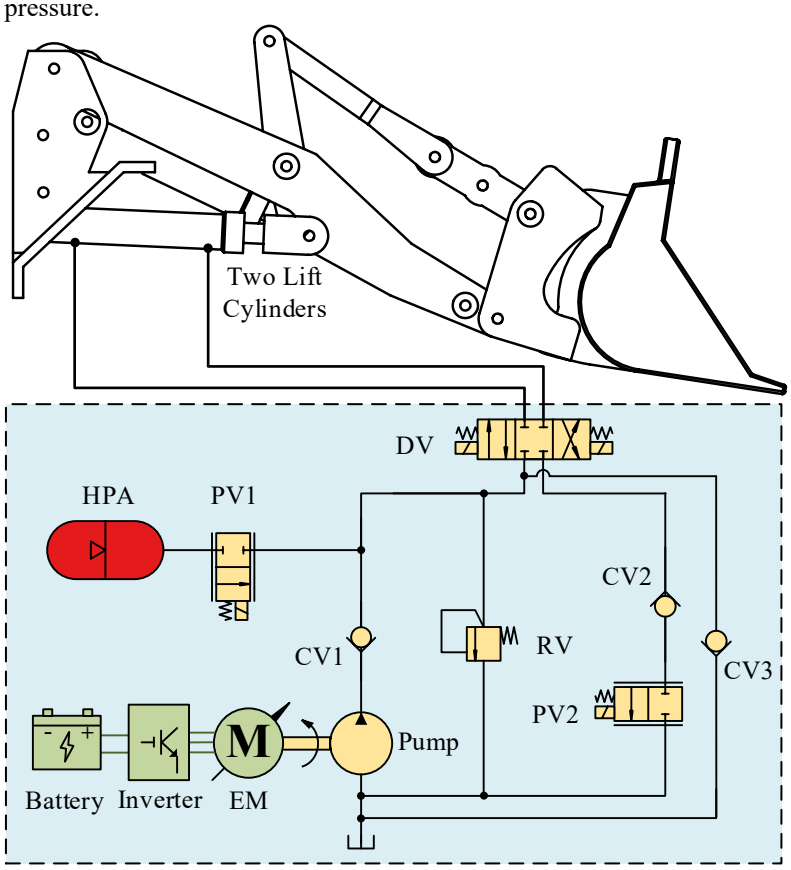


Fig. 1.1. The proposed electro-hydraulic hybrid variable speed drive (EH-VSD) working system

2.2. System Modelling

1) Electric Motor/Generator: The electric motor is modelled using a d–q axis framework, where the direct axis (d) aligns with the rotor flux, and the quadrature axis (q) is orthogonal to it. This empirical model for the permanent magnet synchronous motor (PMSM) captures the motor's dynamic behaviour during rapid acceleration and deceleration and allows for adjustable step sizes in real-time simulations.

According to the generalized form of Ohm's law, the d- axis and q-axis components of the stator voltage space vector, u_d and u_q, are expressed as:

$$\begin{cases} u_d = R_s I_d - \omega_e \psi_q + \dot{\psi}_d \\ u_q = R_s I_q + \omega_e \psi_d + \dot{\psi}_q \end{cases} \tag{4}$$

where R_s is stator winding resistance and ω_e is the angular velocity of the rotor.

The d–q components of the flux linkage, ψ_d and ψ_q, are given by:

$$\begin{cases} \psi_d = L_d I_d + \psi_{pm} \\ \psi_q = L_q I_q \end{cases} \tag{5}$$

where L_d and L_q are stator cyclic inductances along the d-q axis, I_d and I_q are corresponding stator currents, and ψ_{pm} is permanent magnet flux linkage.

The electromagnetic torque, T_{em}, is calculated by:

$$T_{em} = \frac{3}{2} N\left(\psi_d I_q - \psi_q I_d\right) \tag{6}$$

where p is the number of pole pairs.

The dynamic equation of the motor shaft is described by:

$$\left(J_{em} + J_p\right)\dot{\omega_{em}} = T_{em} - T_p - T_f \tag{7}$$

where J_{em} and J_p are the shaft inertia of electric motor and pump, T_p is the torque of pump and T_f is viscous friction torque.

2) Fixed Displacement Pump: The fixed-displacement pump model incorporates the variation of component efficiencies with operating speed and pressure. It determines the output torque and flow rate based on an efficiency map that reflects these dependencies. The shaft torque T_p, and inlet flow rate Q_p, of the pump are given by:

$$\begin{cases} T_p = \dfrac{D_p(p_H - p_L)}{\eta_m^a} \\ Q_p = D_p \omega_p \eta_v^a \end{cases} \tag{9}$$

where D_p is full displacement of pump, p_H and p_L are the high and low pressure sides of the pump, η_m and η_v are mechanical and volumetric efficiencies which are pressure and shaft speed-dependent, a is mode factor of pump, ω_p is the shaft speed of pump and is equal to the electric motor speed $\omega_p = \omega_{em}$.

3) Accumulator: During the loading cycle, both the charging and discharging of the accumulator are relatively short. Therefore, the accumulator is modelled using the isothermal gas law, assuming a constant operating temperature. The pressure of the accumulator, p_{acc}, is given by:

$$p_{acc} = p_0 \left(\frac{V_0}{V_{acc}}\right)^{\gamma} \tag{10}$$

where p_0 is pre-charge pressure, V_0 and V_{acc} are initial and current gas volume and γ is polytropic index of gas.

The current gas volume of the accumulator is determined by the integral of the flow rate over time. The flow rate, Q_{acc}, is considered positive during discharge and negative during charging. The gas volume, V_{gas}, is expressed as:

$$V_{acc} = V_0 + \int Q_{acc} dt \tag{11}$$

The hydraulic state of charge (SOC) of the accumulator, SOC_{acc}, is defined as:

$$SOC_{acc} = \frac{p_{acc} - p_{min}}{p_{max} - p_{min}} \tag{12}$$

where p_{min} and p_{max} are the maximum and minimum allowable accumulator pressure.

4) Proportional Valve: A proportional valve is used to regulate the flow rate from the high-pressure accumulator and another one controls the lowering speed of the lift cylinder. The flow rate across the directional valve, Q_v, is:

$$Q_v = c_q A_v x_v \sqrt{\frac{2|\Delta p|}{\rho}} \tag{13}$$

where c_q is flow coefficient, A_v is throttling area of proportional valve, x_v is the valve opening ratio, Δp is the pressure difference of valve inlet and outlet, ρ is hydraulic fluid density.

5) Hydraulic Cylinder: The dynamic behaviour of cylinder can be described by a set of well-established mathematical equations:

$$\begin{cases} Q_A - A\dot{x_{cyl}} = \frac{V_A}{E_h}\dot{p_A} \\ a\dot{x_{cyl}} - Q_a = \frac{V_a}{E_h}\dot{p_a} \\ p_A A - p_a a - F_{load} - Bv = m\ddot{x_{cyl}} \end{cases} \tag{16}$$

where p_A and p_a are pressure in the piston and rod chambers, A and a are the effective area of piston and rod sides, Q_A and Q_a are flow rate of piston and rod chamber, V_A and V_a are control volumes of piston and rod chambers, E_h is the effective bulk modulus of hydraulic fluid, F_{load} is the external load force on the piston, B is viscous friction coefficient, x_{cyl} is piston displacement and m is equivalent mass of piston.

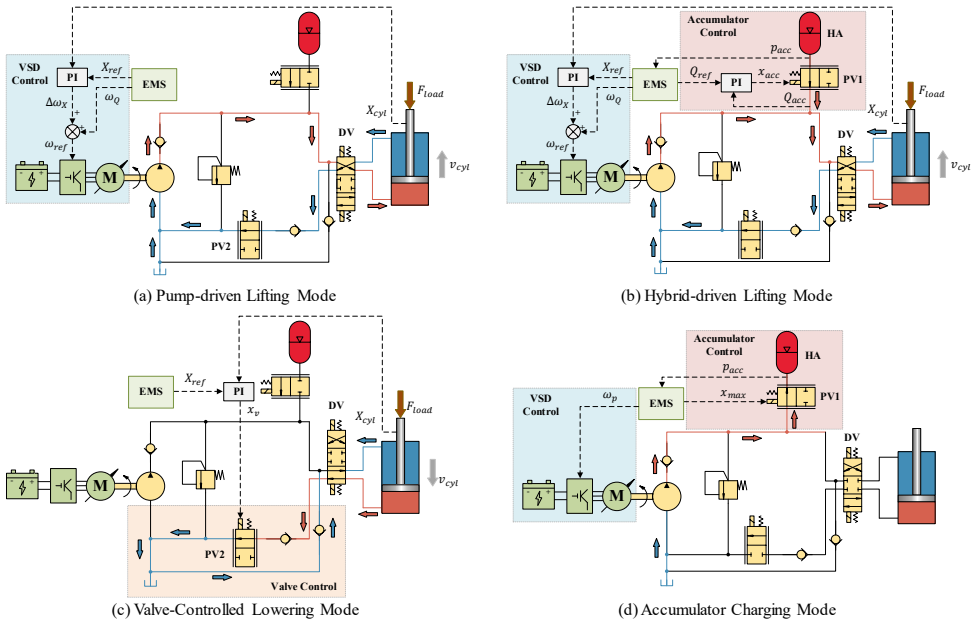
2.3. Working Mode

The proposed EH-VSD system operates in four working modes: pump-driven lifting mode, hybrid-driven lifting mode, valve-controlled lowering mode, and accumulator charging mode. The overall system operation is coordinated by an energy management strategy (EMS), as illustrated in Fig. 2.2.

(a) Pump-driven Lifting Mode: In this mode, the control strategy integrates feedforward speed estimation with position PI feedback to achieve accurate cylinder motion control. Based on the cylinder velocity in the given loading cycle, the EMS calculates the theoretical rotational speed of the VSD unit, which serves as the feedforward term for the electric motor speed command. In parallel, a position PI feedback controller computes a speed compensation from the position difference between the actual cylinder position and reference given by EMS. The feedforward and feedback components are combined to ensure

precise cylinder position control. The baseline VSD system also employs this same control approach, regulating cylinder position through motor speed adjustment.

(b) Hybrid-driven Lifting Mode: In the hybrid mode, the flow rate outputs of the VSD unit and the high-pressure accumulator are determined in advance by the dynamic programming–based optimization method (to be described in the following subsection), and EMS assigns the corresponding reference control signals to each module. The high-pressure accumulator is regulated using proportional control, where the opening of PV1 is determined by the PI controller based on the error between the theoretical and actual flow rates, ensuring accurate control of the accumulator's output power. The VSD unit continues to employ the feedforward and feedback control strategy, with the feedback loop ensuring precise lift cylinder position tracking.



(a) Pump-driven Lifting Mode (b) Hybrid-driven Lifting Mode

(c) Valve-Controlled Lowering Mode (d) Accumulator Charging Mode

Fig. 2.2. Working modes of the EH-VSD working system

(c) Valve-Controlled Lowering Mode: In this mode, a hydraulic cylinder position PI controller is employed. The PI controller regulates the opening of the proportional valve based on the error between the theoretical position provided by the EMS and the actual actuator position, achieving controlled lowering of the cylinder through outlet throttling. Furthermore, not utilizing the VSD unit for energy recovery in this mode offers additional benefits: (i) the hydraulic pump does not need to operate in bidirectional rotation, so a unidirectional pump can be utilized, thereby reducing component cost, (ii) the electric motor avoids operating in generator mode, which would otherwise cause significant power losses and increase motor temperature, negatively affecting its operating condition.

(d) Accumulator Charging Mode: In this mode, the proportional valve of the accumulator module is kept fully open to avoid additional throttling losses. The EMS regulates the VSD unit speed according to the optimization results obtained from the dynamic programming,

ensuring that both the motor and hydraulic pump operate within their high-efficiency regions. Once SOC_acc reaches its maximum value (i.e., the initial state), the proportional valve of the accumulator module is closed, and the VSD unit returns to standby mode.

3. SYSTEM CONTROL STRATEGY

3.1. Overview

Compared with the VSD working system, the proposed EH-VSD system incorporates two power sources: the VSD unit and the high-pressure hydraulic accumulator. The inclusion of the accumulator provides an additional degree of freedom for system control, and certain parts of the system configuration differ from the conventional design. Consequently, different working modes must be assigned according to the wheel loader's loading cycle. Multiple power sources enable power-split operation within the system. The accumulator not only reduces the installed power requirement but also facilitates optimal system-level control aimed at improving energy efficiency. The power distribution between the VSD unit and the accumulator has a significant impact on both installed power and system energy efficiency, making it necessary to perform optimization of power distribution.

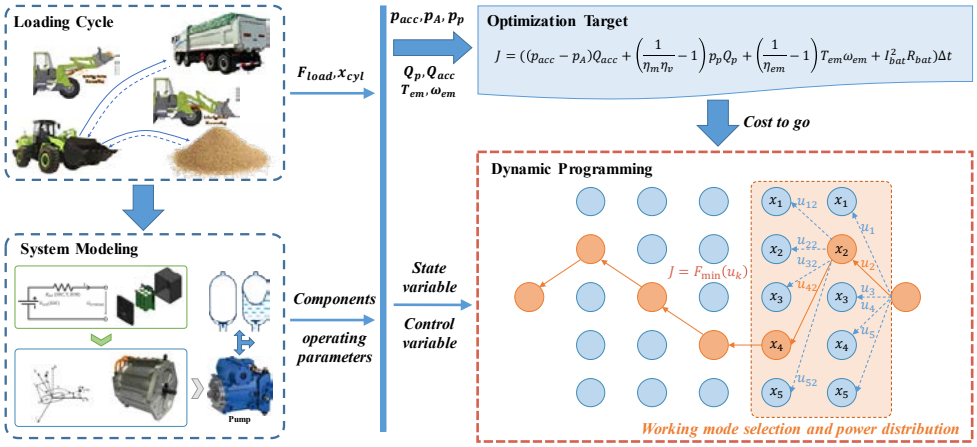


Fig. 3.1. Control schematic of the mode-based control strategy of proposed powertrain

The control strategy and power distribution optimization framework for the proposed EH-VSD system are illustrated in Fig. 3.1. First, the loading cycle of wheel loader are analyzed to determine the control strategy and optimization boundary. Then, based on the loading cycle and the system mathematical model established in section 2, a cost function is formulated to define the optimization objectives. Finally, the dynamic programming is then applied to select the working mode of system and achieve the globally optimal power distribution parameters, thereby maximizing energy efficiency under the minimum installed power constraint.

3.2. Optimization Method

Dynamic programming (DP) is commonly used to solve global optimization problems by decomposing them into a series of simpler sub-problems [18, 19]. Under a known loading

cycle, DP discretizes the cycle and solves the problem step-by-step, reducing computational complexity by storing and reusing solutions of sub-problems [20, 21].

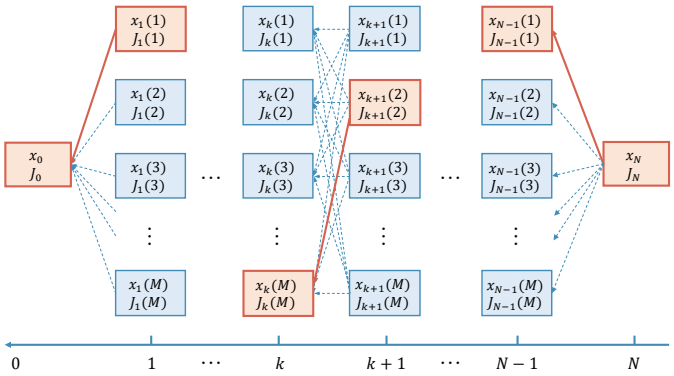


Fig. 3.2. Dynamic programming for the power distribution optimization method

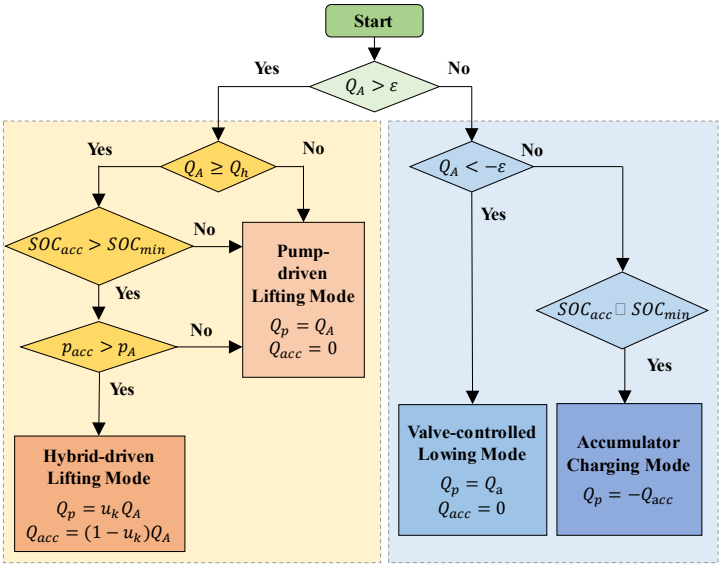


Fig. 3.3. Working mode selection within the dynamic programming

This study focuses on the EH-VSD working system to investigate the optimal power distribution between VSD unit and high-pressure accumulator. The objective is to reduce the installed power while improving overall system energy efficiency. The time horizon is discretized into N equally spaced intervals. Appropriate state and control variables are selected and discretized into an M-dimensional grid. A performance objective function is formulated based on the system's energy losses. Under constraints on initial and terminal states as well as component operational limits, a backward recursion algorithm is used to find the global optimum. Subsequently, a forward pass calculates the power-optimal control inputs.

In addition to the conventional optimal power distribution, a working mode selection strategy is integrated into the dynamic programming framework to ensure proper system

operation and reduce the computational load of the DP process. The control logic of this strategy is consistent with that described in Section 2, where one of four operating modes is selected based on the required flow rate of the working system in the loading cycle.

4. SIMULATION STUDY

4.1. Simulation Parameters

The simulation study is conducted on Matlab and Amesim platforms. Matlab is used for control strategy and optimization method, while Amesim is applied for component model verification and simulation operation. The main parameters of the proposed EH-VSD system are designed according to the existing electric wheel loader working system with all components commercially available. The battery pack is a 456 Ah lithium iron phosphate type, whose open-circuit voltage and internal resistance vary with battery SOC. When the SOC decreases, the battery voltage drops while the internal resistance rises.

Table 3.1

Main parameters of proposed EH-VSD system

	Property	Value	Unit
Battery	Normal capacitance	456	Ah
	Normal voltage	618	V
	Normal Initial resistance	20	mΩ
Electric motor	Max. speed	3000	rpm
	Max. torque	450	Nm
Hydraulic pump	Geometric displacement	100	cc/rev
	Max. speed	3000	rpm
	Max. pressure	30	MPa
Hydraulic accumulator	Volume	25	L
	Max./Min. pressure	32/12	MPa
	Efficiency	85%	\
Proportional valve	Valve coefficient	0.7	\
	Natural frequency	80	Hz
Hydraulic pipeline	Length	3	m
	Diameter	20	mm
	Relative roughness	0.002	\
	Hydraulic fluid viscosity	46	mm^2/s
Hydraulic cylinder	Diameter of piston	150	mm
	Diameter of rod	90	mm
	Piston stroke	815	mm
	Viscous friction coefficient	2000	N/(m/s)

Permanent magnet synchronous motor (PMSM) is adopted, with maximum speed and torque limited to 3000 rpm and 450 Nm respectively. A unidirectional fixed-displacement pump of 100 cc/rev is employed with a maximum pressure of 30 MPa, and a 25 L high-pressure accumulator operates within pressure range of 12–32 MPa.

The electric motor–pump unit (VSD unit) used in this study adopts efficiency maps obtained from manufacturer data and test bench measurements. The operating efficiency depends on motor torque (related to pump pressure) and rotational speed. Therefore, optimizing the operating points of the VSD unit can effectively improve system energy efficiency.

4.2. Simulation Results and Performance Comparison

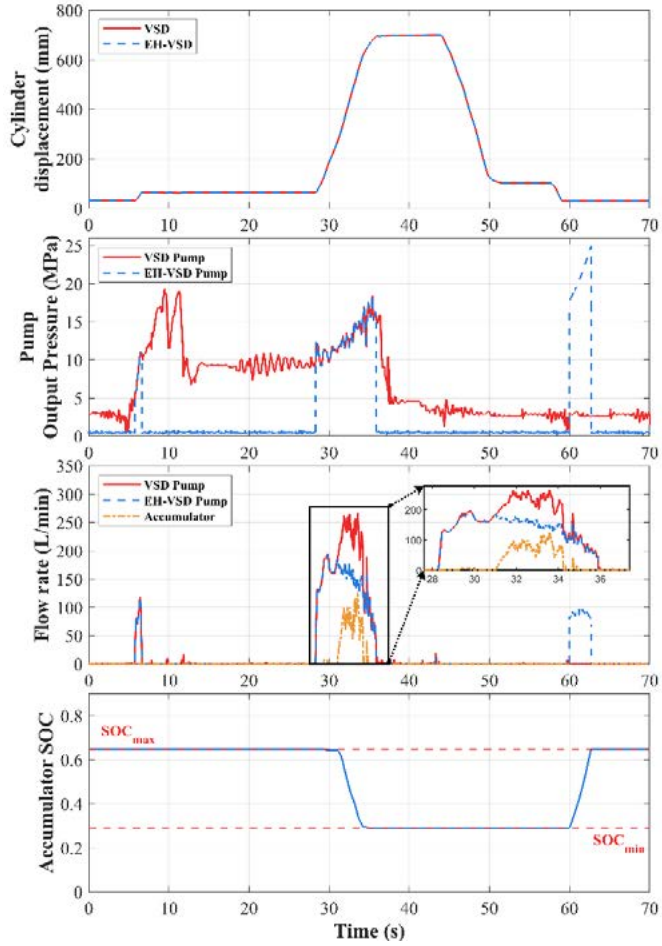


Fig. 4.1. Comparison of hydraulic components performance between systems

The performance of hydraulic components in both the baseline VSD system and the proposed EH-VSD system is shown in Fig. 4.1. In both cases, the hydraulic cylinders accurately follow the reference signal to complete the loading task. In VSD system, the pump outlet pressure remains nearly consistent with the load pressure. In the EH-VSD system, due to the check valve at the pump outlet, pressure is generated only when the pump works, and it closely matches the load pressure. After the loading cycle, the pump also charges the accumulator, during which outlet pressure follows the accumulator pressure.

The flow rate output highlights the differences between the two systems. In the VSD system, the hydraulic pump alone supplies the flow required for cylinder lifting. In contrast, the EH-

VSD system combines the outputs of the pump and accumulator, where the pump provides a nearly constant flow and the accumulator compensates for the remaining demand, thereby reducing the pump's power output. During cylinder lowering, the VSD and EH-VSD system adopt throttling, eliminating the need for pump operation. The accumulator SOC remains within its operating range, starting from a maximum of 0.65, decreasing to 0.3 after the hybrid lifting mode, and then being recharged by the VSD unit to its initial state in preparation for the next cycle.

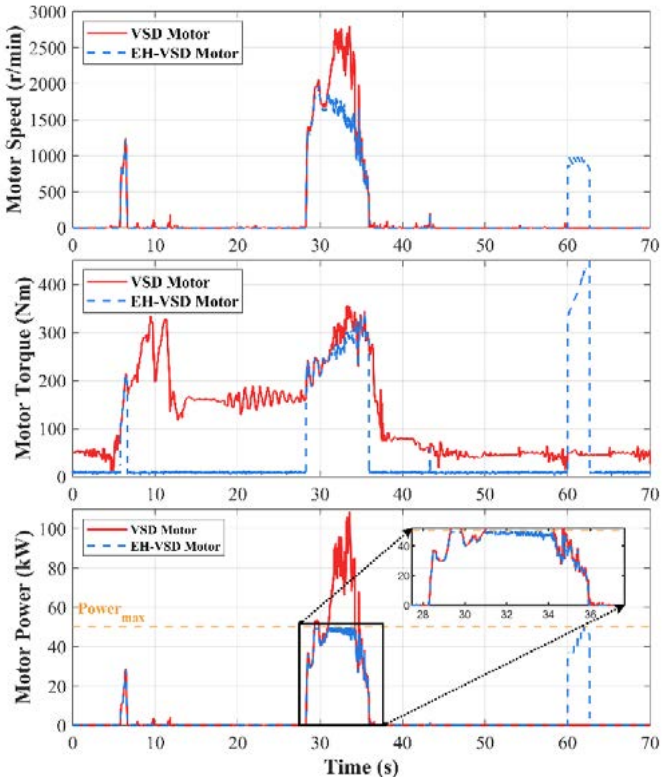
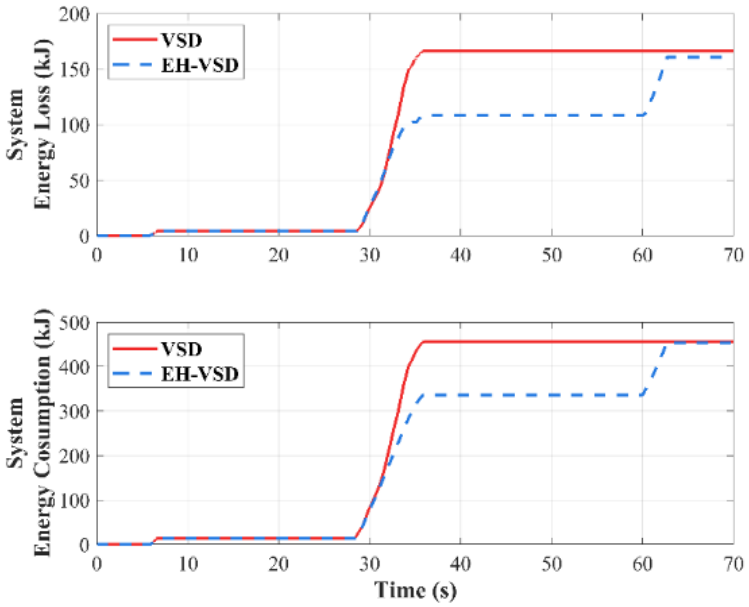


Fig. 4.2. Comparison of electric motor performance between systems

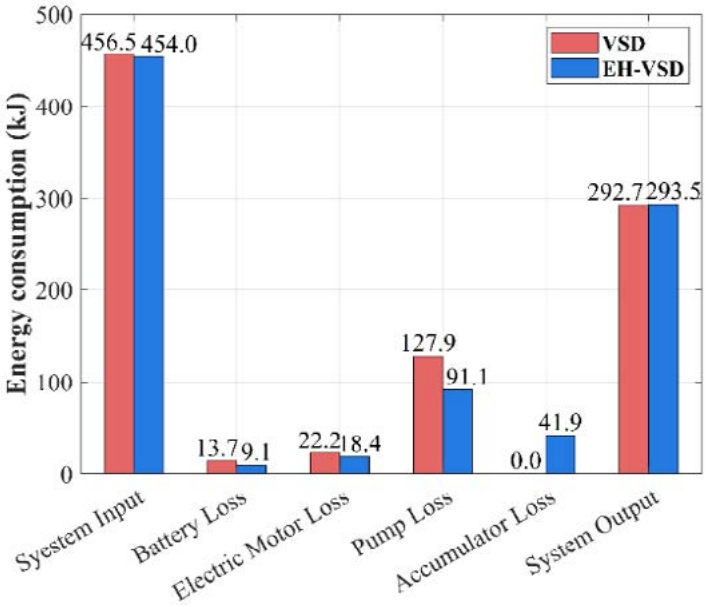
Fig. 4.2. compares the motor performance of the two systems. The motor speeds are nearly identical, except that the EH-VSD shows a lower speed during lifting since part of the flow is supplied by the accumulator. The motor torque follows a similar trend to the pump pressure, consistent with Eq. (7) and (9). The motor power of the EH-VSD system remains below the power limit of 50 kW in both lifting and charging modes, while the VSD system must supply the entire load independently, reaching a peak power of 110 kW. This represents a 54.5% reduction in installed power, confirming the feasibility of the proposed strategy in which the motor provides average power and the accumulator delivers peak power.

Fig. 4.3 compares the energy consumption of the EH-VSD and VSD systems to evaluate energy efficiency. Fig. 4.3 (a) shows system energy loss and consumption during one loading cycle. In the lifting mode, the EH-VSD system exhibits lower energy loss, attributed to the optimized power distribution strategy. During accumulator charging, EH-VSD incurs

additional loss from the VSD unit, bringing the total loss close to that of the VSD system. The final energy consumption of both systems is nearly identical, and VSD system is even higher.



(a) System energy loss and consumption



(b) Energy consumption of components

Fig. 4.3. The energy consumption comparison between VSD and EH-VSD system in one loading cycle

Fig. 4.3 (b) illustrates the component-level energy losses, including battery, motor, pump, accumulator, and other hydraulic elements (merged into pump loss). The VSD system requires higher input energy than the EH-VSD system. The EH-VSD system exhibits lower losses in the battery, motor, and pump, thereby improving overall efficiency. Although EH-VSD has accumulator charging/discharging and throttling loss, these have been minimized through optimization.

Table 4.1

Performance comparison between VSD and EH-VSD systems

	VSD	**EH-VSD**	**Comparison**
Installed Power	110 kW	50 kW	**-54.5%**
Energy Consumption	456.5 kJ	454.0 kJ	**-0.6%**
Energy Efficiency	64.1%	64.6%	**+0.5%**

As shown in Table 4.1, the performance of the two systems is compared for a single loading cycle. The proposed EH-VSD system reduces installed motor power by 54.5%, energy consumption by 0.6%, and improves energy efficiency by 0.5% compared with the VSD system, significantly enhancing overall operational performance. The EH-VSD system achieves a significant reduction in installed power while maintaining a high energy efficiency comparable to that of the VSD system, effectively avoiding substantial throttling losses.

5. CONCLUSIONS

In this paper, an electro-hydraulic hybrid variable speed drive system (EH-VSD) is proposed to downsize the electric motor of electric wheel loader working function. A high-pressure accumulator in parallel configuration assists the VSD unit under high load condition, to reduce system installed power. An energy management strategy based on the dynamic programming is developed to ensures the VSD unit operates within its high-efficiency region, improving the energy efficiency of the system. The comparison between proposed system and baseline system, and comparative simulation of different accumulator parameters are conducted in terms of system performance, installed power and energy efficiency.

Simulation results show that the installed power of EH-VSD system reduces by 54.5% compared to VSD system (from 100 kW to 50 kW). The EH-VSD system achieves a significant reduction in installed power while maintaining a high energy efficiency comparable to that of the VSD system, effectively avoiding substantial throttling losses.

The proposed EH-VSD system significantly reduces the installed power while maintaining energy efficiency. However, the complexity of system will increase due to the addition of components, heightened requirements for electro-hydraulic control precision, and the incorporation of additional control algorithms. Furthermore, this paper has only conducted simulation studies for a single cycle. Although the methodology can be applied to other

scenarios, further experimental validation across additional cycles is required to demonstrate the universality of system and control strategy.

Future work will focus on three aspects: 1) optimizing the system configuration to realize boom energy regeneration; 2) optimizing accumulator parameters to further improve energy efficiency; and 3) conducting experiments on a test bench or machine to evaluate the actual performance.

6. REFERENCES

[1] Huang XH, Huang Q, Cao HJ, Wang QY, Yan WB, Cao L. Battery capacity selection for electric construction machinery considering variable operating conditions and multiple interest claims. Energy. 2023;275:127454. DOI: 10.1016/j.energy.2023.127454.

[2] Lin TL, Lin YZ, Ren HL, Chen HB, Chen QH, Li ZS. Development and key technologies of pure electric construction machinery. Renewable & Sustainable Energy Reviews. 2020;132:110080. DOI: 10.1016/j.rser.2020.110080.

[3] Ma ZY, Liu WW, Li CC, Sang Y, Zhang YZ, Li GF, et al. Research on Energy-Saving Control Strategy of Loader Based on Intelligent Identification of Working Stages. Journal of Construction Engineering and Management. 2024;150(7):04024075. DOI: 10.1061/jcemd4.Coeng-14807.

[4] Qu SY, Fassbender D, Vacca A, Busquets E. A high-efficient solution for electro-hydraulic actuators with energy regeneration capability. Energy. 2021;216:119291. DOI: 10.1016/j.energy.2020.119291.

[5] Quan ZY, Quan L, Zhang JM. Review of energy efficient direct pump controlled cylinder electro-hydraulic technology. Renewable & Sustainable Energy Reviews. 2014;35:336-46. DOI: 10.1016/j.rser.2014.04.036.

[6] Wang F, Qiao YH, Zhou H, Wu JM, Xu B. An Independent Electro-Hydraulic Variable Speed Drive to Improve Energy Efficiency of Electric Wheel Loader Working System. Ieee Transactions on Vehicular Technology. 2024;73(12):18250-65. DOI: 10.1109/tvt.2024.3438268.

[7] Guo T, Wu B, Lin TL, Chen HH, Chen QH. Closed-Circuit Pump-Controlled Electro-Hydraulic Steering System for Pure Electric Wheel Loader. Applied Sciences-Basel. 2022;12(11):5740. DOI: 10.3390/app12115740.

[8] Qu SY, Zappaterra F, Vacca A, Busquets E. An electrified boom actuation system with energy regeneration capability driven by a novel electro-hydraulic unit. Energy Conversion and Management. 2023;293:117443. DOI: 10.1016/j.enconman.2023.117443.

[9] Qiao SF, Hao YX, Quan L, Ge L, Xia LP. A Novel Electro-Hydraulic Compound Driving System With Potential Energy Regeneration Capability for Lifting Device. Ieee Access. 2022;10:18248-56. DOI: 10.1109/access.2022.3149515.

[10] Xue LX, Wu S, Xu YZ, Ma DL. A Simulation-Based Multi-Objective Optimization Design Method for Pump-Driven Electro-Hydrostatic Actuators. Processes. 2019;7(5):274. DOI: 10.3390/pr7050274.

[11] Trinh HA, Truong HVA, Do TC, Nguyen MH, Phan VD, Ahn KK. Optimization-based energy management strategies for hybrid construction machinery: A review. Energy Reports. 2022;8;6035-6057. DOI: 10.1016/j.egyr.2022.04.050.

[12] Yang J, Liu B, Zhang TZ, Hong JC, Zhang HX. Application of energy conversion and integration technologies based on electro-hydraulic hybrid power systems: A

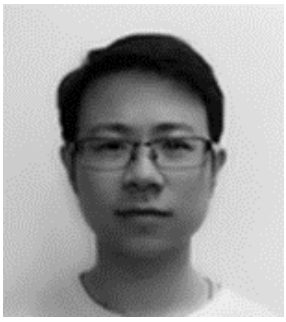
review. Energy Conversion and Management. 2022;272:116372. DOI: 10.1016/j.enconman.2022.116372.
[13] Nguyen TH, Do TC, Nguyen MH, Ahn JH, Ahn KK. Energy consumption reduction technologies in hybrid hydraulic construction machine: a review. 2026;225;116150. DOI: 10.1016/j.rser.2025.116150.
[14] Kang RJ, Jiao ZX, Wang SP, Chen LS. Design and Simulation of Electro-hydrostatic Actuator with a Built-in Power Regulator. Chinese Journal of Aeronautics. 2009;22(6):700-6. DOI: 10.1016/s1000-9361(08)60161-2.
[15] Shang YX, Li XB, Qian H, Wu S, Pan QX, Huang LG, et al. A Novel Electro Hydrostatic Actuator System With Energy Recovery Module for More Electric Aircraft. Ieee Transactions on Industrial Electronics. 2020;67(4):2991-9. DOI: 10.1109/tie.2019.2905834.
[16] Seo HT, Hong YP, Park J, Kim KS. Improving the Response of an Electrohydrostatic Actuation (EHA) System Using a High-Pressure Accumulator. Ieee Access. 2022;10:93353-61. DOI: 10.1109/access.2022.3202933.
[17] Wang F, Wu JM, Lin ZC, Zhang HX, Xu B. A power-sharing electro-hydraulic actuator system to downsize electric motors for electric mobile machines. Energy. 2023;284:129343. DOI: 10.1016/j.energy.2023.129343.
[18] Jiang Y, Jiang ZP. Robust Adaptive Dynamic Programming and Feedback Stabilization of Nonlinear Systems. Ieee Transactions on Neural Networks and Learning Systems. 2014;25(5):882-93. Doi: 10.1109/tnnls.2013.2294968.
[19] Nurmi J, Mattila J. Global energy-optimised redundancy resolution in hydraulic manipulators using dynamic programming. Automation in Construction. 2017;73:120-34. Doi: 10.1016/j.autcon.2016.09.006.
[20] Casoli P, Gambarotta A, Pompini N, Riccó L. Hybridization methodology based on DP algorithm for hydraulic mobile machinery - Application to a middle size excavator. Automation in Construction. 2016;61:42-57. DOI: 10.1016/j.autcon.2015.09.012.
[21] Zeng X, Wang J. Globally energy-optimal speed planning for road vehicles on a given route. Transportation Research Part C-Emerging Technologies. 2018;93:148-60. DOI: 10.1016/j.trc.2018.05.027.

Biographies



Yihan Qiao received B.S. degree in mechanical engineering from Zhejiang University, Hangzhou, China, in 2022. He is currently pursuing his Ph.D. degree in mechanical

engineering at Zhejiang University, Hangzhou, China. His main research focuses on electro-hydraulic hybrid variable speed drive system for wheel loader.



Jiaming Wu received B.S. degree in mechanical engineering from Zhejiang University, Hangzhou, China, in 2019. He is currently pursuing his Ph.D. degree in mechanical engineering at Zhejiang University, Hangzhou, China. His main research focuses on modeling and control of electro-hydrostatic actuator system for off-road vehicles.



Feng Wang received B.S., M.S., and Ph.D. degrees in mechanical engineering from Zhejiang University, Hangzhou, China, in 2003, 2005, and 2009. In 2009 he became a Postdoctoral Associate in the Department of Mechanical Engineering at University of Minnesota, Minneapolis, MN, USA, where he joined the NSF Engineering Research Center for Compact and Efficient Fluid Power (CCEFP). In 2016 he became an Assistant Professor in the School of Mechanical Engineering at Zhejiang University, Hangzhou, China, where he joined the State Key Laboratory of Fluid Power and Mechatronics Systems. His research interests include modeling and control of intelligent hydraulic powertrains for heavy-duty vehicles and large utility wind turbines.



Bing Xu received Ph.D. degree in mechanical engineering from Zhejiang University, Hangzhou, China, in 2001. He is currently a Professor in the school of mechanical engineering and also the Director of the State Key Laboratory of Fluid Power Transmission and Control at Zhejiang University. He has authored or co-authored more than 200 journal and conference papers, and authorized 49 patents. His research interests include fluid power components and systems, mechatronic systems design and motion control for mobile machinery. Prof. Xu is a Chair Professor of the Yangtze River Scholars Program and a science and technology innovation leader of the Ten Thousand Talent Program.

Motion Control of Multi-DOF Hydraulic Manipulator Driven by Independent Metering System with Friction Compensation

Yangxiu Xia[1], Litong Lyu[2,*], Bobo Helian[3], Zheng Chen[1]

[1] *Ocean College, Zhejiang University, 1 Zheda Road, Zhoushan, Zhejiang 316021, P. R. China*
[2] *School of Mechanical Engineering, Shijiazhuang Tiedao University, 17 North Second Ring East Road, Shijiazhuang, Hebei 050043, P. R. China*
[3] *Institute of Mobile Machines, Karlsruhe Institute of Technology, Kaiserstr. 12, 76131 Karlsruhe, Germany*

* Corresponding author: Tel.: +86-15267008148; E-mail address: litong_lyu@stdu.edu.cn

Abstract.

Multi-degree-of-freedom hydraulic manipulators are widely used in heavy-duty machinery, due to the advantages of hydraulic transmission. While previous research focused on output force and power, recent developments have shifted toward control performance and energy efficiency. However, precise control remains challenging for hydraulic manipulators compared to electric manipulators, due to the complexities of hydraulic transmission and multi-body dynamics. Nonlinear model-based control strategies and independent metering techniques have shown promise in overcoming these challenges. This paper investigates the motion control of a multi-DOF hydraulic manipulator driven by an independent metering system. A hydraulic circuit based on independent metering is designed, and the system's dynamics are modeled. To address nonlinearities, uncertainties, and friction in the system, an adaptive robust control strategy is proposed. Friction, which plays a significant role in degrading control performance, is compensated using the LuGre model-based approach. The proposed method is evaluated through simulations and experiments on a 4-DOF hydraulic manipulator test rig. Quantitative analysis, including the Integral of Time-weighted Absolute Error (ITAE) and maximum tracking errors, verifies that the proposed strategy achieves improved tracking performance and energy efficiency compared to traditional control methods.

Keywords. Independent metering system, hydraulic manipulator, motion control, friction compensation

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Multiple degrees of freedom (multi-DOF) hydraulic manipulators are typically installed on heavy-duty machinery like excavators [1] and Autonomous Underwater Vehicles (AUVs) [2] due to the benefits offered by hydraulic transmission systems. Historically, the primary focus regarding hydraulic manipulators has been on high output force/power [3], with limited attention given to studying control performance. This is because most machines were traditionally operated by operators controlling the valve orifices directly, without the application of closed-loop control systems. Nowadays, the control performance has emerged as a primary concern for hydraulic manipulators, given that fully-autonomous or semi-autonomous off-road machinery represents a significant research focus in numerous countries within the framework of intelligent development [4].

Unlike electric manipulators utilized in industrial applications, which employ servo motors to control the joints and therefore achieve higher control performance relatively easily, multi-DOF hydraulic manipulators typically utilize cylinders to drive the rotary joints instead of hydraulic motors [4]. As a result, precise control of hydraulic manipulators is more challenging due to the complex nature of hydraulic transmission and the multi-body mechanism. The dynamics of multi-DOF hydraulic manipulator is highly nonlinear with various uncertainties [5]. To attain high precision, nonlinear model-based control methods such as adaptive control or robust control have been proven effective [6, 7]. These control strategies aim to compensate for the nonlinear dynamics as accurately as possible and incorporate robust feedback mechanisms to address uncertainties and disturbances. Notably, friction represents a key nonlinear property within the dynamics of hydraulic manipulators [8]. Friction exhibits a complex relationship with velocity, particularly at low speeds [9]. Due to the kinematics of the hydraulic manipulator, when the end-effector is moving continuously and relatively fast, some joints may experience frequent stop-and-go or low speed motions. Inadequate compensation for such behaviors can lead to significant peak errors in the joints and degrade overall control performance.

Typically, the friction force is represented as a combination of Coulomb and viscous terms [10]. This model is accurate when the velocity is relatively high and can be easily integrated into model compensation control. However, as the velocity decreases, the discrepancy between this model and the actual behavior can be increased, since the real friction force exhibits a pronounced Stribeck effect at low speeds [11, 12]. Thus, extra attention should be paid to make more effective compensation for the nonlinear friction in order to enhance the control performance of multi-DOF hydraulic manipulator. Among the existing models, LuGre friction model [13] has been demonstrated to be effective in various actuation systems such as linear motors [14], PMSM servo systems [15] and hydraulic cylinders [16]. However, few studies can be found about applying LuGre model in hydraulic manipulators. Considering the complex dynamics of hydraulic manipulator itself, designing a model-based controller is still challenging.

In addition to control precision, energy efficiency is also important for hydraulic manipulators [4]. In most hydraulic manipulators, each hydraulic cylinder is controlled by a single proportional or servo valve. The hydraulic circuit is quite straightforward, and such four-way valve ensures the stability of the internal dynamics within the hydraulic system [10]. However, the inlet and outlets orifices of the actuator is coupled strongly by the mechanical linkage of the valve spool. This linkage restricts the control to a single control input for a given actuator, which makes it unfeasible for energy saving. In the field of electro-hydraulics, independent metering technique has been a solution for addressing energy efficiency concerns [17]. With in independent metering systems, each actuator is

controlled by two independent valves at least, which enables independent metering of the inlet and outlet orifices. This additional control input facilitates simultaneous energy savings through strategic control design. Although in many researches direct pump control becomes a good choice when energy saving is a major concern [18], the control performance of independent metering system can be much better since the actuator is still controlled by the high-response proportional or servo valves [19]. Implementing independent metering systems in hydraulic manipulators could offer both superior control performance and improved energy efficiency. However, the corresponding control flexibility should be handled properly by the controller [20].

To achieve high precise motion tracking performance while improve the energy efficiency, motion control of the hydraulic manipulator driven by an independent metering system is studied. The hydraulic circuit is designed based on the independent metering technique and the dynamics of the system is modelled. Then, the desired load force ensuring the tracking performance is synthesized based on the adaptive robust control theory [10], additionally considering nonlinear friction compensation. To address the flexibility of the independent metering configuration, a mode selection method is proposed to determine the control objective for each cylinder chamber. Motion and pressure controllers are then designed for both chambers. Finally, the control performance is evaluated theoretically and validated through comparative simulations and experiments.

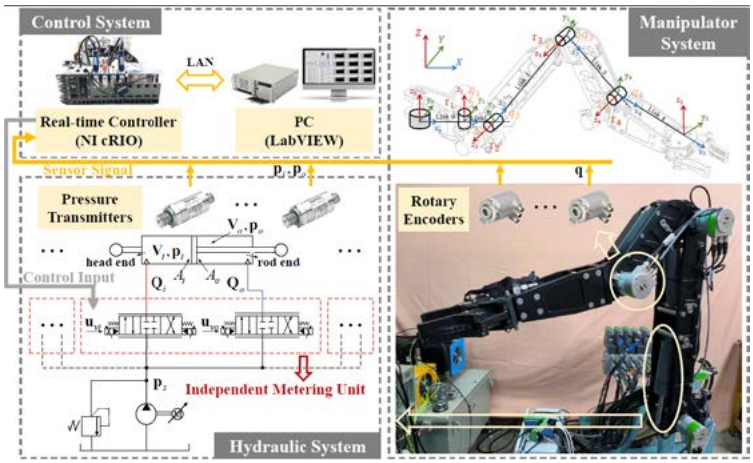


Figure 1: The common mechanism of multi-DOF hydraulic manipulator

2. HARDWARE CONFIGURATION AND PROBLEM FORMULATION

2.1. Hardware Configuration

A hydraulic manipulator with cylinders driving the joints is given in Fig. 1, illustrating the common mechanism of multi-DOF hydraulic manipulator. The corresponding coordinates are defined following the typical processes. Usually, each cylinder is controlled by a single four-way proportional or servo valve [4]. Instead, in this study, the independent metering technique is implemented in the hardware configuration. For example, for the n-th cylinder, two proportional four way valves, namely V_{in} and V_{on} are used coordinately to control the inlet and outlet flows, denoted as Q_{in} and Q_{on}, respectively. These two valves

collectively form an independent metering unit as shown in Fig. 1. It is important to mention that in independent metering systems, controlling the pump to match the flow required by the actuator is also crucial for energy savings. However, in this study, our primary focus will be on the control of independent metering valves.

2.2. Modeling of Hydraulic Manipulator

The model of the hydraulic manipulator is constructed as

$$\begin{aligned} &M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = \tau + D_F \\ &\tau = J_h(q)F_L - T_f, F_L = A_i P_i - A_o P_o \end{aligned} \tag{1}$$

where $q \in \mathbb{R}^n$ is the joint angle. $M(q) \in \mathbb{R}^{n\times n}$, $C(q,\dot{q}) \in \mathbb{R}^{n\times n}$, and $G(q) \in \mathbb{R}^n$ denote the inertia matrix, the centripetal and Coriolis matrix, and the gravity terms, respectively. $D_F \in \mathbb{R}^n$ is the lumped modeling error including external disturbances and other hard-to-model terms, such as the non-uniform mass distribution across each link of the hydraulic manipulator. $J_h(q) \in \mathbb{R}^{n\times n}$ represents the nonsingular joint Jacobian matrix and $T_f \in \mathbb{R}^n$ means the friction torque. $F_L \in \mathbb{R}^n$ denotes the load force, with $A_i \in \mathbb{R}^{n\times n}$ and $A_o \in \mathbb{R}^{n\times n}$ being the head and rod-end ram areas of each cylinder. $P_i \in \mathbb{R}^n$ and $P_o \in \mathbb{R}^n$ represent the pressure of the two chambers, whose dynamics can be further expressed as

$$\begin{aligned} &V_i(x_L)\beta_{ei}^{-1}\dot{P}_i = -A_i\dot{x}_L + Q_i + D_{Qi} \\ &V_o(x_L)\beta_{eo}^{-1}\dot{P}_o = A_o\dot{x}_L - Q_o - D_{Qo} \end{aligned} \tag{2}$$

where $\dot{x}_L = J_h\dot{q}$, with x_L being the cylinder displacement. $\beta_{ei} \in \mathbb{R}$ and $\beta_{eo} \in \mathbb{R}$ denote the effective bulk modulus of the two chambers. $V_i(x_L) \in \mathbb{R}^{n\times n}$ and $V_o(x_L) \in \mathbb{R}^{n\times n}$ are the total compressible volumes of the head and rod end, respectively. $D_{Qi} \in \mathbb{R}^n$ and $D_{Qo} \in \mathbb{R}^n$ denote the lumped modeling errors, which include leakages in the hydraulic circuit and other minor nonlinear characteristics. Owing to the independent metering unit, Q_i and Q_o are two independent control inputs. In practice, Q_i and Q_o need to be reasonably distributed to obtain control signals for each valve core. Note that the dynamic response of the valve is much higher than other parts of the system, so only the static mapping of Q_i and Q_o to the valve control signal is considered. Therefore, this step will be omitted in the modeling process in this paper, and the general problem of controlling the motion of hydraulic manipulator with Q_i and Q_o will be studied.

The friction torque T_f is given by

$$T_f = Z\sigma_s + \dot{Z}\bar{\sigma}_m(\dot{q}) + \breve{\dot{q}}f_v, \dot{Z} = \dot{q} - \left|\dot{q}\right| g^{-1}(\dot{q})Z \tag{3}$$

where $Z \in \mathbb{R}^{n\times n}$ means the internal friction state and $\breve{\dot{q}} = diag(\dot{q})$. $\sigma_s \in \mathbb{R}^n$, $\bar{\sigma}_m(\dot{q}) = h_m(\dot{q})\sigma_m \in \mathbb{R}^n$, and $f_v \in \mathbb{R}^n$ are the stiffness, a velocity dependent damping coefficient, and the viscous friction coefficient, respectively. $g(\dot{q}) \in \mathbb{R}^{n\times n}$ is a positive diagonal matrix describing the Stribeck effect, in which

$$g_i(\dot{q}_i) = \alpha_{ci} + (\alpha_{si} - \alpha_{ci})e^{-(\dot{q}_i/\dot{q}_{si})^2} \tag{4}$$

with $\sigma_{si}\alpha_{si}$ and $\sigma_{si}\alpha_{ci}$ corresponding to the static friction and the Coulomb friction, respectively. $\dot{q}_{si}$ is the Stribeck velocity.

In addition, though the lumped modeling error $D_\bullet$ in (1) and (2) is unknown and time-varying, it can be split into the nominal value $D_{\bullet n}$ with slow variation and the bounded deviation value $\Delta D_\bullet$, i.e., $D_\bullet$ can be further denoted by

$$\begin{aligned} D_F &= D_{Fn} + \Delta D_F \\ D_{Qi} &= D_{Qin} + \Delta D_{Qi} \\ D_{Qo} &= D_{Qon} + \Delta D_{Qo} \end{aligned} \tag{5}$$

Property 1. The inertia matrix $M(q)$ is a positive definite symmetric matrix, and satisfies

$$\underline{\lambda}_M I_n \le M(q) \le \overline{\lambda}_M I_n \tag{6}$$

with $\underline{\lambda}_\bullet$ and $\overline{\lambda}_\bullet$ being the minimum and maximum eigenvalues of $\bullet$, respectively.

Property 1. The matrix $N(q,\dot{q}) = \dot{M}(q) - 2C(q,\dot{q})$ is skew-symmetric and

$$\dot{M}(q) = C(q,\dot{q}) + C^T(q,\dot{q}) \tag{7}$$

2.3. Dynamics parameterizing

For the convenience of subsequent controller derivation, the model in Section 2.1 can be parameterized as

$$\begin{aligned} M\ddot{q} + C\dot{q} &= J_h F_L - G - T_f(Z, \vartheta_{s1}) + \theta_4 + \Delta D_F \\ V_i(x_L)\theta_5 \dot{P}_i &= -A_i \dot{x}_L + Q_i + \theta_6 + \Delta D_{Qi} \\ V_o(x_L)\theta_7 \dot{P}_o &= A_o \dot{x}_L - Q_o - \theta_8 - \Delta D_{Qo} \end{aligned} \tag{8}$$

where $\vartheta_{s1} \triangleq [\theta_1^T, \theta_2^T, \theta_3^T, \theta_4^T]^T$, $\vartheta_{s2} \triangleq [\theta_5^T, \theta_6^T]^T$, and $\vartheta_{s3} \triangleq [\theta_7^T, \theta_8^T]^T$, in which θ_j means the parameter set of the j^{th} dynamics parameter at each joint, that is

$$\begin{aligned} \theta_j &= [\pi_1(j), \cdots, \pi_i(j), \cdots, \pi_n(j)]^T \\ \pi_i &= [\sigma_{si}, \sigma_{mi}, f_{vi}, D_{Fni}, , \beta_{ei}^{-1}, D_{Qini}, \beta_{eo}^{-1}, D_{Qoni}]^T \end{aligned} \tag{9}$$

with $\pi_i(j)$ being the j^{th} element of π_i. $T_f(\dot{q}, \vartheta_{s1}) = \phi_f^T \vartheta_{s1}$, and $\phi_f^T = \left[Z, \dot{Z}\, h_m, \breve{\dot{q}}, 0_{n\times 1} \right]$ is a linear regression matrix.

Assumption 1. The bounds of parametric uncertainties and uncertain nonlinearities are known, i.e.,

$$\vartheta_{si} \in \Omega_{\vartheta_{si}} \triangleq \{\vartheta_{si} : \vartheta_{si\min} \le \vartheta_{si} \le \vartheta_{si\max}\}$$
$$|\Delta D_F| \le \delta_F, |\Delta D_{Qi}| \le \delta_{Qi}, |\Delta D_{Qo}| \le \delta_{Qo} \tag{10}$$

where $\vartheta_{si\min}$, $\vartheta_{si\max}$, δ_F, δ_{Qi}, and δ_{Qo} are known scalars.

2.4. Control Objectives

Since independent metering technique is used for hydraulic manipulator, the controller should achieve high tracking performance while effectively managing the additional flexibility to enhance energy efficiency. The dual design objectives can be specifically described as:

1) High precision motion control: given a set of desired trajectories $q_d(t)$, the primary objective is to synthesize the valve spool voltage u_v, so that each joint angle $q(t)$ can track $q_d(t)$ as closely as possible, especially under low-speed conditions.

2) Low energy consumption: utilize the dual independent control valves for each actuator to maintain chamber pressures at low levels to conserve energy.

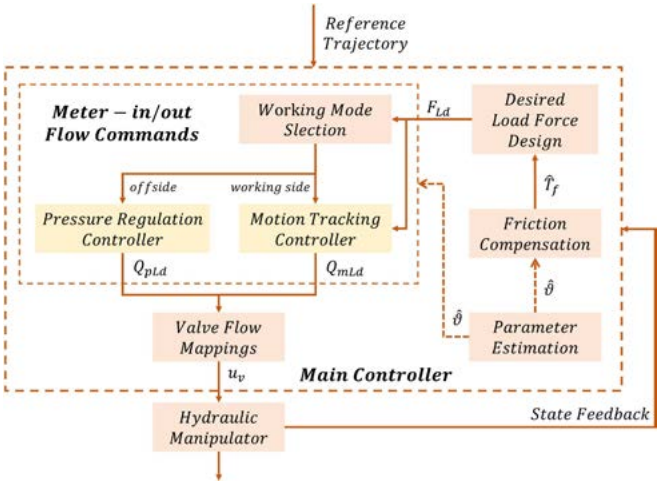


Figure 2: The overall control framework.

3. CONTROLLER DESIGN

In order to achieve the dual control objectives, the working mode selection is applied following the standard procedures proposed in [21]. In the selected working mode, the offside (to keep the chamber at a low pressure) and working side(motion tracking) independently control the pressure of the two chambers based on adaptive robust control theory [10]. The specific control framework is shown in Fig. 2.

The following notations are used throughout this paper: $\hat{\bullet}$ denotes the estimate of $\bullet$ and $\tilde{\bullet} = \hat{\bullet} - \bullet$ denotes the parameter estimation error.

3.1. Desired Load Force Design

This subsection aims to achieve motion tracking of the hydraulic manipulator and calculate the desired load force of the hydraulic cylinder F_{Ld} so that the angle tracking error $e_1 = q - q_d$ can be kept as small as possible.

Define a switching-function-like quantity

$$e_2 = \dot{e}_1 + K_1 e_1 = \dot{q} - \dot{q}_{eq}, \dot{q}_{eq} \square \dot{q}_d - K_1 e_1 \tag{11}$$

with K_1 being a positive gain matrix. As long as the $e_2 \to 0$, the tracking error e_1 will be very small or converge to zero.

Differentiating (11) and noting (8), it has

$$\begin{aligned} M\dot{e}_2 + Ce_2 = J_h F_{Ld} + J_h e_3 - G - T_f(Z, \vartheta_{s1}) \\ +\theta_4 + \Delta D_F - (M\ddot{q}_{eq} + C\dot{q}_{eq}) \end{aligned} \tag{12}$$

with $e_3 = F_L - F_{Ld}$.

Considering that Z is unmeasurable, a dual observer based on mapping functions is designed as

$$\begin{aligned} \dot{\hat{z}}_1 &= \mathrm{Proj}_{\hat{z}_1}\left\{ x_2 - \frac{|x_2|}{g(x_2)}\hat{z}_1 - \gamma_1 e_2 \right\} \\ \dot{\hat{z}}_2 &= \mathrm{Proj}_{\hat{z}_2}\left\{ x_2 - \frac{|x_2|}{g(x_2)}\hat{z}_2 + \gamma_2 \frac{h(x_2)|x_2|}{g(x_2)} e_2 \right\} \end{aligned} \tag{13}$$

The virtual control force F_{Ld} can be designed as

$$\begin{aligned} F_{Ld} &= F_{Lda} + F_{Lds} \\ F_{Lda} &= F_{Lda1} + F_{Lda2},\ F_{Lds} = F_{Lds1} + F_{Ldas2} \\ \cdot F_{Lda1} &= J_h^{-1}\left(M\ddot{q}_{eq} + C\dot{q}_{eq} + G + \hat{T}_f(\square \hat{Z}, \hat{\vartheta}_{s1}) - \hat{\theta}_4 \right)\cdot \\ F_{Lds1} &= -J_h^{-1} K_{2s1} e_2, F_{Lds2} = -J_h^{-1} K_{2s2} e_2 \end{aligned} \tag{14}$$

in which K_{2s1} is a sufficiently large nonlinear gain matrix.

Consolidate the model compensation errors resulting from uncertain nonlinearities and physical parameter estimation discrepancies, and then separate them into a static component d_1 and a high-frequency component $\square d_1^*(t)$

$$d_1 + \square d_1^*(t) \square \Delta D_F - \phi_F^T \tilde{\vartheta}_{s1} \tag{15}$$

Then, design fast dynamics compensation F_{Lda2} as

$$F_{Lda2} = -J_h^{-1}\hat{d}_1, \dot{\hat{d}}_1 = \mathrm{Proj}_{\hat{d}}(\gamma_F e_2) \tag{16}$$

in which γ_F is a positive constant matrix, and $\mathrm{Proj}_{\hat{d}}(\bullet)$ is a projection mapping.

K_{2s2} is a nonlinear robust feedback gain chosen large enough such that F_{Lds2} satisfies the following robust performance conditions:

$$
\begin{aligned}
&(i)\ \ e_2^T J_h F_{Lds2} \le 0 \\
&(ii)\ e_2^T (J_h F_{Lds2} - \tilde{d}_1 + \tilde{d}_1^* \\
&\quad + \theta_1 \bar{Z}_1 - \theta_2 h_m \left|\dot{q}\right| g^{-1} \bar{Z}_2) \le \eta_1
\end{aligned} \tag{17}
$$

with $\eta_1 > 0$ being an arbitrarily small parameter.

Using (14) and (16), the resulting error dynamics of e_2 can be written as

$$
\begin{aligned}
M\dot{e}_2 + Ce_2 = &-K_{2s1}e_2 + \theta_1 \bar{Z}_1 - \theta_2 h_m \left|\dot{q}\right| g^{-1} \bar{Z}_2 \\
&+ J_h e_3 + \left(J_h F_{Lds2} - \tilde{d}_1 + \bar{d}_1^*\right)
\end{aligned} \tag{18}
$$

3.2. Pressure Regulation Controller

Hydraulic cylinders need to generate the desired load force mentioned above to achieve motion control goals, but the independent form of the load port allows for independent control of the pressure in both chambers, resulting in countless combinations of chamber pressures. According to the direction of the load force, when a forward load force is required, the pressure in the rod end chamber is controlled at a small value, while when a reverse load force is required, the pressure in the head chamber is controlled at a small value p_c, which can effectively achieve the goal of maintaining low pressure [21]. This subsection takes the rod end chamber as the offside as an example to design a pressure regulation controller. The pressure design for the head chamber follows the same procedure and is omitted here.

Define the tracking error of pressure as $e_{3p} = P_o - p_c$. The error dynamics would be the same as the pressure dynamics because p_c is constant

$$
\theta_7 \dot{e}_{3p} = V_o^{-1} \left(A_o \dot{x}_L - Q_o - \theta_8 - \Delta D_{Qo}\right) \tag{19}
$$

with Q_o being the control input, the proposed control law is given by

$$
\begin{aligned}
Q_{oLd} &= Q_{oLda} + Q_{oLds} \\
Q_{oLda} &= Q_{oLda1} + Q_{oLda2} \\
Q_{oLds} &= Q_{oLds1} + Q_{oLdas2} \\
Q_{oLda1} &= A_o \dot{x}_L - \hat{\theta}_8, \ \ Q_{oLda2} = V_o \hat{d}_p \\
Q_{oLds1} &= K_{os1} V_o e_{3p}, \ \ Q_{oLds2} = K_{os2} V_o e_{3p}
\end{aligned} \tag{20}
$$

where Q_{oLda1} is the model compensation term and $\hat{d}_p$ is a fast dynamics compensation with $d_p + \bar{d}_p^*(t) \in V_o^{-1}\left(\tilde{\theta}_8 - \Delta D_{Qo}\right)$ and $\dot{\hat{d}}_p = \mathrm{Proj}_{\hat{d}}(\gamma_{Qo} e_{3p})$. $K_{os1} > 0$ and K_{os2} is a nonlinear feedback gain chosen to satisfy the following condition for performance robustness to model uncertainties:

$$
\begin{aligned}
&(i)\ \ -e_{3p}^T V_o^{-1} Q_{oLds2} \le 0 \\
&(ii)\ e_{3p}^T \left(-V_o^{-1} Q_{oLds2} - \tilde{d}_p + \bar{d}_p^*\right) \le \eta_p
\end{aligned} \tag{21}
$$

with $\eta_p > 0$ being an arbitrarily small parameter.

Substituting control law (20) into (19), the error dynamics can be written as

$$\theta_7 \dot{e}_{3p} = -K_{os1} e_{3p} + \left(\bar{d}_p^* - \tilde{d}_p - V_o^{-1} Q_{oLds2} \right) \tag{22}$$

3.3. Motion Tracking Controller

The pressure regulation controller has already calculated the control flow Q_{oLd}. Furthermore, it is necessary to calculate the control flow rate based on the desired load force to achieve accurate motion tracking in the closed-loop system.

The error between the real load force and the desired load force is $e_{3m} = F_L - F_{Ld}$, whose derivative is

$$\begin{aligned} \theta_5 \dot{e}_{3m} &= \theta_5 \dot{F}_L - \theta_5 \dot{F}_{Ld} \\ &= -(V_i^{-1} A_i^2 + V_o^{-1} A_o^2) \dot{x}_L + V_i^{-1} A_i Q_i + V_o^{-1} A_o Q_o \\ &\quad + V_i^{-1} A_i \theta_6 + V_o^{-1} A_o \theta_8 + A_i \Delta D_{Qi} - A_o \Delta D_{Qo} - \dot{F}_{Ld} \end{aligned} \tag{23}$$

where $\dot{F}_{Ld}$ can be further expressed as the incalculable part $\dot{F}_{Ldi}$ and calculable part $\dot{F}_{Ldc}$.

Similar to Q_{oLd}, Q_{iLd} in (23) can be generated as

$$\begin{aligned} &Q_{iLd} = Q_{iLda} + Q_{iLds} \\ &Q_{iLda} = Q_{iLda1} + Q_{iLda2} \\ &Q_{iLds} = Q_{iLds1} + Q_{iLdas2} \\ &Q_{iLda1} = V_i A_i^{-1} [(V_i^{-1} A_i^2 + V_o^{-1} A_o^2) \dot{x}_{\mathrm{L}} - V_o^{-1} A_o Q_o \\ &-V_i^{-1} A_i \hat{\theta}_6 - V_o^{-1} A_o \hat{\theta}_8 + \dot{F}_{\mathrm{Ldc}} + \mathrm{Y}_{Qm}] \\ &Q_{iLda2} = -V_i A_i^{-1} \hat{d}_m \\ &Q_{iLds1} = -k_{\mathrm{is1}} V_i A_i^{-1} e_{3m},\ Q_{iLds2} = -k_{\mathrm{is2}} V_i A_i^{-1} e_{3m} \end{aligned} \tag{24}$$

with k_{is1} being a gain matrix and $\mathrm{Y}_{Qm} = -\omega_1 \omega_2^{-1} J_h e_2$. $\hat{d}_m$ is a fast dynamics compensation with $\dot{\hat{d}}_m = \mathrm{Proj}_{\hat{d}}(\gamma_{Qi} e_{3m})$. k_{is2} is a nonlinear robust feedback gain chosen large enough such that Q_{Lds2} satisfies the following robust performance conditions:

$$\begin{aligned} &(i)\ \ e_{3m}^T Q_{iLds2} \le 0 \\ &(ii)\ e_{3m}^T \left(V_o^{-1} A_o Q_{iLds2} - \tilde{d}_2 + \bar{d}_2^* \right) \le \eta_m \end{aligned} \tag{25}$$

with $\eta_m > 0$ being an arbitrarily small parameter.

Substituting control law (24) into (23), the error dynamics can be written as

$$\dot{e}_{3m} = -k_{\mathrm{is1}} z_3 + \left(V_o^{-1} A_o Q_{Lds2} - \tilde{d}_2 + \bar{d}_2^* \right) + \mathrm{Y}_{Qm} \tag{26}$$

3.4. Parameter Estimation

The parameter estimation law is given as

$$\dot{\hat{\vartheta}}_{si} = \mathrm{sat}_{\hat{\vartheta}}(\mathrm{Proj}_{\hat{\vartheta}}(\Gamma_i \tau_i)) \tag{27}$$

The projection function $\mathrm{Proj}_{\hat{\vartheta}}(\bullet)$ is defined as

$$\mathrm{Proj}_{\hat{\vartheta}}(\bullet)=\begin{cases}\bullet, & \hat{\vartheta}\in\Omega_{\vartheta}\ \text{or}\ n_{\hat{\vartheta}}^{T}\bullet\le 0\\ \left(I-\Gamma\dfrac{n_{\hat{\vartheta}}n_{\hat{\vartheta}}^{T}}{n_{\hat{\vartheta}}^{T}\Gamma n_{\hat{\vartheta}}}\right)\bullet, & \hat{\vartheta}\in\partial\Omega_{\vartheta}\ \text{and}\ n_{\hat{\vartheta}}^{T}\bullet\ge 0\end{cases} \tag{28}$$

where Ω_{ϑ}, $\partial\Omega_{\vartheta}$, and $n_{\hat{\vartheta}}$ are the interior and the boundary of Ω_{ϑ} and the outward unit normal vector at $\hat{\vartheta}\in\partial\Omega_{\vartheta}$.

Define regressors as

$$\begin{aligned}\phi_1^T &= \left[-Z_d, -\dot{Z}_d\ h_m, -\dot{q}, I_{n\times 1}\right]\\ \phi_2^T &= [-\dot{P}_i, V_i^{-1}]^T\\ \phi_3^T &= [-\dot{P}_o, -V_o^{-1}]^T\\ u_1 &= J_h F_L - M\ddot{q} - C\dot{q} - G\\ u_2 &= V_i^{-1}(-A_i\dot{x}_L + Q_i)\\ u_3 &= V_o^{-1}(A_o\dot{x}_L - Q_o)\end{aligned} \tag{29}$$

A stable filter $H_f(s)$ with a relative degree no less than one is used to handle the sensor noise. Noting (29) we can get

$$u_{if} = -\phi_{if}^T\vartheta_{is} \tag{30}$$

Define the prediction out and prediction error as

$$\begin{aligned}\hat{u}_{if} &= \phi_{if}^T\hat{\vartheta}_{is}\\ \dot{\varsigma}_i &= \hat{u}_{if} - u_{if} = \phi_{if}^T\tilde{\vartheta}_{is}\end{aligned} \tag{31}$$

$\Gamma_i(t)$ is updated by

$$\dot{\Gamma}_i=\begin{cases}\alpha\Gamma_i-\Gamma_i\phi_{if}\phi_{if}^T\Gamma_i, & \text{if } \lambda_{\max}(\Gamma_i(t))\le\rho_M\\ 0, & \text{else}\end{cases} \tag{32}$$

in which ρ_M is the upper bound of $\Gamma_i(t)$, α is the forgetting factor, and $\nu\ge 0$. The adaptation function $\tau_i=\phi_{if}\dot{\varsigma}_i$.

Theorem 1. All signals in the system controller are bounded. Furthermore, the positive-definite function V_s defined by

$$V_s=\frac{1}{2}e_1^Tw_1K_1^2e_1+\frac{1}{2}e_2^Tw_1Me_2+\frac{1}{2}e_{3m}^T\theta_5e_{3m}+\frac{1}{2}e_{3p}^Tw_2\theta_7e_{3p} \tag{33}$$

is bounded above by

$$V_s(t)\le\exp(-\lambda_s t)V_s(0)+\frac{\eta_s}{\lambda_s}\left[1-\exp\left(-\lambda_s t\right)\right] \tag{34}$$

with $\lambda_s=\min\left\{\underline{\lambda}_{K_1}, 2\underline{\lambda}_{K_{2s1}}/\overline{\lambda}_M, 2\underline{\lambda}_{K_{is1}}/\theta_{5max}, 2\underline{\lambda}_{K_{os1}}/\theta_{7max}\right\}$ and $\eta_s=w_1\eta_1+w_2\eta_m+\eta_p$. In addition, if there is only parameter uncertainties in the system after a finite time, the

closed-loop system can achieve asymptotic tracking performance, i.e., the tracking errors $e_1 \to 0 \quad \text{as} \, t \to \infty$.

4. SIMULATIONS AND EXPERIMENTS

To verify the feasibility and effectiveness of the proposed control strategy, both simulations and experiments were conducted. The simulations were designed to evaluate the controller performance under controlled and reproducible conditions, while the experiments further validated its applicability in real hardware environments.

4.1. Simulations

In the simulation, the multi-DOF coupling effect, sensor measurement noise, and the dynamic process of the valve simulated by a second-order system are all considered, to make it more practical. The noises of the angle encoder is set as $1\times10^{-4} rad$, which is consistent with reality.

Two controllers were compared to highlight the advantages of the proposed method as follows:.

C1: the proposed controller with the LuGre model compensation, enabling better handling of nonlinear friction, particularly at low velocities. The control law can be represented by (14), (20), (24).

C2: the proposed controller with the Coulomb and viscous friction model, which is widely used [22]. The only difference from C1 is that the friction compensation term is composed of $\hat{T}_f = S_a(\dot{q})\hat{\theta}_3 + \dot{q}\,\hat{\theta}_4$.

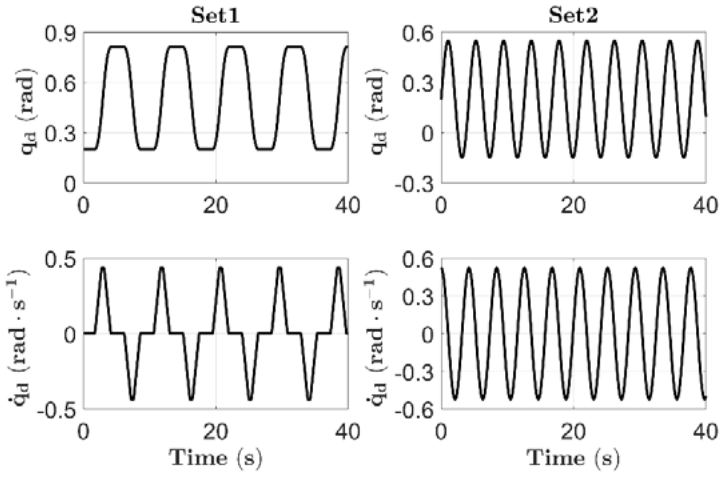


Figure 3: The reference trajectories.

Two sets of experiments, as show in Fig.3, were set up as follows:

Set1: Slow point-to-point(P2P) tracking experiment. The desired trajectory was from 0.2rad to 0.81rad, with the maximum angular velocity 0.43rad/s, which includes various motion states such as starting, accelerating, constant speed, and decelerating.

Set2: Slow SINE curve tracking experiment. The desired trajectory was set as $q_d(t) = 20\sin(1.5t)\,\text{rad}$.

In the simulation, all joints are controlled by the proposed method to track the reference trajectory simultaneously. For clarity, the analysis here focuses solely on the swing joint. To rigorously evaluate the tracking accuracy, quantitative metrics including the maximum tracking error e_{max} and the Integral of Time-weighted Absolute Error (ITAE) were introduced. The ITAE index is particularly significant as it penalizes long-duration errors, thereby reflecting the steady-state accuracy and convergence speed.

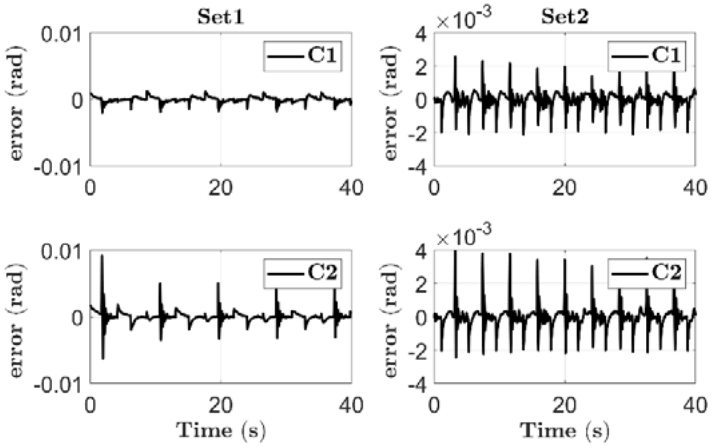


Figure 4: The tracking errors of C1 and C2.

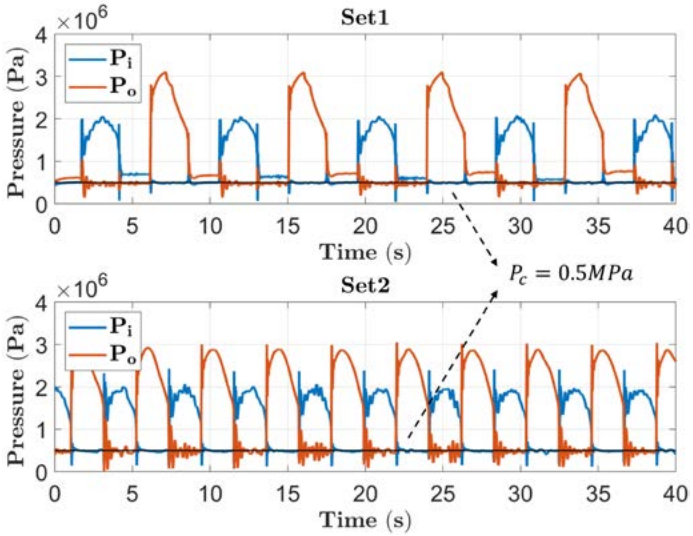


Figure 5: The pressure regulation results of C1.

The experiment results are shown in Fig. 4, and Table 1 is a summary of the comparison results. In the Point-to-Point motion (Set1), which involves frequent acceleration and deceleration phases, the proposed adaptive robust controller (C1) demonstrated remarkable performance. Specifically, C1 reduced the maximum tracking error e_{max} by 77.73% (from 0.0092 rad to 0.0020 rad) compared to the traditional controller (C2). This substantial reduction indicates that the proposed LuGre model-based friction compensation effectively

suppresses the error spikes caused by the Stribeck effect during zero-speed crossings. Furthermore, the ITAE index was reduced by 27.44%, confirming improved global tracking capability. Similarly, in the continuous Sine wave tracking (Set2), C1 consistently outperformed C2, achieving a 41.94% reduction in peak error and a 13.87% reduction in ITAE. These quantitative improvements validate that the proposed strategy not only handles nonlinear friction more effectively but also achieves higher precision across different motion trajectories. In addition, the pressure regulation results of C1 are show in Fig. 5, where the pressures of both chambers are regulated at a low level, which means the energy efficiency can be further improved by the independent metering technique.

Table 1: Quantitative comparison of tracking performance metrics.

Motion Set	Metric	C1	C2	Improvement
Set 1 (P2P)	e_{max} (rad)	0.0020	0.0092	77.73%
	ITAE ($rad \cdot s$)	0.2313	0.3188	27.44%
Set 2 (Sine)	e_{max} (rad)	0.0025	0.0044	41.94%
	ITAE ($rad \cdot s$)	0.2537	0.2946	13.87%

4.2. Experiments

The experimental validation was conducted on a customized 4-DOF hydraulic manipulator test rig (as shown in Fig. 1). The system is powered by a variable displacement pump station with a supply pressure of 10 MPa. The joint angles are measured by high-resolution rotary encoders, and the cylinder pressures are monitored by pressure sensors installed at both chambers to provide feedback for the pressure regulation loop. Real-time control algorithms are implemented on a NI cRIO system with the control frequency being 300 Hz. Compared with the simulation, additional real-world effects such as valve hysteresis, signal noise, and load fluctuations were present. The experimental results, as shown in Fig. 6, confirm that the proposed method effectively improves trajectory tracking accuracy, validating its potential for practical implementation in hydraulic manipulators.

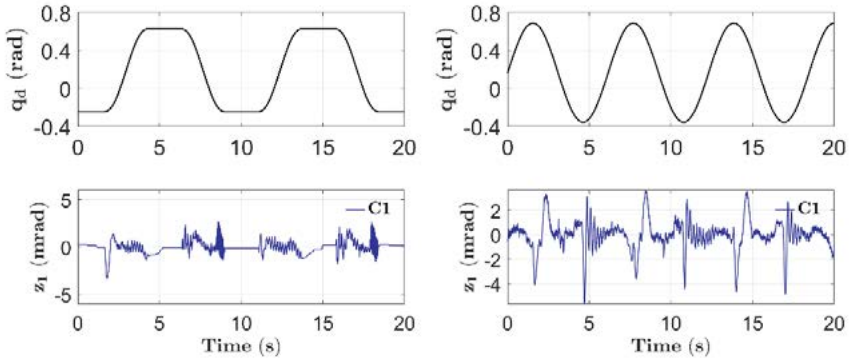


Figure 6: The experiment results of C1.

5. CONCLUSION

This paper presented a motion control strategy for a multi-DOF hydraulic manipulator driven by an independent metering system. An adaptive robust controller with LuGre

friction compensation was developed to handle system nonlinearities and uncertainties while improving energy efficiency. Both simulation and experimental results verified that the proposed method achieves high tracking accuracy and maintains low chamber pressures, demonstrating its potential for efficient and precise hydraulic manipulation.

6. REFERENCES

[1] Lee, M., Choi, H., Kim, C., Moon, J., Kim, D., and Lee, D., "Precision motion control of robotized industrial hydraulic excavators via data-driven model inversion," Proc. in IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 1912–1919, 2022.

[2] Zhou, S., Shen, C., Xia, Y., Chen, Z., and Zhu, S., "Adaptive robust control design for underwater multi-DOF hydraulic manipulator," Proc. in Ocean Engineering, vol. 248, p. 110822, 2022.

[3] Sivcev, S., Coleman, J., Omerdic, E., Dooly, G., and Toal, D., "Underwater manipulators: A review," Proc. in Ocean Engineering, vol. 163, pp. 431–450, 2018.

[4] Mattila, J., Koivumäki, J., Caldwell, D. G., and Semini, C., "A survey on control of hydraulic robotic manipulators with projection to future trends," Proc. in IEEE/ASME Transactions on Mechatronics, vol. 22, no. 2, pp. 669–680, 2017.

[5] Zhang, J., Zhang, F., Cheng, M., Ding, R., Xu, B., and Zong, H., "Parameter identification of hydraulic manipulators considering physical feasibility and control stability," Proc. in IEEE Transactions on Industrial Electronics, vol. 71, no. 1, pp. 718–728, 2024.

[6] Mohanty, A., and Yao, B., "Integrated direct/indirect adaptive robust control of hydraulic manipulators with valve deadband," Proc. in IEEE/ASME Transactions on Mechatronics, vol. 16, no. 4, pp. 707–715, 2010.

[7] Deng, W., Zhou, H., Zhou, J., and Yao, J., "Neural network-based adaptive asymptotic prescribed performance tracking control of hydraulic manipulators," Proc. in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, no. 1, pp. 285–295, 2022.

[8] Liang, X., Yao, Z., Deng, W., and Yao, J., "Adaptive control of n-link hydraulic manipulators with gravity and friction identification," Proc. in Nonlinear Dynamics, vol. 111, no. 20, pp. 19093–19109, 2023.

[9] Feng, H., Qiao, W., Yin, C., Yu, H., and Cao, D., "Identification and compensation of non-linear friction for an electro-hydraulic system," Proc. in Mechanism and Machine Theory, vol. 141, pp. 1–13, 2019.

[10] Yao, B., Bu, F., Reedy, J., and Chiu, G. T. C., "Adaptive robust motion control of single-rod hydraulic actuators: Theory and experiments," Proc. in IEEE/ASME Transactions on Mechatronics, vol. 5, no. 1, pp. 79–91, Mar. 2000.

[11] Pan, Q., Li, Y., and Huang, M., "Control-oriented friction modeling of hydraulic actuators based on hysteretic nonlinearity of lubricant film," Proc. in Mechatronics, vol. 53, pp. 72–84, 2018.

[12] Helian, B., Chen, Z., Yao, B., Lyu, L., and Li, C., "Accurate motion control of a direct-drive hydraulic system with an adaptive nonlinear pump flow compensation," Proc. in IEEE/ASME Transactions on Mechatronics, vol. 26, no. 5, pp. 2593–2603, 2020.

[13] Chen, P., Liu, X., and Yan, Q., "Adaptive friction compensation for a class of mechanical systems based on LuGre model," Proc. in International Journal of Robust and Nonlinear Control, vol. 32, no. 7, pp. 4510–4534, 2022.
[14] Lu, L., Yao, B., Wang, Q., and Chen, Z., "Adaptive robust control of linear motors with dynamic friction compensation using modified LuGre model," Proc. in Automatica, vol. 45, no. 12, pp. 2890–2896, 2009.
[15] Zhang, W., Li, M., Gao, Y., and Chen, Y., "Periodic adaptive learning control of PMSM servo system with LuGre model-based friction compensation," Proc. in Mechanism and Machine Theory, vol. 167, p. 104561, 2022.
[16] Dai, K., Zhu, Z., Shen, G., Tang, Y., Li, X., Wang, W., and Wang, Q., "Adaptive force tracking control of electrohydraulic systems with low load using the modified LuGre friction model," Proc. in Control Engineering Practice, vol. 125, p. 105213, 2022.
[17] Abuowda, K., Okhotnikov, I., Noroozi, S., Godfrey, P., and Dupac, M., "A review of electrohydraulic independent metering technology," Proc. in ISA Transactions, vol. 98, pp. 364–381, 2020.
[18] Chen, Z., Helian, B., Zhou, Y., and Geimer, M., "An integrated trajectory planning and motion control strategy of a variable rotational speed pump-controlled electro-hydraulic actuator," Proc. in IEEE/ASME Transactions on Mechatronics, vol. 28, no. 1, pp. 588–597, 2022.
[19] Lyu, L., Chen, Z., and Yao, B., "Advanced valves and pump coordinated hydraulic control design to simultaneously achieve high accuracy and high efficiency," Proc. in IEEE Transactions on Control Systems Technology, vol. 29, no. 1, pp. 236–248, 2020.
[20] Lyu, L., Chen, Z., and Yao, B., "Energy saving motion control of independent metering valves and pump combined hydraulic system," Proc. in IEEE/ASME Transactions on Mechatronics, vol. 24, no. 5, pp. 1909–1920, 2019.
[21] Liu, S., and Yao, B., "Coordinate control of energy saving programmable valves," Proc. in IEEE Transactions on Control Systems Technology, vol. 16, no. 1, pp. 34–45, 2008.
[22] Guo, Q., Wang, Q., and Liu, Y., "Antiwindup control of an electrohydraulic system with load disturbance and modeling uncertainty," Proc. in IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3097–3108, 2018.

Biographies



Yangxiu Xia received the B.Eng. degree in ocean engineering and technology from Zhejiang University, Hangzhou, China, in 2022. He is currently working toward the Ph.D. degree in ocean technology and engineering with the Ocean College, Zhejiang University, Zhoushan, China.



Litong Lyu received the B.Eng. and Ph.D. degrees in mechatronics engineering from Zhejiang University, Hangzhou, China, in 2015 and 2020, respectively.

Since 2020, he has been with the School of Mechanical Engineering, Shijiazhuang Tiedao University, Hebei, China, where he is currently an Associate Professor.



Bobo Helian received the B.Eng. degree in mechanical engineering from Yanshan University, Qinhuangdao, China, in 2015, and the Ph.D. degree in mechatronic engineering from Zhejiang University, Hangzhou, China, in 2021. Since 2021, he has been a Postdoctoral Researcher with the Institute of Mobile Machines, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. His research interests include advanced controls of electro-hydraulic systems (e.g., motion control, machine learning, trajectory planning, constrained control, nonlinear adaptive robust control, energy-saving control, and independent metering control).



Zheng Chen received the B.Eng. and Ph.D. degrees in mechatronic control engineering from Zhejiang University, Hangzhou, China, in 2007 and 2012, respectively.

From 2013 to 2015, he was a Postdoctoral Researcher with the Department of Mechanical Engineering, Dalhousie University, Halifax, NS, Canada. Since 2015, he has been an Associated Professor with the Ocean College, Zhejiang University, Hangzhou. His research interests mainly focus on advanced control of robotic and mechatronic systems (e.g., nonlinear adaptive robust control, motion control, trajectory planning, telerobotics, exoskeleton, mobile manipulator, precision mechatronic systems, and underwater robots).

Direct model reference adaptive control (MRAC) augmented frameworks for load pressure and position control of servo-valve controlled hydraulic cylinders

Marcel Steidle

Haeusler AG Duggingen, msteidle@haeusler.com

Patrick Stadler

Haeusler AG Duggingen, pstadler@haeusler.com

Jürgen Weber

Chair of Fluid Mechatronic Systems (Fluidtronic), TU Dresden, juergen.weber@tu-dresden.de

Abstract

Servo-valve controlled hydraulic cylinders are widely used in various industrial applications. These systems show highly nonlinear dynamic behavior which poses significant challenges for control system design. This paper develops two novel adaptive control frameworks for cylinder position and load pressure control to enhance the robustness and stability of these systems. The frameworks are specifically designed to address common issues such as parametric uncertainties, unexpected external disturbances and component wear. Key contributions are the development of a single, adaptive control design for the entire operation range as well as the introduction of an algorithm for seamless transition between adaptive position and load pressure control which improves the frameworks' overall performance and practicality. The proposed control frameworks are validated and tested through both numerical simulations and experiments conducted on a 4-roll plate bending machine. Finally, an experimental comparison between the proposed adaptive controllers and conventional linear and nonlinear controllers is presented. The results demonstrate an improvement in accuracy for both position and load pressure control.

Keywords: industrial hydraulics, 4-roll plate bending machine, adaptive control, Model Reference Adaptive Control (MRAC), valve control, position control, load pressure control

1. Introduction

Hydraulic drives are key components in many industries including robotics, manufacturing, forestry and earthwork. Due to their excellent energy- and power-density, they are often deployed for heavy-duty tasks like metal forming or excavation. These systems commonly operate under harsh and varying environmental conditions over extended periods, leading to significant changes in system behavior throughout a machine's lifetime. This variability presents a considerable challenge for control system design.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

To accomplish common control tasks like position and load pressure regulation, linear controllers are predominantly used. Their widespread adaptation is due to their conceptual simplicity, ease of implementation and sole reliance on readily available pressure and position measurements [1, 2]. While manually tuned PID based controllers remain common, more sophisticated linear control techniques, such as state-space control, are becoming more popular. For instance, [3] presents methods for system modeling and their application to linear control design, while [4] describes a method to improve feedforward control by combining physical modeling with measurement-based system identification during commissioning.
Additionally, extensive research has been conducted to address the challenges posed by the highly nonlinear dynamics of hydraulic systems. Different applications of nonlinear control techniques can be found for example in [5, 6, 7, 8, 9, 10]. However, these control designs suffer from drawbacks like the reliance on high-fidelity system models [5, 6] or require a fixed, repeatable process [10]. The development of high-fidelity system models is often resource- and time-intensive, making them less accessible to small and medium sized enterprises and manufacturers of special-machinery. This disadvantage is further exacerbated by the aforementioned parametric uncertainties, which reduce model accuracy and, consequently, the performance of model based control designs. The type and magnitude of these parameter variations are difficult to estimate during the design phase, thus motivating the application of robust and adaptive control techniques.
Since the 1950s, adaptive control has been an ongoing subject of research due to its capabilities in rejecting unforeseen disturbances such as structural damage, component wear or changes in the machine configuration. Today, a wide range of different approaches exist which can be broadly categorized into model-based and data-driven techniques [11]. Purely data-driven methods show increasingly promising results in hydraulic system control as seen in [12] and [13]. However, they often lack physical interpretability which poses a challenge for industrial applications as commissioning and maintenance tasks become increasingly complex and opaque. Therefore, despite this extensive research, a control architecture which simultaneously addresses the challenges of parametric uncertainty and external disturbances while maintaining physical interpretability remains a research gap. This work thus aims to develop two novel, adaptive control frameworks that address this gap. A particular emphasis is placed on the practical integration and applicability of these methods within existing industrial control frameworks.

2. System model

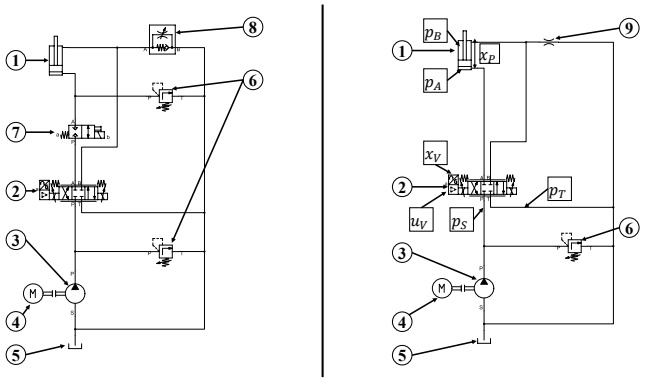
Mathematical system models simplify the development of controller designs and enable an assessment of open- and closed-loop behavior and performance. Especially state-space models are widely used in control theory as they provide an intuitive and physically meaningful system description. These models combine a set of coupled, first order differential equations to describe system dynamics with algebraic equations to determine the system outputs [14]:

$$\begin{aligned} \dot{\underline{x}} &= \underline{f}(\underline{x}, \underline{u}, \underline{\Theta}, t) \\ \underline{y} &= \underline{g}(\underline{x}, \underline{u}, \underline{\Theta}, t) \end{aligned} \tag{2.1}$$

The determination of required system equations is demonstrated using the hydraulic plant shown in Table 2.1. First, the original system is considered. It includes a pressure relief valve on the piston side and a check valve on the rod side of the cylinder. While the pressure relief

valve protects the cylinder from overpressure, the check valve prevents underpressure. Both phenomena are typically caused by external disturbances. However, if the control system performs adequately, both components remain inactive and are therefore irrelevant for the control design. Furthermore, the seat valve can also be neglected because it remains open during the entire control cycle. This leads to the simplified hydraulic scheme shown on the right side of Table 2.1.

Table 2.1: Schematic representation of the hydraulic drive used for bottom roller movement in a 4-roll plate bending machine.
Left: Original system; **Right**: Simplified model for control design



Position	Designation	Position	Designation
1	Differential cylinder	6	Pressure relief valve
2	4/3 proportional valve	7	2/2 seat valve
3	Hydraulic pump	8	Throttle and check valve
4	Electric motor	9	Throttle
5	Hydraulic tank		

To further simplify the model, it is assumed that the supply and tank pressure, p_{S} and p_{T}, are piecewise constant. The resulting model thus requires mathematical descriptions of a 4/3 proportional valve, a differential cylinder and an orifice.
Applying Newtons second law to the cylinder piston one obtains [2]:

$$\ddot{x}_{\mathrm{P}} = \frac{1}{m} \cdot (A_{\mathrm{A}} \cdot p_{\mathrm{A}} - A_{\mathrm{B}} \cdot p_{\mathrm{B}} - F_{\mathrm{f}}(\dot{x}_{\mathrm{P}}) - m \cdot g - F_{\mathrm{ext}}) \tag{2.2}$$

The friction force is given by:

$$F_{\mathrm{f}}(\dot{x}_{\mathrm{P}}) = \sigma \cdot \dot{x}_{\mathrm{P}} + \mathrm{sign}(\dot{x}_{\mathrm{P}}) \cdot \left[F_{\mathrm{c0}} + F_{\mathrm{s0}} \cdot \exp\left(-\frac{\|\dot{x}_{\mathrm{p}}\|}{c_{\mathrm{s}}} \right) \right] \tag{2.3}$$

The external forces vary between different applications of the hydraulic cylinder. For the controller design of a 4-roll plate bending machine, the external forces can be approximated as a mechanical hard stop:

$$F_{\mathrm{ext}}\left(x_{\mathrm{P}}\right) = \begin{cases} 0, & x_{\mathrm{P}} \leq a \\ k \cdot (x_{\mathrm{P}} - a), & x_{\mathrm{P}} > a \end{cases} \tag{2.4}$$

Using the continuity equation, the pressure dynamics of the cylinder chambers are obtained:

$$\dot{p}_{\mathrm{A}} = \frac{E\left(p_{\mathrm{A}}\right)}{V_{\mathrm{A}}\left(x_{\mathrm{P}}\right)} \cdot \left(Q_{\mathrm{A}}\left(p_{\mathrm{A}}, x_{\mathrm{V}}\right) - Q_{\mathrm{L,i}}\left(p_{\mathrm{A}}, p_{\mathrm{B}}\right) - Q_{\mathrm{L,e,A}}\left(p_{\mathrm{A}}\right) - A_{\mathrm{A}} \cdot \dot{x}_{\mathrm{P}}\right) \tag{2.5}$$

$$\dot{p}_{\mathrm{B}} = \frac{E\left(p_{\mathrm{B}}\right)}{V_{\mathrm{B}}\left(x_{\mathrm{P}}\right)} \cdot \left(Q_{\mathrm{B}}\left(p_{\mathrm{B}}, x_{\mathrm{V}}\right) + Q_{\mathrm{L,i}}\left(p_{\mathrm{A}}, p_{\mathrm{B}}\right) - Q_{\mathrm{L,e,B}}\left(p_{\mathrm{B}}\right) + A_{\mathrm{B}} \cdot \dot{x}_{\mathrm{P}}\right) \tag{2.6}$$

The chamber volumes and leakage flow rates are defined as:

$$V_{\mathrm{A}}\left(x_{\mathrm{P}}\right) = V_{0,\mathrm{A}} + A_{\mathrm{A}} \cdot x_{\mathrm{P}} \quad \text{and} \quad V_{\mathrm{B}}\left(x_{\mathrm{P}}\right) = V_{0,\mathrm{B}} + A_{\mathrm{B}} \cdot \left(S - x_{\mathrm{P}}\right)$$
$$Q_{\mathrm{L,i}}\left(p_{\mathrm{A}}, p_{\mathrm{B}}\right) = k_{\mathrm{L,i}} \cdot \left(p_{\mathrm{A}} - p_{\mathrm{B}}\right) \quad \text{and} \quad Q_{\mathrm{L,e,A/B}}\left(p_{A/B}\right) = k_{\mathrm{L,e,A/B}} \cdot \left(p_{A/B} - p_0\right)$$

The flow rates for both cylinder chambers are given by:

$$Q_{\mathrm{A}}\left(p_{\mathrm{A}}, x_{\mathrm{V}}\right) = Q_{\mathrm{V,A}}\left(p_{\mathrm{A}}, x_{\mathrm{V}}\right) \quad \text{and} \quad Q_{\mathrm{B}}\left(p_{\mathrm{B}}, x_{\mathrm{V}}\right) = Q_{\mathrm{O}}\left(p_{\mathrm{B}}\right) + Q_{\mathrm{V,B}}\left(p_{\mathrm{B}}, x_{\mathrm{V}}\right)$$

With the flow rates through the orifice and 4/3 proportional valve:

$$Q_{\mathrm{O}}\left(p_{\mathrm{B}}\right) = k_{\mathrm{O}} \cdot \operatorname{sign}\left(p_{\mathrm{T}} - p_{\mathrm{B}}\right) \cdot \sqrt{\left\|p_{\mathrm{T}} - p_{\mathrm{B}}\right\|} \tag{2.7}$$

$$Q_{\mathrm{V,A}}\left(p_{\mathrm{A}}, x_{\mathrm{V}}\right) = k_{\mathrm{V,A}} \cdot \begin{cases} \operatorname{sign}\left(p_{\mathrm{S}} - p_{\mathrm{A}}\right) \cdot \sqrt{\left\|p_{\mathrm{S}} - p_{\mathrm{A}}\right\|} \cdot x_{\mathrm{V}}, & x_{\mathrm{V}} \geq 0 \\ \operatorname{sign}\left(p_{\mathrm{T}} - p_{\mathrm{A}}\right) \cdot \sqrt{\left\|p_{\mathrm{T}} - p_{\mathrm{A}}\right\|} \cdot \left(-x_{\mathrm{V}}\right), & x_{\mathrm{V}} < 0 \end{cases} \tag{2.8}$$

$$Q_{\mathrm{V,B}}\left(p_{\mathrm{B}}, x_{\mathrm{V}}\right) = k_{\mathrm{V,B}} \cdot \begin{cases} \operatorname{sign}\left(p_{\mathrm{T}} - p_{\mathrm{B}}\right) \cdot \sqrt{\left\|p_{\mathrm{T}} - p_{\mathrm{B}}\right\|} \cdot x_{\mathrm{V}}, & x_{\mathrm{V}} \geq 0 \\ \operatorname{sign}\left(p_{\mathrm{S}} - p_{\mathrm{B}}\right) \cdot \sqrt{\left\|p_{\mathrm{S}} - p_{\mathrm{B}}\right\|} \cdot \left(-x_{\mathrm{V}}\right), & x_{\mathrm{V}} < 0 \end{cases} \tag{2.9}$$

To further simplify the model for control design, the following assumptions are made:

- No leakage flows $Q_{\mathrm{L,i}} = Q_{\mathrm{L,e,A/B}} = 0$
- The bulk modulus is constant $E\left(p_{\mathrm{A}}\right) = E\left(p_{\mathrm{B}}\right) = \mathrm{cst}.$
- The dynamic behavior of the valve spool is neglected $x_{\mathrm{V}} = u_{\mathrm{V}}$
- The valve has no overlap
- Only viscous friction is considered $F_{\mathrm{f}}\left(\dot{x}_{\mathrm{P}}\right) = \sigma \cdot \dot{x}_{\mathrm{P}}$

Defining the state vector $\underline{x} = \begin{bmatrix} x_{\mathrm{P}} & \dot{x}_{\mathrm{P}} & p_{\mathrm{A}} & p_{\mathrm{B}} \end{bmatrix}^{\intercal}$, the disturbance as $d = F_{\mathrm{ext}}$ and the input as $u = u_{\mathrm{V}}$ and combining equations 2.2, 2.5 and 2.6 one obtains the state-space representation:

$$\dot{\underline{x}} = \begin{pmatrix} x_2 \\ \frac{1}{m} \cdot \left(A_{\mathrm{A}} \cdot x_3 - A_{\mathrm{B}} \cdot x_4 - m \cdot g - \sigma \cdot x_2\right) \\ \frac{E}{V_{\mathrm{A}}(x_1)} \cdot \left(Q_{\mathrm{A}}\left(x_3, u\right) - A_{\mathrm{A}} \cdot x_2\right) \\ \frac{E}{V_{\mathrm{B}}(x_1)} \cdot \left(Q_{\mathrm{B}}\left(x_4, u\right) + A_{\mathrm{B}} \cdot x_2\right) \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{1}{m} \\ 0 \\ 0 \end{pmatrix} \cdot d \tag{2.10}$$

While measurements of the cylinder position and both chamber pressures are available, the output equation differs based on the current control task:

$$y_{\mathrm{Pos}} = x_1 \qquad y_{\mathrm{Load}} = x_3 - \frac{A_{\mathrm{B}}}{A_{\mathrm{A}}} \cdot x_4 \tag{2.11}$$

3. Controller design

This section presents two different methods to design adaptive controllers for cylinder and load pressure control. Both designs are based on a direct MRAC approach.

3.1. Direct model reference adaptive control (MRAC)

Direct MRAC aims at preserving desirable system behavior in the presence of unexpected parametric variations or disturbances. Therefore, it compares the behavior of a reference model against the real plant. Controller parameters are directly adapted based on the observed error between reference model and plant. Figure 3.1 shows the core components of a direct MRAC approach which are the aforementioned reference model, the adaptation law, the adaptive controller and the base controller.

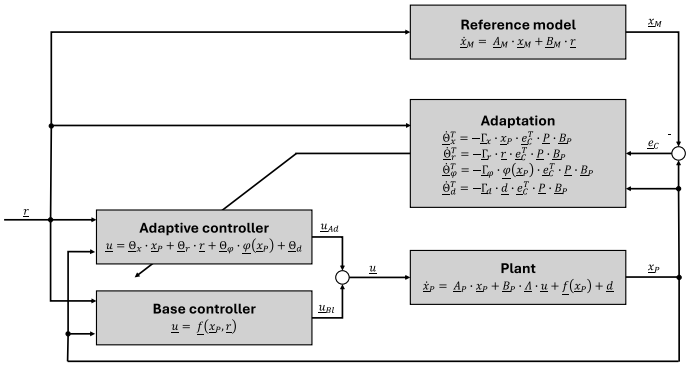


Figure 3.1: Block diagram of direct-MRAC based on the work of [15]

The base controller is designed to stabilize the plant during nominal conditions. If the plant is expected to be linear, the base controller can be designed as [16]:

$$\underline{u} = \underline{M}_{\mathrm{u}} \cdot \underline{r} + \underline{K} \cdot (\underline{M}_{\mathrm{x}} \cdot \underline{r} - \underline{x}) \tag{3.1}$$

Applying this control input to the nominal plant dynamics results in a design method for the reference model:

$$\dot{\underline{x}}_{\mathrm{M}} = \underbrace{\left(\underline{\hat{A}}_{\mathrm{P}} - \underline{\hat{B}}_{\mathrm{P}} \cdot \underline{K}\right)}_{\underline{A}_{\mathrm{M}}} \cdot \underline{x}_{\mathrm{M}} + \underbrace{\underline{\hat{B}}_{\mathrm{P}} \cdot (\underline{M}_{\mathrm{u}} + \underline{K} \cdot \underline{M}_{\mathrm{x}})}_{\underline{B}_{\mathrm{M}}} \cdot \underline{r} \tag{3.2}$$

If the real system deviates from the reference model, the adaptation mechanism becomes active. The choice of the adaptation mechanism and the appropriate control law depends on the nature of the expected disturbances. A general approach can be seen in Figure 3.1. However, if the disturbance is expected to be highly nonlinear, the following control law is sufficient:

$$\underline{u}_{\mathrm{Ad}} = \underline{\Theta}_{\varphi} \cdot \underline{\varphi}\left(\underline{x}_{\mathrm{P}}\right) + \underline{\Theta}_{\mathrm{d}} \tag{3.3}$$

The regressor vector $\underline{\varphi}\left(\underline{x}\right)$ can either be derived from physical system knowledge or by approximation, commonly using a single hidden layer neural network [17].

3.2. Type I: Feedback linearization

For the application of direct MRAC, models with a nearly linear behavior are preferred. Furthermore, control design for this type of models is comparatively straightforward as a broad variety of well-founded control design methods is available. While methods for linear systems distinguish themselves with verifiable, global stability and robustness measures [16], design methods for nonlinear systems are often cumbersome, require simplifications or provide only limited insight into the closed-loop system behavior. This motivates the application of feedback linearization, a technique that transforms a nonlinear system into a linear one, thereby enabling the use of well established linear control methods for nonlinear systems. This is achieved by a state transformation $\underline{z} = \underline{\Phi}(\underline{x})$. If the state transformation $\underline{\Phi}(\underline{x})$ is a local diffeomorphism around a workpoint $\underline{x}_0$ and the system description is affine in $\underline{u}$, application of the state transformation renders the system linear in the new state description $\underline{z}$ [18].
In order to use feedback linearization, the nonlinear system has to be expressed in the form:[1]

$$\begin{aligned} \dot{\underline{x}} &= \underline{f}(\underline{x}) + \underline{g}(\underline{x}) \cdot u \\ y &= h(\underline{x}) \end{aligned} \tag{3.4}$$

To determine the new state variables $\underline{z}$, the output equation is differentiated in time until the r-th derivative of y shows a direct dependency on u. Defining the states $\underline{\xi} = \begin{bmatrix} y & \dot{y} & \dots & y^{(r-1)} \end{bmatrix}^\mathsf{T}$ and the new input $v = y^{(r)}$, a linear model is obtained:

$$\begin{aligned} \dot{\underline{\xi}} &= \underbrace{\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \ddots & 0 \\ 0 & & \dots & 0 & 1 \\ 0 & & \dots\dots & & 0 \end{pmatrix}}_{\tilde{\underline{A}}} \cdot \underline{\xi} + \underbrace{\begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix}}_{\tilde{\underline{b}}} \cdot v \\ y &= \underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}}_{\tilde{\underline{c}}} \cdot \underline{\xi} \end{aligned} \tag{3.5}$$

Controller design for the system given by Equation 3.5 can be done for example by using pole-placement or linear quadratic regulation (LQR). To obtain the input to the nonlinear system, following transformation based on the Lie-derivatives has to be applied [18]:

$$u = \frac{v - L_{\underline{f}}^{r} h(\underline{x})}{L_{\underline{g}} L_{\underline{f}}^{r-1} h(\underline{x})} \tag{3.6}$$

However, Equation 3.5 only represents the entire, original system if the relative degree r equals the number of states. Otherwise, the system possesses $(n_x - r)$ states $\underline{\eta}$ which are not influenced by the control input and thus can not be stabilized by the control system. Therefore, these states have to be inherently stable to guarantee stability of the entire system. Determination of the zero dynamics can be achieved through two distinct approaches [18]:

[1] The following discussion focuses on SISO-systems. For discussions of the MIMO case refer e.g. to [18] or [19]

1. Solving the partial differential equations $\frac{\partial \eta_i}{\partial t} = \frac{\partial \Phi_{r+i}(\underline{x})}{\partial \underline{x}} \cdot \left(\underline{f}(\underline{x}) + \underline{g}(\underline{x}) \cdot u\right)$
2. Selecting them arbitrarily and checking if $\det\left(\frac{\partial \underline{\Phi}(\underline{x})}{\partial \underline{x}}\right) \neq 0$

Table 3.1: Results of feedback linearization for position and load pressure control applied to the state space system described by equations 2.10 and 2.11

	Position	**Load pressure**
Controllable states	$\xi_1 = x_1$ $\xi_2 = x_2$ $\xi_3 = \frac{1}{m} \cdot (A_\mathrm{A} \cdot x_3 \ - A_\mathrm{B} \cdot x_4 - m \cdot g - \sigma \cdot x_2)$	$\xi_1 = x_3 - \frac{A_\mathrm{B}}{A_\mathrm{A}} \cdot x_4$
Zero dynamics	$\eta_1 = x_3 - \frac{A_\mathrm{B}}{A_\mathrm{A}} \cdot x_4$ $\eta_2 = x_3$	$\eta_1 = x_1$ $\eta_2 = x_2$ $\eta_3 = \frac{1}{m} \cdot (A_\mathrm{A} \cdot x_3 \ - A_\mathrm{B} \cdot x_4 - m \cdot g - \sigma \cdot x_2)$ $\eta_4 = x_3$
Linear system	$\dot{\underline{\xi}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \cdot \underline{\xi} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot v$	$\dot{\xi} = v$
Lie-derivatives	$L_{\underline{f}}^r h(\underline{x}) = \frac{1}{m} \cdot$ $\big[- \frac{E}{V_A(x_1)} \cdot A_A^2 \cdot x_2$ $- \frac{E}{V_B(x_1)} \cdot A_B \cdot (Q_O(x_4) + A_B \cdot x_2)$ $- \frac{\sigma}{m} \cdot (A_A \cdot x_3 - A_B \cdot x_4 - mg - \sigma \cdot x_2) \big]$ $L_{\underline{g}} L_{\underline{f}}^{r-1} h(\underline{x}) = \frac{1}{m} \cdot$ $\big[\frac{E}{V_\mathrm{A}(x_1)} \cdot A_\mathrm{A} \cdot Q_{\mathrm{V,A,nom}}(x_3)$ $- \frac{E}{V_\mathrm{B}(x_1)} \cdot A_\mathrm{B} \cdot Q_{\mathrm{V,B,nom}}(x_4) \big]$	$L_{\underline{f}}^r h(\underline{x}) =$ $- \frac{E}{V_A(x_1)} \cdot A_A \cdot x_2$ $- \frac{A_B}{A_A} \cdot \frac{E}{V_B(x_1)} \cdot [Q_O(x_4) + A_B \cdot x_2]$ $L_{\underline{g}} L_{\underline{f}}^{r-1} h(\underline{x}) =$ $\frac{E}{V_\mathrm{A}(x_1)} \cdot Q_{\mathrm{V,A,nom}}(x_3)$ $- \frac{A_\mathrm{B}}{A_\mathrm{A}} \cdot \frac{E}{V_\mathrm{B}(x_1)} \cdot Q_{\mathrm{V,B,nom}}(x_4)$
Interpretation	• States represent position, velocity and acceleration • Control input represents jerk • Cylinder pressures are not controllable. They result from the current velocity and external load.	• System is scalar and represents load pressure dynamics. • Other states are only considered during input transformation

Unfortunately, the system description in Equation 2.10 is not affine in the control variable as the flow rate through the proportional valve is more involved. However, equations 2.8 and 2.9 can be rewritten as:

$$Q_{\mathrm{V,A}}(x_3, u) = Q_{\mathrm{V,A,nom}}(x_3, u) \cdot u \quad \text{and} \quad Q_{\mathrm{V,B}}(x_4, u) = Q_{\mathrm{V,B,nom}}(x_4, u) \cdot u \quad (3.7)$$

With:

$$\begin{aligned} Q_{\mathrm{V,A,nom}}\left(x_3,u\right) &= k_{\mathrm{V,A}}\cdot\begin{cases}\mathrm{sign}\left(p_{\mathrm{S}}-x_3\right)\cdot\sqrt{\|p_{\mathrm{S}}-x_3\|}, & u\geq 0\\ -\mathrm{sign}\left(p_{\mathrm{T}}-x_3\right)\cdot\sqrt{\|p_{\mathrm{T}}-x_3\|}, & u<0\end{cases}\\ Q_{\mathrm{V,B,nom}}\left(x_4,u\right) &= k_{\mathrm{V,B}}\cdot\begin{cases}\mathrm{sign}\left(p_{\mathrm{T}}-x_4\right)\cdot\sqrt{\|p_{\mathrm{T}}-x_4\|}, & u\geq 0\\ -\mathrm{sign}\left(p_{\mathrm{S}}-x_4\right)\cdot\sqrt{\|p_{\mathrm{S}}-x_4\|}, & u<0\end{cases}\end{aligned} \tag{3.8}$$

Using Equation 3.7, the system becomes affine in the control variable if $u>0$ or if $u<0$. Thereby, if the system description is separated around $u=0$, it satisfies Equation 3.4 which justifies the usage of feedback linearization.
An overview of the feedback linearized systems can be found in Table 3.1. It can be seen that the zero dynamics of the position and load pressure controller are selected in a way that both scenarios use the same transformed state representation.
Combining feedback linearization with model reference adaptive control leads to the control scheme shown in Figure 3.2a. In this context, the adaptive controller acts on the transferred states. Thereby, it needs to compensate for uncertainties and deviations in the state and input transformation. These errors are usually nonlinear. Therefore, the control law given by Equation 3.3 is used for MRAC. As the proposed position and load pressure controller share the same state representation and act on the same system, both control systems are affected by the same disturbances and uncertainties. For this reason, a transfer of adapted controller gains between both controllers is desirable.
To transfer the gains between both controllers, it is further assumed that they share the same regressor vector $\underline{\varphi}\left(\underline{x}\right)$. Based on the idea of bumpless transfer, it is assumed that at the time of controller switching t_s both controllers should produce the same output:

$$u_{\mathrm{Pos}}\left(t_{\mathrm{s}}\right)\overset{!}{=}u_{\mathrm{Load}}\left(t_{\mathrm{s}}\right) \tag{3.9}$$

Separating the output of the base and adaptive controller in Equation 3.6 one obtains:

$$u_i\left(t_{\mathrm{s}}\right)=\underbrace{\frac{v_{\mathrm{Bl},i}-L_{\underline{f}}^{r_i}h_i\left(\underline{x}\right)}{L_{\underline{g}}L_{\underline{f}}^{r_i-1}h_i\left(\underline{x}\right)}}_{\text{Base controller}}+\underbrace{\frac{v_{\mathrm{Ad},i}}{L_{\underline{g}}L_{\underline{f}}^{r_i-1}h_i\left(\underline{x}\right)}}_{\text{Adaptive controller}} \tag{3.10}$$

Combining equations 3.3, 3.9 and 3.10 and solving for the adaptive control input of the load pressure controller results in:

$$\begin{aligned}&\underline{\Theta}_{\varphi,\mathrm{Load}}\cdot\underline{\varphi}\left(\underline{x}\right)+\Theta_{\mathrm{d,Load}}=\\ &\frac{L_{\underline{g}}L_{\underline{f}}^{r_{\mathrm{Load}}-1}h_{\mathrm{Load}}\left(\underline{x}\right)}{L_{\underline{g}}L_{\underline{f}}^{r_{\mathrm{Pos}}-1}h_{\mathrm{Pos}}\left(\underline{x}\right)}\cdot\underline{\Theta}_{\varphi,\mathrm{Pos}}\cdot\underline{\varphi}\left(\underline{x}\right)+\Delta u_{\mathrm{Bl}}+\frac{L_{\underline{g}}L_{\underline{f}}^{r_{\mathrm{Load}}-1}h_{\mathrm{Load}}\left(\underline{x}\right)}{L_{\underline{g}}L_{\underline{f}}^{r_{\mathrm{Pos}}-1}h_{\mathrm{Pos}}\left(\underline{x}\right)}\cdot\Theta_{\mathrm{d,Pos}}\end{aligned} \tag{3.11}$$

With:

$$\Delta u_{\mathrm{Bl}}=L_{\underline{g}}L_{\underline{f}}^{r_{\mathrm{Load}}-1}h_{\mathrm{Load}}\left(\underline{x}\right)\cdot\left[\frac{v_{\mathrm{Bl,Pos}}-L_{\underline{f}}^{r_{\mathrm{Pos}}}h_{\mathrm{Pos}}\left(\underline{x}\right)}{L_{\underline{g}}L_{\underline{f}}^{r_{\mathrm{Pos}}-1}h_{\mathrm{Pos}}\left(\underline{x}\right)}-\frac{v_{\mathrm{Bl,Load}}-L_{\underline{f}}^{r_{\mathrm{Load}}}h_{\mathrm{Load}}\left(\underline{x}\right)}{L_{\underline{g}}L_{\underline{f}}^{r_{\mathrm{Load}}-1}h_{\mathrm{Load}}\left(\underline{x}\right)}\right]$$

Using coefficient comparison results in:

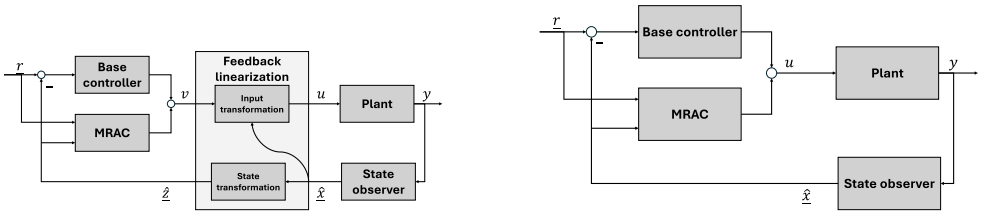
$$\underline{\Theta}_{\varphi,\mathrm{Load}} = \frac{L_{\underline{g}} L_{\underline{f}}^{r_{\mathrm{Load}}-1} h_{\mathrm{Load}}\left(\underline{x}\right)}{L_{\underline{g}} L_{\underline{f}}^{r_{\mathrm{Pos}}-1} h_{\mathrm{Pos}}\left(\underline{x}\right)} \cdot \underline{\Theta}_{\varphi,\mathrm{Pos}} \tag{3.12}$$

$$\Theta_{\mathrm{d,Load}} = \Delta u_{\mathrm{Bl}} + \frac{L_{\underline{g}} L_{\underline{f}}^{r_{\mathrm{Load}}-1} h_{\mathrm{Load}}\left(\underline{x}\right)}{L_{\underline{g}} L_{\underline{f}}^{r_{\mathrm{Pos}}-1} h_{\mathrm{Pos}}\left(\underline{x}\right)} \cdot \Theta_{\mathrm{d,Pos}} \tag{3.13}$$

Equations 3.12 and 3.13 represent the exchange of obtained knowledge based on adaptation. As these mappings are bijective, transfer in both directions is possible by inverting.

3.3. Type II: Linear replacement model

The main purposes of Type II adaptive control are the preservation of controller performance for industrial applications while simultaneously relying on minimum model knowledge. Therefore, it incorporates arbitrary, existing control frameworks by identification of the closed loop system after commissioning and uses the identified system as reference model. The block diagram of the resulting control scheme can be seen in Figure 3.2b. Compared to the scheme of Type I, Type II is distinguished by a reduced complexity as it avoids the additional state transformation.



(a) Adaptive control type I

(b) Adaptive control type II

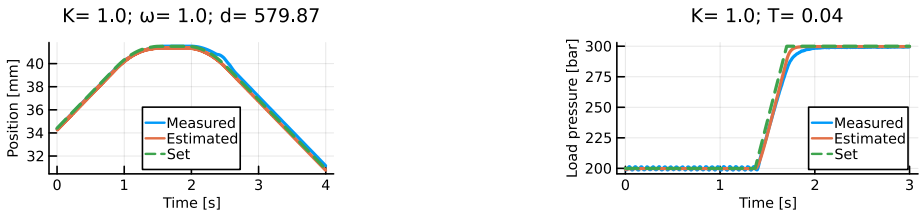
Figure 3.2: Block diagrams of adaptive control frameworks

Additionally, Type II adaptive control integrates smoothly into existing commissioning routines as can be seen by the following, recommended sequence:

1. Commissioning of the known control framework
2. Collecting data from the closed-loop system
   - The collected data should represent system behavior over the entire operating range.
   - Data collection requires movement in closed loop mode.
   - The set and actual values of the controlled output should be recorded.
3. Identification of a black box model using for example least squares or maximum likelihood estimation
4. Defining the identified model as reference model for direct MRAC
5. Parametrization of the adaptation behavior

Examples for the results of system identification can be seen in Figure 3.3. Hereby, the positioning behavior of the actuator is modeled as a PT2 element while the pressure dynamics are approximated as PT1 model. Since the identified model does not incorporate

any information about the underlying system physics, all deviations from the nominal performance are compensated by adaptation of the controller. This complicates error tracking as intended system variations like changing supply pressures have to be compensated by the base controller. Otherwise they appear as disturbance to the adaptive controller and lead to adaptation of the control law.



(a) Identified PT2 position model

(b) Identified PT1 load pressure model

Figure 3.3: Results of closed-loop system identification

4. Simulation results

Simulation based design is nowadays standard in research and industry as it provides a cost and time efficient way to validate and test new concepts. Furthermore, it allows for system evaluation under wide parametric variations or under safety critical conditions. For this work, an open-source simulation environment developed in *julia* [20] is used, leveraging the *DifferentialEquations.jl* [21] and *ModellingToolkit.jl* [22] packages for acausal, symbol-based modeling and control design.
For evaluation of the control performance the following test scenarios are evaluated:

1. Load pressure control with suddenly increasing cylinder leakage
2. Position and load pressure control with different supply pressure control schemes
3. Switching between adaptive position and load pressure control

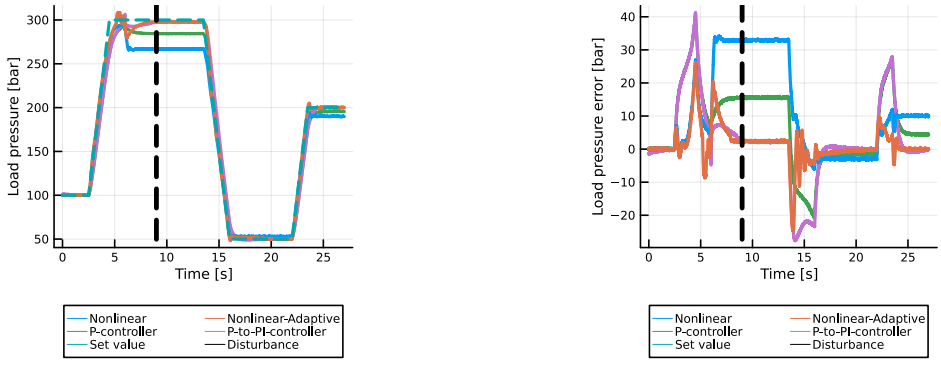
All scenarios use a common system model which is based on the experimental setup presented in Section 5. The nominal parameter values are listed in Table 4.1.

Table 4.1: Nominal parameters of the simulated cylinder

Parameter	**Unit**	**Value**	**Parameter**	**Unit**	**Value**
Piston side area A_A	mm^2	3848.5	Piston rod side area A_B	mm^2	1885.0
Piston stroke S	mm	85.0	Viscous friction coefficient σ	$\frac{N}{\frac{mm}{s}}$	192.0
Piston mass m_P	kg	2.96	Load mass m_L	kg	475.6
Flow rate coefficient valve k_V	$\frac{l}{min \cdot \sqrt{bar}}$	0.338	Flow rate coefficient orifice k_O	$\frac{l}{min \cdot \sqrt{bar}}$	0.0
Flow rate coefficient internal lekage $k_{L,i}$	$\frac{l}{min \cdot bar}$	0.0	Flow rate coefficient external lekage $k_{L,e}$	$\frac{l}{min \cdot bar}$	0.0
Supply pressure p_S	bar	310.0	Tank pressure p_T	bar	2.0
Bulk modulus E	bar	18000.0			

The first scenario examines the control performance of Type I adaptive control under unexpected disturbances. Therefore, an additional 2-way valve is introduced between the piston

and ring side of the cylinder. Activating the 2-way valve leads to an unexpected flow rate between both piston chambers which has to be compensated by the controller. To evaluate the performance of the adaptive controller, it is compared against a P-controller, a P-to-PI switching control structure, where the integral action is added in proximity of the target value, and a nonlinear controller based on feedback linearization. Figure 4.1 shows the results. All designs show a good tracking behavior during the first 9 s. At this point, the disturbance is introduced by opening the 2-way valve. This leads to an immediate decline of control performance for all four designs. The nonlinear controller, in particular, shows the strongest reaction to the disturbance and fails to recover which leads to a sustained steady-state error for the rest of the simulation run. As the adaptive controller is based on the nonlinear controller, it also shows a strong initial reaction to the disturbance. However, due to the adaptation mechanism, it quickly recovers its performance. During the following setpoint changes, the system with adaptive control shows the smallest control deviations. Especially during the transition phases, nonlinear control designs show a significantly improved performance compared to the linear controllers.



(a) Trajectory

(b) Tracking error

Figure 4.1: Performance of different load pressure controller designs under the influence of an unexpected disturbance caused by a cylinder leakage

In summary, it can be said that nonlinear control designs improve the control performance during nominal conditions. However, their dependence on a sophisticated system model makes them vulnerable against unexpected disturbances or model errors. These disadvantages can be compensated by a nonlinear controller with adaptation, as described in Section 3.2, resulting in the increased performance shown in Figure 4.1.
To further investigate the applicability of adaptive controllers in industrial settings, the second scenario is considered. For this scenario, a total of six different control designs are benchmarked. Similar to the designs of scenario 1, feedback control is achieved using a P-controller, a P-to-PI switching controller and a nonlinear controller based on feedback linearization. Additionally, all three foundational controllers are enhanced by an adaptive layer and the linear position controllers are augmented by a model-based feedforward structure which inverts equations 2.8 and 2.9 to calculate the necessary control input.
During this scenario, the system is disturbed by a varying supply pressure. This represents common practical applications where systems lack a pressure accumulator or utilize load-

sensing, which actively modulates the supply pressure. To quantify the influence of changing supply pressures, two control schemes are evaluated:

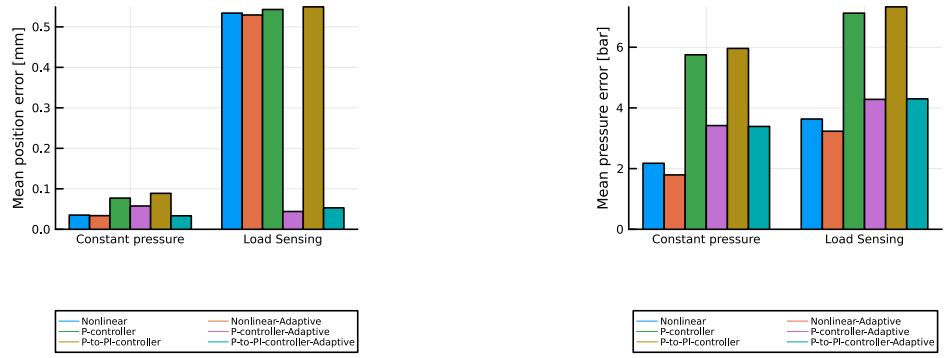
1. **Constant pressure**: The supply pressure is constantly hold at $310\,\mathrm{bar}$.
2. **Load-Sensing**: Supply pressure is adapted based on the highest moving actuator pressure.

The setpoint trajectories are generated to replicate the operation of the CNC-interface of the test machine in Section 5 and are illustrated in Figure 4.2. Notably, the position setpoint is designed to induce an external force on the cylinder at a position of $40\,\mathrm{mm}$, simulating the contact force between bottom and top rollers of the machine.



(a) Position (b) Load pressure

Figure 4.2: Setpoint trajectories used for controller comparison

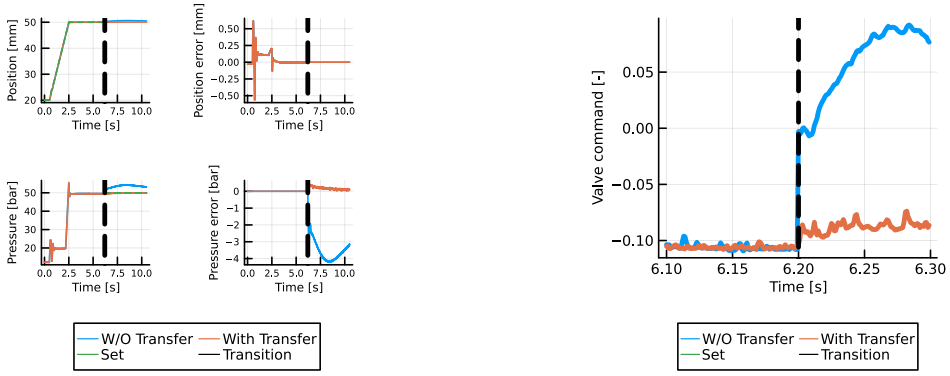


(a) Position (b) Load pressure

Figure 4.3: Simulated performance of different control designs

Figure 4.3 presents the mean absolute control deviations for both the load pressure and position controllers. It is evident that supply pressure variations significantly degrade control performance. This effect is especially pronounced in the linear controllers without adaptation as they lack an effective mechanism to compensate for supply pressure variations. Consequently, extending these controllers with a Type II adaptive framework leads to a substantial improvement in accuracy by compensating for these variations. The performance of the adaptive position controllers under load-sensing distinctively highlights this effect. Because

load-sensing stabilizes the difference between supply and cylinder pressure, the valve flow rate becomes primarily dependent on the control input alone. This reduces system complexity which in turn facilities faster adaptation and improved convergence, leading to enhanced control performance. In contrast, the adaptive enhancement has a reduced influence on the nonlinear controller's performance, as the adaptation process is already significantly affected by the underlying structure of feedback linearization.
In summary, adaptive and nonlinear controllers are both capable of improving accuracy, particularly under varying supply pressures. The Type II adaptive controller design is especially promising, as it significantly enhances the performance of conventional linear controllers.
Finally, the third scenario is examined. As discussed in Section 3.2, transfer of adaptation gains between position and load pressure control is a common and essential task. To validate the increase in controller performance by introduction of the transfer algorithm, the same control task is simulated with and without the switching algorithm. The 2-way valve between both cylinder chambers is open for the entire simulation. During the first $6.2\,\mathrm{s}$, position control is active and adapts for the unexpected flow rate between the cylinder chambers. Thereafter, the controller mode switches from position to pressure control. Results can be seen in Figure 4.4. As expected, the transfer of controller gains significantly reduces the initial control transient after switching, as the previously acquired knowledge about the system's behavior is leveraged by the new controller.



(a) Trajectories and tracking errors

(b) Control input

Figure 4.4: Trajectories and control deviations using adaptation transfer

5. Experimental results

An experimental campaign was designed and executed to validate the novel control approaches by comparing different controller designs. All experiments were conducted on a **4-roll plate bending machine** (HAEUSLER, VRM-2000x23) which can be seen in Figure 5.1. Both side rollers and the bottom roller can be moved linearly using hydraulic actuators while the top and bottom roller are rotary driven by electric motors. Linear movement is conducted via two servo-valve controlled differential cylinders per roller which operate in leader-follower control to synchronize their movements. The required oil for all hydraulic actuators is provided by a **gear pump** (Bosch Rexroth, PGH2-2X/005RE07VU2). As the

pump provides a constant flow rate, supply pressure is adjusted between 20 bar to 310 bar via a proportional pressure relief valve. For process and CNC-control, the machine is equipped with a state of the art PLC developed at HAEUSLER AG. The hardware consists of a **4/3-way proportional valve** (Bosch Rexroth, 4WRPEH6C3B02L-3X/M/24A1), an **absolute position encoder** (Baumer, EAL580-B) in combination with a **draw-wire sensor** (MICRO-EPSILON, WDS-2000-P60-M-SO-HAE) and two **pressure transducers** (IFM, PA 9020) per linear axis. Thereby, the machine is capable to perform high precision interpolation-movements between all axes, allowing for the production of complex work pieces with high radius transition requirements.

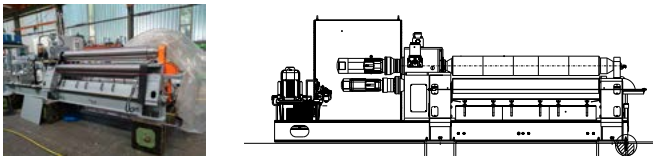
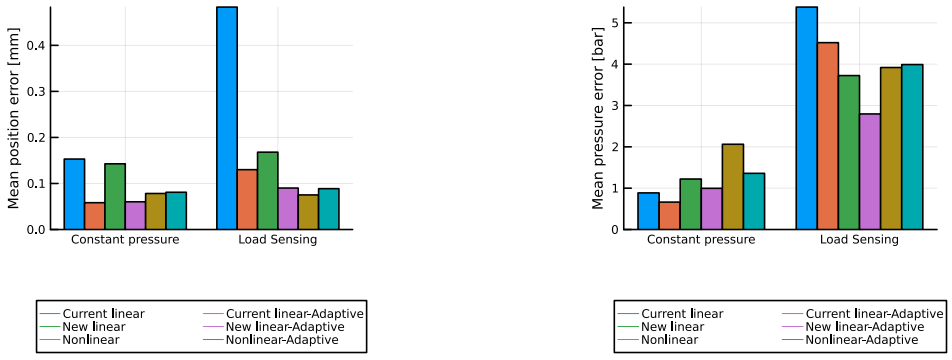


Figure 5.1: Test machine (HAEUSLER, VRM-2000x23) at HAEUSLER AG

For the experimental survey, a total of six different control approaches were benchmarked. These were based on three foundational controllers: the **currently implemented machine controller**, a **newly designed linear switching P-to-PI controller** (derived from scenarios 1 and 2 in Section 4), and a **nonlinear controller based on feedback linearization**. The primary difference between the default machine controller and the new linear controller lies in their feedforward structures. Where the new linear controller employs a model-based feedforward structure described in Section 4, the default machine controller utilizes a one-dimensional characteristic curve that maps the cylinder velocity to the control input. All three foundational controllers were enhanced by an adaptive layer. The two linear controllers were extended using the Type II adaptive framework (see Section 3.3), while the nonlinear controller was augmented with the Type I adaptive framework (see Section 3.2).



(a) Position

(b) Load pressure

Figure 5.2: Controller performance metrics of the experimental campaign

To validate controller robustness and performance, each control approach was tested under two supply pressure conditions —constant pressure and an active load-sensing system—

mirroring the simulation setup of scenario 2 in Section 4. The experiments were repeated three times per controller and supply pressure scheme, with results averaged to mitigate the influence of outliers.
A comparative evaluation of the controller performances is shown in Figure 5.2. The experimental results generally concur with the simulation outcomes presented in Figure 4.3, thereby, validating the conclusion that both the supply pressure and the adaptive control scheme have a significant influence on performance. A notable discrepancy between the experimental and simulated results is the reduced performance improvement of Type II adaptive position control. This reduction is likely attributable to non-modeled nonlinearities present in the physical system.
In summary, the experimental campaign successfully validated the capabilities of both adaptive and nonlinear controllers to improve accuracy and robustness under real-world conditions.

6. Conclusion

This work presented two novel approaches for the adaptive control of servo-valve controlled hydraulic cylinders. Performance of these designs was analyzed and validated through both numerical simulations and an extensive experimental campaign on a production-grade 4-roll plate bending machine.
The results demonstrate that the proposed adaptive control frameworks significantly enhance accuracy and robustness against a wide range of external disturbances and parametric uncertainties, including external forces, increasing leakages or varying supply pressures. The adaptive gain transfer mechanism, a key contribution of this work, was also validated through simulation, showing improved accuracy and a smooth transition during the crucial switching phase between position and load pressure control.
The experimental campaign, comparing the novel adaptive controllers against conventional linear and nonlinear designs, underscores the practical applicability of the proposed methods. Notably, it was shown that an adaptive enhancement of existing, industrial linear controllers, specifically using the Type II approach, can lead to a significant and verifiable improvement in control accuracy. This finding highlights a clear path for integrating advanced adaptive control into real-world industrial applications without requiring a complete overhaul of existing control systems.

7. References

[1] Gebhardt Norbert; Weber, Jürgen, *Hydraulik – Fluid-Mechatronik: Grundlagen, Komponenten, Systeme, Messtechnik und virtuelles Engineering*, Springer Berlin Heidelberg, 2020.

[2] Jelali, Mohieddine and Kroll, Andreas, *Hydraulic Servo-systems*, Springer London, 2003.

[3] Lohse, Harald, *Modellierung hydraulischer Tiefziehpressen für Prozesskopplung, Reglerauslegung und energetische Bilanzierung*, 1. Aufl., Fluidmechatronische Systeme, Aachen: Shaker, 2015, 186 pp.

[4] Helmke, Marcus et al., *Computer-assisted modeling and automatic controller adjustment for hydraulic drives based on an innovative nonparametric identification method*, Universitätsbibliothek der RWTH Aachen, 2018.

[5] Borrelli, Mauro et al., “MPC load control for aircraft actuator testing”, in: *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, IEEE, June 2020, pp. 587–592.
[6] Gücker, Jan Ulrich, ed., *Experimentelle Identifikation und nichtlineare Regelung eines einachsigen servohydraulischen Antriebs*, Download lizenzpflichtig, Kassel: Kassel Univ. Press, 2006.
[7] Komsta, Jan, Oijen, Nils van, and Antoszkiewicz, Peter, “Integral sliding mode compensator for load pressure control of die-cushion cylinder drive”, in: *Control Engineering Practice* 21.5 (May 2013), pp. 708–718.
[8] Mitov, Alexander et al., “Comparison between Model Predictive (MPC) and Model Reference Adaptive Controllers (MRAC) for Electrohydraulic Steering System Implemented as Real-Time Simulink® Program”, in: *IOP Conference Series: Materials Science and Engineering* 1002.1 (Dec. 2020), p. 012034.
[9] Prabel, Robert and Aschemann, Harald, “Nonlinear adaptive backstepping control of two coupled hydraulic servo cylinders”, in: *2014 American Control Conference*, IEEE, June 2014, pp. 1310–1315.
[10] Radermacher, Tobias, “Modellbasierte Vorsteuerungskonzepte für drehzahlvariable hydraulische Antriebe am Beispiel der Kunststoff-Spitzgießmaschine”, PhD thesis, 2022.
[11] Benosman, Mouhacine, “Model-based vs data-driven adaptive control: An overview”, in: *International Journal of Adaptive Control and Signal Processing* 32.5 (Mar. 2018), pp. 753–776.
[12] Lee, Minhyeong et al., “Precision Motion Control of Robotized Industrial Hydraulic Excavators via Data-Driven Model Inversion”, in: *IEEE Robotics and Automation Letters* 7.2 (Apr. 2022), pp. 1912–1919.
[13] Nan, Fang and Hutter, Marco, “Learning Adaptive Controller for Hydraulic Machinery Automation”, in: *IEEE Robotics and Automation Letters* 9.4 (Apr. 2024), pp. 3972–3979.
[14] Föllinger, Otto, *Nichtlineare Regelungen I*, München: De Gruyter Oldenbourg, 1998.
[15] Bierling, Thomas et al., “Comparative Analysis of MRAC Architectures in a Unified Framework”, in: *AIAA Guidance, Navigation, and Control Conference*, 2010.
[16] Föllinger, Otto, *Regelungstechnik, Einführung in die Methoden und ihre Anwendung*, ed. by Konigorski, Ulrich et al., 13., überarbeitete Auflage, Berlin: VDE Verlag GmbH, 2022, 1452 pp.
[17] Holzapfel, Florian, “Nichtlineare adaptive Regelung eines unbemannten Fluggerätes”, de, PhD thesis, Technische Universität München, 2004, p. 282.
[18] Isidori, Alberto, *Nonlinear Control Systems*, 3. ed., [Nachdr.], vol. [1], Berlin: Springer, 2002, 549 pp.
[19] Khalil, Hassan K., *Nonlinear systems*, 2. ed., Upper Saddle River, NJ [u.a.]: Prentice Hall, 1996, 734 pp.
[20] Bezanson, Jeff et al., “Julia: A Fresh Approach to Numerical Computing”, in: *SIAM Review* 59.1 (Jan. 2017), pp. 65–98.
[21] Rackauckas, Christopher and Nie, Qing, “DifferentialEquations.jl–a performant and feature-rich ecosystem for solving differential equations in Julia”, in: *Journal of Open Research Software* 5.1 (2017).
[22] Ma, Yingbo et al., *ModelingToolkit: A Composable Graph Transformation System For Equation-Based Modeling*, 2021.

Biographies



Marcel Steidle received his bachelor's and master's degrees in Mechanical Engineering from the Technical University of Munich (TUM) in 2017 and 2019, respectively. He is currently an automation engineer at Haeusler AG Duggingen, where he is also pursuing his doctoral degree in cooperation with the the Chair of Fluid Mechatronic Systems at TU Dresden. His research focuses on adaptive control and methods for system and parameter identification in industrial hydraulic systems. Additionally, he explores opportunities to improve the energy efficiency of plate roll bending machines.



Patrick Stadler earned a Bachelor's degree in 2016 and a Master's degree in 2018 in Mechanical Engineering from ETH Zürich, with a focus on robotics and control system design. Since 2019, he has been working at Haeusler AG in Duggingen, applying scientific knowledge to enhance industrial applications.



Jürgen Weber received his Dipl.-Ing. and Dr.-Ing. in Mechanical Engineering from the Technical University of Dresden (TU Dresden) in 1986 and 1991, respectively. His early career was spent in academia, serving as a scientific assistant and senior engineer at TU Dresden's Institute for Machine Tools and Institute for Hydraulics and Pneumatics until 1997.

Following a 13-year period in the industry, where he held various leadership roles in hydraulic systems and advanced development, he returned to academia in 2010. He was appointed an University Professor and Director of the Institute of Fluid Power at TU Dresden. In 2018, he took on an additional leadership role as the Director of the Institute of Mechatronic Engineering. Since 2022, he has been the CEO of Construction Future Lab gGmbH, a startup focused on the digitization of construction machinery, further bridging his academic and industrial expertise.

Design Overview and Performance Assessment of an Integrated Electro-Hydraulic Actuator (iEHA)

Seshan Calapatti Suresh [1], Hassan Assaf [1], Andrea Vacca [1], Howard Zhang [2]

1. Maha Fluid Power Research Centre, Purdue University, West Lafayette, IN, USA, scalapat@purdue.edu; Hassan.Assaf@cat.com; avacca@purdue.edu

2. Parker Hannifin Corporation, Macedonia, OH, USA, hzhang@parker.com

Abstract.

The recent electrification trend in mobile and aerospace applications is increasing interest in electro-hydraulic actuators (EHAs), which are essentially primary-controlled hydraulic cylinders with an electric prime mover. EHAs can achieve high energy efficiency, a desirable feature particularly for electrified vehicles. Moreover, when implemented in a self-contained arrangement, EHAs are particularly attractive, as they can serve as an alternative to electromechanical actuators, leveraging the typical advantages of hydraulics. However, existing EHA designs have a structure unsuitable for most mobile applications and do not fully leverage the benefits of component integration.

This paper proposes an innovative design for an integrated Electro-Hydraulic Actuator (iEHA) based on aggressive component integration. The electric prime mover and the hydraulic machine are encapsulated in the cylinder head, while the reservoir is integrated alongside the walls of the hydraulic cylinder. The internal flow connections allow the differential flow between the cylinder and reservoir to circulate through the electric machine, thereby enabling optimal heat removal from the electric machine. The paper details the design morphology, the hydraulic and cooling circuit, and the sizing of the internal components, with supporting simulation results.

Keywords. EHA, iEHA, Electrification, Thermal Analysis.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

NOMENCLATURE

ϕ_A - Piston diameter [m]	ϕ_a - Rod diameter [m]
γ – Polytropic index of the charge gas for the Accumulator [-]	n_s - Rotary speed [rev/min]
A_A - Piston chamber area [m^2]	A_a - Rod chamber area [m^2]
A_{ce} – Accumulator heat exchange surface area [m^2]	dh – Accumulator heat exchange [W]
hgf – Thermal heat exchange between liquid and gas [W]	h_{conv} – Accumulator heat exchange coefficient [-]
k_{heat} – Accumulator heat exchange gain [-]	η_v - Volumetric efficiency [-]
n_s – Pump/ E-pump speed [RPS]	$n_{s,max}$ – Maximum pump/ E-pump speed [RPS]
p_0, p_1, p_2 – Accumulator pressures at corresponding pressures [bar]	P_{frict}^{P} – Frictional heat generated in pumping mode [kW]
P_{frict}^{M} – Frictional heat generated in motoring mode [kW]	Δp – Pressure drop across the high-pressure and low-pressure ports of the HM [bar]
Q_A – Piston chamber flowrate [m^3/s]	Q_a – Rod chamber flowrate [m^3/s]
Q_{in} – HM inlet flowrate [m^3/s]	Q_{out} – HM outlet flowrate [m^3/s]
S_l – Cylinder stroke [m]	τ – HM Torque [N/m]
T_g – Accumulator gas temperature [k].	T_l – Accumulator liquid temperature [k]
T_w – Accumulator wall temperature [k]	$tcgf$ – Thermal conductivity of the gas/.liquid [W/m k]
V_a – Rod chamber volume of the cylinder [m^3]	V_A – Piston chamber volume of the cylinder [m^3]
V_D - Pump/motor displacement [m^3 /rev]	ΔV – Change in volume of the Accumulator [Liters]
V_0, V_1, V_2 – Accumulator volume filled at 0%, 10%, 90% respectively [m^3]	$\dot{x}$ - Linear velocity [m/s]

Abbreviations

A1: Manifold - accumulator port 1	A2: Manifold - accumulator port 1
BPV: Bypass valve	C1: Manifold - cylinder piston chamber port
C2: Manifold - cylinder rod chamber port	CFD: Computational Fluid Dynamics
DRV: Dual Acting Pressure Relief Valve	DV1 and DV2: 2/2 direction control valves
EM: Electric Motor	EHA: Electro-Hydraulic Actuator

E-pump: Electro-hydraulic Pump	EHU: Electro Hydraulic Unit
iEHA: integrated Electro-Hydraulic Actuator	IGM: Internal Gear Machine
INV: Inverter	Pacc: Accumulator pressure
PMSM: Permanent Magnet Synchronous Motor	PRV: Pressure Relief Valve
Qacc: Accumulator flow	SHV: Shuttle Valve
THA: Thermo-Hydraulic Accumulator	THC: Thermo-Hydraulic Chamber
THFHE: Thermo-Hydraulic Flow with Heat Exchange	THPM: Thermo-Hydraulic Pump/ Motor

1. INTRODUCTION

Self-contained EHAs have been introduced decades ago for critical controls in aerospace applications, with clear reliability advantages over electromechanical actuators, EMAs [1]. Their usage can eliminate long hydraulic lines and reduce leakages. Most importantly, their high energy efficiency, comparable to that of electromechanical actuators [2], makes their use particularly attractive for electrified off-road vehicles. EHAs also consistently reduce the space-claims actuation system, even compared to EMAs [3]. In EHA, high efficiency is achieved through a throttle-less implementation of the primary control concept [4, 5], in which the electric machine and hydraulic machines (i.e., E-Pump) are regulated to precisely match the hydraulic cylinder's flow demand.

Additionally, several EHA concepts also enable energy recuperation during overrunning loads [6]. These positive, energetic features of EHAs have promoted the implementation of "distributed EHA", as conceptually represented in Fig. 1.1, in which all components of the EHA system are distributed across different locations within the machine. Examples include the systems by Qu et al. in [7, 8], which also demonstrate the flexibility of the distributed arrangement, enabling both open-circuit and closed-circuit primary-controlled configurations. However, these implementations do not exploit the potential advantages of component integration, such as reduced overall space requirements and leakage. Such benefits are clearly desirable for off-road machinery and clearly motivated efforts such as the implementation proposed by Pietrzyk et al. [9].

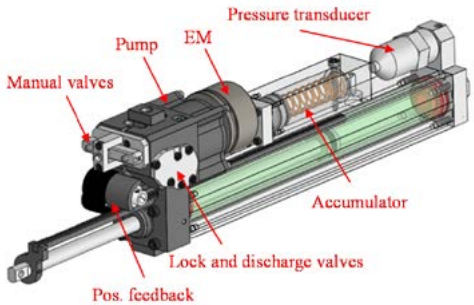


Figure 1.1. – Self-contained Electro Hydraulic Actuator designed by Altare et al. [10]

A clear challenge in deploying self-integrated EHAs in mobile applications, such as off-road machinery, is related to the design morphology. In fact, the typical design of a compact EHA resembles that of Fig. 1.1, where all the critical EHA elements are widely exposed to the surrounding environment. This arrangement is clearly unfavourable in harsh applications with a high risk of impact, such as construction equipment.

Integrated EHA typically require closed-circuit configurations of the hydraulic system, as in Fig. 1.2, that leverage pressurized reservoirs (usually low-pressure accumulators) for a flexible installation arrangement. In such a circuit, the cylinder extension and retraction always require flow corresponding to the displaced rod volume (usually referred to as differential flow). Creating control complexities and increasing the requirements on the hydraulic control system [11]. State-of-the-art EHAs do not leverage this differential flow for useful functions, such as heat dissipation, as in the proposed solution.

The present paper proposes a novel integrated design methodology for EHAs, particularly suitable for mobile applications. The study covers the design and the integration of the EHA, including component sizing and thermal behaviour analysis through simulations. The paper also discusses practical aspects of design integration and concludes with a summary of findings and conclusions.

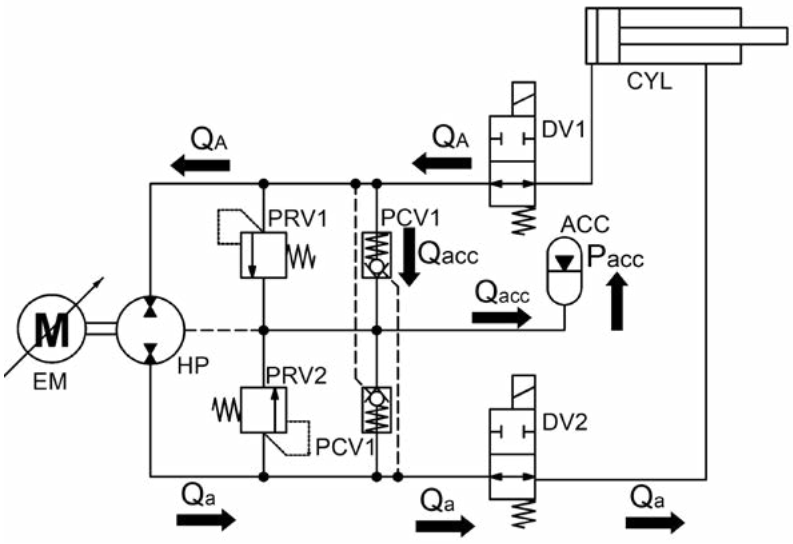


Figure 1.2. Closed-circuit EHA

The remainder of this paper is divided into the following sections. Section 2 presents the proposed electro-hydraulic actuator, including an overview and rationale, as well as its basic functioning, with isometric views and CAD details such as the cooling path. Section 3 describes the reference application and the choice of points for EHA functioning. Section 4 covers the basic component sizing, detailing the methods and final parameters for the electric machine, hydraulic machine, and Accumulator, along with the implications of the bypass valve. Section 5 provides practical information on design integration. Section 6 assesses the thermal behaviour through numerical simulations for the iEHA and cooling analysis for the electric motor. Finally, Section 7 concludes the paper by summarising the findings and conclusions.

2. INTEGRATED ELECTRO-HYDRAULIC ACTUATOR DESIGN

The principle of the proposed design is to maximize component integration, bringing the conventional architecture conceptually represented in Fig. 2.1(a) to a resulting architecture of Fig. 2.1(b). The Accumulator is concentric to the cylinder, and the E-pump is placed behind the cylinder on the same axis. Finally, the valve manifold is placed between the cylinder and the E-pump. The valve manifold not only serves as the point where all the valves are mounted but also as the part that provides hydraulic connections to all the components of the actuator. The proposed architecture leverages the known advantages for integrating electric and hydraulic machines (E-Pumps), in terms of magnet mass of the electric machine and overall dimension [12, 13,14], with additional advantages of integrating the pressurized reservoir within the hydraulic actuator. Additionally, an internal manifold valve integrates the hydraulic circuit, allowing for recirculation of the differential flow within the E-Pump. These features will be further detailed in the following subsections.

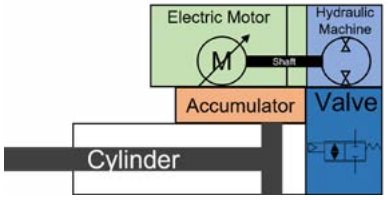


Figure 2.1(a). Traditional EHA design

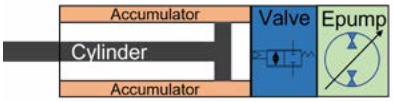


Figure 2.1(b). Proposed iEHA design

2.1. *Accumulator Design*

The low-pressure Accumulator is arranged concentric to the cylinder, similarly to a multi-chamber cylinder arrangement as commercialized by [15]. A ring-type piston is used to implement an original form of piston-type gas accumulator as shown in Fig. 2.2. Such a design not only reduces spatial requirement, but it is also adaptable to different piston strokes, as the volume requirement of the Accumulator linearly scales with the stroke length. The Accumulator uses the outer surface of the cylinder as one running surface, and a concentric cylinder running the entire length of the cylinder serves as the outer running surface for the Accumulator.

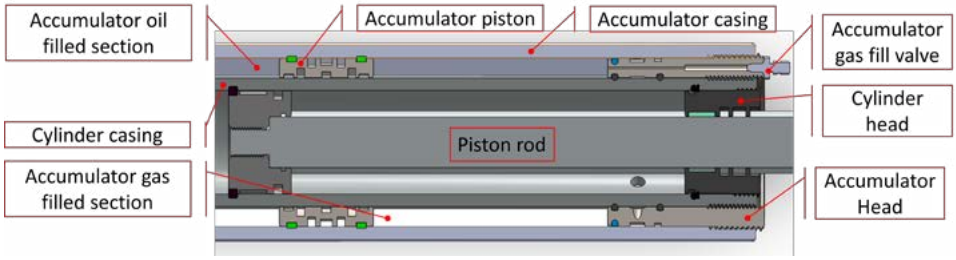


Figure 2.2: Accumulator design

2.2. *E-pump Design*

This section covers the design of the E-pump. The E-pump is designed based on the previous work by Pan and Zappaterra [13]. To enable energy recovery during assistive loading conditions, a four-quadrant hydraulic pump is required. Due to space constraints, a fixed-displacement pump is preferred, which also allows us to maintain a simple, compact design.

Considering the requirement of noise reduction, an internal gear pump is selected. And since internal gear pumps provide better filling and less noise [16] than external gear machines, a more compact design is possible. Moreover, using an internal gear pump offers a key advantage: the external gear can be directly driven, as illustrated in Fig. 2.3, resulting in a more compact E-pump design. In addition to its compactness, the internal gear pump can be designed so that the journal bearing supporting the external gear of the pump can also support the rotor of the electric motor [17]. Finally, the pump sizing is determined by the required flow rate derived from the cylinder specifications and the electric motor's maximum operating speed. Based on this, an internal gear machine with a 13-tooth pinion and a 19-tooth external gear was chosen.

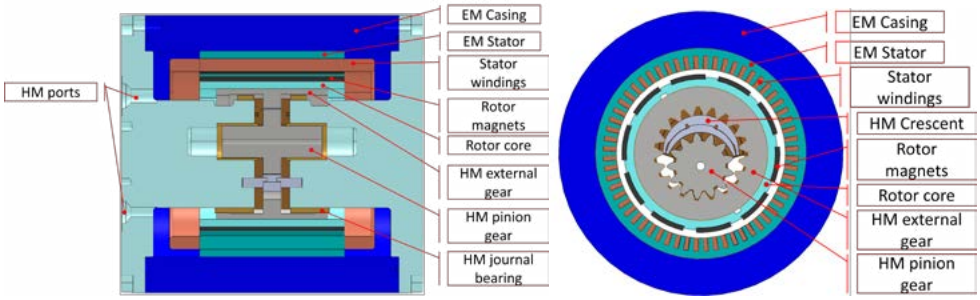


Figure 2.3(a). Radial cross-section Figure 2.3(b). Axial cross-section

Figure 2.3: E-pump design

A novel design feature of the E-pump design is using the differential flow to and from the Accumulator to cool the electric motor, as illustrated in Fig.2.5. The E-pump is designed so that the accumulator flow passes over the stator of the electric motor, thus cooling it. The flow path of the oil has been optimized using simulation to make sure that there are no hot spots in the stator. The heat energy transferred into the oil is later radiated into the atmosphere by the accumulator [18]. The E-pump design also discharges the drain flow into the differential flow.

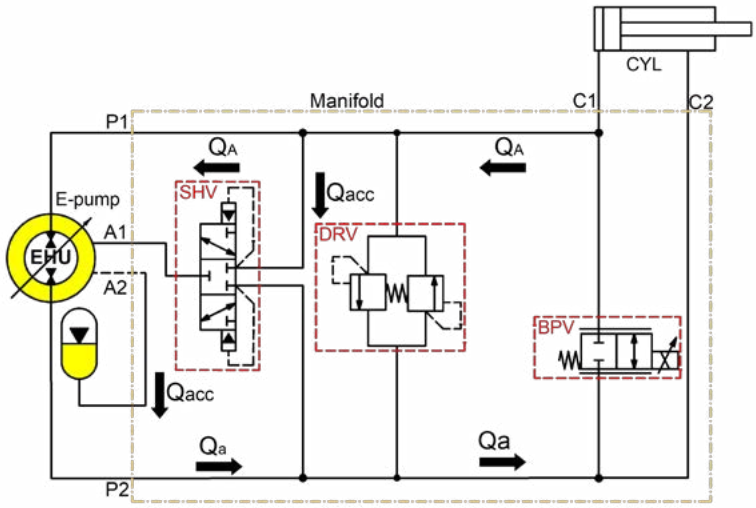


Figure 2.4 ISO schematic of the proposed EHA

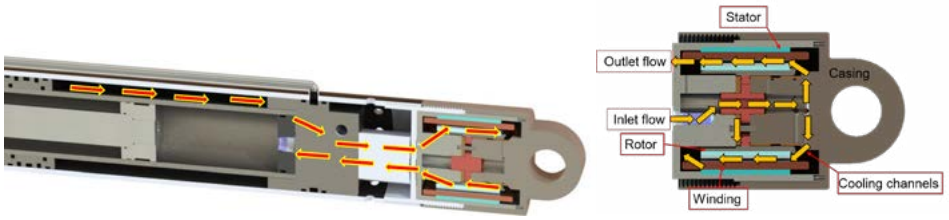


Figure 2.5. EHA Cooling flow path

2.3. Hydraulic circuit and valve manifold design

As illustrated in Fig. 2.4, the EHA was designed with three valves, of which one is a solenoid-operated proportional valve. This architecture represents the most minimalist implementation of a primary controlled circuit, where the low-pressure circuit is handled by an inverse shuttle valve (SHV) similar to that discussed in [19]. The circuit also includes a bypass valve (BPV) to improve control at low actuator velocities without running the E-pump at low RPMs. Moreover, as described in [20], the BPV also allows achieving high retraction velocities without increasing the E-pump's displacement, so that its displacement can be optimised to meet the maximum extension velocity requirement. To prevent any over-pressurisation events, a dual-acting version of an adjustable pressure relief valve (PRV) called the DRV was used. The DRV consists of 2 antiparallelly arranged pressure relief valves that are adjusted using a single screw. The final of the manifold design is shown in Fig. 2.6.

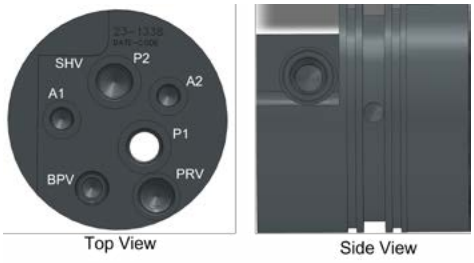


Figure 2.6: Manifold design

3. SIZING CONSIDERATIONS

The design sizing was based on considering the boom function of a mini excavator available in the market. The design data were obtained based on the cylinder size, velocity, and force required for the boom operation of a reference machine operated by an expert operator. Using this known data, a drive cycle was generated (shown in Fig. 3.3), and a basic assumption was made on the speed and torque requirements of the electric motor. Here, a design process, as shown in Fig. 3.1, was used to design and optimize the sizes of the electric motor, hydraulic pump, and manifold.



Figure 3.1: Designing procedure

After the initial design considerations were completed, particular attention to the packing of the integrated cylinder was taken into account. The valve locations impact how much they stick out for the manifold; this directly impacts the overall length of the cylinder. This is further complicated by the fact that the BPV, which requires exposure to air flow as it is solenoid-operated and requires cooling.

3.1. EHA operating points

As illustrated in Fig. 3.2, the design of the EHA components is based on a reference cycle to which the reference machine will be regularly subjected, as shown in Fig. 3.3. The electric motor power was selected based on the peak power in the reference drive cycle being run for a sustained period of time without leading to overheating.

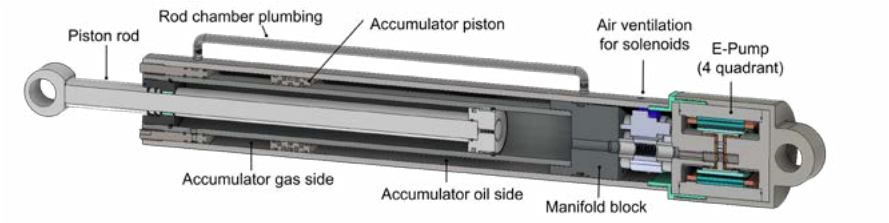


Figure 3.2 Design Overview

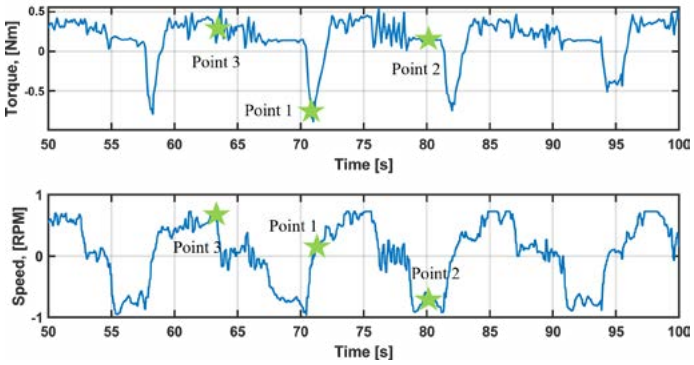


Figure 3.3 Operating points

4. BASIC COMPONENT SIZING

As detailed in Chapter 3, sizing begins with selecting a reference machine. Thus, the cylinder details are driven by the cylinder on the reference machine. The cylinder size then drives the volume of the Accumulator. The max velocity and the details of the cylinder drive the HM sizing. The HM size then determines the EM size. The sizing methodology used ensures that the EM's power rating accounts for all losses of earlier stages and can sustainably provide the power required for the reference drive cycle. Table 4.1 provides the normalized and relative cylinder details.

Parameters	*Symbol*	*Value*
Normalized Stroke length	S_l	1
Relative Piston Diameter	-	$S_l/7.1$
Relative Rod Diameter	-	$S_l/13.4$

Table 4.1 Normalized/ relative cylinder details

4.1. Electric motor sizing methodology & Final parameters

A Permanent Magnet Synchronous Motor (PMSM) was selected, as it offers very high power density. Furthermore, since the rotor of a PMSM comprises a set of magnets, the rotor core can be made hollow, allowing the hydraulic machine to nest inside the rotor.

The EM design was performed using GOSET software from Sudhoff [21]. This software is written in MATLAB and uses a multi-objective design approach to generate the EM design parameters based on the selected input parameters. The software generates a Pareto-optimal curve to select the optimal design. The software then uses the chosen design to develop the final EM design.

Parameters	*Symbol*	*Value*	*Units*
Nominal DC supply voltage	V_{DC}	800	[V]
Maximum motor speed	n_s	9500	[RPM]
Nominal motor torque	$\tau_{m,nom}$	25	[Nm]
Maximum motor torque	$\tau_{m,max}$	40	[Nm]

Table 4.2 EM input parameters

The input parameters are provided in Table 4.2. The input parameters are based on a drive cycle generated (Fig. 3 .3) using the methodology detailed earlier in this chapter, and on the choice of the battery and the inverter. Three points of interest from the drive cycle were selected based on the action being performed and the electric motor's instantaneous requirements. These 3 points cover the three most common modes of operation at regular loading conditions. Points 1 & 3 are a normal raising or quadrant one in Fig. 4.1. At point 1, the cylinder is starting to accelerate, and thus the EM speed is low, but the torque is at $\tau\ m$, max, But, at point 1, the cylinder is extending at high velocity and at no appreciable load. This requires a high EM speed and low torque. Point 2 is assistive retraction or quadrant four in Fig. 4.1. Here, the boom is lowering due to gravity, and thus energy recovery is possible. The final design parameters for the EM as detailed in Table 4.3

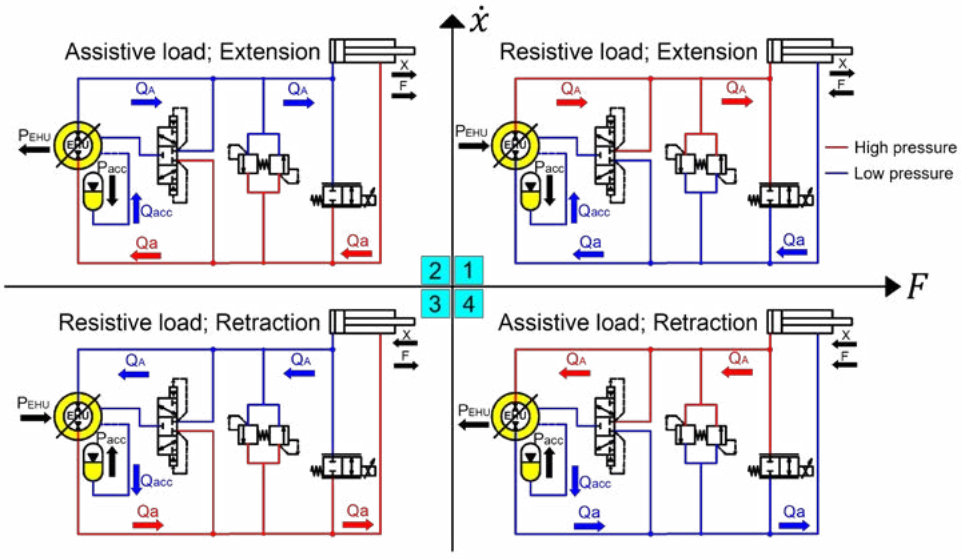


Figure 4.1 Operating modes of the EHA

Parameters	*Symbol*	*Value*	*Units*
Maximum motor speed	$n_{s,\,max}$	9500	[RPM]
Maximum motor torque	$\tau_{m,max}$	40	[Nm]

Table 4.3 Final EM parameters

4.2. Hydraulic machine sizing & Final parameter

The primary factor driving the design of the hydraulic machine is the maximum operating speed. The hydraulic machine is integrated into the rotor of the electric machine, resulting in a 1:1 gear ratio. Based on the conclusion from Zappaterra et al.[8] and Pietrzyk et al.[7]. An internal gear machine was determined to be suitable for this purpose.

The size of the hydraulic machine is determined by using Eqns.4.1 & 4.2. Based on the max velocity of the actuator during extension and the max rotational speed of the electric motor ($n_{s,\,max}$).

$$Q_A = \dot{x} \cdot A_A \tag{4.1}$$

$$V_D = \frac{Q_A \cdot \eta_v}{n_{s,max}} \tag{4.2}$$

In addition, the pump diameter is constrained by the electric motor's rotor diameter. This means that beyond a certain point, the length of the pump's gear set must increase to increase the displacement. Table 4.4 shows the final parameter for the hydraulic machine.

Parameters	*Symbol*	*Value*	*Units*
HM displacement	V_D	8	[cc]

Table 4.4. Final HM parameter

4.3. Accumulator sizing & Final parameters

Since a closed-circuit design was selected, an accumulator provides the differential flow of the cylinder. Hence, this is the primary driving factor for sizing the Accumulator. The cylinder's annular volume is derived from Eqn 4.3

$$V_a = A_a \cdot S_l \tag{4.3}$$

Based on the annular volume and the minimum and maximum accumulator pressures, the accumulator size is calculated using Eqns 4.4 & 4.5

$$\Delta V = V_0 \cdot p_0^{\frac{1}{\gamma}} \cdot \frac{p_2^{\frac{1}{\gamma}} - p_1^{\frac{1}{\gamma}}}{p_2^{\frac{1}{\gamma}} * p_1^{\frac{1}{\gamma}}} \tag{4.4}$$

$$V_0 = \frac{(\frac{p_2}{p_1})^{\frac{1}{\gamma}} \cdot \Delta V}{(\frac{p_0}{p_1})^{\frac{1}{\gamma}} \cdot \left[(\frac{p_2}{p_1})^{\frac{1}{\gamma}} - 1\right]} \tag{4.5}$$

The input parameters for the sizing of the accumulator are derived based on the cylinder details and are detailed in Table 4.5, and the final accumulator sizing is detailed in Table 4.6.

Parameters	*Symbol*	*Value*	*Units*
Charge gas	-	Nitrogen	[-]
Polytropic index	γ	1.4 @ 20°C	[-]
Precharge pressure	p_0	5.2	[bar]
Minimum working pressure	p_1	6.0	[bar]
Required volume	ΔV	3.44	[Liters]

Table 4.5 Accumulator input parameters

Parameters	*Symbol*	*Value*	*Units*
Accumulator volume	V_0	3.6	[Liters]
Maximum accumulator pressure	p_2	8.92	[bar]

Table 4.6 Accumulator final parameters

4.4. Bypass valve implications

The bypass valve is used in two scenarios. The primary use is to reduce the size of the hydraulic machine while achieving the same fast retraction velocity. The bypass valve is opened in quadrant 4 of Fig. 4.1(assistive retraction), which diverts some flow from the piston side chamber to the rod side of the cylinder. The secondary use is to allow high-velocity operation, especially when the hydraulic machine's speed falls below the minimum required to achieve this velocity.

5. DESIGN INTEGRATION

As mentioned, in the existing EHA designs, the components are externally connected to each other. This design aims to integrate all the components and place them on the same axis as the cylinder. For this reason, the majority of the components are cylindrical and are stacked. The Accumulator was designed to be concentric to the cylinder to increase the integration. The concentric cylinder would also allow the outer surface of the cylinder to be used as the inner running surface for the accumulator piston. Stacking/placing the manifold behind the cylinder will enable it to act as both the piston end cap of the cylinder and one of the end caps for the Accumulator, while providing fluid paths to both. The E-pump is placed/ stacked behind the manifold, and a connector connects the E-pump to the manifold. This connector also provides the space required for the valves. Since the outer running surface of the Accumulator is exposed to the environment, it can be used as a radiator for the heat generated. This eliminates the need for an additional external cooling system. Fig. 3.1 shows a near-complete design of the iEHA.

6. SIMULATION – THERMAL BEHAVIOUR & RESULTS

To access the thermal behaviour of the EHA, thermal simulations were conducted using two software packages. Simcenter's Amesim and Dassault Systèmes's SolidWorks were used to run thermal simulations of the complete EHA and the electric machine, respectively.

These simulations were performed using drive cycles generated to mimic the three loading points from Fig 3.3. The thermal simulation time period was set to a minimum of 9000 seconds for each drive cycle.

To ensure consistency between the two simulation methods. The boundary conditions

(mainly the oil flowrate and oil pressure used to cool the E-Pump) and other parameters were obtained by first running the Amesim simulations using the generated drive cycles, capturing the data, and then using this data as inputs to the SolidWorks simulations.

6.1. Hydrothermal simulation using Amesim

As illustrated in Fig. 6.1, a thermal simulation model that considered both electrical and hydraulic characteristics was built and simulated, using the hydraulic equations from Chapter 4 for the hydraulic system. Further equations shown in this chapter were used to simulate thermal energy in electrical and hydraulic systems. Specifically, Eqs. 6.1 & 6.2 were used to build a discretized thermal model of the EM, accounting for energy loss due to electrical resistance and Eddy currents. The heat exchange between the oil and the EM is also simulated using a thermohydraulic flow model with heat transfer, and the contact areas were determined from the EM's 3D CAD. In addition, the accumulator oil flow path through the EM for cooling purposes was considered, along with the flow path lengths and diameters. Furthermore, the accumulator components used consider the design's features, and the design's dimensions are used to calculate the external surface area of the Accumulator in contact with the environment. A conservative wind speed and temperature were also considered.

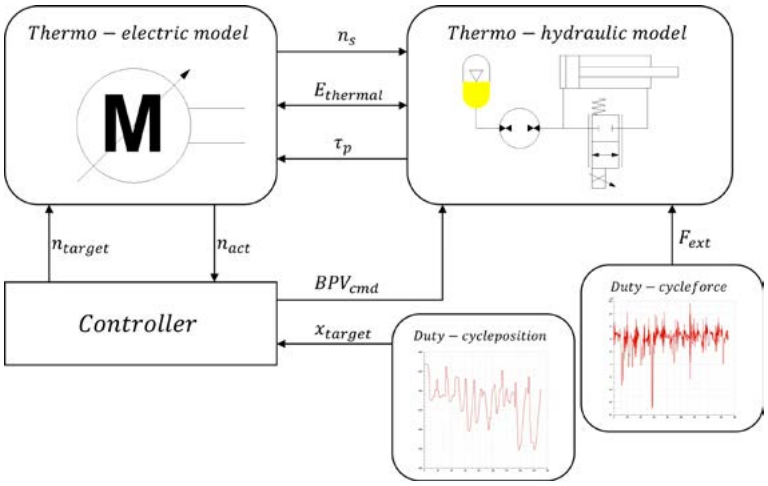


Figure 6.1 Simulation block diagram

$$dh_1 = \frac{t_2 - t_1}{TR} \tag{6.1}$$

$$\frac{dT}{dt} = \frac{\sum_{i=1}^{N} dh_i}{TC} \tag{6.2}$$

For the thermo-electric motor simulation, a lumped parameter RC model was built in Amesim. Fig. 6.2 illustrates an equivalent RC lumped parameter model. The thermal resistances (TR; illustrated by the black resistors in Fig. 6.2) and their heat flow rates are calculated using Eqn. 6.1. The thermal capacitances (TC) and the rate of change of temperature are computed using Eqn. 6.2. The EM lumped parameter model also accounts for the cooling oil flow and models the heat exchange using thermo-hydraulic flow with a heat exchange component in Amesin (THFHE; illustrated as red resistors in Fig. 6.2).

$$P_{frict}^{P} = (\tau \cdot \omega) - (Q_{in} \cdot \Delta p) \quad (6.3)$$

$$P_{frict}^{M} = (Q_{out} \cdot \Delta p) - (\tau \cdot \omega) \quad (6.4)$$

$$hgf = tcgf \cdot (T_g - T_l) \quad (6.5)$$

$$dh = k_{Heat} \cdot h_{conv} \cdot A_{ce} \cdot (T_l - T_w) \quad (6.6)$$

Furthermore, the thermo-hydraulic components we used for the hydraulic machine, cylinder and the Accumulator. As such, the friction in the pump/ motor is calculated using Eqn. 6.3 in pumping mode and Eqn. 6.4 in motoring mode. The heat exchange in the Accumulator is calculated using Eqn. 6.5. Finally, the heat exchange between the air and the Accumulator is calculated using Eqn. 6.6.

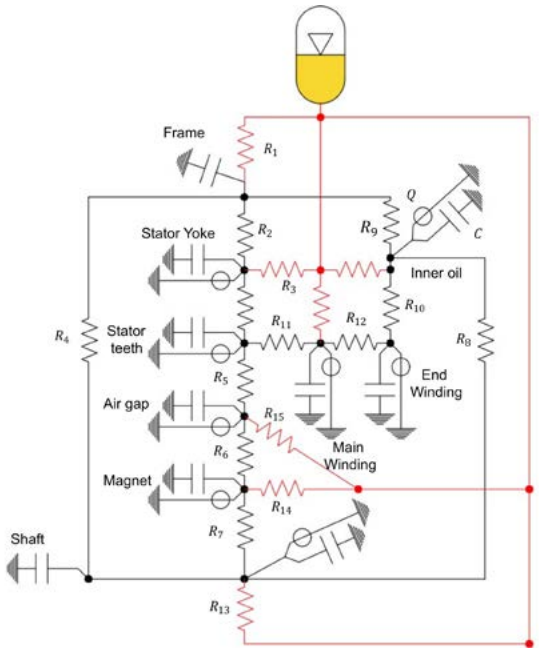


Figure 6.2 Equivalent EM lumped parameter model

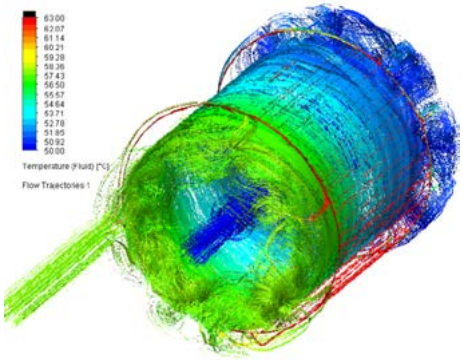
6.2. Thermal behaviour simulation using Amesim & SolidWorks

In addition to the hydrothermal simulation in Amesim. A thermal simulation in SolidWorks was performed to verify the proper/ complete cooling of the EM. This simulation was performed on the standalone E-pump design with properties identical to those of the integrated E-pump design. The primary objective of simulating EM cooling using both a lumped-parameter Amesim model and a SolidWorks model was to verify that the lumped-parameter model accurately simulates the losses and temperatures of the EM. The secondary objective of the flow simulation was to verify the even cooling of the EM stator. The results and the subsequent comparison of the temperatures are shown in the Table. 6.1. The average temperature difference between the models was 1.6°C, and the maximum temperature difference was 4.0°C. Since the windings generate most of the heat and are the main priority, a total temperature reduction of 4.0°C (5.27%) is promising.

Location	*Amesim*	*SolidWorks*	*ΔT*
Winding	76.0	72.0	4.0
Stator	71.0	71.5	0.5
Magnet	54.0	54.0	0
Rotor core	53.6	54.2	0.6
Fluid at the outlet	54.1	57.0	2.9
Average deviation			1.6

Table 6.1 Temperature comparison

The secondary reason for this simulation was to ensure even & sufficient cooling of the EM. To access this, a further simulation was performed at multiple cylinder velocities to account for the varying cooling oil flow rates. As illustrated in Fig. 6.3, the fluid can evenly cool the EM, resulting in a maximum stator temperature of 76 °C. Fig. 6.4 also illustrates the temperature distribution in the EM. It can be observed that the maximum temperature deviation in the windings is 9.29°C.



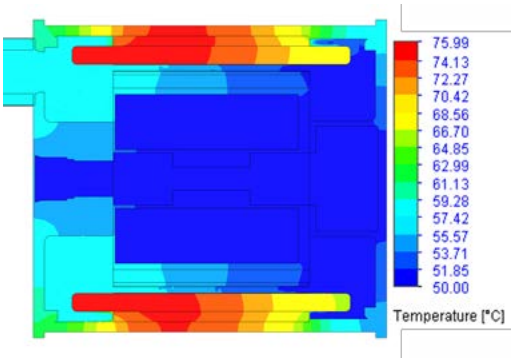


Figure 6.3 SolidWorks thermal & flow simulation Figure 6.4 SolidWorks – Simulated temperature distribution

7. CONCLUSION

In conclusion, this paper proposes an innovative design for an integrated Electro-Hydraulic Actuator (iEHA) by aggressively integrating the components of a linear hydraulic actuation system. The paper focuses on the design challenges of integrating the electric prime mover (EM), hydraulic pump, valve manifold, and Accumulator into a single-cylinder unit. Particular attention is given to key design areas. Such as the integration of the EM and the hydraulic pump (the E-pump), the concentric design approach for the cylinder and the Accumulator, and the utilization of differential flow to cool the E-pump.

Furthermore, the system's sizing is supported by two thermal simulation methods. A lumped-parameter model built in Siemens Amesim to verify the iEHA's design approach, with a particular focus on the E-pump's thermals. Additionally, a CFD simulation in Dassault Systèmes' SolidWorks confirmed the feasibility of the E-pump's cooling approach. These results demonstrate that the system temperature remains within acceptable limits and that the cooling architecture provides adequate flow to cool the E-pump evenly.

Though this paper focuses on simulation, the current work involves testing prototypes of the E-pump, the manifold, and the control system. In the future, the plan is to build a complete

iEHA prototype based on lessons learned from both simulations and the prototyping phase. Once proven, the proposed iEHA concept could find wide application in several electrified areas, such as aerospace and electrified mobile machines, where efficiency is critical.

8. REFERENCES

[1] Huang, H., 2015, Challenges in More Electric Aircraft (MEA), IEEE Transportation Electrification, Community newsletter July/August 2015.

[2] Haack S., Flaig A., 2022, Sustainable Hydraulics for Industrial and Mobile Applications, 13th Int. Fluid Power Conference, Germany, June 2022. https://www.doi.org/10.18154/RWTH-2023-05258.

[3] D. Hagen, D. Padovani and M. Choux, "Guidelines to Select Between Self-Contained Electro-Hydraulic and Electromechanical Cylinders," *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Kristiansand, Norway, 2020. https://www.doi.org/10.1109/ICIEA48937.2020.9248373.

[4] Vacca, Andrea, and Germano Franzoni. Hydraulic fluid power: Fundamentals, applications, and circuit design. John Wiley & Sons, 2021; ISBN: 978-1-119-56910-7.

[5] Li, Bo, Liu, Yongteng, Tan, Cao, Qin, Qijing and Lu, Yingtao. "Review on electro-hydrostatic actuator: System configurations, design methods and control technologies." International Journal of Mechatronics and Manufacturing Systems Vol. 13 No. 4 (2020): pp. 323–346. https://www.doi.org/10.1504/IJMMS.2020.112356

[6] Qu S., Fassbender D., Vacca A., Busquets E., 2021, A High-Efficient Solution for Electro-Hydraulic Actuators with Energy Regeneration Capability, Energy, Vol 2016, February 2021. https://www.doi.org/119291. 10.1016/j.energy.2020.119291.

[7] Shaoyang Qu, Federico Zappaterra, Andrea Vacca, Enrique Busquets, An electrified boom actuation system with energy regeneration capability driven by a novel electro-hydraulic unit, Energy Conversion and Management, Volume 293, 2023. https://www.doi.org/10.1016/j.enconman.2023.117443

[8] Qu, S., Zappaterra, F., Vacca, A., Liu, Z., & Busquets, E. (2023). Experimental Verification of An Electro-Hydraulic Actuation System Driven by An Integrated Electro-Hydraulic Unit. International Journal of Fluid Power. https://www.doi.org/10.13052/ijfp1439-9776.2427

[9] Pietrzyk, Tobias et al. "Design study of a high speed power unit for electro hydraulic actuators (EHA) in mobile applications." (2018). https://www.doi.org/10.18154/RWTH-2018-224632

[10] G. Altare, A. Vacca and C. Richter, "A novel pump design for an efficient and compact Electro-Hydraulic Actuator IEEE aerospace conference," 2014 IEEE Aerospace Conference, Big Sky, MT, USA, 2014. https://www.doi.org/10.1016/j.proeng.2015.06.003

[11] Rahmfeld, Robert, Monika Ivantysynova, and Jürgen Weber. "Displacement controlled wheel loader–a simple and clever solution." Proceedings of the 4th International Fluid Power Conference, Dresden, Germany. 2004.

[12] Federico Zappaterra, Andrea Vacca, Scott D. Sudhoff, A compact design for an

electric driven hydraulic gear machine capable of multiple quadrant operation, Mechanism and Machine Theory, Volume 177, 2022. https://www.doi.org/10.1016/j.mechmachtheory.2022.105024

[13] Federico Zappaterra, Dinghao Pan, Thomas Ransegnola, Andrea Vacca, Scott D. Sudhoff, Enrique Busquets, A novel electro-hydraulic unit design based on a shaftless integration of an internal gear machine and a permanent magnet electric machine, Energy Conversion and Management, Volume 310, 2024. https://doi.org/10.1016/j.enconman.2024.118432

[14] Andrea Vacca, Hassan Assaf, Hao Zhang, Federico Zappaterra, Integrated electrohydraulic actuator, 2023, Patent application: WO/2024/205693.

[15] Norrhydro, NorrDigi® EMA - Complete electromechanical solutions for superior energy-efficiency & reliable motion control, https://www.norrhydro.com/en/norrdigi_ema, Last accessed: 10/25/2025

[16] Ivantysyn, J. and Ivantysynova, M. 2001. Hydrostatic Pumps and Motors, Principles, Designs, Performance, Modelling, Analysis, Control and Testing. New Delhi. Academia Books International. ISBN 81-85522-16-2

[17] Hao Zhang, Federico Zappaterra, Satish Kumar Raju Kalidindi, Andrea Vacca, Hassan Assaf, Assemblies for an internal gear pump with hydrostatic support features, 2023, Patent application: WO/2024/205694.

[18] Andrea Vacca, Hassan Assaf, Satish Kumar Raju Kalidindi, Federico Zappaterra, Cooling configuration for an electric motor driving an electrohydraulic actuator, 2023, Patent application: WO/2024/205692.

[19] Ketelsen, S.; Padovani, D.; Andersen, T.O.; Ebbesen, M.K.; Schmidt, L. Classification and Review of Pump-Controlled Differential Cylinder Drives. Energies 2019. https://doi.org/10.3390/en12071293

[20] Shaoyang Qu, Federico Zappaterra, Andrea Vacca, Enrique Busquets, An electrified boom actuation system with energy regeneration capability driven by a novel electro-hydraulic unit, Energy Conversion and Management, Volume 293, 2023. https://doi.org/10.1016/j.enconman.2023.117443

[21] Sudhoff, Scott D. Power magnetic devices: a multi-objective design approach. John Wiley & Sons (2021). ISBN: 978-1-119-67463-4

Biographies



Seshan Calapatti Suresh received his bachelor's degree in Mechatronics Engineering from Acharya Institutes in 2019. He then pursued a master's degree in Mechatronics Engineering from Politecnico Di Torino, Italy. In 2022, he had the opportunity to complete his master's thesis at the Maha Fluid Power Research Centre (Maha), Purdue University, as a visiting scholar. This visiting scholarship led to my candidacy for a PhD at Maha.



Dr. Andrea Vacca is the director of the Maha Fluid Power Research Centre, the largest academic centre dedicated to fluid power and motion control research in the U.S. His research focuses on components and systems design to promote electrification, low noise emissions, high productivity, and concepts and techniques for condition monitoring. Andrea Vacca is the editor in chief for the International Journal of Fluid Power, a fellow of the American Society of Mechanical Engineers, and a director of the Global Fluid Power Society. In 2019, he also received the J. Bramah medal from the Institution of Mechanical Engineers. In 2025, he was inducted into the Fluid Power Hall of Fame by the International Fluid Power Society.



Dr. Hassan is Control Systems Senior Engineer at Caterpillar Inc. He specializes in development and tuning of electro-hydraulic systems and components. He completed his PhD at Purdue University in 2023 through the Maha Fluid Power Research Center. His doctoral research focused on developing Innovative Methods for Electro-Hydraulic Actuation and Fluid Power Education.



Dr. Zhang is a chief engineer and Parker Engineering Fellow at Parker Motion Systems Group, with over 20 years of experience in the motion control industry. He has led multiple market-driven technology roadmap establishments and directed the execution of technology development for mobile and industrial applications. He has led and coordinated internal and external resources in various advanced mobile electro-hydraulic components/systems development and commercialization activities. His current focus is on intelligent mobile hydraulic system controls with power-dense and efficient E-Pumps.

Non-destructive Detection of Fatigue Cracks in Hydraulic Components

Boden, Lukas[1*]; Blatter, Fabian[2]; Marcus Geimer [2], Katharina Schmitz [1]

1. RWTH Aachen University – Institute for Fluid Power Drives and Systems (ifas),
2. Institute of Mobile Machines, Karlsruhe Institute of Technology, Karlsruhe, German

lukas.boden@ifas.rwth-aachen.de

Abstract.

As hydraulic components continue to evolve toward higher power densities and efficiencies, their structural design must meet increasingly demanding performance and reliability requirements. This often requires a targeted, sustainable use of material, particularly in highly stressed regions. This demands the assessment of the components' load limits, which are determined by cyclic pressure impulse tests.

During operation most components fail due to fatigue cracks, which have severe consequences. Thus, detailed knowledge of crack-growth behaviour is essential. Crack-propagation depends on the material properties and, above all, on the local stress amplitudes at the crack tip. The latter is changed in fluid power applications since the viscous oil interacts with the crack faces. This study aims to investigate the effect on a large scale. To make experimental investigation more efficient and economical, non-destructive, online monitoring of the crack development is employed. This work focuses on crack detection by acoustic emission (AE) methods, including the measurement setup and the evaluation of the acquired AE signals, while addressing common challenges such as background noise and data handling.

Keywords. Fatigue cracks, non-destructive testing, acoustic emission, pressure peaks

1. INTRODUCTION

As hydraulic components continue to evolve toward higher power densities and efficiencies, their structural design must meet increasingly demanding performance and reliability requirements. Achieving high power density often involves operating at elevated pressures or exploiting a weight-optimized design. Both require a targeted, sustainable use of material, particularly in highly stressed regions, which demands a precise understanding of the components' load limits.

A key failure mechanism is the initiation of fatigue cracks, which can have severe consequences, especially under cyclic internal pressure loading. Consequently, detailed knowledge of crack-growth behaviour is essential. Crack-propagation rates depend on the material properties and, above all, on the local stress amplitudes at the crack tip. [1] Furthermore, in oil environments, two opposing effects alter the stress amplitude compared with classical fracture-mechanics tests in air. On one hand, the crack faces are additionally loaded by the hydrostatic pressure of the oil, which increases the stress amplitude. Therefore, the crack propagates faster than without oil. However, there is an effect known as oil-

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

induced crack closure [2], which is illustrated in **Figure 1-1**. This effect is due to the viscous behaviour of the oil, which is particularly noticeable at high pressure drop rates. In this case, the oil cannot flow out of the crack sufficiently fast after the end of the load. The trapped oil prevents the crack faces from fully reclosing, sustaining residual internal pressure that lowers the effective stress amplitude (while the peak stress remains unchanged) and consequently extends component life. [3]

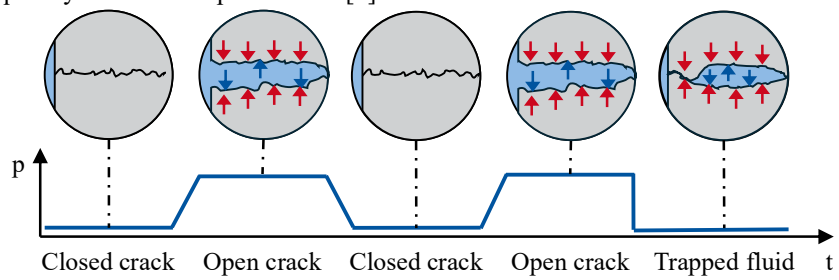


Figure 1-1: Oil-induced crack-closure caused by steep pressure drop rate.

Traditionally, hydraulic components are tested to failure under cyclic loading on pressure-impulse rigs, where cycle frequency significantly affects test duration and thus operating cost. Raising cycle frequency can substantially shorten operating time, in turn, steepen the load gradients, favouring oil-induced crack closure. Consequently, current testing relies on the incorrect assumption that component life is independent of cycle frequency, resulting in discrepancies between theoretical and experimental lifetime predictions [4].

To assess crack propagation during pressure-impulse testing, crack length must be evaluated at distinct intervals. Due to the large scope of the planned measurement series, manual measurement of the forcefully opened crack faces is not feasible. Therefore, this work develops a non-destructive, online monitoring method for hydraulic components to fundamentally investigate the influence of load frequency and pressure drop rates on crack propagation. For this purpose, samples are tested that are designed to initiate a well-defined crack quickly and sustain a steady propagation rate. The detection method involves using acoustic emission (AE) sensors, which are placed at suitable locations on the surface of the samples.

2. THEORY / STATE OF THE ART

The following sections provide the theoretical background and current state of research relevant to this work. They cover the fundamentals of fatigue crack growth, the influence of viscous fluids, and non-destructive testing techniques with a focus on acoustic emission methods.

2.1. Fatigue crack growth

Fatigue cracks are a critical failure mechanism in mechanical components, because they often develop unnoticed and do not cause a significant reduction in mechanical properties until critical failure. [5]

The process leading to final failure in metals, steels, and engineering alloys typically progresses through three distinct stages: nucleation, propagation, and breakthrough. Crack nucleation is the initial phase where microscopic damage begins to accumulate within the

material. This usually occurs under cyclic loading, even when the applied stress is below the material's yield strength. Localized plastic deformation occurs at microstructural imperfections due to dislocation activity, as the material seeks a lower energy configuration. In metals and alloys, persistent slip bands often form within grains and interact with grain boundaries, resulting in surface intrusions and extrusions that serve as early crack precursors. These imperfections can be intrinsic, such as inclusions, grain boundaries, or phase interfaces, or extrinsic, including surface defects like machining marks or scratches. Both types act as stress concentrators, accelerating local damage accumulation and promoting crack nucleation. Once nucleated, cracks begin to grow slowly and stably under continued loading, often referred to as the initiation phase. Initiation typically marks the onset of measurable crack growth from an initial micro-defect. [6]

To describe crack tip propagation, one must first address the mathematical singularity in the stress field at the crack tip [7]. Hence, the stress intensity factor (SIF) K is introduced to characterize this field in a finite and quantifiable manner, enabling calculation of crack growth by means of linear elastic fracture mechanics. It is dependent on the far-field applied stress σ, the current crack length a, and a geometric correction factor f, which accounts for geometry and orientation of the crack, with respect to the stress field. [1]

$$\Delta K = \Delta\sigma\sqrt{\pi a} \cdot f \tag{2-1}$$

Observing the change in SIF ΔK during cyclic loading of stress amplitude $\Delta\sigma$, one can describe the crack growth per cycle da/dN by the Paris-Erdogan equation [8]

$$\frac{\mathrm{da}}{\mathrm{dN}} = \mathrm{C} \cdot \Delta K^{m}, \tag{2-2}$$

with C and m being parameters determined by experimental material tests. The propagation phase can be divided into three phases, as can be seen in **Figure 2-1** in which the crack propagation rate per cycle is plotted against the change in SIF on logarithmic axes. Phase one characterizes the SIF region up to a threshold value ΔK_{th}. Up to this value, no crack propagation is expected. Phase two represents the stable crack growth region, characterized by a linear relationship between $log(da/dN)$ and $log(\Delta K)$, typically described by the Paris Law, extending up to the critical SIF range ΔK_c. Beyond this, in phase three, the crack grows rapidly and becomes unstable, leading to final fracture or breakthrough of the remaining cross-section. Thus, crack growth rate is influenced by the SIF range as well as the stress ratio R, which relates the maximum SIF to the minimum SIF. For a fixed ΔK, da/dN increases with R. Conversely, in the near-threshold regime, the threshold stress intensity factor ΔK_{th} decreases with increasing R [9].

The crack can propagate in relation to the direction of the applied far-field stress, as can be seen in Figure 2-1 on the right side, with three distinct modes identified: Mode I (opening mode), Mode II (sliding mode), and Mode III (tearing mode). In practical applications, a combination of these modes typically occurs. [10] However, this study specifically focuses on the effect of fluid on crack propagation under uniaxial loading of the crack faces subjected to normal stress, which corresponds exclusively to Mode I.

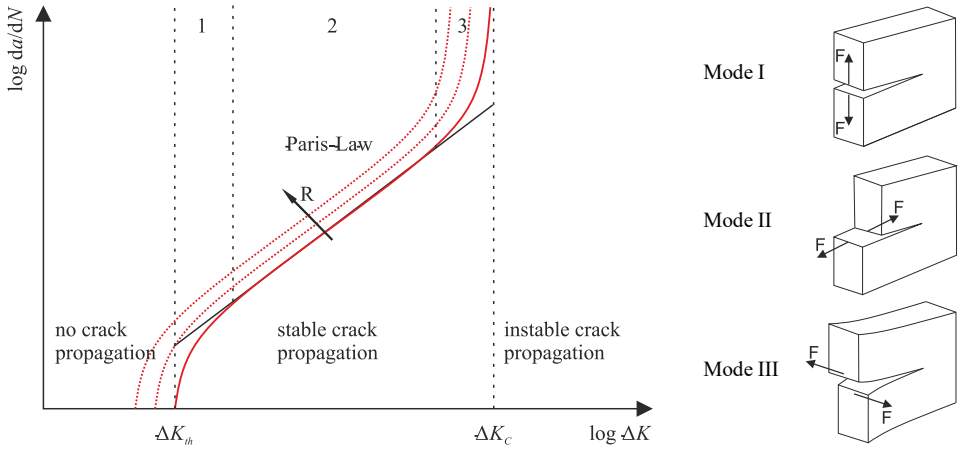


Figure 2-1: Crack growth rate with respect to the stress intensity range (left). Crack modes (right).

Little research has been conducted on the influence of fluid on the crack propagation, since typical metallurgical investigations are standardized and carried out in controlled environments [11]. In most tests, load is applied externally by special test rigs. However, these tests don't represent the real-world loading of cracked hydraulic components. Here, the load is applied directly on the crack faces as hydrostatic pressure. This is why most hydraulic components are subject to pressure impulse testing during development to determine load limits and lifetime.

Early research was conducted by Davis et al., who investigated the effect of oil under hydrostatic pressure on crack propagation in 1989. For this purpose, the specimen was subjected to cyclic mechanical loads in a pressure vessel. Depending on the type of loading, an improvement (tensile loading) or a deterioration (tensile-compressive alternating loading) of fatigue life was observed. [12]

In 2023, Michiels et al. examined both the influence of the pressure gradient [13] during cyclic pressure loading tests on crack growth and the effect of pulse frequency [14]. Additionally, they conducted and validated simulations to evaluate fluid behavior within a crack [3]. It was found that at very high pressure drop rates, oil can be trapped within the crack, as illustrated in Figure 1-1. This leads to a significant change in stress at the crack tip region, especially lowering the effective stress amplitude and thus extending fatigue lifetime. This phenomenon is known as oil-induced crack closure, where the presence of oil within the crack inhibits the crack faces from forming contact. A similar effect can occur due to roughness-induced crack closure, where surface irregularities cause early contact between crack faces, and oxide-induced crack closure, where oxidation products accumulate at the crack tip and impede full closure. [15]

To assess crack propagation rates, the crack length must be measured at defined intervals. In classical fatigue testing, cracks initiate and propagate along the surface and can be monitored using optical or mechanical methods, or through integrated measurement systems within the test rig. In contrast, cracks in hydraulic components, as well as in the specimens investigated in this study, form on the inner surfaces, making direct observation and

measurement significantly more challenging. Moreover, the pressure impulse test rig used in this work does not include built-in capabilities for crack length measurement. Consequently, non-destructive testing (NDT) methods must be employed.

The field of NDT for crack detection is vast, encompassing techniques based on electrical resistivity [16], ultrasonic energy [17], acoustic emission [18]–[21], and electromagnetic impedance [22], as well as digital image correlation [23]. Among these, the present work focuses on the application of AE testing, which enables the online and non-invasive detection of crack initiation and propagation processes during cyclic loading.

2.2. Acoustic emission testing

AE testing is a passive non-destructive testing technique utilized for monitoring the integrity and performance of solid materials, components and structures under stress. This method is based on the detection of high-frequency sound waves generated by the rapid release of energy from localized sources within a material, such as crack propagation, fiber breakage, or other forms of deformation. As these events occur, they emit elastic waves that can be captured by sensitive acoustic sensors such as piezoelectric sensors. In contrast to ultrasonic testing, which requires the active introduction of external acoustic energy, AE testing is considered passive, as it records signals that are naturally emitted by the material itself during loading. The sensors utilize a piezoelectric material that generates an electrical charge in response to mechanical stress. Sensor coupling to the specimen, as well as the setup of the amplification, filtering, and digitalizing devices, is essential for obtaining high quality signals with minimal influence of background noise. The schematics of AE sensor placement on the specimen and the acquired signal are shown in **Figure 2-2**. [24, 25]

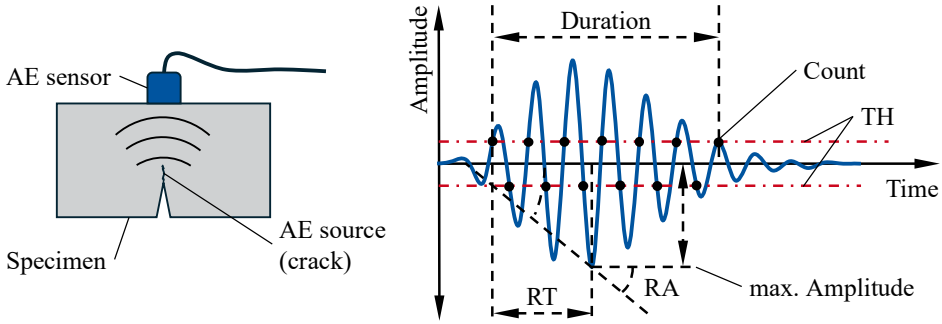


Figure 2-2: Illustration of AE sensor placement on specimen and AE source (left). Schematic AE signal amplitude over time (right) with signal parameters: Risetime (RT), riseangle (RA), duration, maximum amplitude, count of threshold (TH) crossing.

AE signals can be divided into continuous signals (e.g., process-related noise) and transient signals (signals with a recognizable beginning and end), often referred to as bursts [24]. While burst signals can be used to detect specific events such as crack growth, evaluation of continues signals at different time intervals allows for assessing structural changes or degradation over time. In fatigue and material testing burst signals are of primary importance. The raw signal is evaluated in the time and frequency domains to assess the presence of a burst and the corresponding crack propagation event [24]. **Figure 2-3** lists common signal processing approaches in the time domain. The raw signal (top) was created by a pencil lead break on the surface of an aluminium specimen.

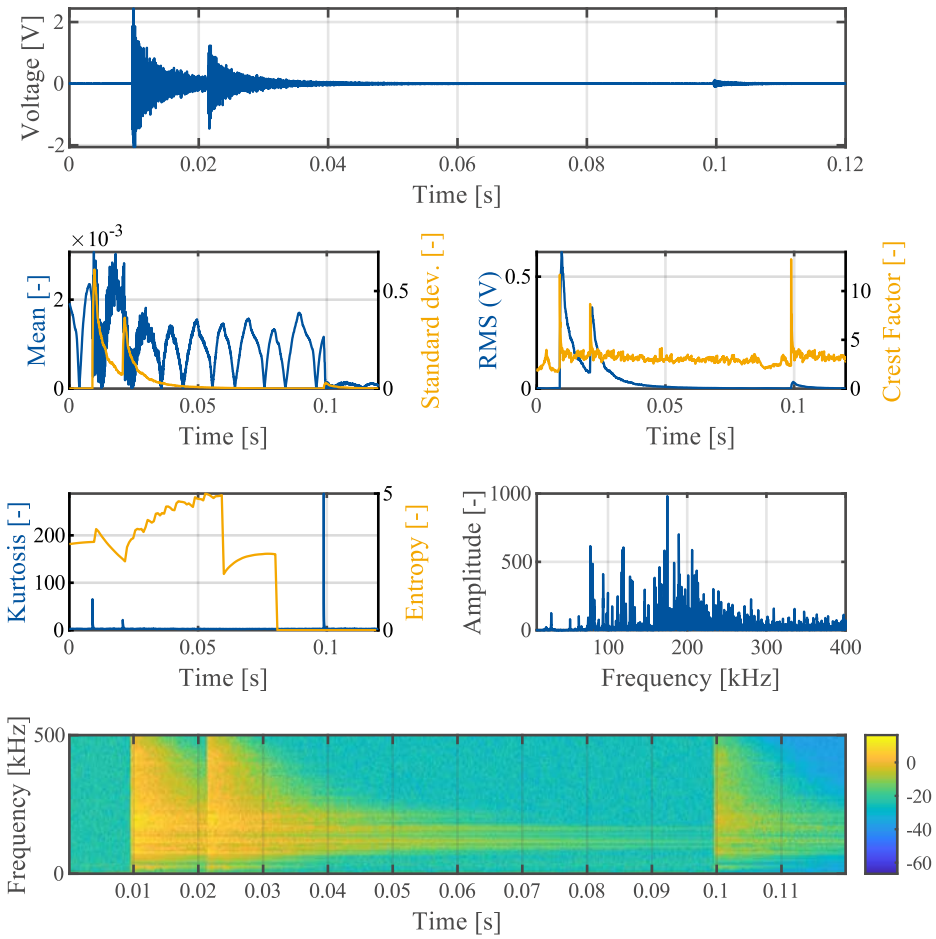


Figure 2-3: Time and frequency domain evaluation of raw signal (top).

Due to the axial symmetry of AE signals, the arithmetic mean often approaches zero, limiting its usefulness for analysis but forming the basis for further calculation. The root mean square (RMS) is the square root of the time-averaged squared measurements, allowing for burst signal evaluation and serving as a measure of signal energy. Variance and standard deviation indicate the dispersion of measurements relative to the mean, providing insights into burst presence: low variance suggests no bursts, while higher variance indicates their occurrence. The crest factor describes the ratio of peak value to RMS, with high values indicating isolated peaks with low energy. Finally, kurtosis measures the deviation from a Gaussian distribution; it equals 3 for normally distributed data and increases in the presence of bursts. In addition to analyzing time-domain parameters, AE signals can also be examined in the frequency domain. Two relevant methods for this study are the Fast Fourier Transform (FFT) and Short-Time Fourier Transform (STFT).

The presented time-domain approaches yield scalar values for a given signal. By applying windowing techniques, it becomes possible to track the temporal evolution of these signal properties, as summarized in Table 1. In the corresponding figures, the response of these

properties to transient signal components, such as bursts, can be clearly observed. Utilizing these processed signals together with appropriately defined threshold values, individual bursts can be detected and characterized. The resulting burst counts and their characteristic parameters can then be employed to evaluate the relationship between AE activity and the propagation of fatigue cracks.

3. METHOD

In the following, the test design, focussing on cyclic hydraulic loading, specimen design, and measurement strategy, is presented.

3.1. Test Design

The experiments are conducted on a pressure impulse test rig from Maximator [26], featuring a hydraulic power of 120 kW and maximum pressures of 1200 bar. This test rig generates a trapezoidal pressure profile with a frequency of 3 Hz, an upper pressure level of 400 bar, and a lower pressure level of 20 bar. The pressure increases at a rate of 20 kbar/s and decreases at 40 kbar/s. HLP46 serves as the pressure medium. **Figure 3-1** illustrates both the specified pressure cycle and the actual pressure during the cycle.

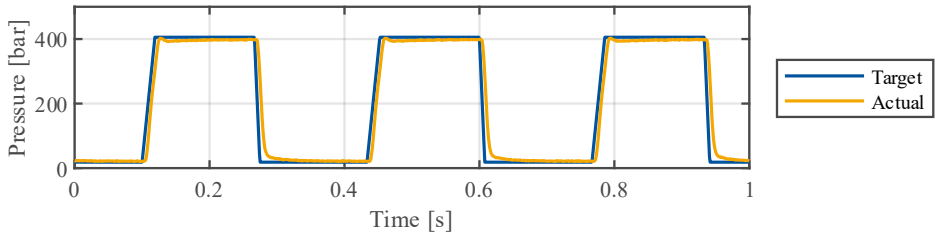


Figure 3-1: Target and actual pressure cycle at 3 Hz.

The geometry of the specimen made from steel S235 JR and the connection to the distribution manifold of the pressure impulse test rig is shown in **Figure 3-2**. The cavity inside the specimen is cyclically subjected to pressure during the experiment. At the base of the cavity, three holes are machined (see detail B) to facilitate fast crack formation due to local stress concentration. These specimens are designed to allow for stable crack growth under uniaxial hydraulic loading of the crack faces. Furthermore, a bleed plug is installed to allow for gas evacuation of the specimen. Bleeding is performed prior to each test run. Sealing the interface between the specimen and adapter is provided by a rubber O-ring.

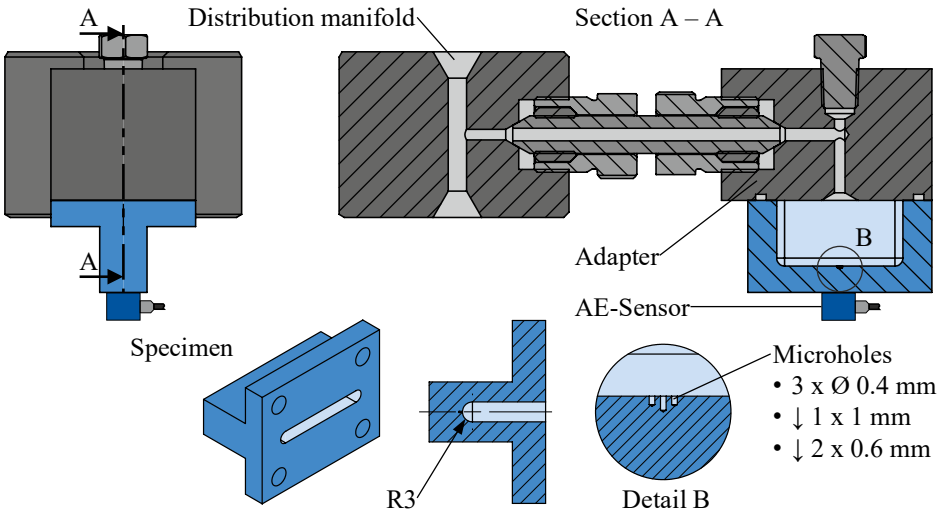


Figure 3-2: Specimen design and connection to pressure impulse test rig. Crack formation and growth is suspected to occur in section A-A of the specimen.

3.2. AE - Measurement setup

The AE sensor is mounted on the end face of the specimen using a magnetic holder, with wax serving as the coupling medium. A second AE sensor is placed at the corresponding position on a specimen without internal holes inside the cavity, hereafter referred to as the dummy specimen. Pressure sensors are installed at the distribution manifold, which are used to monitor and control the pressure profile (see **Figure 3-3).** During the experiment, signals from both AE sensors and a pressure sensor are recorded to evaluate AE activity in relation to the pressure cycle. To ensure temporal synchronization of the results, the signals are digitized using the same A/D converter. Piezotron® AE sensor type 8152C0 from Kistler is utilized. This sensor features an integrated preamplifier and has a sensitivity of 57 dB (ref. 1V) in the frequency range of 50 to 400 kHz. To amplify the AE signals further, Piezotron coupler type 5125C from Kistler is employed, which includes a tenfold amplification along with a 50 kHz high-pass filter and a 1000 kHz low-pass filter. The signals are digitized by the PCIe-6374 analog/digital converter from National Instruments at a maximum sampling rate of 3.5 MHz before being transmitted to the measurement computer. Both data acquisition and analysis are performed using MATLAB R2023b [27].

The high sampling rates required for the reliable detection of AE events result in a significant demand for storage space to capture the measurement data. One AE signal digitized at 3.5 MS/s with a sample size of 16 bits generates approximately 25 GB of measurement data per hour. Depending on the chosen loading frequency, test durations of several hours are expected. These large data volumes pose a challenge for continuous data acquisition, as the memory of the measurement computer will quickly become saturated. A similar issue also affects the evaluation process, as MATLAB requires all necessary variables and data to be present in memory. To address this, the data acquisition concept involves writing data in blocks of fixed sample size from memory to disk. A separate file is created for each block, and these files are numbered in the order of their creation. Specifically, data is acquired every 1000 load cycles over a period of 1 minute. Each acquisition sequence is further

divided into six consecutive blocks of 10 seconds each. This results in a large number of files, each containing a manageable amount of data, and prevents memory overflow during measurements, allowing for gradual evaluation of the measurement data. Figure 3-3 schematically illustrates the implemented data acquisition strategy, showing the periodic measurement intervals and the subdivision into time segments, as well as the sensor and specimen placement in the test chamber.

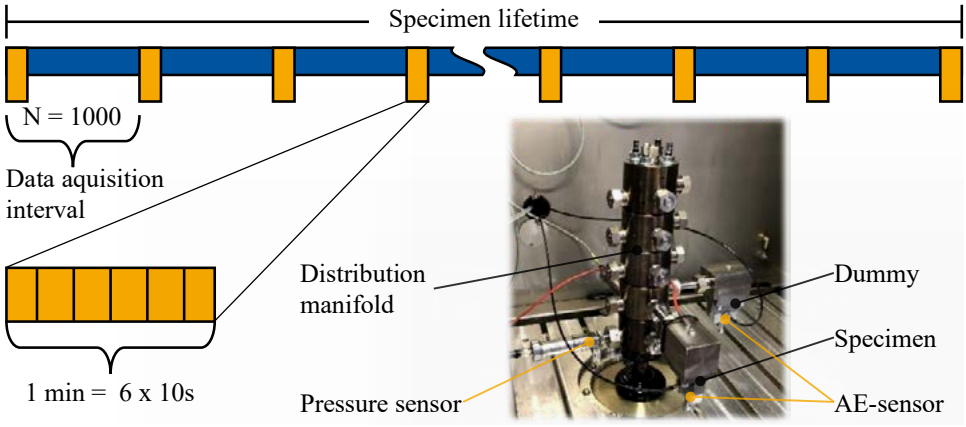


Figure 3-3: Schematic data acquisition strategy showing the periodic measurement intervals and the arrangement of sensors and specimens within the test chamber.

4. RESULTS

Final crack breakthrough of the investigated specimen occurred after 555.427 load cycles. AE and pressure data were acquired at intervals of 1000 cycles for both the test specimen and the dummy specimen, as shown in Figure 3-3. In the following, the raw AE signals of both specimens are presented for selected successive cycle numbers throughout the specimens' lifetime. Subsequently, AE bursts from both specimens are evaluated and characterized according to the methods described above. Finally, the bursts are analyzed in conjunction with the corresponding pressure signals.

Figure 4-1 shows AE signals of both specimens at the beginning of the test (N = 1k) as well as at intermediate cycle numbers in intervals of 100k cycles starting at N = 100k for the duration of 3 s or 10 load cycles, respectively. AE activity can be observed for the dummy specimen from the beginning of the test up to 500k cycles, exhibiting periodic bursts (3 Hz) with amplitudes of approximately ±0.011 V compared to a background noise level of ±0.003 V. This corresponds to a signal-to-noise ratio (SNR) of 3.7, which remains nearly constant throughout the test duration.

In contrast, the AE signal of the investigated specimen shows distinctly different behavior. No notable AE activity is detected during the initial 1000 cycles up to 300k cycles. A slight increase in amplitude is observed at 400k cycles (signal ±0.008 V, noise ±0.003 V, SNR = 2.7). At 500k cycles, the AE bursts become clearly distinguishable from the background noise, reaching amplitudes of around ±0.023 V (SNR = 7.7), indicating the onset of significant crack growth activity.

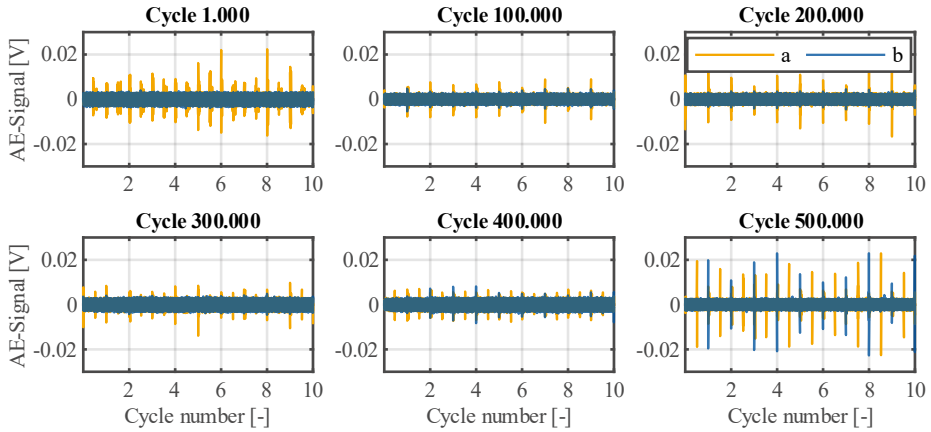


Figure 4-1: Raw AE signals at succeeding cycles for dummy (a) and specimen(b).

The increase in AE activity of the specimen between 400k and 500k cycles necessitates a more detailed examination of the period leading up to the end of the specimen's lifetime. Therefore, **Figure 4-2** focuses on the AE signals for selected cycle numbers closer to the specimen's final failure, shown in comparison with the dummy specimen.

At N = 545k, the specimen exhibits AE burst amplitudes ranging from approximately ±0.015 V to a maximum of ±0.030 V, with a background noise level of ±0.003 V, corresponding to a maximum SNR of 10. In the same interval, the dummy specimen shows a comparable maximum amplitude of ±0.021 V at similar noise levels.

At N = 550k, the specimen signal increases further, with amplitudes between ±0.018 V and ±0.037 V (SNR ≈ 12). The dummy specimen, however, shows only a single pronounced peak of about ±0.040 V, while the remaining signal remains around ±0.011 V with comparable noise (±0.003 V).

Shortly before final failure (N = 555k), the specimen signal exhibits a pronounced increase in AE activity, reaching amplitudes between ±0.050 V and up to ±0.140 V, resulting in a maximum SNR of 47. In contrast to previous intervals, the AE bursts no longer occur strictly at the load-cycle frequency. Instead, alternating high and lower peaks appear, effectively doubling the apparent signal frequency. The lower peaks reach amplitudes of approximately ±0.010 V (SNR ≈ 3). The dummy specimen signal remains largely unchanged during this stage.

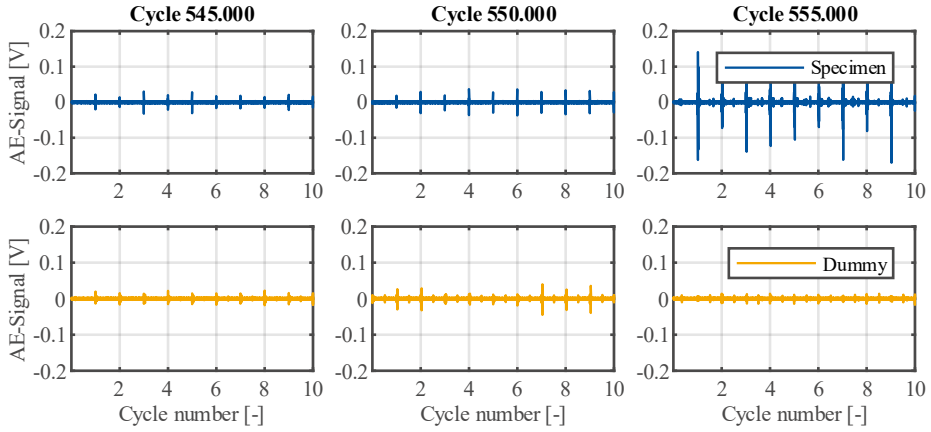


Figure 4-2: Raw AE signals at final succeeding cycles for both specimens.

As those results show distinct cases and a manual evaluation of the bursts of the time signals is not feasible, the method described above is applied to detect and characterize bursts. **Figure 4-3** presents the detected bursts of the specimen over the total test duration, obtained using different detection approaches. The methods include a static threshold applied to the raw signal, burst detection based on calculations in the time and frequency domains, and an entropy-based detection approach. From the bursts detected by those methods, mean values are calculated and compared with those of the dummy specimen, which were determined accordingly. The different detection methods exhibit varying sensitivities. When ranked by their overall sensitivity throughout the test, the methods can be ordered as follows: time-domain analysis (most sensitive), frequency-domain analysis, entropy-based detection, and finally, static thresholding.

The qualitative evolution of the detected burst rate shows no burst activity up to approximately 50k cycles, which aligns with the observations from the raw AE signals. From 50k to 300k, the detection rate remains nearly constant, with an average of around 100 bursts per minute. Between 300k and 400k cycles, a slight increase in detected bursts can be observed, followed by a more pronounced rise between 430k cycles and the end of the test, where all methods yield very similar burst rates of roughly 200 bursts per minute. Shortly before the final fracture, all methods indicate a sharp increase in AE activity. In contrast, the dummy specimen shows almost no detected bursts throughout the entire duration of the test, apart from a few sporadic detections.

It should be noted that data are missing for approximately 5k cycles at 100k, 200k, 300k, 400k, and 500k cycles. During these interruptions, the pressure profile was adjusted by changing the stress ratio from R = 0.05 to 0.6 at a maximum pressure of 400 bar. These steps were necessary to introduce optical beach marks on the fracture surface, allowing the crack growth path to be visually correlated with the corresponding cycle counts. After each change in stress ratio, the test system was restarted. In nearly all cases, a gradual increase in the number of detected bursts was observed immediately following each restart.

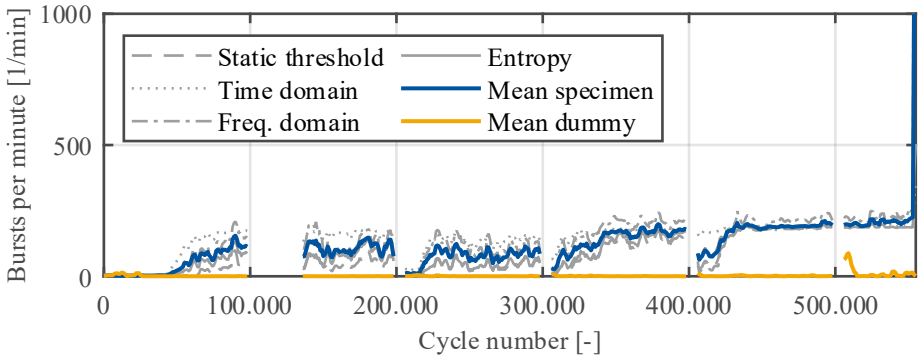


Figure 4-3: Results of burst detection with different approaches for both specimens.

Next, the focus is placed on the characterization of the detected bursts. For each identified burst, its amplitude, duration, and the number of threshold crossings are determined. During each acquisition phase (every 1k cycles for a period of 60 seconds), these parameters are averaged to obtain representative values. The normalized results are presented in **Figure 4-4** for the investigated and the dummy specimen.

The few bursts detected for the dummy specimen do not exhibit any consistent pattern or correlation, and there is almost no variation in their characteristic features from the beginning to the end of the test. In contrast, the burst features of the investigated specimen show a distinct evolution over time. The maximum values for amplitude, duration, and threshold crossings are observed near the end of the test, close to the point of final crack breakthrough. Similar to the trend in the overall burst count, these features remain nearly constant up to approximately 300k cycles, after which a gradual increase is observed, culminating in a pronounced rise shortly before failure.

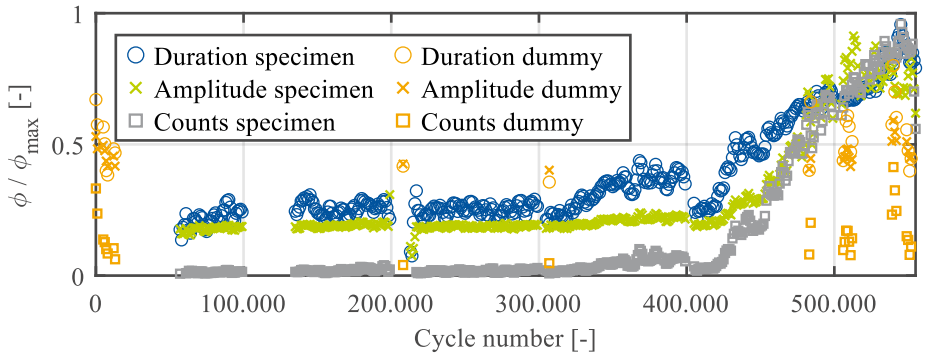


Figure 4-4: Characteristic features of detected bursts for both specimens.

To get a better understanding of the burst activity in relation to the load cycle, **Figure 4-5** illustrates the timing of cumulative burst occurrences within defined cycle intervals over one averaged pressure cycle. During the first 500k cycles, almost all bursts appear right at the onset of pressure increase (approximately 5.5 % of the cycle). As shown for the intervals between 510k to 550k cycles, both the timing of the bursts relative to the pressure rise and their count remain largely unchanged. Only during the last 5k cycles bursts also appear throughout the high-pressure phase.

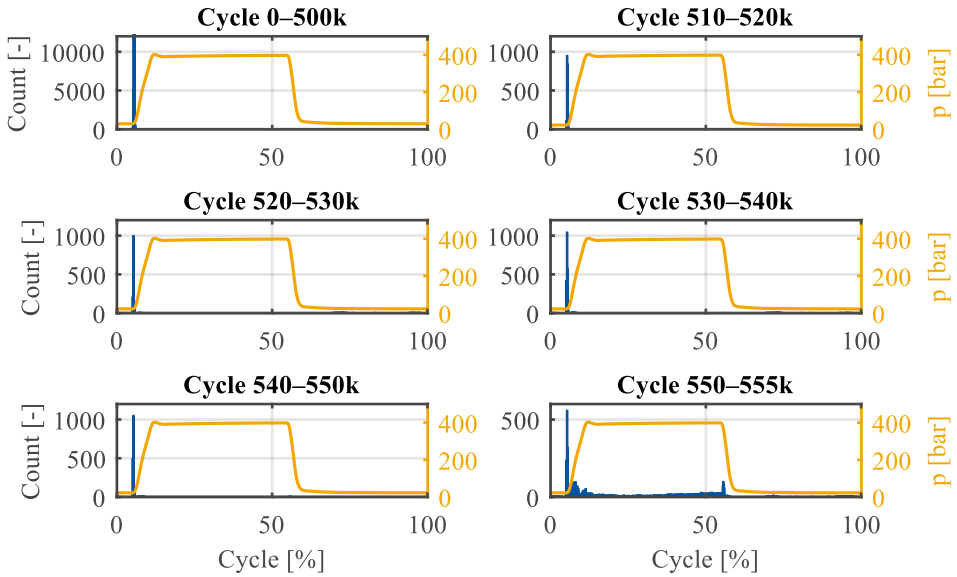


Figure 4-5: Cumulative cycle count for distinct intervals over pressure cycle.

5. DISCUSSION

The observed AE activity correlates strongly with the occurrence of cracking in the investigated specimen. Since the dummy specimen shows no notable change in signal characteristics, the AE sources in the specimen are likely associated with crack propagation. During the crack initiation phase, almost no AE activity is detected. This is followed by a long period of stable crack growth, characterized by nearly constant burst counts and signal features. Toward the end of the test, a phase of unstable crack growth is observed, during which both the number of detected bursts and their characteristic features (amplitude and duration) increase significantly.

However, it is unlikely that all AE activity originates solely from crack propagation. Other potential AE sources include plastic deformation, erosion, corrosion, impacts, and leakage. In addition, the pressure transducer and other hydraulic components in the pressure impulse test rig can generate spurious signals that interfere with AE detection and are audible to the human ear.

Interestingly, the highest AE activity is observed at the beginning of the pressure increase, when the maximum stress intensity at the crack tip has not yet been reached. This suggests that these detections are not only related to crack growth. Instead, they are likely associated with the sealing system of the specimen. The test specimen is sealed using an O-ring, which, due to design, exhibits slight play within its groove. Post-test inspection revealed some abrasion of the O-ring, indicating movement and potential gap extrusion, which could generate AE signals.

To mitigate this effect, a revised sealing concept is proposed, for example, using copper or flat gaskets to achieve a more stable and rigid seal. Additionally, flange deformation was observed during testing, which likely reduced the preload of the sealing assembly and further

contributed to AE noise. Future test setups will employ a stiffer specimen geometry and more uniform clamping with six instead of four bolts. These measures aim to minimize external interferences and enhance the reliability of AE-based crack detection.

To unambiguously attribute AE activity to the source of crack propagation, the implementation of a source localization strategy is suitable. By mounting two AE sensors on the same specimen, the origin of AE events can be determined from the arrival time difference of the signals in combination with the known acoustic wave velocity of the material [24].

To further improve signal quality, a noise-canceling approach can be implemented by using one AE sensor on the specimen and a second on the dummy specimen. The signal from the dummy specimen serves as a reference for background and system noise and can be subtracted from the specimen signal, for example, by applying methods such as continuous wavelet transform (CWT) or cross-correlation. This approach allows for the suppression of noise originating from the test setup, hydraulic system, or other external sources.

For signal evaluation, a manual analysis approach quickly becomes impractical due to the large volume of acquired data. Therefore, machine learning techniques are particularly suitable for the classification and analysis of AE signals. With the existing feature extraction (amplitude, duration, counts, entropy, etc.), the fundamental basis for applying such algorithms is already established. Future work may thus focus on developing automated workflows that use these features for clustering, pattern recognition, or predictive modeling of crack propagation behavior.

6. SUMMARY

This work demonstrates the application of AE techniques for the detection of fatigue cracks in hydraulic components during cyclic pressure impulse testing to investigate the influence of viscous fluids on crack propagation. The study introduces the experimental setup for both pressure impulse and AE measurements and addresses common challenges associated with AE measurements, such as the influence of background noise and the handling of large data volumes.

A dedicated data acquisition and evaluation framework was developed to capture and subsequently analyze AE signals during pressure impulse testing. The result shows crack propagation until final fracture. By comparing AE signals from the test specimen with those from a dummy specimen, which did not form cracks, AE activity related to crack growth could be identified.

Nevertheless, some uncertainty remains regarding the unambiguous attribution of AE activity to the crack propagation source, as AE signals are strongly affected by operational conditions, environmental influences, and especially sealing system behavior. Future work will focus on minimizing these uncertainties through optimized specimen and sealing design, as well as by advancing data analysis techniques using machine learning for automated signal classification and source localization.

ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (DFG) in the scope of the Project „Einfluss der Belastungsgradienten auf das Ermüdungsverhalten hydraulischer Komponenten“ (449676223). The authors would like to thank DFG for its support.

7. REFERENCES

[1] Paris, P., Gomez, M., Anderson, W. ‘A rational analytic theory of fatigue’, The Trend in Engineering, pp. 9–14, 1961.

[2] Tzou, J.-L., Hsueh, C. H. et al. ‘Fatigue crack propagation in oil environments— II. A model for crack closure induced by viscous fluids’, Acta Metallurgica, Vol. 33, Nr. 1, pp. 117–127, 1985. DOI: 10.1016/0001-6160(85)90225-1.

[3] Michiels, L. ‘A simulation-based approach to the fluid-structure interaction inside fatigue cracks in hydraulic components’, Dissertation, Karlsruhe, 2024. DOI: 10.5445/IR/1000171515.

[4] Michiels, L., Geimer, M. ‘On the frequency dependency of fatigue damage caused by viscous fluid-structure interaction in hydraulic components’, Proceedings of the 18th Scandinavian International Conference on Fluid Power, 2023. DOI: 10.5445/IR/1000160350.

[5] Suresh, S. ‘Fatigue of materials’, 2. ed., reprint, Cambridge Univ. Pr, Cambridge, 2004. DOI: 10.1017/CBO9780511806575.

[6] Sangid, M. D. ‘The physics of fatigue crack initiation’, International Journal of Fatigue, Vol. 57, pp. 58–72, 2013. DOI: 10.1016/j.ijfatigue.2012.10.009.

[7] Sangid, M. D. ‘The physics of fatigue crack propagation’, International Journal of Fatigue, Vol. 197, pp. 108928, 2025. DOI: 10.1016/j.ijfatigue.2025.108928.

[8] Paris, P., Erdogan, F. ‘A Critical Analysis of Crack Propagation Laws’, Journal of Basic Engineering, Vol. 85, Nr. 4, pp. 528–533, 1963. DOI: 10.1115/1.3656900.

[9] Forman, R. G., Mettu, S. R. ‘Behavior of surface and corner cracks subjected to tensile and bending loads in Ti-6Al-4V alloy’, National Aeronautics and Space Administration, Houston, TX (USA). Lyndon B. Johnson Space Center, https://www.osti.gov/biblio/5555321, 1990.

[10] Irwin, G. R. ‘Analysis of Stresses and Strains Near the End of a Crack Traversing a Plate’, Journal of Appl. Mech., Vol. 24, Nr. 3, pp. 361–364, 1957. DOI: 10.1115/1.4011547.

[11] ASTM International: ‘ASTM E647-24 - Standard Test Method for Measurement of Fatigue Crack Growth Rates’, ASTM International, West Conshohocken, PA, 2024.

[12] Davis, F. H., Ellison, E. G., Plumbridge, W. J. ‘Effects of Hydrostatic Pressure on the Rate of Fatigue Crack Growth’, Fatigue & fracture of engineering materials & structures, Vol. 12, Nr. 6, pp. 511–525, 1989. DOI: 10.1111/j.1460-2695.1989.tb00560.x.

[13] Michiels, L., Geimer, M. ‘Influence of High Pressure Drop Rates on Fatigue Crack Growth’, Chemical Engineering & Technology, Vol. 46, Nr. 1, pp. 45–52, 2023. DOI: 10.1002/ceat.202200385.

[14] Michiels, L., Geimer, M. ‘Non-Destructive Pressure Pulse Examination for Fatigue Crack Detection in Hydraulic Components’, Proceedings of the ASME/BATH 2023 Symposium on Fluid Power and Motion Control, Sarasota, 16th-18th October 2023, 2023. DOI: 10.1115/FPMC2023-110728.

[15] Pippan, R., Hohenwarter, A. ‘Fatigue crack closure: a review of the physical phenomena’, Fatigue & fracture of engineering materials & structures, Vol. 40, Nr. 4, pp. 471–495, 2017. DOI: 10.1111/ffe.12578.

[16] Ritchie, R. O., Garrett, G. G., Knott, J. P. ‘Crack-growth monitoring: Optimisation of the electrical potential technique using an analogue method’, International Journal of Fracture Mechanics, Vol. 7, Nr. 4, pp. 462, 1971. DOI: 10.1007/BF00189118.

[17] Papazian, J., Nardiello, J., Silberstein, R., Welsh, G., Grundy, D., Craven, C., Evans, L., Goldfine, N., Michaels, J., & Michaels, T. ‘Sensors for monitoring early stage fatigue cracking’, International Journal of Fatigue, Vol. 29, 9-11, pp. 1668–1680, 2007. DOI: 10.1016/j.ijfatigue.2007.01.023.

[18] Deschanel, S., Ben Rhouma, W., Weiss, J. ‘Acoustic emission multiplets as early warnings of fatigue failure in metallic materials’, Scientific reports, Vol. 7, Nr. 1, pp. 13680, 2017. DOI: 10.1038/s41598-017-13226-1.

[19] Chai, M., Hou, X. et al. ‘Identification and prediction of fatigue crack growth under different stress ratios using acoustic emission data’, International Journal of Fatigue, Vol. 160, pp. 106860, 2022. DOI: 10.1016/j.ijfatigue.2022.106860.

[20] La Selle, T. de, Réthoré, J. et al. ‘Signatures of fatigue crack growth from acoustic emission repeaters’, Engineering Fracture Mechanics, Vol. 309, pp. 110388, 2024. DOI: 10.1016/j.engfracmech.2024.110388.

[21] Cui, J., Qu, X. et al. ‘Multi-parameter acoustic emission analysis for fatigue crack evaluation in structural health monitoring’, Measurement, Vol. 256, pp. 118529, 2025. DOI: 10.1016/j.measurement.2025.118529.

[22] Zilberstein, V., Schlicker, D. et al. ‘MWM eddy current sensors for monitoring of crack initiation and growth during fatigue tests and in service’, International Journal of Fatigue, Vol. 23, pp. 477–485, 2001. DOI: 10.1016/S0142-1123(01)00154-2.

[23] Gehri, N., Mata-Falcón, J., Kaufmann, W. ‘Automated crack detection and measurement based on digital image correlation’, Construction and Building Materials, Vol. 256, pp. 119383, 2020. DOI: 10.1016/j.conbuildmat.2020.119383.

[24] Grosse, C. U., Ohtsu, M. et al. ‘Acoustic Emission Testing - Basics for Research – Appl. in Engineering’, Springer eBook Collection, 2nd ed. 2022, Springer Int. Publishing; Imprint Springer, Cham, 2022. DOI: 10.1007/978-3-030-67936-1.

[25] The Japanese Society for Non-Destructive: ‘Practical Acoustic Emission Testing’, Springer, Tokyo, 2016. DOI: 10.1007/978-4-431-55072-3.

[26] Maximator GmbH: ‘Maximator GmbH’, https://www.maximator.com, 2025.

[27] The MathWorks Inc. ‘MATLAB Version 23.2 R2023b’, Massachusetts, USA, 2023.

Biographies



Lukas Boden obtained his master's degree in mechanical engineering at RWTH Aachen University in 2023. Since then, he is working as a research assistant at the Institute for Fluid Power Drives and Systems (ifas), focussing on identifying the influence of viscous fluid on fatigue crack propagation in fluid power components.



Fabian Blatter studied mechanical engineering at the Karlsruhe Institute of Technology. He is currently working as a Research Associate at the Department of Mobile Machines at the Karlsruhe Institute of Technology. His research areas include simulation of fluid-structure-interaction and hydraulic applications.



Prof. Dr.-Ing. Marcus Geimer received the Diploma degree in mechanical engineering from the RWTH Aachen University, Germany, in 1990, and the Ph.D. degree from the Institute of Hydraulics and Pneumatics, today named the Institute for Fluid Power Drives and Systems, RWTH Aachen University. Since 2005, he has been a Full Professor and the Director of the Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology

(KIT), Germany. He started his industrial career in 1995 in the field of construction at the company Krupp Berco Bautechnik GmbH, Germany, where he was the Leader of the research group for hydraulic breakers. In 2000, he was with Bucher Hydraulics GmbH, Germany, where he led the construction and customer development for mobile hydraulics.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Investigation of Fatigue Crack Growth under high Pressure Gradients with a Fluid-Structure-Interaction Model

Blatter, Fabian[1*]; Boden, Lukas[2]; Schmitz, Katharina[2]; Geimer, Marcus[1]

1. *Institute of Mobile Machines, Karlsruhe Institute of Technology, Germany*
2. *Institute for Fluid Power Drives and Systems, RWTH Aachen University, Germany*

* fabian.blatter@kit.edu

Abstract.

Hydraulic components in mobile machinery are increasingly exposed to dynamic high-pressure gradients, which can accelerate fatigue crack growth and affect component reliability. This study investigates fluid-induced fatigue crack propagation by implementing a newly developed test specimen within an existing fluid–structure interaction (FSI) simulation framework. The specimen was designed to enable reproducible crack initiation, measurable crack propagation, and efficient manufacturability for experimental validation. FSI simulations were conducted across multiple crack lengths and pressure loading configurations to examine the influence of pressure drop and build-up rates. The results indicate that fluid effects, such as hydrostatic pressure and fluid-induced crack closure, modify the effective stress intensity factor (effective SIF) and the corresponding crack growth rates. While the observed trends are generally consistent with Paris' law, the influence of pressure gradients appears limited within the investigated range. Overall, the study highlights existing gaps in understanding fluid-induced crack propagation in hydraulic components and underscores the need for further experimental and numerical investigations, particularly given contradictory findings in previous work.

Keywords. Fluid-structure interaction, fatigue crack propagation, fluid-induced crack closure, pressure gradients

1. INTRODUCTION

Hydraulic systems are frequently exposed to highly dynamic load changes during operation that result in significant and constant changes of pressure. High pressure gradients can occur in different components of a hydraulic system such as axial piston pumps or throttle and orifice geometries of valves. Pressure gradients of up to 1.000 kbar/s have been reported for some of these components [1]. Previous studies have shown that drop rates of pressure in particular have an influence on the propagation rate of fatigue cracks [2, 3]. Under such

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

conditions, fatigue strength becomes a more decisive criterion. Although design strategies exist that successfully increase the service life of critical regions, they cannot fully eliminate fatigue-related failures. As a result, material fatigue remains a fundamental challenge.

The challenge intensifies in the context of mobile hydraulic applications and increasing power densities, where lightweight components offer clear advantages. However, the reduced and targeted material use required for lighter constructions also increases the likelihood of crack failure. Therefore, a strong need exists to better understand and predict crack growth mechanisms in hydraulic components.

The present work investigates fluid-induced fatigue crack growth using a FSI approach. A new test specimen is developed and implemented within an existing FSI simulation framework originally proposed by Michiels [4]. The paper focuses on the design of this test specimen and on the presentation of simulation results obtained with the updated test specimen. The simulation framework has so far been tested with only one specimen and has shown limitations. These limitations motivate further investigations, particularly with regard to the influence of pressure gradients on fatigue crack growth. The simulation results presented in this paper serve as reference data for future comparisons, including pressure distributions within the crack, calculated SIFs, and resulting crack propagation rates. The updated specimen will be used for the development and validation of a new FSI approach that addresses the limitations of the current framework. The overarching goal of this research is to enable a reliable estimation of crack growth rates of fluid-induced fatigue cracks, which could be used to define inspection intervals for hydraulic components or to replace critical parts prior to failure.

2. Effects on Crack Propagation of Fluid-filled Cracks

This research focuses on fatigue crack propagation in hydraulic components under cyclic loading. Specifically, it investigates fluid-filled cracks, where crack growth is driven by pressurized fluid both inside and outside the crack. In fracture mechanics, several parameters are essential to describe crack behaviour under cyclic loading. The most fundamental loading case mode I (opening mode) is shown in Figure 2.1, where the load acts perpendicular to the x-z-plane and drives the crack faces apart [5].

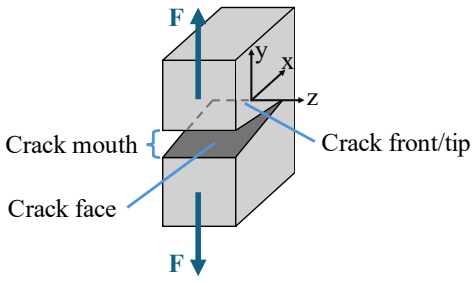


Figure 2.1. Description of the crack mode I (opening mode) and components of the crack

The governing parameter for characterizing the stress field at the crack tip is the stress intensity factor (SIF) K. It is described by the far-field applied stress σ, the current crack length a in x-direction and the geometric correction factor f as [5]:

$$K = \sigma\sqrt{\pi a} \cdot f \qquad (2.1)$$

The R-ratio defines the ratio of minimum to maximum stress σ_{min} and σ_{max} in a load cycle [6]:

$$R = \frac{\sigma_{min}}{\sigma_{max}} \qquad (2.2)$$

The SIF range ΔK is the difference between the minimum and maximum SIF [6]:

$$\Delta K = K_{max} - K_{min} \qquad (2.3)$$

Paris' law from 1963 shows that a decreasing SIF range ΔK reduces the crack propagation. It expresses the relationship between the crack propagation rate da/dN (N is the number of applied cycles) and the SIF range ΔK together with the constants C_p and m_p that depend on the material, temperature, environmental medium and load as [7]:

$$\frac{da}{dN} = C_p \cdot \Delta K^{m_p} \qquad (2.4)$$

Typically, no load is applied to the crack faces [5]. If hydrostatic fluid pressure acts on the crack faces it increases the crack propagation rate [3]. The penetration of fluid also influences the contact behaviour of the crack surfaces. When a pressure drop occurs rapidly in a fluid-filled crack, the fluid cannot completely escape and remains trapped inside the crack. The fluid remaining in the crack indirectly connects the crack faces and prevents a complete return to the original state. When the load is increased again, the crack remains closed to a certain load level. As a result, the crack is not open during the entire loading cycle. Crack closure was first discovered by Elber [8]. Figure 2.2 shows the curve of the SIF over time with and without crack closure effect. In the following, the SIFs are denoted with the subscript I, since the loading discussed in this paper is predominantly mode I.

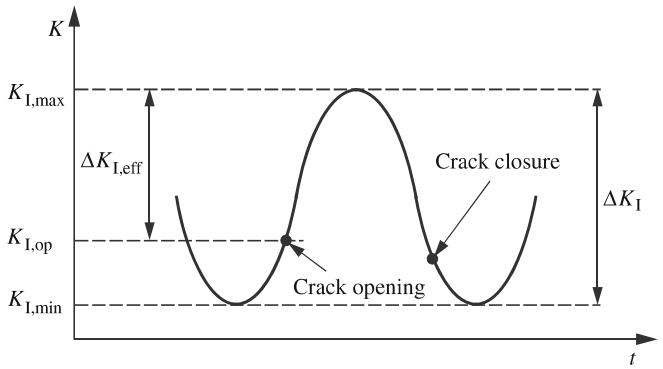


Figure 2.2. Comparison of the SIF with and without crack closure effect [9]

With crack closure effects the effective SIF range $\Delta K_{I,eff}$ is reduced. According to Paris' law it can extend a component's lifetime compared to a crack without entrapped fluid [9, 10]. The classical computation of the SIF ΔK_I doesn't cover shielding effects like crack closure

[11]. Therefore, the effective SIF range $\Delta K_{I,eff}$ covering crack closure is calculated with the maximum applied SIF $K_{I,max}$ and the crack opening SIF $K_{I,op}$ [9, 10]:

$$\Delta K_{I,eff} = K_{I,max} - K_{I,op} \qquad (2.5)$$

Other known crack closure phenomena are plasticity induced, roughness induced, and oxide induced crack closure. Plasticity-induced closure originates from residual deformation in the crack wake, roughness-induced closure from mismatch of the fracture surfaces, and oxide-induced closure from the buildup of corrosion products [5].

Investigations of fluid-induced crack closure are primarily found in related research fields, for example on railway wheel–rail systems and submerged fatigue specimens. For submerged specimens, it has been demonstrated that fluid-filled fatigue cracks exhibit a reduced crack propagation rate compared to air-filled cracks. This is due to fluid-induced crack closure and the reduced effective SIF ΔK_{eff}. [12]

Similar mechanisms have been identified in railway tracks, where fatigue cracks propagate mainly due to entrapped fluid. Under rolling contact, surface cracks open due to traction forces, allowing fluid to penetrate the crack. As the wheel passes and forces the crack mouth to close, the trapped fluid becomes pressurized, generating tensile and shear stresses on the crack faces that accelerate crack growth in the direction of wheel motion. [13, 14, 15] In the absence of fluid, such crack growth is largely suppressed [15].

Numerical and experimental investigations of pressurized cracks have shown partially contradictory results. In [4], simulations of fully fluid-filled cracks indicated that neither the pressure build-up rate nor the pressure drop rate significantly affect the stress amplitude at the crack tip and therefore the crack propagation rate. In contrast, corresponding experiments showed a significant influence of the pressure drop rate on the crack propagation rate.

Despite these studies, the mechanisms behind fluid penetration into fatigue cracks in hydraulic components are not fully understood. There is a lack of systematic investigation into how externally pressurized fluids penetrate and interact with fatigue cracks under cyclic loading. Furthermore, how the interaction of the fluid with the mechanical structure affects the effective SIF of fatigue cracks in hydraulic components has not been fully quantified. Addressing these questions constitutes a critical research gap, especially for the reliable lifetime prediction of hydraulically loaded components under cyclic loading.

3. INTRODUCTION TO FSI APPROACH OF MICHIELS

The interaction between solids and fluids is a multi-physics problem, which is addressed in this study through a two-way fluid–structure coupling. The FSI framework is proposed by Michiels [4] and is summarized in this chapter with a focus on its most relevant characteristics. For a detailed description, the reader is referred to the original publication. The FSI approach avoids fully fluid and structural simulations (e.g. through the finite element method) and instead employs reduced-order models for both the structural and fluid domain. Input parameters for the FSI model include the applied load, weight functions, and flow factors. Weight functions correlate the load to the structural deformation, while flow factors account for roughness effects in narrow crack flows. Both were determined prior in pre-processing steps. The simulation is conducted for a fixed crack length. The resulting outputs of the coupled FSI are the pressure field and fluid flow inside the crack and its crack

opening displacement (COD). The crack is discretized into N stripes with a constant length Δx, shown in Figure 3.1. This model is referred to as strip model and uses the same discretization for the fluid inside the crack.

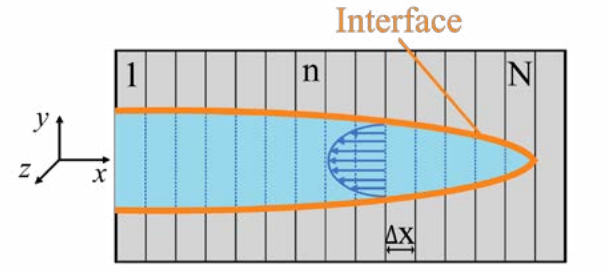


Figure 3.1. Discretization of the crack and the fluid into N stripes. The structural domain is shown in grey, the fluid domain in blue and the interface between both in orange.

Fluid pressure acting on the crack faces is transferred to the structural domain, while the resulting structural displacements are transferred back into the fluid domain, shown in the following Figure 3.2.

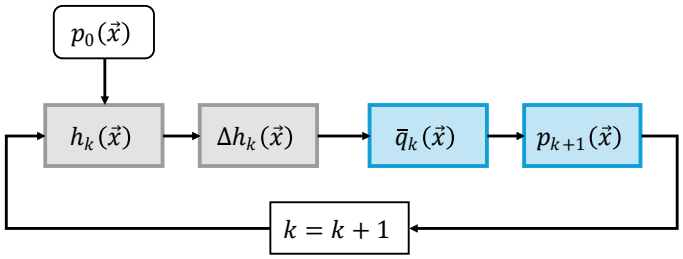


Figure 3.2. Iteration through the two-way coupled FSI with the structural domain in grey and the fluid domain in blue during a timestep k.

At the first timestep k = 0, the initial pressure p_0 is given as 0 bar and transferred to the mechanical domain. The crack height h is calculated with the given arbitrary pressure p and the weight functions, in the following given in matrix notation as W, while ξ describes the centre of a stripe.

$$h(\vec{x} = \vec{\xi}) = \mathbf{W}p(\vec{x}) \qquad (3.1)$$

The used weight functions depend only on the geometry including the crack and are independent from the applied load. They are calculated for the examined specimen with discrete crack lengths via finite element method in a pre-processing step. The detailed calculation procedure of the weight functions can also be found in [4]. Assuming linear elastic material behaviour, the total deformation is obtained by superposition of the deformations induced by individual load contributions. The pressure force acting from the fluid on the crack surface is decomposed into N+1 subparts. One associated with the displacement caused by external forces, like a pressure chamber, and N corresponding to the individual crack sections called stripes.

The resulting structural displacement Δh calculated from equation 3.1 is transferred to the fluid domain. The fluid domain describes the fluid flow inside the crack as a planar flow, neglecting the crack width, and is modelled using an adapted thin-film flow formulation, which is appropriate due to the small crack opening relative to its length. The governing equations are derived from the Navier–Stokes equations under laminar flow assumptions and discretized along the crack using the same spatial resolution as the structural strip model. In addition, the fluid flow is assumed to be fully developed, which is consistent with the neglect of fluid inertia effects. To account for the influence of surface roughness on flow resistance, flow factors are introduced. They represent the reduction in flow capacity caused by rough crack surfaces. The use of flow factors $\theta \in [0,1]$ enables the model to capture roughness-induced flow restriction without explicitly resolving surface topography. They compare the average pressure flow through rough gaps relative to that through smooth gaps. The average Reynolds equation using flow factors was introduced by Patir and Cheng in 1978 [16, 17]. The averaged volume flow per unit length $\bar{q}$ along the crack in x-direction is given by

$$\bar{q} = -\theta \frac{h^3}{12\eta} \frac{\partial p}{\partial x} \qquad (3.2)$$

where h denotes the crack opening and η the dynamic viscosity. For a perfectly smooth surface θ equals 1. The pressure p of the following time step k = k + 1 is calculated by the following equation with the density ρ, reference density ρ_0 and the bulk modulus K:

$$\frac{\partial p}{\partial t} = -\left(\frac{\partial(\rho \bar{q})}{\partial x} + \rho \frac{\partial h}{\partial t}\right) \frac{K}{\rho_0 h} \qquad (3.3)$$

4. TEST SPECIMEN GEOMETRY

A test specimen geometry was proposed by Michiels in [4], which is shown in Figure 4.1. It contains an internal pressure chamber with a notch. Under pressurization, mechanical deformation and stress are induced. Due to the increased stress concentration at the notch root, a crack initiates from this location and subsequently propagates. Since the crack originates from the interior pressure chamber, it remains in contact with the pressurized fluid, that can flow into and out of the crack. The fracture process is predominantly Mode I.

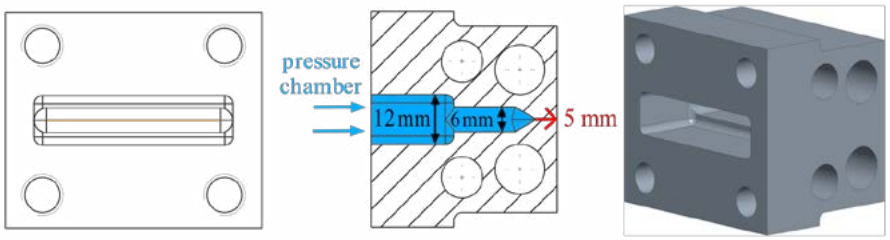


Figure 4.1. Proposed test specimen in [4]. Test specimen contains four threaded holes, a pressure chamber (blue), a notch (orange) and four relief holes. The ligament of the uncracked specimen with a length of 5 mm is shown in red.

According to [4], the experimental configuration enables a stable crack growth regime without causing catastrophic rupture of the structure, allowing the influence of fluid-induced crack closure and fluid flow within the crack to be investigated. Furthermore, it provides a basis for assessing the transferability between planar two-dimensional simulations and three-dimensional test specimens. However, several limitations are associated with this specimen design for experiments on a hydraulic test rig:

- The specimen geometry is complex (threads, geometry of pressure chamber, relief holes) and therefore requires considerable time to manufacture. For future investigations a broader validation campaign is planned with a larger number of specimens to investigate several load configurations (see chapter 4).
- Manufacturing constraints, as a sharp notch cannot be produced with conventional machining like milling. A residual radius remains and varies due to tool wear, resulting in geometric inconsistencies.
- Crack initiation can occur at any location along the notch shown in Figure 4.1, which reduces reproducibility. Also, several cracks can be initiated along the notch during testing.
- The remaining ligament is short, allowing a maximum crack extension of only 5 mm. A longer ligament provides more opportunities to monitor relevant crack growth effects and allows observations over a larger crack length.
- The internal volume is comparatively large. Since the objective is not to measure compression-induced volumetric flow, a minimal dead volume is required.
- Crack breakthrough was observed only after averaged 812k load cycles, representing an excessively high cycle count.

Therefore, a new test specimen is proposed. The following additional requirements are defined for its geometry:

- Less complex test specimen geometry to ensure efficient and reproducible production as well as economical manufacturability.
- Sufficient remaining ligament to allow measurable crack growth and data acquisition over a range of crack lengths.
- Reproducible crack initiation at a defined location within the specimen with only one crack developing per test specimen.
- Rapid crack initiation to minimize the number of load cycles required in total.
- Consistent crack length through minimal geometric variation between specimens in the region from the notch to the opposite side.
- Minimal plastic deformation, as excessive plastic strain may alter the material properties within the plastic zone and distort the fracture behaviour.
- Reduced internal volume to minimize the influence of compression-induced flow.
- Surfaces parallel and perpendicular to crack growth for sensor installation (strain gauge, temperature sensor, acoustic emission sensor).

According to those defined by Michiels and the additional requirements, a new specimen was developed (see Figure 4.2). A parameter study was conducted for the final geometric dimensions, varying the geometric properties of the specimen to achieve a favourable stress distribution for an assumed pressure of 400 bar.

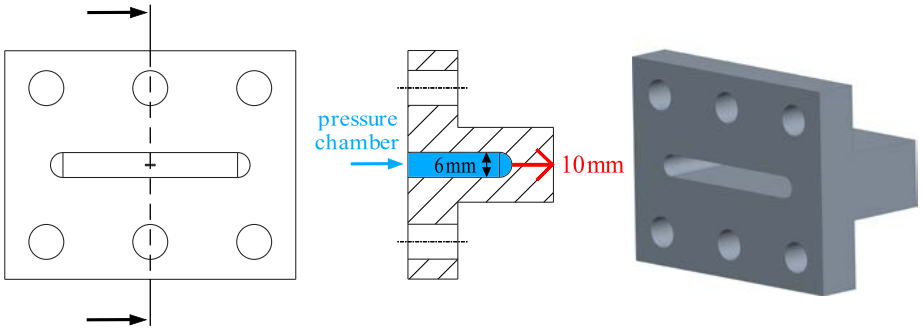


Figure 4.2. Updated test specimen according to the defined requirements. It contains a pressure chamber (blue) and six through holes for connecting the specimen to an adapter of the test stand. The ligament of the uncracked specimen with a length of 10 mm is shown in red.

In the revised specimen design, through holes were implemented instead of threaded holes, eliminating one manufacturing step. The threaded holes were relocated to the adapter, and due to the more slender geometry no relief holes were required. The pressure chamber was designed with a simplified geometry and its dead volume was reduced from 16221.3 mm^3 to 6816.35 mm^3, corresponding to a reduction of 58 %. Furthermore, the ligament thickness was increased from 5 to 10 mm in order to extend the achievable crack length. To ensure consistency across all manufactured test bodies, a more precise tolerance was introduced, resulting in a ligament length with a total deviation of 0.25 mm.

A considerable amount of lifetime is spent in crack initiation [9, 18, 19]. To shorten this phase and define a starting point for crack growth as defined above, micro-notches are introduced in the form of five micro-drilled holes. Cracks usually propagate from discontinuities in geometry like holes, notches or shoulders, where stress concentrations occur. The use of micro-drilled holes for this purpose is a common practice in fracture mechanics testing, as shown in [20, 21]. In the section view of the test specimen in Figure 4.3 the micro-notches are implemented at the center of the inner cavity. This study also focuses on crack propagation rather than initiation, as effects such as crack closure mainly occur during the propagation phase.

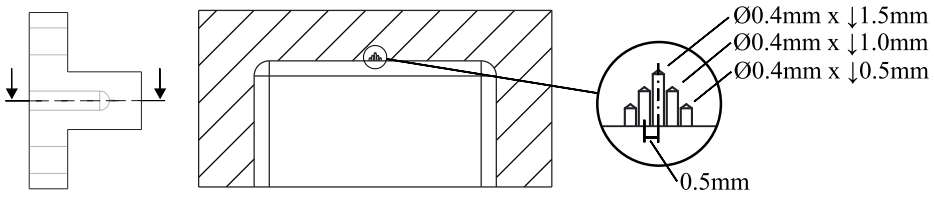


Figure 4.3. Sectional view of the test specimen showing the five micro-drilled holes inserted as micro-notches for a defined crack initiation.

To evaluate the stability of crack growth, a fatigue crack simulation was carried out using the SMART simulation tool within the FEM software Ansys [22]. The initial crack geometry was derived by the outer contour of the five micro-notches. For the crack propagation

simulation Paris' law and the material parameters $C_p = 1.25 \cdot 10^{-9}$ and m_p = 3.38 were used [9]. A maximum pressure of 400 bar and an R-ratio of 0.1 were applied. Also, no FSI inside the crack was considered for this step. Figure 4.4 shows the simulated crack front geometries for different crack lengths.

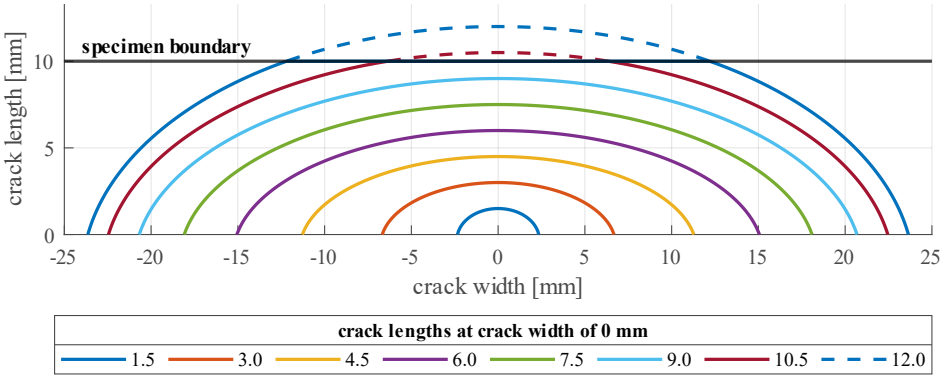


Figure 4.4. Approximated crack front geometries for several crack lengths

The calculated ΔK_I values for crack lengths between 1.5 and 10 mm are presented in Figure 4.5, starting at 1.5 mm due to the micro-notches. The values of ΔK_I remain above the threshold for fatigue crack growth of S235 for R=0.1, $\Delta K_{I,th}$ = 10.2 MPa√m [9]. For structural steels in general, ΔK_{IC}-values between 50 and 150 MPa√m are common [23]. Consequently, the results indicate that stable fatigue crack growth is expected under the investigated loading conditions. The final steep increase in ΔK_I is caused by the reduction in remaining ligament length.

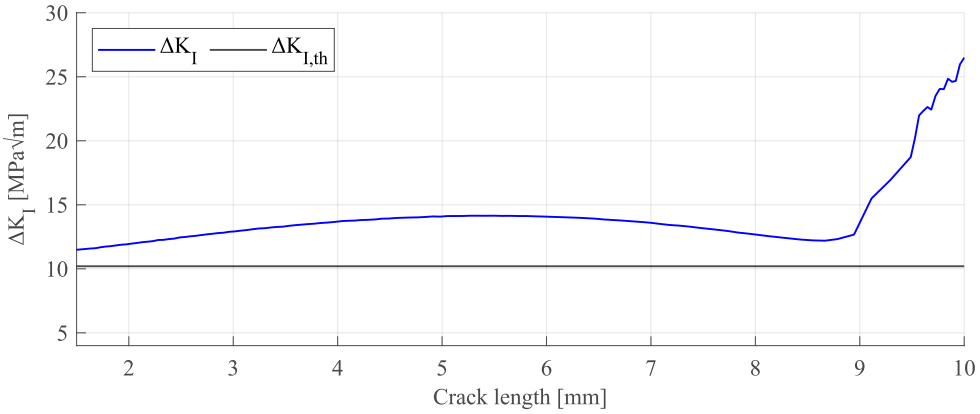


Figure 4.5. Simulated ΔK_I without considering FSI inside the crack shown for crack lengths of 1.5 to 10 mm.

5. RESULTS OF THE FSI APPROACH

The load profile under investigation (Figure 5.1) consists of the low-pressure level p_{min}, the pressure build-up dp_{bu}/dt , the high-pressure level p_{max} and the pressure drop dp_{dr}/dt.

The partitioning of the load cycle ensures that the individual load components are clearly separated and can therefore be investigated in a targeted manner. Furthermore, the interactions of pressure levels, holding times and pressure gradients can be examined. The low-pressure duration t_{LP} includes the period at low pressure and the pressure build-up, while the high-pressure duration t_{HP} includes the period at high pressure and the pressure drop. From these durations, a frequency f can be derived to further describe the cyclic load:

$$f = \frac{1}{t_{LP} + t_{HP}} \qquad (5.1)$$

Additionally, a high-to-low-pressure-duration ratio R_t is introduced:

$$R_t = \frac{t_{HP}}{t_{LP}} \qquad (5.2)$$

The load profile was also used in Michiels [4] and is inspired by pressure impulse tests, e.g. described in EN ISO 6803:2017.

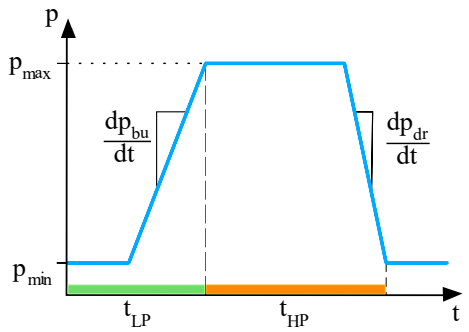


Figure 5.1. Schematic pressure load profile of the test specimen with the duration

The pressure build-up gradient was varied at three levels: 5, 12.5 and 20 kbar/s. The pressure drop gradient was varied at 5, 25 and 45 kbar/s. As this study focuses on pressure gradients, the maximum and minimum pressures were kept constant at p_{max} = 400 bar and p_{min} = 40 bar, respectively. The high-to-low pressure duration ratio was set to R_t = 1 and the excitation frequency to 6 Hz. The FSI simulations were performed for six discrete crack lengths a ∈ {1.5, 3, 4.5, 6, 7.5, 9} mm. Figure 5.2 shows the pressure profile over two load pulses within a crack of length a = 4.5 mm at different positions, for a pressure drop gradient of dp_{dr}/dt = 45 kbar/s and a pressure build-up gradient of dp_{bu}/dt = 12.5 kbar/s.

During pressure build-up, the pressure inside the crack increases with a noticeable delay compared to the system pressure in the pressure chamber. This delay is greater at positions deeper inside the crack. The duration of the investigated high-pressure phase is sufficient to approach the system pressure in the entire crack volume. After the system pressure drops, pressure peaks occur at various positions along the crack. This behaviour suggests that fluid is being entrapped during crack closure, leading to localized pressure increases. The highest pressure of 1440 bar is observed at the crack tip. In contrast, close to the crack mouth at x = 0.25 mm, the pressure decreased without returning to the initial low-pressure level p_{min}. The pressure at the crack mouth equals the system pressure during the whole load cycle due to the set boundary condition.

As the system pressure remains at a low-pressure level following the pressure drop, fluid can gradually escape the crack, resulting in decreasing pressure. A renewed pressure increase in the pressure chamber initially causes a pressure release at the observed positions. Close to the crack mouth (at x = 0.25 mm), the pressure drops shortly before increasing to the system pressure again. Deeper inside the crack the pressure drops to 0 bar and remains at this pressure level for a short period. While a low pressure inside the crack can be explained by the rapid opening of the crack in response to the pressure increase in the pressure chamber, the recorded value of 0 bar is still unrealistically low. This discrepancy is likely caused by insufficient volume flow in the model. Pressure close to the crack mouth does not drop to 0 bar, because of the resupply from the pressure chamber. The 0 bar pressure level is again followed by a subsequent increase to the system pressure that is delayed the deeper inside the crack. The high pressure duration is long enough to allow pressure equalization to p_{max} throughout the crack. In comparison to the former specimen with a shorter ligament of 5 mm, [4], the current simulated pressure profiles follow an equal pattern for a similar crack length.

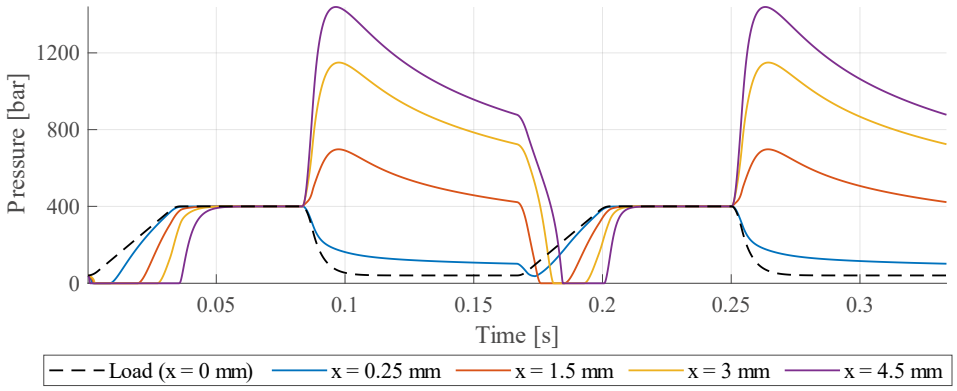


Figure 5.2. Pressure profile along the crack length a = 4.5 mm at various positions over time. The coordinate x = 0 mm corresponds to the crack mouth connected to the pressure chamber, while x = 4.5 mm represents the crack tip. The pressure distribution results from a load applied in two successive pulses.

Figure 5.2 displays the Mode I SIFs together with the corresponding system pressure over time for a crack length of 4.5 mm and the described load configuration. To evaluate the influence of the FSI, the SIF was calculated both with and without considering the fluid in the crack. The FSI accounts for hydrostatic pressure acting on the crack faces as well as fluid-induced crack closure effects. Without FSI, the SIF shows a strong correlation with the system pressure and the SIF range ΔK_I spans a wide interval. When fluid effects are considered $\Delta K_{I,eff}$ is reduced, indicating a lower crack propagation rate according to Paris' law (equation 2.4). Although higher peak values of the SIF are observed, the SIF does not drop significantly, presumably due to crack closure mechanisms. Overall, the comparison of SIF with and without FSI is in good agreement with the theoretical predictions of chapter 2, confirming that the influence of fluid on crack propagation is captured by the model.

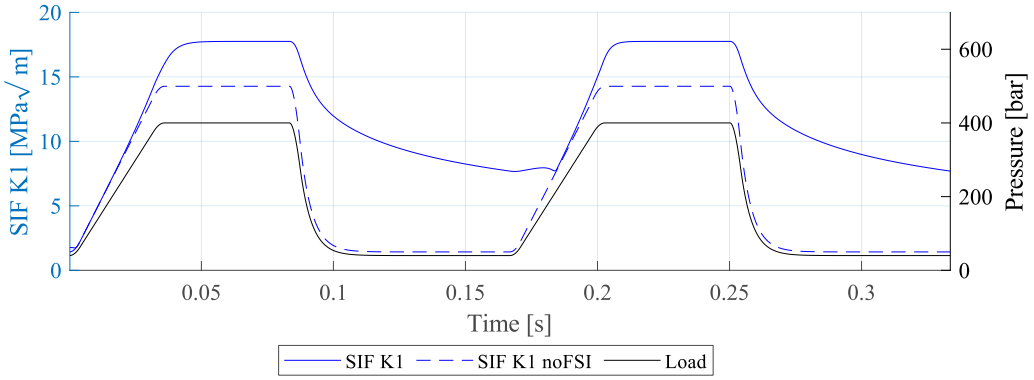


Figure 5.3. Comparison of the SIFs with and without considering the FSI (hydrostatic pressure on crack faces and fluid-induced crack closure). Left axis displays the mode I SIF, the right axis the corresponding system pressure.

According to Paris' law and the above-mentioned material parameters, this load case leads to a crack propagation rate of 0.729 µm per 1000 cycles for the evaluated crack length and load configuration, considering the FSI. Without the hydrostatic pressure and crack closure effects a crack growth rate of 1.5 µm per 1000 cycles was calculated, indicating a faster crack propagation due to a higher effective SIF range $\Delta K_{I,eff}$ at the crack tip for cracks without FSI. Figure 5.4 compares the mean crack propagation rate as a function of pressure drop rate and pressure build-up rate for cases with and without FSI effects averaged over the crack lengths. Crack propagation increases slightly as the pressure drop rate rises, with even higher growth rates predicted when FSI is neglected. In contrast, the influence of pressure build-up rate is comparatively weak, with only minor variations in crack propagation observed.

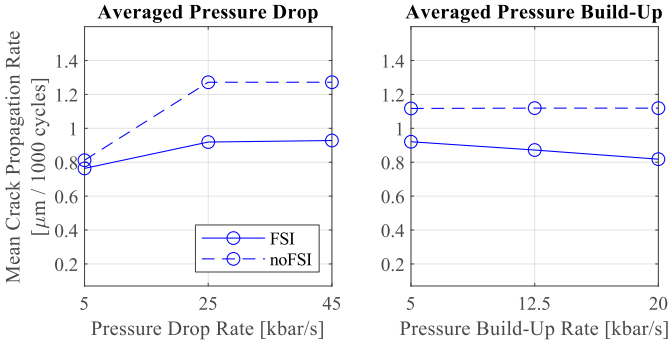


Figure 5.4. Comparison of the average crack propagation rate with and without FSI as a function of pressure drop rate and pressure rise rate. The values shown are averaged over the crack lengths (main effects).

Figure 5.5 shows the crack propagation rates of the discretely simulated crack lengths as a function of the pressure drop and build-up rates. These values primarily serve as a reference for comparison in future investigations. A clear trend can be observed indicating that the

crack propagation rate increases with increasing crack length, corresponding with the introduced crack propagation law by Paris (equation 2.4).

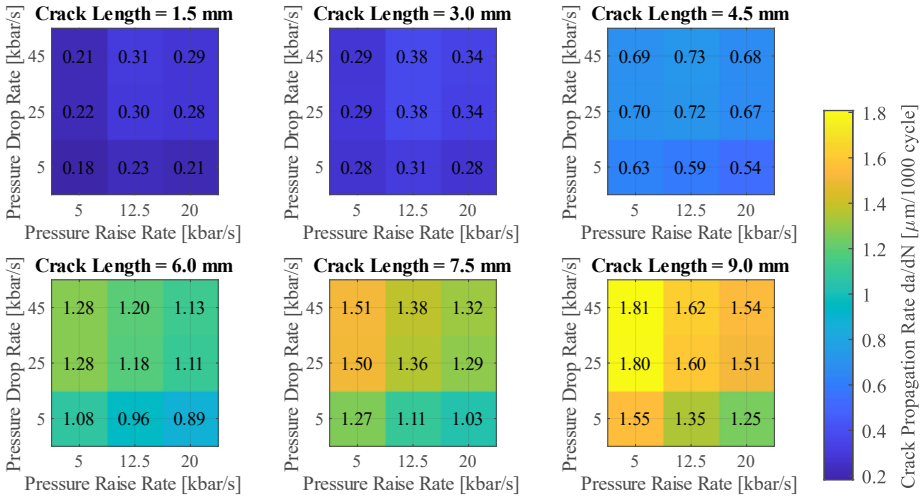


Figure 5.5. Crack propagation rates of discretely simulated crack lengths as a function of pressure drop and build-up rates and considering FSI effects.

6. Discussion and Outlook

In this work, a new version of the test specimen was developed according to extended and refined requirements to enable reproducible crack initiation, longer measurable crack propagation, and efficient manufacturability for a larger test campaign. Simulations showed that the test specimen experiences stable crack propagation without FSI under cyclic loading. The new specimen was implemented into the existing FSI framework. A plausible pressure profile for the first pressurization is achieved. However, during the second pressurization an unexpected depressurization down to 0 bar is observed, motivating a refinement of the simulation in the future. Furthermore, crack propagation rates for all configurations of pressure build-up and drop rate have been simulated. Within the considered interval, an increase in pressure drop rate shows a small influence on fatigue, resulting in a slightly shorter lifetime, while the build-up rate does not have a significant effect. The calculated values serve primarily as a reference point for future improvements.

In future work, the fluid domain should be extended to account for the crack width in a two-dimensional formulation. Fluid flow within the crack is expected to occur not only in the primary x-direction, but also in multiple directions within the crack plane. Introducing an additional spatial dimension would require the definition of an associated flow factor for this direction. As flow factors are defined along main roughness directions [17], this introduces a strong directional dependence of the surface roughness representation and consequently, of the volume flow. For non-isotropic roughness, this assumption is only valid if defined main roughness directions exist and agree with the directions for which the flow factors are defined. Otherwise, the resulting flow description becomes inaccurate. Alternatively, a different approach for modelling surface roughness should be employed. In [24], Rom et al.

compared the average flow model proposed by Patir and Cheng with a homogenization-based approach using a journal and slider bearing. They showed that the homogenization approach provides higher accuracy in predicting averaged pressure distributions while maintaining a comparable computational cost. Furthermore, the homogenization method can account for surface roughness of arbitrary anisotropy. In contrast to the flow factor approach, it offers greater flexibility when applied to non-isotropic surfaces.

Further limitations arise from the simplified fluid modelling assumptions. In particular, the assumptions of laminar, fully developed flow are insufficient to adequately represent the strongly transient flow conditions occurring within propagating cracks. In addition, possible cavitation effects resulting from the low pressure levels observed inside the crack after the pressure drop should be investigated.

Pulsation experiments will be used to compare and calibrate the current approach and following improvements. Furthermore, while the microholes have been considered in the calculation of the weight functions, their influence on the fluid flow, whose influence on fluid flow has not yet been considered will be addressed.

ACKNOWLEDGEMENTS

This work was funded by the German Research Foundation (DFG) in the scope of the project „Einfluss der Belastungsgradienten auf das Ermüdungsverhalten hydraulischer Komponenten" with the number 449676223. The authors would like to thank DFG for its support.

7. REFERENCES

[1] L. Brinkschulte, 'Assistenzsysteme zur Reduktion des Schädigungsverhaltens von Komponenten einer mobilen Arbeitsmaschine', Karlsruher Institut für Technologie, Karlsruhe, 2021. doi: 10.5445/KSP/1000130176.

[2] F. H. Davis and E. G. Ellison, 'Hydrodynamic Pressure Effects of Viscous Fluid Flow in a Fatigue Crack', Fatigue Fract. Eng. Mater. Struct., vol. 12, no. 6, pp. 527–542, Nov. 1989, doi: 10.1111/j.1460-2695.1989.tb00561.x.

[3] F. H. Davis, E. G. Ellison, and W. J. Plumbridge, 'Effects of Hydrostatic Pressure on the Rate of Fatigue Crack Growth', Fatigue Fract. Eng. Mater. Struct., vol. 12, no. 6, pp. 511–525, Nov. 1989, doi: 10.1111/j.1460-2695.1989.tb00560.x.

[4] L. Michiels, 'A simulation-based approach to the fluid-structure interaction inside fatigue cracks in hydraulic components', Karlsruher Institut für Technologie, Karlsruhe, 2024. Accessed: Feb. 06, 2025. [Online]. Available: https://publikationen.bibliothek.kit.edu/1000171515

[5] D. Gross and T. Seelig, Bruchmechanik. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016. doi: 10.1007/978-3-662-46737-4.

[6] M. Kuna, Finite Elemente in der Bruchmechanik: Theorie – Numerik – Anwendungen. Wiesbaden: Springer Fachmedien Wiesbaden, 2025. doi: 10.1007/978-3-658-46192-8.

[7] P. Paris and F. Erdogan, 'A Critical Analysis of Crack Propagation Laws', J. Basic Eng., vol. 85, no. 4, pp. 528–533, Dec. 1963, doi: 10.1115/1.3656900.

[8] W. Elber, 'Fatigue crack closure under cyclic tension', Eng. Fract. Mech., vol. 2, no. 1, pp. 37–45, Jul. 1970, doi: 10.1016/0013-7944(70)90028-7.

[9] H. A. Richard and M. Sander, Fatigue Crack Growth, vol. 227. in Solid Mechanics and Its Applications, vol. 227. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-32534-7.

[10] S. Ismonov and S. R. Daniewicz, 'Simulation and comparison of several crack closure assessment methodologies using three-dimensional finite element analysis', Int. J. Fatigue, vol. 32, no. 8, pp. 1322–1329, Aug. 2010, doi: 10.1016/j.ijfatigue.2010.01.016.

[11] E08 Committee, Test Method for Measurement of Fatigue Crack Growth Rates. doi: 10.1520/E0647-24.

[12] F. H. Davis and E. G. Ellison, 'Hydrodynamic Pressure Effects of viscous Fluid Flow inside a fatigue crack', Fatigue Fract. Eng. Mater. Struct., vol. 12, no. 6, pp. 527–542, Nov. 1989, doi: 10.1111/j.1460-2695.1989.tb00561.x.

[13] M. Akama and T. Mori, 'Boundary Element Analysis of Effects of Crack Face Friction and Trapped Fluid on Rolling Contact Fatigue Cracks', Q. Rep. RTRI, vol. 46, no. 4, pp. 231–237, 2005, doi: 10.2219/rtriqr.46.231.

[14] D. I. Fletcher, P. Hyde, and A. Kapoor, 'Investigating fluid penetration of rolling contact fatigue cracks in rails using a newly developed full-scale test facility', Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit, vol. 221, no. 1, pp. 35–44, Jan. 2007, doi: 10.1243/09544097JRRT63.

[15] A. F. Bower, 'The Influence of Crack Face Friction and Trapped Fluid on Surface Initiated Rolling Contact Fatigue Cracks', J. Tribol., vol. 110, no. 4, pp. 704–711, Oct. 1988, doi: 10.1115/1.3261717.

[16] N. Patir and H. S. Cheng, 'An Average Flow Model for Determining Effects of Three-Dimensional Roughness on Partial Hydrodynamic Lubrication', J. Lubr. Technol., vol. 100, no. 1, pp. 12–17, Jan. 1978, doi: 10.1115/1.3453103.

[17] N. Patir and H. S. Cheng, 'Application of Average Flow Model to Lubrication Between Rough Sliding Surfaces', J. Lubr. Technol., vol. 101, no. 2, pp. 220–229, Apr. 1979, doi: 10.1115/1.3453329.

[18] G.-J. Deng, S.-T. Tu, X.-C. Zhang, Q.-Q. Wang, and C.-H. Qin, 'Grain size effect on the small fatigue crack initiation and growth mechanisms of nickel-based superalloy GH4169', Eng. Fract. Mech., vol. 134, pp. 433–450, Jan. 2015, doi: 10.1016/j.engfracmech.2015.01.002.

[19] D. L. McDowell and F. P. E. Dunne, 'Microstructure-sensitive computational modeling of fatigue crack formation', Int. J. Fatigue, vol. 32, no. 9, pp. 1521–1542, Sep. 2010, doi: 10.1016/j.ijfatigue.2010.01.003.

[20] K. Kida, T. Yamazaki, M. Shibata, N. Oguma, and H. Harada, 'Crack initiation from micro surface holes in bearings under rolling contact fatigue', Fatigue Fract. Eng. Mater. Struct., vol. 27, no. 6, pp. 481–493, Jun. 2004, doi: 10.1111/j.1460-2695.2004.00771.x.

[21] Y. Murakami and M. Endo, 'Effects of defects, inclusions and inhomogeneities on fatigue strength', Int. J. Fatigue, vol. 16, no. 3, pp. 163–182, Apr. 1994, doi: 10.1016/0142-1123(94)90001-9.

[22] ANSYS Mechanical. ANSYS, Inc., Canonsburg, Pennsylvania, USA. [Online]. Available: https://www.ansys.com/

[23] T. L. Anderson, Fracture mechanics: fundamentals and applications, 2nd ed. Boca Raton Ann Arbor London [etc.]: CRC press, 1995.

[24] M. Rom, F. König, S. Müller, and G. Jacobs, 'Why homogenization should be the averaging method of choice in hydrodynamic lubrication', Appl. Eng. Sci., vol. 7, p. 100055, Sep. 2021, doi: 10.1016/j.apples.2021.100055.

Biographies



Fabian Blatter studied mechanical engineering at the Karlsruhe Institute of Technology. He is currently working as a Research Associate at the Department of Mobile Machines at the Karlsruhe Institute of Technology. His research areas include simulation of fluid-structure-interaction and hydraulic applications.



Lukas Boden obtained his master's degree in mechanical engineering at RWTH Aachen University in 2023. Since then, he is working as a research assistant at the Institute for Fluid Power Drives and Systems (ifas), focussing on identifying the influence of viscous fluid on fatigue crack propagation in fluid power components.



Prof. Dr.-Ing. Marcus Geimer received the Diploma degree in mechanical engineering from the RWTH Aachen University, Germany, in 1990, and the Ph.D. degree from the Institute of Hydraulics and Pneumatics, today named the Institute for Fluid Power Drives and Systems, RWTH Aachen University. Since 2005, he has been a Full Professor and the Director of the Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), Germany. He started his industrial career in 1995 in the field of construction at the company Krupp Berco Bautechnik GmbH, Germany, where he was the Leader of the research group for hydraulic breakers. In 2000, he was with Bucher Hydraulics GmbH, Germany, where he led the construction and customer development for mobile hydraulics.



Prof. Dr.-Ing. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Feasibility of Electro-Hydrostatic Actuators in Wheel Loader Application: A Case Study Focusing on Energy Efficiency, Controllability and Thermal Behaviour

Vili Vironmäki*, Ville Raunio, Mikko Huova and Petteri Multanen

Innovative Hydraulics and Automation, Tampere University, Tampere, Finland
* Corresponding author: vili.vironmaki@tuni.fi

Abstract.

Hydraulic actuators are extensively applied in non-road mobile machinery (NRMM) owing to their high power-to-weight ratio, robustness, and compact design. Conventional systems, however, exhibit throttling losses in control valves, which limit efficiency. While drivetrain electrification has been widely studied, the electrification of working hydraulics with electro-hydrostatic actuators (EHAs) remains less explored. EHAs eliminate throttling losses and enable energy regeneration, yet their suitability for wheel loaders has not been thoroughly assessed. This study investigates the feasibility of EHAs in wheel loaders through MATLAB/Simulink simulations using measured load cycle data of a typical short Y-cycle. EHA-based and conventional systems are compared with respect to energy efficiency, while thermal behaviour under simulated power losses is analysed using a lumped-parameter model. The results indicate that EHAs exhibit satisfactory position tracking accuracy and markedly improve energy efficiency, while thermal management remains an aspect that must be considered for reliable operation. Thus, EHAs show considerable promise for NRMM applications, provided that they are implemented properly, including adequate cooling.

Keywords. Electro-hydrostatic actuator, electrification, wheel loader, energy efficiency, thermal management.

1. INTRODUCTION

Hydraulic actuators are widely used in non-road mobile machinery (NRMM), such as wheel loaders and excavators, due to their high power density, robustness, and compact design [1, 2]. However, conventional valve-controlled hydraulics suffer from low overall efficiency because of throttling losses in directional control valves [3]. Meanwhile, stricter emission regulations and the growing demand for energy efficiency highlight the need for alternative hydraulic architectures.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Electrification offers a promising way to meet these requirements. While electric and hybrid drivetrains are already commercially available in NRMM, the electrification of working hydraulics remains less developed [4]. Electro-hydrostatic actuators (EHAs), in which an electric motor drives a hydraulic pump that directly supplies the actuator without control valves, can eliminate throttling losses and enable energy recovery [5, 6]. EHAs have been investigated in several NRMM applications, including mining loaders and excavators, focusing on efficiency improvements and alternative EHA concepts, such as high-speed implementations [7–11]. In wheel loader studies, EHAs have demonstrated significant efficiency gains over load-sensing (LS) systems but also indicated a need for thermal analysis [12].

Thermal behaviour is critical for closed-circuit EHAs, as they lack a reservoir that plays an important role in heat dissipation in conventional hydraulic systems. Elevated oil temperature reduces viscosity, accelerates component wear, and shortens the lifespan of hydraulic oil, thereby making temperature prediction essential during system design. High-fidelity thermal models exist but are computationally demanding for long duty cycles and often unnecessary in early-stage studies [13]. Lumped-parameter models provide a practical compromise, reducing data requirements and computational effort while maintaining sufficient accuracy [14]. Although thermal simulation studies of EHAs have been reported in the literature, for example by Michel et al. [15, 16], these focus on actuator sizes and load cycles that differ substantially from those encountered in wheel loader applications. Consequently, their results cannot be directly extrapolated to the present case, motivating a dedicated thermal simulation tailored to wheel loader operating conditions.

Most prior work addresses efficiency, controllability, or thermal behaviour in isolation and often relies on single-actuator setups, while also overlooking real load cycle data. However, trade-offs may arise between controllability, losses, and temperature rise, while position tracking accuracy may degrade in multi-actuator systems. This paper evaluates EHAs for wheel loader work functions holistically. Using measured duty-cycle and load data [17], a MATLAB/Simulink model with a high-fidelity mechanical boom representation compares an EHA system with an LS baseline. The analysis focuses on energy efficiency, controllability, and thermal behaviour under realistic load cycles to assess the feasibility of EHA integration in wheel loader application.

The remainder of this paper is organized as follows. Section 2 presents the system overview, including the boom structure and EHAs. Section 3 describes the system modelling. Simulation results are given in Section 4, followed by thermal modelling and analysis in Section 5. Section 6 concludes the paper.

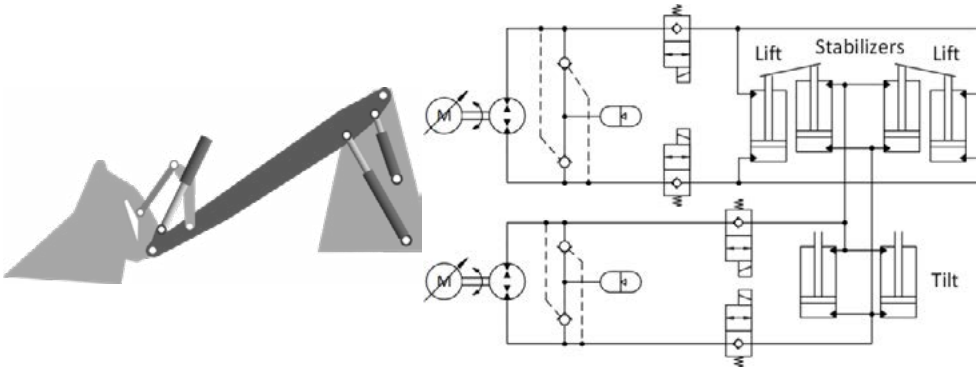
2. OVERVIEW OF THE SYSTEM

The system is a 6-tonne wheel loader performing the short Y-cycle of gravel loading. The boom (Fig. 2.1) includes two primary working actuators: lift cylinders and tilt cylinders, and a pair of stabilising (self-levelling) cylinders that are kinematically linked to the lift mechanism. When the lift cylinders extend, the stabilising cylinders move accordingly and drive the tilt cylinders to maintain bucket orientation. In the proposed architecture, the working actuators are implemented as EHAs.

Each EHA employs single-rod (asymmetric) hydraulic cylinders. The unequal chamber volumes are compensated by differential-flow compensation implemented with check valves, which allow flow from a low-pressure accumulator to meet the additional demand on the rod side. Each actuator is isolated by a normally closed 2/2 poppet valve functioning as an on/off load-holding valve (LHV), included for both safety and efficiency. Relief valves protect the system against overload conditions.

The cylinders are supplied by bidirectional pump/motor units capable of four-quadrant operation. The pumps are driven by permanent magnet synchronous motors (PMSMs), enabling high efficiency and precise control. During motion commands, the load-holding valves are energised (open), and otherwise they hold the load.

In the present work, conventional electric motors and hydraulic pumps with a maximum speed of 4000 rpm are employed. However, decentralized EHA architectures generally require more installation space than centralised hydraulic systems, which must be considered during system design. This limitation may be mitigated by using high-speed drives, as demonstrated in [10], which enable higher power density.



(a) (b)

Figure 2.1. (a) Wheel loader boom with the lift, tilt, and stabilising cylinders highlighted. (b) Simplified hydraulic schematic of the proposed EHA architecture.

3. SYSTEM MODELS

This section presents the system modelling. Section 3.1 develops the EHA model. Section 3.2 derives the electric power of the baseline LS system from measured data. Section 3.3 presents the EHA control architecture with PMSM current and speed control in the inner loops and position control in the outer loop.

3.1. EHA Model

The electro-hydrostatic actuators are modelled in MATLAB/Simulink, while the multibody dynamics of the wheel loader boom are represented in Simscape Multibody. The model incorporates boom and bucket inertias, gravitational effects, and linkage geometry. External loads such as payloads are not explicitly modelled but applied as an external force term derived from measured data. Actuator displacements are kinematically coupled to the linkage, and end-stop impacts are neglected as duty cycles remain within the stroke range.

The EHAs consist of a PMSM, an axial piston pump, hydraulic cylinders, accumulators, and auxiliary valves. The PMSM is described in the rotating d–q reference frame as

$$u_d = R_s i_d + L_d \frac{di_d}{dt} - P\omega L_q i_q \qquad u_q = R_s i_q + L_q \frac{di_q}{dt} + P\omega\left(L_d i_d + \phi_f\right) \tag{3.1}$$

$$T_e = \frac{3}{2} P\left(\phi_f i_q + \left(L_d - L_q\right) i_d i_q\right) \tag{3.2}$$

where u_d, u_q are stator voltages, i_d, i_q currents, R_s stator resistance, L_d, L_q inductances, P pole pairs, ω rotor speed, ϕ_f flux linkage (constant), and T_e electromagnetic torque. Motor parameters are taken from the Bosch Rexroth MS2N10-F0BHA (22.4 kW) datasheet [18].

The axial piston pump/motor is first expressed ideally as

$$Q_t = D_p \omega_p \qquad T_t = D_p \Delta p \tag{3.3}$$

where ω_p [rad/s] is the shaft speed, D_p [m^3/rad] the displacement, and $\Delta p = p_A - p_B$ the differential pressure over the pump/motor. Volumetric and mechanical losses are modelled using the analytical formulation of Huova et al. [19]. The actual flow rates to the A and B sides, and the shaft torque, are given by

$$\begin{aligned} Q_A &= Q_t - Q_{il} - Q_{el,A} \\ Q_B &= -Q_t + Q_{il} - Q_{el,B} \end{aligned} \qquad T_p = T_t + T_l \tag{3.4}$$

where Q_{il} and Q_{el} denote the internal and external leakages, and T_l the torque loss. The internal leakage Q_{il} is defined positive from the A-side to the B-side and may reverse direction depending on the pressure difference. External leakages depend on the pressure differences between the line and the accumulator.

Hydraulic cylinder chamber pressures are governed by

$$\frac{dp}{dt} = \frac{B_{eff}}{V}\left(\sum Q - \frac{dV}{dt}\right) \tag{3.5}$$

where B_{eff} is the effective bulk modulus, V the chamber volume, and Q the chamber flows. The actuator force is

$$F = A_a p_a - A_b p_b - F_\mu - F_{ext} \tag{3.6}$$

where A_a, A_b are piston areas, p_a, p_b chamber pressures, F_μ friction, and F_{ext} the measured external load. Friction is represented using a Coulomb–viscous–Stribeck model.

The accumulator pressure follows a polytropic gas relation,

$$p_{acc} = \left(\frac{V_0}{V_0 - \int Q_{acc} dt}\right)^n p_0 \qquad Q_{acc} = Q_{el,A} + Q_{el,B} + Q_{RV} - Q_{CV} \tag{3.7}$$

where V_0 is the nominal volume, p_0 the pre-charge pressure, and $n = 1.4$. The accumulators are pre-charged to 4.5 bar, with 6 L for the lift and 10 L for the tilt actuator.

Flow through the three types of orifices is modelled using the same orifice law in (3.8), with component-specific definitions for the opening area A:

- Pilot-operated check valve (POCV): the opening area A depends on the control pressure p_c as defined in (3.9)–(3.10).
- Load-holding valve: the area A is set by the on/off signal u_{LH} according to (3.11).
- Cylinder orifices: the area A is constant.

$$Q = AC_{q,max}\tanh\left(\frac{2\lambda}{\lambda_{crit}}\right)\sqrt{\frac{2|\Delta p|}{\rho}}\,\mathrm{sign}(\Delta p) \qquad \lambda = \frac{D_h}{\nu}\sqrt{\frac{2|\Delta p|}{\rho}} \tag{3.8}$$

where ρ is the fluid density, $\Delta p = p_1 - p_2$ the pressure difference across the orifice, and $\lambda_{crit} = 1000$ [20]. The flow number λ characterises the laminar–turbulent transition and modifies the discharge coefficient via the hyperbolic tangent term.

For the POCV, the control pressure and area–pressure relationships are defined as

$$p_c = (p_x - p_1)k_p + (p_1 - p_2) \tag{3.9}$$

$$A(p_c) = \begin{cases} A_{leak}, & p_c \le p_{cr} \\ A_{leak} + k(p_c - p_{cr}), & p_{cr} < p_c < p_{max} \\ A_{max}, & p_c \ge p_{max} \end{cases} \qquad k = \frac{A_{max} - A_{leak}}{p_{max} - p_{cr}} \tag{3.10}$$

The opening area A of both the POCV and LHV follows a first-order lag with a time constant of 0.1 s. The LHV is normally closed and energised according to

$$u_{LH} = \begin{cases} 1, & |\dot{x}_{ref}| > 0\ \mathrm{m/s} \text{ or } |e_c| > 1\ \mathrm{mm} \\ 0, & \text{otherwise} \end{cases} \tag{3.11}$$

where $\dot{x}_{ref}$ is the commanded actuator velocity and e_c the position error. Simulation parameters are listed in Table 1.

Table 1: Parameters of lift and tilt EHAs

Parameter	Lift	Tilt	Parameter	Lift	Tilt
D_p [cc/rev]	50.7	45	J_{PM} [kg·m^2]	0.0035	0.0025
A_a [cm^2]	100.5	127.2	A_b [cm^2]	68.7	88.0
$C_{q,max}$ [–]	0.7	0.7	s [mm]	640	400
p_{cr} [bar]	0.5	0.5	k_p [–]	3	3
V_0 [L]	6	10	p_0 [bar]	4.5	4.5
f_{sw} [Hz]	2000	2000	ϵ [–]	0.5	0.5

3.2. LS Model

The LS system is used as a reference for comparison with the EHA-based system. Measurement data were obtained from an LS configuration originally driven by a

combustion engine [17], but for consistency, the LS system is assumed to be driven by an electric motor operating at constant speed.

Measured pump flow Q_p and pressure p_p are converted into mechanical shaft power using the pump model defined in (3.3)–(3.4). The mechanical shaft power and the corresponding electrical input power are expressed as

$$P_{shaft} = T_p \omega_p \qquad P_{el} = \frac{P_{shaft}}{\eta_{em}(T_p, \omega_p)} \tag{3.12}$$

where $\eta_{em}(T_p, \omega_p)$ is obtained from the manufacturer's efficiency map [21].

3.3. Controller Design

The PMSM in the EHA is controlled by a cascade structure with dq-axis current loops and outer speed loop, all implemented with PI controllers tuned using the method in [22]. The dq-axis currents are regulated by PI controllers with decoupling terms, according to

$$u_q = K_{pc}(i_{q,ref} - i_q) + K_{ic}\int (i_{q,ref} - i_q)dt + \omega_e(L_d i_d + \phi_f) \tag{3.13}$$

$$u_d = K_{pc}(i_{d,ref} - i_d) + K_{ic}\int (i_{d,ref} - i_d)dt - \omega_e L_q i_q \tag{3.14}$$

where $i_{d,ref} = 0$ and $\omega_e = P\omega$. The controller gains are

$$K_{pc} = L_s \omega_c \qquad K_{ic} = R_s \omega_c \tag{3.15}$$

with the current control bandwidth ω_c set one decade below the inverter switching frequency f_{sw}.

The electromagnetic torque is generated by the q-axis current i_q, which serves as the control variable in the outer speed loop. The PI speed controller generates the current reference as

$$i_{q,ref} = K_{ps}(\omega_{ref} - \omega) + K_{is}\int (\omega_{ref} - \omega)dt \tag{3.16}$$

$$K_{ps} = \frac{J_{tot}\omega_s}{K_t} \qquad K_{is} = K_{ps}\frac{\omega_s}{5} \tag{3.17}$$

where J_{tot} is the total inertia, K_t the torque constant, and $\omega_s = \omega_c/5$.

The motor speed reference combines position feedback, velocity feedforward, and dynamic pressure feedback, conditioned by the LHV logic:

$$\omega_{ref} = \begin{cases} K_\omega(u_{fb} + u_{ff} - u_{pf}), & u_{LH} = 1 \\ 0, & \text{otherwise} \end{cases} \tag{3.18}$$

This ensures that the motor is commanded only when the LHV is energised, preventing line pressurisation when the valves are closed.

The velocity conversion ratio between cylinder and motor is defined as

$$K_\omega = \begin{cases} A_a/D_p, & \Delta p_L > 0 \\ A_b/D_p, & \text{otherwise} \end{cases} \tag{3.19}$$

where D_p is in [m^3/rad]. The position feedback is implemented using proportional control, while the velocity feedforward term directly applies the commanded velocity as

$$u_{fb} = K_p(x_{ref} - x) \qquad u_{ff} = K_{ff} v_{ref} \tag{3.20}$$

Damping is increased using high-pass filtered load-pressure feedback:

$$u_{pf} = \frac{K_{pf} s}{s + \omega_{pf}} \Delta p_L \qquad \Delta p_L = p_a - \frac{A_b}{A_a} p_b \tag{3.21}$$

This control approach follows [23], with the addition of the LHV logic. The overall controller is illustrated in Fig. 3.1, and the controller parameters are listed in Table 2.

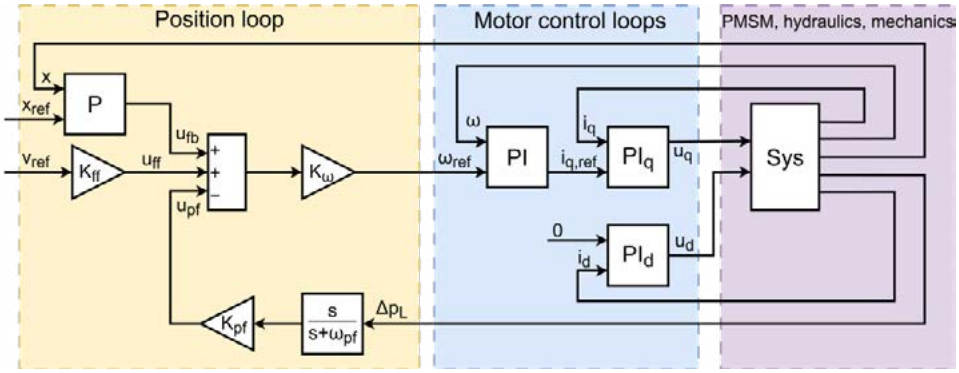


Figure 3.1. Control structure of the EHA showing the position control loop and the motor speed and current control loops. The decoupling terms of the inner PI_q and PI_d current controllers are omitted for clarity.

Table 2: Controller parameters

Parameter	Lift	Tilt	Parameter	Lift	Tilt
K_p [1/s]	2	4	K_{ff} [–]	1	0.5
ω_{pf} [rad/s]	6.60	8.70	K_{pf} [m·s^{-1}·bar^{-1}]	$1.5 \cdot 10^{-4}$	$2 \cdot 10^{-4}$
K_{ps} [A·s/rad]	2.28	2.15	K_{is} [A/rad]	115	108
K_{pc} [V/A]	3.15	3.15	K_{ic} [V·A^{-1}·s^{-1}]	106	106

4. SIMULATION RESULTS

This section presents simulation results of the proposed EHA architecture for wheel loader application. Position references for the lift and tilt cylinders were derived from Y-cycle measurements on a real machine. The analysis focuses on actuator controllability and energy efficiency, while thermal behaviour is discussed separately in Section 5.

4.1. Controllability

Tracking performance is evaluated using the short Y-cycle reference (Fig. 4.1). The actuators follow the position references with satisfactory accuracy. The maximum tracking errors are below 4 mm for the lift actuator and 29 mm for the tilt actuator, with corresponding root mean square (RMS) errors of 1.3 mm and 5.4 mm, respectively. During the load-holding phase, errors remain below 1 mm due to the control rule defined in equation (3.11). The tilt EHA exhibits higher position errors than the lift actuator because of the lift–tilt coupling through the stabilising cylinder. This coupling introduces disturbances in the position loop, requiring the controller to prioritise feedback compensation over faster velocity feedforward action, for which $K_{ff} = 0.5$ is applied in the tilt actuator. Despite these larger deviations, the achieved accuracy can be considered sufficient for typical load-handling tasks.

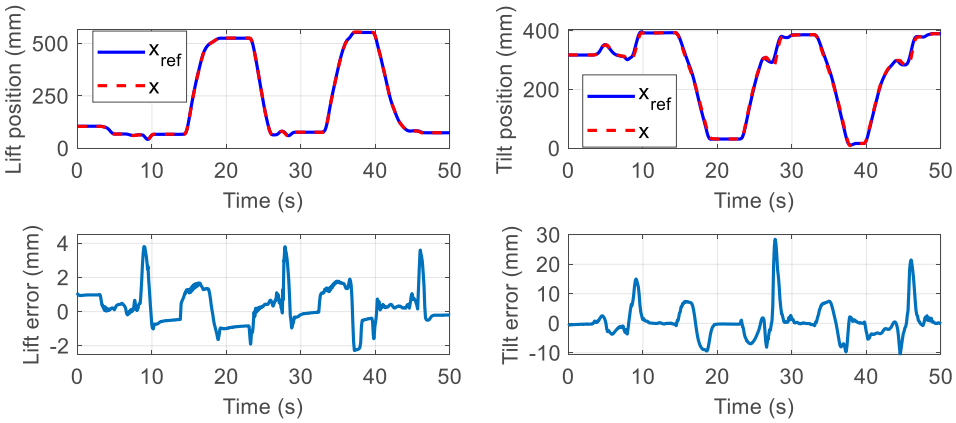


Figure 4.1. Position-tracking performance of lift and tilt actuators during the Y-cycle.

The dynamic behaviour during load-holding valve operation is illustrated in Fig. 4.2. When the LHV is closed, the actuator chamber is isolated from the line, allowing it to hold the load while a pressure difference of approximately 7.6 MPa gradually builds up between the chamber (p_a) and the connected line (p_{la}). When the valve opens, this pressure imbalance induces a brief cylinder motion as the pressures equalise. The resulting position deviation is minor and does not significantly affect overall performance.

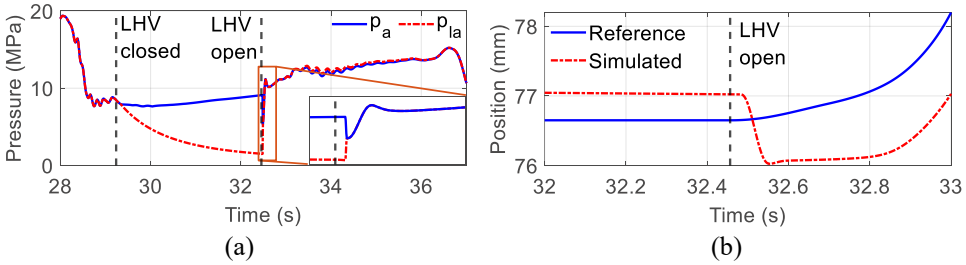


Figure 4.2. (a) Pressure dynamics in the A-chamber and connected line during load-hold valve operation. (b) Position tracking after valve opening.

4.2. *Energy Analysis*

Energy efficiency over the load cycle is evaluated and compared with the LS baseline. The instantaneous mechanical power of the cylinder and the electrical power of the PMSM are

$$P_{cyl} = F_{cyl}\dot{x} \qquad P_{el} = T_e\omega + P_{Cu} + P_{Fe} \tag{4.1}$$

where F_{cyl} is the cylinder force and $\dot{x}$ the piston velocity. Copper and iron losses are

$$P_{Cu} = \frac{3}{2}R_S\left(i_d^2 + i_q^2\right) \qquad P_{Fe} = \frac{K_{Fe}m_{EM}\epsilon}{1000\pi^{3/2}}|P\omega|^{3/2} \tag{4.2}$$

with $K_{Fe} = 4$ W/kg as the core loss per unit mass of sheet metal, m_{EM} the PMSM mass, and ϵ the sheet metal fraction [24]. The iron loss model provides only a simplified approximation. The EHA losses are primarily determined by the component efficiencies, unlike in conventional systems where throttling losses dominate. Therefore, Fig. 4.3 shows the efficiency maps of the main components together with their operation points over the lift actuator load cycle, providing an indication of the achievable overall efficiency. The motor efficiency map in Fig. 4.3a is based on the loss calculations in (4.2), while the pump map in Fig. 4.3b is derived from the model described in (3.3)–(3.4).

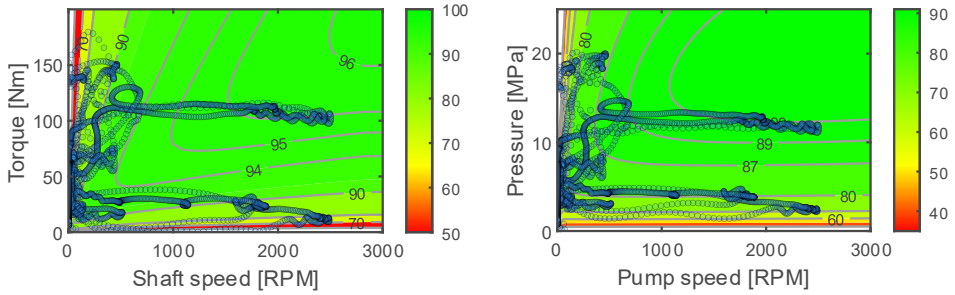


Figure 4.3. (a) Electric motor efficiency map and (b) lift pump efficiency map. Blue circles denote the operation points over the lift actuator load cycle.

Inverter losses are considered under the assumption of a constant inverter efficiency:

$$P_{DC} = \begin{cases} \dfrac{P_{el}}{\eta_{inv}}, & P_{el} \geq 0 \\ \eta_{inv}P_{el}, & P_{el} < 0 \end{cases} \qquad \eta_{inv} = 0.97 \tag{4.3}$$

The cycle energies and efficiency are defined as

$$E_{in} = \int P_{DC}dt \qquad E_{out} = \int P_{cyl}dt \qquad \eta = E_{out}/E_{in} \tag{4.4}$$

Positive P_{cyl} corresponds to load-resistive phases, whereas negative values indicate load-assistive phases. As the boom motion is periodic, the net work on the structure's own mass over a cycle is zero, and only the useful payload work is considered as output.

Figure 4.4 illustrates motor and cylinder powers for lift and tilt EHAs during the short Y-cycle. In load-resistive phases the motor receives electrical input ($P_{el} > 0$) and the cylinder

outputs mechanical power ($P_{cyl} > 0$). In load-assistive phases the load drives the actuator ($P_{cyl} < 0$) and the motor operates regeneratively ($P_{el} < 0$). For the tilt actuator, brief resistive segments with $P_{out} > P_{in}$ occur because the lift actuator momentarily drives the tilt cylinders through the stabilising cylinders.

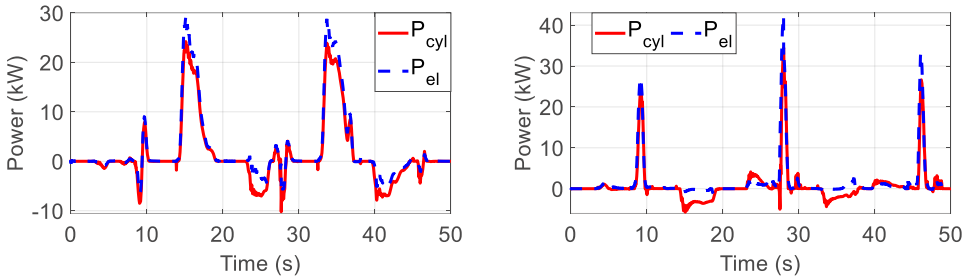


Figure 4.4. Motor electrical power P_{el} and cylinder mechanical power P_{cyl} for lift (a) and tilt (b) EHAs during the Y-cycle.

Figure 4.5 compares lift–tilt powers for LS and EHA systems, and Fig. 4.6 presents the corresponding cumulative energies. For the Y-cycle, the EHA requires 31.9 % of the electrical input of the LS system, yielding efficiencies of $\eta_{LS} = 0.17$ and $\eta_{EHA} = 0.55$. These results demonstrate that the EHA is substantially more efficient than the LS system.

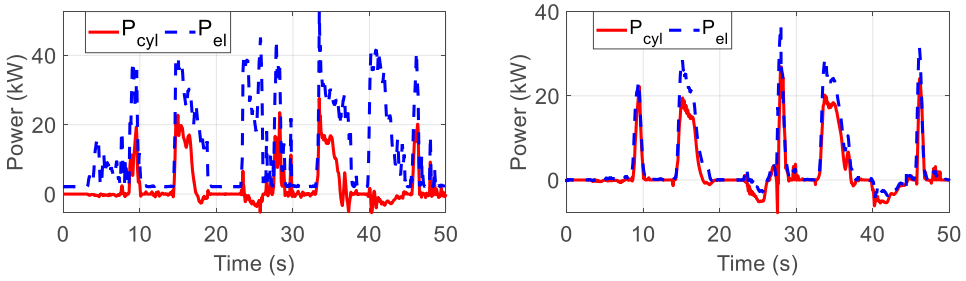


Figure 4.5. Motor electrical power P_{el} and cylinder mechanical power P_{cyl} for LS (a) and EHA (b). Powers are sums over lift and tilt actuators.

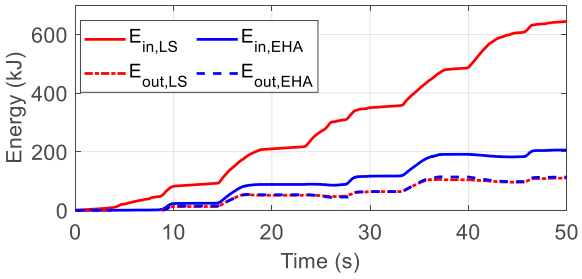


Figure 4.6. Cumulative electrical input energies E_{in} and useful cylinder outputs E_{out} for LS and EHA systems. Energies are sums over lift and tilt actuators.

Component-level losses are obtained by separating resistive and assistive intervals (Fig. 4.7). From the PMSM electrical input of 2.89 kW, 77.0 % is converted into useful mechanical output at the cylinder (2.22 kW). The remainder is dissipated in the motor (6.0%), pump/motor (10.9%), throttling (1.9%), and cylinder friction (5.6%). During assistive phases, the boom's potential energy provides 1.13 kW of mechanical input, of which 47.7 % (538 W) is transferred back to the inverter, while the rest is lost due to electrical, hydraulic, and mechanical losses as well as accumulator charging.

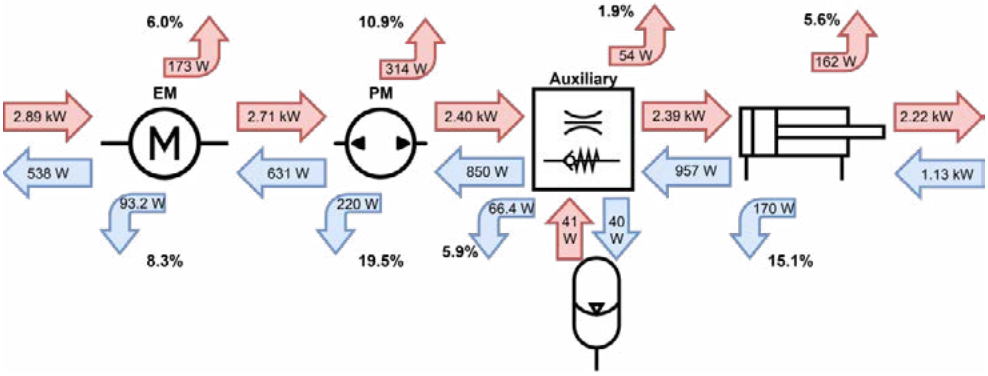


Figure 4.7. Component-level power flows of the lift actuator (EHA). Red = load-resistive (lifting), blue = load-assistive (lowering, regeneration).

Over a full Y-cycle, the mean loss powers are

$$P_{cyl,loss} = 162\ \text{W} + 170\ \text{W} = 332\ \text{W}$$
$$P_{aux,loss} = 54\ \text{W} + 66.4\ \text{W} = 120.4\ \text{W} \quad (4.5)$$
$$P_{PM,loss} = 314\ \text{W} + 220\ \text{W} = 534\ \text{W}$$

These losses are modelled as heat sources in the hydraulic system, while the electric motor losses are excluded from the thermal model (Section 5) since the PMSM is not in direct thermal contact with the hydraulic oil.

Overall, the results demonstrate that the EHA achieves more than three times the efficiency of the LS system under the Y-cycle, primarily due to reduced throttling losses and its capability for regenerative operation. In the present study, the LS system's variable displacement pump is operated at constant speed, which leads to increased losses under partial-load conditions, as the efficiency of the LS pump decreases at lower displacement settings. Since a PMSM is employed, a variable-speed LS implementation is technically feasible and could mitigate these losses, as demonstrated in [12]. Nevertheless, the constant-speed LS configuration represents the traditional load-sensing system architecture and is therefore an appropriate baseline for the present comparison.

The energy breakdown identifies the pump/motor unit as the dominant source of losses, while the auxiliary components contribute only marginally. These findings confirm the superior energy efficiency of the EHA, and the simulated power losses serve as an input in the subsequent thermal analysis.

5. THERMAL ANALYSIS

Thermal management is critical for EHAs, as their closed hydraulic circuit restricts heat dissipation. Unlike centralised hydraulic systems, where oil continuously returns to a reservoir providing effective cooling, the closed-circuit design of an EHA limits heat rejection and leads to higher temperatures. Elevated temperature lowers viscosity, weakens lubrication, and accelerates component wear. This analysis predicts system temperature over the continuous Y-cycle and evaluates the need for additional cooling.

A lumped-parameter model with a single combined control volume, representing the thermal capacities of the fluid and surrounding solids, is used. It defines a single system temperature T and an ambient temperature $T_{sur} = 20$ °C. The model combines convective cooling described by Newton's law of cooling and radiative heat exchange based on the Stefan–Boltzmann law. The oil is assumed to be well mixed, and all losses identified in Section 4.2 are represented as a constant cycle-averaged total heat input P_{loss}.

The system's thermal dynamics are

$$C\frac{dT}{dt} = P_{loss} - Q \qquad Q = \sum_i \frac{T - T_{sur}}{R_{th,i}} \tag{5.1}$$

The lumped heat capacity of the oil and the components is

$$C = c_{p,oil}m_{oil} + \sum_i c_{p,i}m_i \tag{5.2}$$

where c_p denotes specific heat capacity and m mass. Thermal resistances are

$$R_{th,i} = \frac{1}{(h_{conv,i} + h_{rad,i})A_{s,i}} \tag{5.3}$$

where $A_{s,i}$ is the surface area of the component.

The radiative and convective heat transfer coefficients are expressed as

$$h_{rad} = \varepsilon\sigma(T^2 + T_{sur}^2)(T + T_{sur}) \qquad h_{conv} = \frac{Nu \cdot k}{l_c} \tag{5.4}$$

where ε is the surface emissivity, $\sigma = 5.67 \cdot 10^{-8}\ \mathrm{W/(m^2K^4)}$ is the Stefan-Boltzmann constant [25], k is the thermal conductivity of air evaluated at the reference temperature $\bar{T} = (T + T_{sur})/2$, and l_c is the characteristic length of the component. The Nusselt number is geometry-dependent, and simple geometries such as horizontal cylinders and cubes are therefore adopted. The correlations used to calculate the Nusselt numbers for air free convection for the considered geometries are given in Table 3. Heat transfer is assumed to be governed by free convection and radiation, while forced convection resistances associated with internal pipe flow are neglected.

The dimensionless parameters are

$$Nu = \frac{hl_c}{k} \qquad Pr = \frac{\rho\nu c_p}{k} \qquad Gr = \frac{gl_c^3\beta_{air}\Delta T}{\nu^2} \tag{5.5}$$

where Pr is the Prandtl number, Gr the Grashof number, ρ the density, ν the kinematic viscosity, $g = 9.81\ \mathrm{m/s^2}$, and β_{air} the thermal expansion coefficient of air.

The oil heat capacity is updated with temperature following [26]:

$$c_p(T) = c_0 + \alpha_c(T - 273.15) \qquad c_0 = 1807\ \frac{\mathrm{J}}{\mathrm{kg \cdot K}} \qquad \alpha_c = 4.21\ \frac{\mathrm{J}}{\mathrm{kg \cdot K^2}} \tag{5.6}$$

Air is modelled as an ideal gas for evaluating density and thermal expansion:

$$\rho_{air} = p/(R_{air}\bar{T}) \qquad \beta_{air} = 1/\bar{T} \tag{5.7}$$

All other air properties are interpolated from tabulated data from [27]. Thermophysical properties are updated continuously using these relations.

Table 3: Nusselt correlations for free convection geometries [27, 28]. Air properties evaluated at the reference temperature.

Geometry	Nusselt number
Horizontal cylinder	$Nu = \left[0.6 + 0.387(Gr\ Pr)^{1/6}\left(1 + (0.559/Pr)^{9/16}\right)^{-8/27}\right]^2$ $Gr = Gr(T - T_{sur}),\ l_c = d$
Cube	$Nu = 5.748 + 0.752(Pr\ Gr)^{0.252}\left[1 - (0.492/Pr)^{9/16}\right]^{-0.448}$ $Gr = Gr(T - T_{sur}),\ l_c = A/D,\ D = \sqrt{4A_{proj}/\pi}$

Figure 5.1 shows the simulated temperature rise under continuous load-cycle operation. The system temperature approaches 82 °C after 350 min. The hydraulic system temperature exceeds the typical operating range of hydraulic systems (50–60 °C) [29] after about 85 min, indicating that additional cooling is required for long-term reliability. Internal forced convection resistances are omitted, which overestimates heat rejection. The actual cooling capacity may therefore be lower in practice. Furthermore, component power losses depend on the temperature in the real system, so the heat generation in the model may be slightly underestimated.

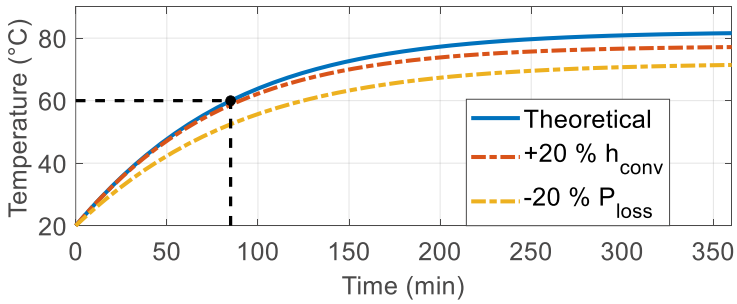


Figure 5.1. Simulated system temperature rise under continuous operation.

Jungnickel (2010, as cited in [14]) reported that natural convection coefficients may exceed theoretical estimates by up to 20 %. Moreover, Ketelsen et al. [14] showed that the results can be highly sensitive to power loss modelling errors. The present sensitivity analysis indicates that increasing the natural convection heat transfer coefficient by 20 % lowers the steady-state temperature by roughly 5 °C, whereas a 20 % reduction in input heat lowers it by about 10 °C. Even in the most favourable case, the temperature remains above the recommended range, indicating that active cooling is required for reliable long-term operation.

Thermal resistances of the components at the steady state temperature are listed in Table 4. These values are updated during the simulation, as they vary with temperature.

Table 4: Heat transfer coefficients and thermal resistances of system components at the steady-state temperature.

Component	Geometry	A_s [cm^2]	l_c [mm]	ε	h_{conv} [W/m^{2}K^{-1}]	h_{rad} [W/m^{2}K^{-1}]	R_{th} [K/W]
Pump	Cube	1350	798	0.90	12.7	7.00	0.376
Hoses	Cylinder	3970	16.2	0.93	8.59	7.24	0.159
Cylinder	Cylinder	4182	104	0.90	6.01	7.00	0.184
Accumulator	Cylinder	1297	200	0.90	5.59	7.00	0.612

6. CONCLUSION

This study evaluated the performance of electro-hydrostatic actuators for wheel loader work functions through simulations based on measured load cycle and external load data. The analysis covered controllability through position tracking precision, energy efficiency in comparison with an LS system, and thermal behaviour using a lumped-parameter model suitable for early-stage evaluation.

The main findings are:

- EHAs achieved satisfactory tracking accuracy, although the load-holding valve caused transient errors. The maximum tracking errors were below 4 mm for the lift actuator and 29 mm for the tilt actuator, with RMS errors of 1.3 mm and 5.4 mm, respectively.
- EHAs required only 31.9 % of the LS system's electrical input for the same work cycle, corresponding to efficiencies of 55.1 % and 17.0 % for the EHA and LS systems, respectively.
- Thermal simulations showed that the EHA exceeded typical hydraulic operating temperatures, indicating the need for additional cooling. The steady-state temperature reached 82 °C with theoretically calculated heat transfer coefficients, and ±20 % variations in these coefficients or heat inputs did not change this conclusion.

These results demonstrate that EHAs provide significant efficiency improvements compared with the LS system, but thermal management must be addressed to ensure reliable operation in such applications. In the present study, active cooling of the EHAs would have been required, which may compromise the efficiency benefits.

ACKNOWLEDGMENTS

This work was supported by Business Finland's Drive Forward project [ref. 428/31/2024].

REFERENCES

[1] Pustavrh, J., Hočevar, M., Podržaj, P., Trajkovski, A., Majdič, F., 'Comparison of hydraulic, pneumatic and electric linear actuation systems', Scientific Reports, 13(1), Article 20938, 2023, https://doi.org/10.1038/s41598-023-47602-x.

[2] Wiens, T., Deibert, B., 'A Low-Cost Miniature Electrohydrostatic Actuator System', Actuators, 9(4):130, 2020, https://doi.org/10.3390/act9040130.

[3] Padovani, D., Rundo, M., Altare, G., 'The Working Hydraulics of Valve-Controlled Mobile Machines: Classification and Review', Journal of Dynamic Systems, Measurement, and Control, 142(7), Article 070801, 2020, https://doi.org/10.1115/1.4046334.

[4] Niemelä, J., Vuorinen, E., Orifjonov, A., et al., Machinopedia—Review of Electric and Hybrid Off-Road Mobile Machines, Tampere University, 2024. Available: https://trepo.tuni.fi/bitstream/handle/10024/159633/978-952-03-2980-8.pdf. Accessed: Sep. 30, 2025.

[5] Minav, T., Bonato, C., Sainio, P., Pietola, M., 'Efficiency of direct driven hydraulic drive for non-road mobile working machines', 2014 International Conference on Electrical Machines (ICEM), Article 6960527, 2014, https://doi.org/10.1109/ICELMACH.2014.6960527.

[6] Agostini, T., De Negri, V., Minav, T., Pietola, M., 'Effect of Energy Recovery on Efficiency in Electro-Hydrostatic Closed System for Differential Actuator', Actuators, 9(1), Article 12, 2020, https://doi.org/10.3390/act9010012.

[7] Minav, T., Lehmuspelto, T., Sainio, P., Pietola, M., 'Series Hybrid Mining Loader with Zonal Hydraulics', IFK 2016, Dresden, Germany, 8-10 Mar. 2016.

[8] Zhang, S., Minav, T., Pietola, M., 'Decentralized Hydraulics for Micro Excavator', 15th Scandinavian International Conference on Fluid Power (SICFP'17), Linköping, 2017, pp. 187–195.

[9] Casoli, P., Scolari, F., Minav, T., Rundo, M., 'Comparative Energy Analysis of a Load Sensing System and a Zonal Hydraulics for a 9-Tonne Excavator', Actuators, 9(2), Article 39, 2020, https://doi.org/10.3390/act9020039.

[10] Pietrzyk, T., Roth, D., Schmitz, K., Jacobs, G., 'Design study of a high speed power unit for electro hydraulic actuators (EHA) in mobile applications', 11th International Fluid Power Conference, Aachen, Germany, 2018, pp. 233–245.

[11] Figge, F., Schmitz, K., 'Simulative Comparison Approach of Electro-Hydrostatic Actuator Circuits in Excavators', ASME/BATH 2023 Symposium on Fluid Power and Motion Control, Florida, Oct. 2023, https://doi.org/10.1115/FPMC2023-111798.

[12] Wang, F., Qiao, Y., Zhou, H., Wu, J., Xu, B., 'An Independent Electro-Hydraulic Variable Speed Drive to Improve Energy Efficiency of Electric Wheel Loader Working System', IEEE Transactions on Vehicular Technology, 73(12), 18250–18265, 2024, https://doi.org/10.1109/TVT.2024.3438268.

[13] Minav, T., Papini, L., Järf, A., Tammi, K., Pietola, M., 'Direct Driven Hydraulics: What can possibly go wrong? –A thermal analysis', 2016 XXII International Conference on Electrical Machines (ICEM), Lausanne, Switzerland, pp. 1618-1624, 2016, https://doi.org/10.1109/ICELMACH.2016.7732740.

[14] Ketelsen, S., Michel, S., Andersen, T., Ebbesen, M., Weber, J., Schmidt, L., 'Thermo-Hydraulic Modelling and Experimental Validation of an Electro-Hydraulic Compact Drive', Energies, 14(9), Article 2375, 2021, https://doi.org/10.3390/en14092375.

[15] Michel, S., Schulze, T., Weber, J., 'Energy-efficiency and thermo energetic behaviour of electrohydraulic compact drives', 9th International Fluid Power Conference, Aachen, Germany, 24–26 March 2014.

[16] Michel, S., Weber, J., 'Prediction of the thermo-energetic behaviour of an electrohydraulic compact drive', 10th International Fluid Power Conference, Dresden, Germany, 8–10 March 2016, pp. 219–234.

[17] Linjama, M., Huova, M., Tammisto, J., Tervonen, J., Ahola, V., '6-ton wheel loader short Y-cycle data for 440 seconds loading of gravel', Zenodo, 2024, https://doi.org/10.5281/zenodo.10639199.

[18] Bosch Rexroth AG, MS2N Synchronous Servomotors: Project Planning Manual, Document R911347583, Edition 07, 06/2021.

[19] Huova, M., Tammisto, J., Linjama, M., Tervonen, J., 'Fuel Efficiency Analysis of Selected Hydraulic Hybrids in a Wheel Loader Application', BATH/ASME 2018 Symposium on Fluid Power and Motion Control, Bath, UK, Sept. 12–14, 2018, V001T01A035, https://doi.org/10.1115/FPMC2018-8869.

[20] Vasiliu, N., Vasiliu, D., Călinoiu, C., Puhalschi, R., Simulation of Fluid Power Systems with Simcenter Amesim, 1st ed., CRC Press, 2018, ISBN 978-1482253559.

[21] ENGIRO GmbH, Data sheet 260W-08058-ABC, ver. 14.01, July 2025. Available: https://www.engiro.com/fileadmin/pdfs/Datenblaetter_engiro/260W/ENGIRO_Datasheet_260W_08058_ABC_V14_01.pdf. Accessed: Aug. 25, 2025.

[22] Kim, S.-H., Electric motor control: DC, AC, and BLDC motors, Elsevier, 2017.

[23] Jensen, K.J., Ebbesen, M.K., Hansen, M.R., 'Novel Concept for Electro-Hydrostatic Actuators for Motion Control of Hydraulic Manipulators', Energies, 14(20), 6566, 2021, https://doi.org/10.3390/en14206566.

[24] Schmidt, L., Ketelsen, S., Mommers, R., Achten, P., 'Analogy between hydraulic transformers & variable-speed pumps', BATH/ASME 2020 Symposium on Fluid Power and Motion Control, 2020, Article V001T01A007, https://doi.org/10.1115/FPMC2020-2719.

[25] Çengel, Y., Cimbala, J., Turner, R., Fundamentals of Thermal-Fluid Sciences, 5th ed., McGraw-Hill Education, 2016, ISBN 978-0078027680.

[26] Busquets, E., Ivantysynova, M., 'Temperature Prediction of Displacement Controlled Multi-Actuator Machines', International Journal of Fluid Power, 14(1), 25-36, 2013, https://doi.org/10.1080/14399776.2013.10781066.

[27] VDI-GVC, VDI Heat Atlas, 2nd ed., Springer, Berlin, 2010, https://doi.org/10.1007/978-3-540-77877-6.

[28] VDI-GVC, VDI-Wärmeatlas, 11th ed., Springer, Berlin, 2013, https://doi.org/10.1007/978-3-642-19981-3.

[29] Ponomarev, P., Åman, R., Handroos, H., Immonen, P., Pyrhönen, J., Laurila, L., 'High power density integrated electro-hydraulic energy converter for heavy hybrid off-highway working vehicles', IET Electrical Systems in Transportation, 4(4), 114–121, 2014, https://doi.org/10.1049/iet-est.2013.0009.

Biographies



Vili Vironmäki received the bachelor's degree in automation engineering from Tampere University in 2025. He is currently pursuing the master's degree in automation engineering at Tampere University, where he also works as a Research Assistant in the Automation Technology and Mechanical Engineering Unit, Faculty of Engineering and Natural Sciences. His research interests include electro-hydrostatic actuator modelling and control.



Ville Raunio received the bachelor's degree in automation engineering from Tampere University in 2023. He is currently pursuing the master's degree in automation engineering at Tampere University. He is currently working as a Model-Based Design Trainee at Epec Oy. His areas of interest include model-based design and electrification of mobile machinery.



Mikko Huova received the M.Sc. in automation engineering in 2008 and completed his Ph.D. in the same field in 2015, both from Tampere University of Technology. He is currently working as a Senior Research Fellow at the Automation Technology and Mechanical Engineering unit at Tampere University. The areas of interest include energy efficient actuator systems, mobile machinery and model-based control.



Petteri Multanen received the master's degree in mechanical engineering from Tampere University of Technology in 1996, and the philosophy of doctorate degree in Electrical-Electronics & Computer Engineering from Tampere University of Technology in 2002. He is currently working as a University Lecturer at the Automation Technology and Mechanical Engineering Unit, Faculty of Engineering and Natural Sciences in Tampere University. His research areas include hybrid drives and energy efficient mobile machines.

Automation Framework for Constrained Optimal Trajectory Planning and Adaptive Robust Motion Control of Electrohydraulic Mobile Machines

Bobo Helian*, Christoph Keller, and Marcus Geimer

Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

**Corresponding author: bobo.helian@kit.edu*

Abstract.

This study presents an integrated automation framework that unifies constrained optimal trajectory planning and adaptive robust motion control for electrohydraulic mobile machines. The proposed multi-layer architecture consists of command, planning, control, and actuation levels, ensuring seamless coordination between task execution and dynamic motion adaptation. A nonlinear filter-based time-optimal trajectory planner is developed to generate real-time feasible trajectories under kinematic and dynamic constraints, while an adaptive robust backstepping controller is designed to compensate for nonlinearities, parameter uncertainties, and external disturbances in hydraulic actuation. Additionally, a simplified control strategy is introduced for pressure-compensated systems to enhance industrial applicability. Comparative validation on an unmanned wheel loader demonstrates the proposed framework's effectiveness, achieving high tracking accuracy, strong robustness, and real-time constraint satisfaction. This work contributes to advancing autonomous and intelligent operation of electrohydraulic mobile machines by bridging planning and control in a cohesive system framework.

Keywords. Electrohydraulic systems, Constrained trajectory planning, Adaptive Robust Control, Mobile working machines, Automation control framework

1. INTRODUCTION

1.1 Motivation

The autonomy of mobile working machines [1], such as wheel loaders and excavators, has become a key research direction in both academia and industry. It aims to reduce dependency on skilled operators, improve operational safety, and enable continuous, autonomous operation in complex environments. Reliable automation of such machines fundamentally depends on two capabilities—trajectory planning that respects system constraints, and motion control that ensures precise and robust execution.

Electrohydraulic systems, though offering high power density, present major challenges due to nonlinearities, parameter uncertainties, and high-order hydraulic dynamics, which degrade tracking accuracy and robustness under varying loads and disturbances. Conventional linear controllers struggle to maintain performance in such conditions. Moreover, unlike electric drives, hydraulic systems are limited by flow saturation and slow responses due to fluid dynamics. Achieving time-optimal and constraint-satisfied trajectories while ensuring precise and robust execution remains a key challenge. These factors motivate the development of a unified control framework that integrates constrained trajectory planning with adaptive robust control for safe and effective autonomous operation of electrohydraulic mobile machines.

1.2 Related works

1) Trajectory planning strategies of hydraulic machines

Constraints pose a major challenge to achieving reliable motion control in mobile working machines, particularly in systems with multiple hydraulic actuators interacting with their environment. These constraints encompass both kinematic and dynamic aspects, such as input and hydraulic force limitations [2], and are influenced by factors including hydraulic configuration, safety requirements, and payload conditions specific to each machine [3]. Moreover, constraint boundaries may vary over time due to changes in operating conditions. Violating these constraints can cause actuator damage (e.g., cavitation), safety risks, and motion instability. Therefore, a robust and real-time constrained control strategy is crucial for enabling autonomous operation.

Various constrained control strategies have been proposed to address the limitations of electro-hydraulic systems. In [4], a motion tracking controller with full-state constraints was developed using a barrier Lyapunov function to ensure constraint satisfaction. The study [5] introduced an input-constrained control scheme that incorporates a bounded smooth hyperbolic tangent function to achieve asymptotic tracking. Furthermore, [6] presented a robust nonlinear control method for electro-hydraulic force systems considering output constraints. Although these works have advanced the understanding of constrained control in electro-hydraulic systems, most approaches focus on a single type of constraint and neglect the time-varying characteristics and interdependence among multiple constraints. Consequently, existing research still faces several open challenges: (1) Low-pass filters are often applied to handle aggressive reference inputs, but they introduce steady-state deviations, and the resulting trajectories are typically suboptimal—failing to reach the desired reference in minimum time while respecting the constraints. (2) The tracking accuracy of synthesized constrained trajectories is not guaranteed due to the nonlinearities and uncertainties inherent in electro-hydraulic systems, limiting the realization of optimal constrained motion control.

Integrating a precise motion tracking controller with real-time constrained trajectory planning provides a promising pathway to achieving both robust control performance and constraint satisfaction [3]. In [2] and [7], constrained adaptive robust motion tracking controllers were developed for linear motors and hydraulic actuators by combining a model-based motion controller with a nonlinear filter-type trajectory planner, yielding satisfactory results. The nonlinear filter planner, first proposed in [8], enables real-time generation of time-optimal trajectories that respect motion constraints. While [2] and [7] successfully

demonstrated this approach in simple mechatronic systems, its application to complex electrohydraulic systems of mobile working machines remains largely unexplored. The trajectory planning of such machines introduces greater challenges due to multiple interacting constraints and strong coupling effects, motivating further development of advanced constrained trajectory planning strategies.

2) Precise motion control strategies of electrohydraulic actuators

While trajectory planners can generate optimal constrained trajectories for operations, tracking the desired trajectory may not be executed accurately. This is due to the nonlinear and uncertain nature of electro-hydraulic actuators [9], as well as external disturbances during mobile machine operation [10]. Therefore, a precise and robust motion control is essential for autonomous operation of mobile machines.

To achieve high-precision and robust control of electro-hydraulic systems, adaptive and nonlinear model-based strategies have been extensively explored [11]. Several methods have been proposed to handle uncertain disturbances and nonlinear effects. For instance, [10] developed an active disturbance rejection adaptive control scheme for hydraulic servo systems, while [12] proposed an H_∞-based control method with feedforward compensation to improve terrain disturbance rejection in mobile manipulators. To mitigate nonlinear friction, a modified LuGre model was introduced in [13], and a virtual decomposition control approach based on subsystem dynamics was presented in [14].

Among these approaches, Adaptive Robust Control (ARC) has shown outstanding potential in compensating uncertain loads, enhancing both accuracy and robustness [15]. As illustrated in Fig. 3, ARC integrates parameter adaptation with nonlinear feedback. Considering the high-order dynamics of hydraulic systems, an ARC backstepping scheme has been developed to ensure precise motion tracking under nonlinearities and uncertainties.ARC theory has been successfully extended to various electro-hydraulic configurations, such as electro-hydraulic actuators [9] and electro-variable pumps [16], and shows strong promise for multi-actuator mobile machines, where coupling dynamics remain an open challenge.

Furthermore, the integration and communication of these methodologies are essential for the robust and autonomous operation of mobile machines. However, such integrated approaches remain relatively underexplored, and further research is required to establish cohesive frameworks that seamlessly bridge planning and control for electro-hydraulic mobile machines [17].

1.3 Contributions

This study proposes a unified framework that integrates constrained trajectory planning and robust nonlinear motion control for autonomous operation of electrohydraulic mobile machines. An unmanned wheel loader is used as a representative case study, which is a highly automated unmanned system. The main contributions of this study are:

- A multi-level integrated control architecture is introduced, consisting of command-level: task command, mid-level: constrained trajectory generation, and lower-level: precise motion tracking. This structure ensures seamless coordination between planning and execution, contributing to robust control performance and flexible application.

- A time-optimal constrained trajectory planning algorithm is developed, utilizing a nonlinear filter-based planner with high-order of integrators, which explicitly considers both the kinematic structure and dynamic limitations of the wheel loader's electrohydraulic actuating system. This ensures feasibility and safety in real-world machines' automated operations.
- An Adaptive Robust Control (ARC) is designed for motion control of proportional valve-control hydraulic system, as the underlying controller. It effectively mitigates the impact of system nonlinearities and uncertainties by robust feedback and uncertain parameter online adaptation, ensuring accurate tracking of planned trajectories and overcoming limitations of traditional linear control methods.
- In addition to the advanced ARC method, a simplified controller was also designed, combining linear feedback with dynamic feedforward compensation. This lightweighted controller is specially designed for pressure-compensated hydraulic systems, offering a practical and easily implementable solution for industrial applications (e.g., wheel loads or excavators).
- The above proposed framework was validated on an unmanned wheel loader, a representative mobile machine with a standard electrohydraulic system and multiple actuators. Results demonstrate high trajectory-tracking accuracy and strong robustness against system uncertainties.

By bridging trajectory planning and adaptive robust control within a cohesive framework, the study offers a scalable and effective solution for future autonomous mobile machines.

2. CONTROL FRAMEWORK

The complete architecture integrates the above components into a hierarchical structure, as shown in Figure 2.1, explained as below:

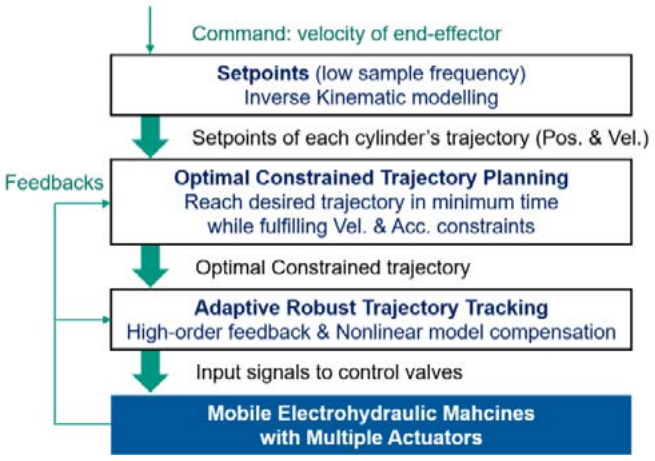


Figure 2.1. Control framework

1) **Command Layer:** The operator or task planner provides the desired end-effector motion (e.g., velocity), as command to the control system. The position & velocity of each actuator can be obtained in the form of discrete setpoints via inverse

kinematic model. This layer can be performed by autonomously or by operator remotely.

2) **Trajectory Planning Layer:** The trajectory planner generates an optimal constrained trajectory based on the setpoints. The planned trajectory reaches the desired setpoints in minimum time with high sampling frequency while simultaneously satisfying velocity and acceleration constraints.

3) **Motion Control Layer:** The motion tracking controller executes the planned trajectories with real time feedback and feedforward compensation. This controller design ensures robustness and precision despite the nonlinearities and uncertainties of the electrohydraulic system.

4) **Actuation Layer:** Control signals are transmitted to the hydraulic valves of the load arm actuators. The feedback from sensors of the electrohydraulic system is given to Trajectory Planning and Motion Control Layer in real-time.

This multi-layered structure ensures high adaptability. Feedback loops across multiple levels allow the system to dynamically compensate for disturbances and operational variations.

3. OPTIMAL CONSTRAINED TRAJECTORY PLANNING

Constraints of the motion control of the hydraulic actuator include the constraints of velocity (i.e., $\dot{x}^+$ and $\dot{x}^-$) and acceleration (i.e., $\ddot{x}^+$ and $\ddot{x}^-$), which are described in (3.1), while the stroke (position) bounds are static constraints that should be addressed at the task-planning level (i.e., setpoints in Figure 2.1) and not the focus of this section.

$$\begin{aligned} \dot{x}^- \le \dot{x} \le \dot{x}^+ \\ \ddot{x}^- \le \ddot{x} \le \ddot{x}^+ \end{aligned} \tag{3.1}$$

The assignments of the preset constraints for trajectory planning depend on the parameters and working conditions of the electro-hydraulic systems.

In this section, a two-integrator based nonlinear filter [8] is detailed to realize the online time-optimal trajectory planning to reach the desired reference while fulfilling the constraints on the velocity and acceleration (6). The two integrators chain is given by

$$\begin{bmatrix} x_p(i+1) \\ \dot{x}_p(i+1) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_p(i) \\ \dot{x}_p(i) \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix} \ddot{x}_p(i) \tag{3.2}$$

where T demotes the sampling time; i demotes the number of sampling periods. When $i = 0$, $\left[x_p(0), \dot{x}_p(0), \ddot{x}_p(0)\right]^T = [x_d(0), \dot{x}_d(0), \ddot{x}_d(0)]^T$. The state equation of the two integrators chain

$$\boldsymbol{x_p}(i+1) = \boldsymbol{A_c} x_p(\mathrm{i}) + \boldsymbol{B_c} \ddot{x}_p(i) \tag{3.3}$$

where $\boldsymbol{x_p}(\mathrm{i}) = \left[x_p(\mathrm{i}), \dot{x}_p(\mathrm{i})\right]^T$ is the system state when the time $t = i \times T$; $\ddot{x}_p$ here serves as the system input.

The error between the planned trajectory x_p and the originally expected command trajectory x_d is defined by $\boldsymbol{y}$, which is given as

$$\begin{aligned} y(i) &= x_p(i) - x_d(i) \\ \dot{y}(i) &= \dot{x}_p(i) - \dot{x}_d(i) \end{aligned} \tag{3.4}$$

where $\boldsymbol{x_d}$ is the given reference, indicating that when $y(i)$ tends to zero, $x_p(i)$ converges to $x_d(i)$. The error dynamics of the transformed system (10) is given by

$$\boldsymbol{y}(i+1) = \boldsymbol{A_c}\boldsymbol{y}(i) + \boldsymbol{B_c}v(i) \tag{3.5}$$

where $v(i) = \ddot{x}_p - \ddot{x}_d$.

In the error dynamics (3.4), the matrix $\boldsymbol{A_c}$ and $\boldsymbol{B_c}$ are both directly related to the sampling time T. To avoid the system error dynamics explicitly relying on the sampling time T, the following transformation is applied [11].

$$\boldsymbol{y}(i) = \boldsymbol{W}\boldsymbol{\eta}(i) \tag{3.6}$$

where

$$\boldsymbol{W} = \begin{bmatrix} T^2 & -\dfrac{T^2}{2} \\ 0 & T \end{bmatrix} \tag{3.7}$$

and $\boldsymbol{\eta}(i) = [\eta_1(i), \eta_2(i)]^T$. Matrix $\mathbf{W}$ is nonsingular, it implies that inverse transformation

$$\boldsymbol{\eta}(i) = \boldsymbol{W}^{-1}\boldsymbol{y}(i). \tag{3.8}$$

Combining the transformed formulation (3.8) and the error dynamics (3.5), it can be obtained that

$$\boldsymbol{\eta}(i+1) = \boldsymbol{A_d}\boldsymbol{\eta}(i) + \boldsymbol{B_d}v(i) \tag{3.9}$$

where

$$\boldsymbol{A_d} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{B_d} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{3.10}$$

In the above formulation (3.9), the system state $\mathbf{x_r}$ is transformed into an equivalent state $\boldsymbol{\eta}$. Since $\boldsymbol{W}$ is nonsingular, the eventual convergence to the origin of $\boldsymbol{\eta}$ implies the convergence of $\boldsymbol{y}$ to zero and the tracking of $\mathbf{x}_d$ is achieved.

To fulfil the requirements of the time-optimal trajectory planning objective and the given constraints (3.1), the control input of (3.9) is given by

$$v = \begin{cases} -\ddot{x}^{-}\mathrm{sat}\left(\dfrac{\delta}{\ddot{x}^{-}}\right), & \text{if } \delta \geq 0 \\ -\ddot{x}^{+}\mathrm{sat}\left(\dfrac{\delta}{\ddot{x}^{+}}\right), & \text{if } \delta < 0 \end{cases} \tag{3.11}$$

where $\mathrm{sat}(\bullet)$ is a saturation function; δ is obtained by,

$$\delta \coloneqq \eta_2 + \frac{\gamma}{m} + \frac{m-1}{\alpha}. \tag{3.12}$$

The detailed design of γ, α, m, and further variables for constraints transformation are given in [7, 8], in which theoretical proof has been rigorously performed that the state point is

forced to the origin of the sliding surface σ in minimum time, meanwhile, it also consequences on the fulfillment of the constraints on the velocity and acceleration.

As defined in (3.5) , the optimized acceleration of the trajectory can be written as

$$\ddot{x}_p = v + \ddot{x}_d \tag{3.13}$$

The optimized position x_p and velocity $\dot{x}_p$ can be calculated through differential equation (3.2). This optimized trajectory is given to motion controller in real time, which is detailed in the next chapter.

4. MOTION CONTROL

4.1. Adaptive Robust Controller Design

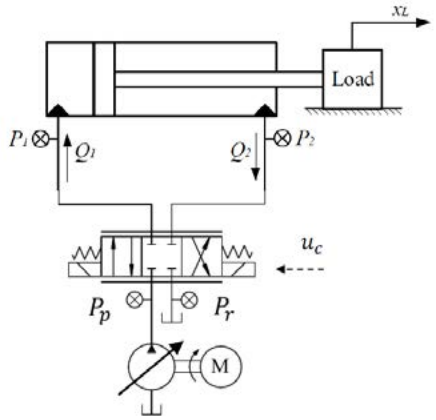


Figure 4.1 A simplified proportional valve-controlled cylinder

1) Dynamics modelling

A simplified general hydraulic system is presented in Figure 4.1, in which the motion of the hydraulic actuator is controlled by a proportional valve. The descriptions of symbols are listed in Table 1 in the Appendix. The dynamics of the cylinder motion can be modelled as

$$m\ddot{x}_L = P_1A_1 - P_2A_2 - b_v\dot{x}_L - b_c \cdot S_f(\dot{x}_L) + F_{\mathrm{d}}(x_L, \dot{x}_L, t) \tag{4.1}$$

where m represents the integrated mass of the inertia load; x_L represents the inertia load displacement; P_1 and P_2 represent the pressures of the head chamber and rod chamber; A_1 and A_2 represent areas of the two sides of the piston; b_v and b_c represent a viscous friction coefficient and a coulomb friction coefficient, respectively; $S_f(\bullet)$ represents a continuous function approximating the $\mathrm{sign}(\bullet)$. In addition, F_{d} represents a lumped uncertain nonlinearity, described as

$$F_{\mathrm{d}}(x_L, \dot{x}_L, t) = F_{\mathrm{dn}} + \tilde{F}_{\mathrm{d}} \tag{4.2}$$

where F_{dn} represents the nominal value of F_{d}.

The pressure dynamics of two chambers of the cylinder actuator can be modelled as

$$\begin{cases} \dfrac{V_1(x_L)}{\beta_e}\dot{P}_1 = -A_1\dot{x}_L - C_i(P_1 - P_2) + Q_1 \\ \dfrac{V_2(x_L)}{\beta_e}\dot{P}_2 = A_2\dot{x}_L + C_i(P_1 - P_2) - Q_2 \end{cases} \tag{4.3}$$

where C_i is the inner leakage coefficient. The flow rate linked to each chamber of the cylinder, controller by the proportional valve, can be modeled as

$$\begin{cases} Q_1 = k_{\mathrm{q1}}\sqrt{|\Delta P_1|}\,u_c \\ Q_2 = k_{\mathrm{q2}}\sqrt{|\Delta P_2|}\,u_c \end{cases} \tag{4.4}$$

where k_{q1} and k_{q2} are the throttling coefficients, $|\Delta P_i|$ (i=1,2) is the pressure drop due to throttling, and u_c denotes the effective valve control, which is proportional to the opening of valve spool.

The state-space variables of the pump control electro-hydraulic actuator are defined as $[x_1, x_2, x_3, x_4]^T = [x_L, \dot{x}_L, P_1, P_2]^T$. The uncertain parameters include m, b_v, b_c, β_e, C_i, and d_n. The uncertain parameter set is defined as $\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6]^T$ in which $\theta_1 = 1/m,\ \theta_2 = b_v/m,\ \theta_3 = b_c/m,\ \theta_4 = F_{dn}/m,\ \theta_5 = \beta_e,\ \theta_6 = \beta_e C_i$.

The dynamics of the actuator motion control system can be parameterized as

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \theta_1(A_1x_3 - A_2x_4) - \theta_2x_2 - \theta_3S_f(x_2) + \theta_4 + \Delta \\ \dot{x}_3 &= \frac{1}{V_1(x_1)}(-\theta_5A_1x_2 - \theta_6(x_3 - x_4) + \theta_5g_1u_c) \\ \dot{x}_4 &= \frac{1}{V_2(x_1)}(\theta_5A_2x_2 + \theta_6(x_3 - x_4) + \theta_5g_2u_c) \end{aligned} \tag{4.5}$$

where $g_i = k_{qi}\sqrt{|\Delta P_i|},\ \Delta = \tilde{F}_{\mathrm{d}}/m$.

2) Controller design

This motion controller is a nonlinear model-based controller, based on an adaptive robust backstepping control structure, to accurately track the planned trajectory x_{p}. As shown in Figure 4.2, the controller includes two steps: motion tracking step and hydraulic force tracking step. Define $\hat{\theta}$ as the estimate of θ with $\hat{\theta}_{\min} < \hat{\theta} < \hat{\theta}_{\max}$ and $\tilde{\theta}$ as the estimation error, (i.e., $\tilde{\theta} = \hat{\theta} - \theta$).

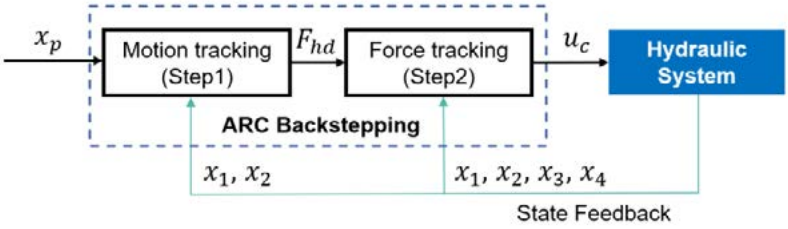


Figure 4.2 ARC Backstepping control strategy

The adaptation law is given by

$$\dot{\hat{\theta}} = \mathrm{Proj}_{\hat{\theta}}(\Gamma\tau) \tag{4.6}$$

where τ is an adaptation function introduced by Eq. (4.10) and (4.13); Γ is an adaptation rate matrix; $\mathrm{Proj}(\bullet)$ is a discontinuous projection [9].

Step 1: Motion tracking

The virtual control input F_h (i.e., hydraulic force) is designed to make the actuator position track the planned trajectory x_r. The output tracking error of the planned trajectory is defined as $z_1 = x_1 - x_r$. A switch function is then defined by

$$z_2 = \dot{z}_1 + k_1 z_1 = x_2 - x_{2eq}, \;\; x_{2eq} = \dot{x}_r - k_1 z_1 \tag{4.7}$$

where $x_r(t)$ represents the planned trajectory to be tracked by output x_1 and k_1 is a positive feedback gain. It is straightforward to see that making z_1 converging to zero implies making z_2 converging to zero [31].

Considering the hydraulic force $F_h = P_1 A_1 - P_2 A_2$, the error dynamics is given by

$$\dot{z}_2 = \dot{x}_2 - \dot{x}_{2eq} = \theta_1 F_h - \theta_2 x_2 - \theta_3 S_f(x_2) + \theta_4 + \Delta - \dot{x}_{2eq}. \tag{4.8}$$

The desired hydraulic force F_{Ld} is defined as the control input of Step 1. Then, the ARC control input law is designed to minimize z_2, which is given by

$$\begin{aligned} &F_{hd}\left(x, \hat{\theta}, t\right) = F_{hda} + F_{hds} \\ &F_{hda} = \frac{1}{\hat{\theta}_1}\left(\hat{\theta}_2 x_2 + \hat{\theta}_3 S_f(x_2) - \hat{\theta}_4 + \dot{x}_{2eq}\right) \\ &F_{hds} = F_{Lds1} + F_{Lds2}, \;\; F_{hds1} = -\frac{1}{\theta_{1min}} k_2 z_2 \end{aligned} \tag{4.9}$$

where F_{hda} demotes an adaptive model compensation; F_{hds} demotes the robust feedback term; k_2 demotes a positive stabilizing feedback gain. Details on how to choose the nonlinear feedback (for F_{hds2}) are proposed in the literatures [15]. The adaptation function of Step 1 is given by

$$\tau_2 = \phi_2 z_2, \; \phi_2 = \left[F_{Lda}, -x_2, -S_f(x_2), 1, 0, 0\right]^T. \tag{4.10}$$

Step 2: Hydraulic force tracking

Control input u_c is synthesized to make the actual hydraulic force track the desired hydraulic force F_{Ld} (i.e., the input of Step 1). The input discrepancy of Step 1 is defined as $z_3 = F_h - F_{hd}$. The error dynamics is modelled as

$$\begin{aligned} \dot{z}_3 = \;& \dot{F}_h - \dot{F}_{hd} \\ = \;& u_c \theta_5 \left(\frac{g_1 A_1}{V_1} + \frac{g_2 A_2}{V_2}\right) - \theta_5 x_2 \left(-\frac{A_1^2}{V_1} - \frac{A_2^2}{V_2}\right) - \theta_6 (x_3 - x_4)\left(-\frac{A_1}{V_1} - \frac{A_2}{V_2}\right) \\ & -\left(\dot{F}_{hdc} + \dot{F}_{hdu}\right) \end{aligned} \tag{4.11}$$

in which $\dot{F}_{hdc}$ and $\dot{F}_{hdu}$ represent the calculable part and uncalculable part of $\dot{F}_{Ld}$, respectively.

The ARC control law u_c, the valve control input, is given by

$$u_{\mathrm{c}}\left(x,\hat{\theta},t\right)=u_{\mathrm{ca}}+u_{\mathrm{cs}}$$
$$u_{\mathrm{ca}}=\frac{1}{\hat{\theta}_5\left(\frac{g_1A_1}{V_1}+\frac{g_2A_2}{V_2}\right)}\left(\theta_5x_2\left(-\frac{A_1^2}{V_1}-\frac{A_2^2}{V_2}\right)+\theta_6(x_3-x_4)\left(-\frac{A_1}{V_1}-\frac{A_2}{V_2}\right)+\dot{F}_{hdc}-\hat{\theta}_1z_2\right)$$
$$u_{\mathrm{cs}}=u_{\mathrm{cs1}}+u_{\mathrm{cs2}},\ \ u_{\mathrm{cs2}}=-\frac{1}{\hat{\theta}_{5\min}\left(\frac{g_1A_1}{V_1}+\frac{g_2A_2}{V_2}\right)}k_3z_3 \tag{4.12}$$

in which u_{ca} functions is an adaptive model compensation term; u_{cs} is the robust feedback term; k_3 is a positive stabilizing feedback gain; u_{cs2} is a nonlinear feedback term.

The adaptation function is given by

$$\tau=\tau_2+\phi_3z_3 \tag{4.13}$$

where ϕ_3 is defined as

$$\begin{aligned}\phi_3 \ &=[z_2-\frac{\partial F_{Ld}}{\partial x_2}F_h,-\frac{\partial F_{hd}}{\partial x_2}x_2,-\frac{\partial F_{hd}}{\partial x_2}S_f(x_2),\frac{\partial F_{hd}}{\partial x_2},\\ &\quad u_{\mathrm{ca}}\left(\frac{g_1A_1}{V_1}+\frac{g_2A_2}{V_2}\right)-x_2\left(-\frac{A_1^2}{V_1}-\frac{A_2^2}{V_2}\right),-(x_3-x_4)\left(-\frac{A_1}{V_1}-\frac{A_2}{V_2}\right)]^T.\end{aligned} \tag{4.14}$$

The theoretical proof of the two-step ARC backstepping controller has been rigorously performed in the literatures [9, 15] and it is omitted in this section. It is proved that the proposed controller guarantees the transient performance and asymptotical tracking in the presence of uncertainties only.

4.2. Lightweighted Controller Design for Pressure-Compensated Systems

In pressure-compensated hydraulic systems, the valve input can be represented as either the desired flow rate (for upstream compensation) or the desired actuator velocity (for downstream compensation). The pressure compensation mechanism effectively decouples the valve flow from load pressure variations, thereby simplifying the control design. As a result, a lightweight linear controller can achieve stable and accurate motion tracking in applications with moderate dynamic requirements.

$$v_{\mathrm{input}}=\dot{x}_p+k_{\mathrm{p}}(x_{\mathrm{d}}-x_1)+k_{\mathrm{v}}(\dot{x}_{\mathrm{d}}-x_2) \tag{4.15}$$

where k_{p} and k_{v} are positive feedback gains for position and velocity; the planed velocity $\dot{x}_p$ servers as the feedforward term for dynamic compensation. Thereby, the flow rate input can be calculated by

$$Q_{\mathrm{input}}=v_{\mathrm{input}}\boldsymbol{A} \tag{4.16}$$

where $\boldsymbol{A}$ denotes the effective piston area of the corresponding chamber (rod or head side), depending on the actuator's moving direction.

5. VERIFICATION ON UNMANNED WHELL LOADER

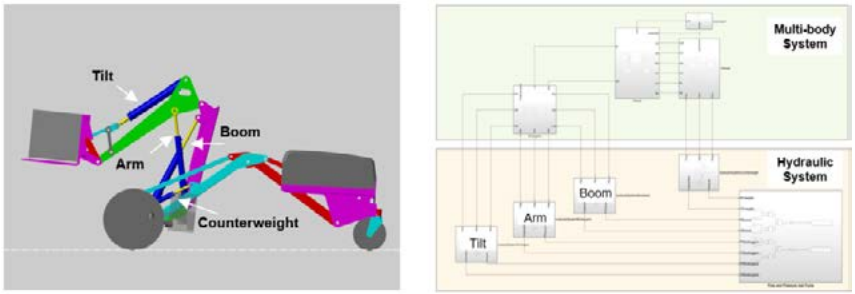
5.1. System Formulation



Figure 5.1. Two-wheel wheel loader by Quimo (picture from [18])

The machine used for a case study in this paper is an unmanned wheel loader, named two-wheel wheel loader, designed by Quimo as shown in Figure 5.1. The working arm of the wheel loader comprises three hydraulic cylinders, each named according to its function: the boom cylinder, arm cylinder, and tilt cylinder, as shown in Figure 5.2 (a). In addition, a counterweight cylinder is used to adjust the counterweight and ensure stability of the machine. All cylinders are equipped with linear position sensors, enabling the measurement of position and velocity. The hydraulic system adopts a load-sensing system with upstream pressure compensators. The valve block used in this study is an M4-12 load-sensing valve block coupled with an A10VO variable-displacement pump (both from Bosch Rexroth).

A co-simulation platform was developed in MATLAB/Simulink using the Simscape "Multibody" and "isothermal Fluid" library, integrating the multi-body dynamics of the wheel loader, the simplified upstream pressure-compensated hydraulic circuit, and their coupled dynamic interactions, as illustrated in Figure 5.2 (b). The density of the steel used for the working arm is set to 7800 kg/m³ in the simulation program.



(a) Multi-body system (b) Co-simulation system

Figure 5.2. Simulation platform

5.2. Single Cylinder Control Result

Before machine-level verification, the proposed controller was first applied to a proportional valve–controlled single-cylinder system. Case 1 was conducted to validate the optimal constrained control performance of the planning strategy integrated with the proposed Adaptive Robust Controller (ARC) described in Section 4.1, referred to as C1. The maximum velocity limit $\dot{x}^+$ in Eq. (3.1) was set to 0.20 m/s for the planner, while the

desired trajectory x_d had a maximum velocity of 0.23 m/s. As shown in Figure 5.3, controller C1 accurately tracked the planned constrained trajectory x_p that reached x_d at steady state while maintaining a smooth transient response. The velocity of the planned trajectory remained strictly within the prescribed constraint boundary throughout the motion, and the actual actuator velocity x_2 demonstrate the accurate tracking of the planned actuator velocity $\dot{x}_p$.

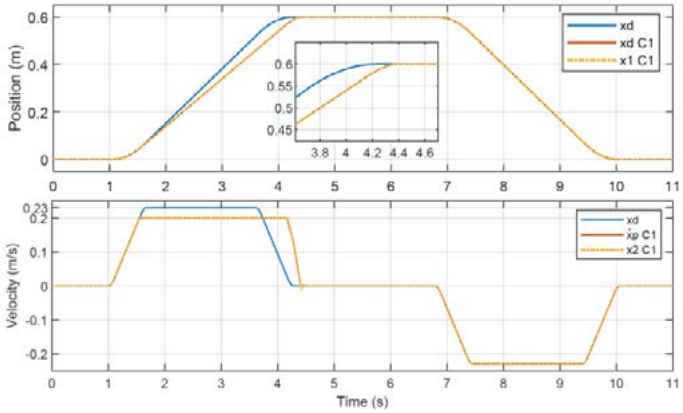


Figure 5.3. Positions and velocities in Case 1

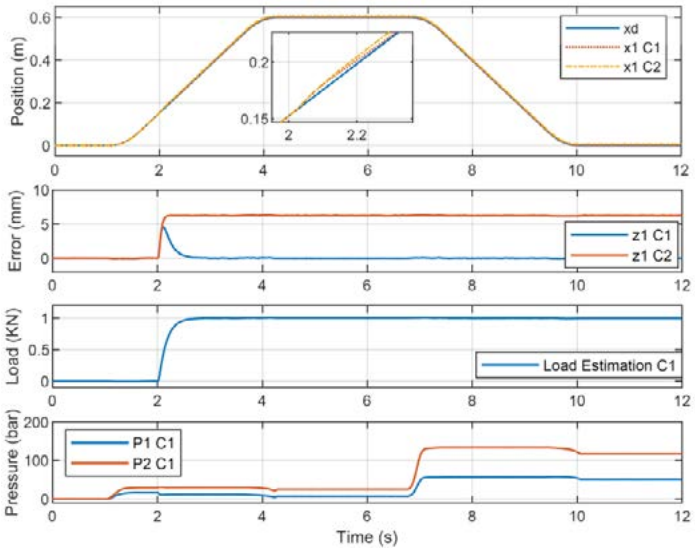


Figure 5.4. Position, tracking errors, load estimation in Case 2

Case 2 was performed to evaluate the anti-disturbance capability of the proposed controller. For comparison, another model-based controller with the same structure but without parameter adaptation was implemented as a baseline, denoted as C2. An external disturbance force of 1 kN was applied to the actuator at t = 2 s. As shown in Figure 5.4, both controllers experienced an immediate rise in tracking error following the disturbance. However, C1 rapidly suppressed the error and restored accurate tracking, demonstrating its superior

disturbance rejection. This improvement is attributed to the effective online parameter adaptation in C1, which also provided accurate load estimation results, as shown in Figure 5.4. The pressure signals of both chambers of the hydraulic cylinder are also presented to provide additional insight into the internal force states of the actuator.

5.3. Working Arm Control Result

In this section, the proposed control framework (as illustrated in Figure 2.1) is fully applied to the working arm of the unmanned wheel loader, with an external mass load of 50 kg.

In this case, since the upstream pressure compensator defines the control input as the desired flow rate, the lightweight controller introduced in Section 4.2 is employed for the motion control layer. For comparison, a baseline controller (C3) is implemented—an open-loop controller with feedforward compensation but without feedback—to demonstrate the significance of closed-loop control in multi-actuator hydraulic systems.

Figure 5.5 shows the control results of the working arm, where the trajectories of all three cylinders are presented. The desired trajectories x_d are discrete setpoints generated at a low sampling frequency of 20 Hz. The proposed controller C1 generates real-time constrained trajectories x_p at 100 Hz, guiding the system smoothly toward x_d while optimally satisfying all constraints. The measured position signals x_1 demonstrate that all cylinders accurately track the planned trajectories under C1. In contrast, C3 exhibits poor tracking accuracy, with cumulative position deviations over time due to the absence of feedback control.

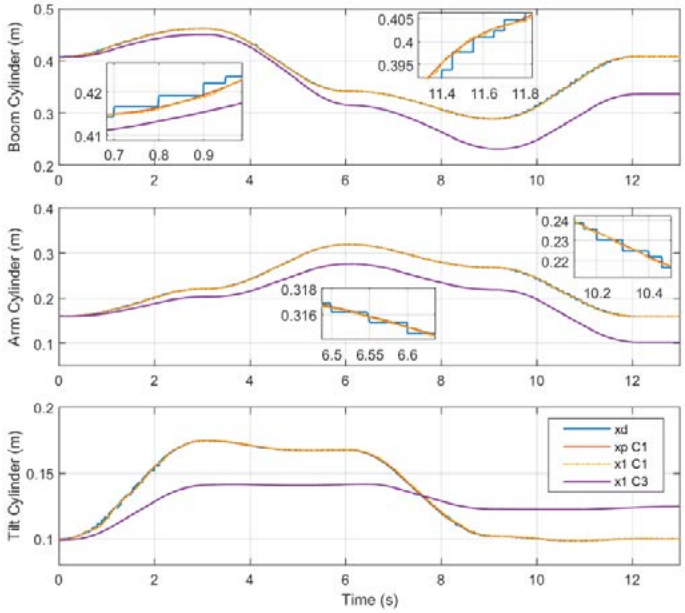


Figure 5.5. Results of working arm control

A comparison of C1 and C3 on the machine level is shown in Figure 5.6. The full operation cycle was designed to drive the end effector along a square trajectory in 12 seconds. Controller C1 maintained high tracking accuracy throughout both transient and steady-state phases, successfully accomplishing the desired motion. In contrast, C3 resulted in significant

deviations at both the cylinder and end-effector levels. Demonstration videos corresponding to Figure 5.6 are provided in the **Appendix**.

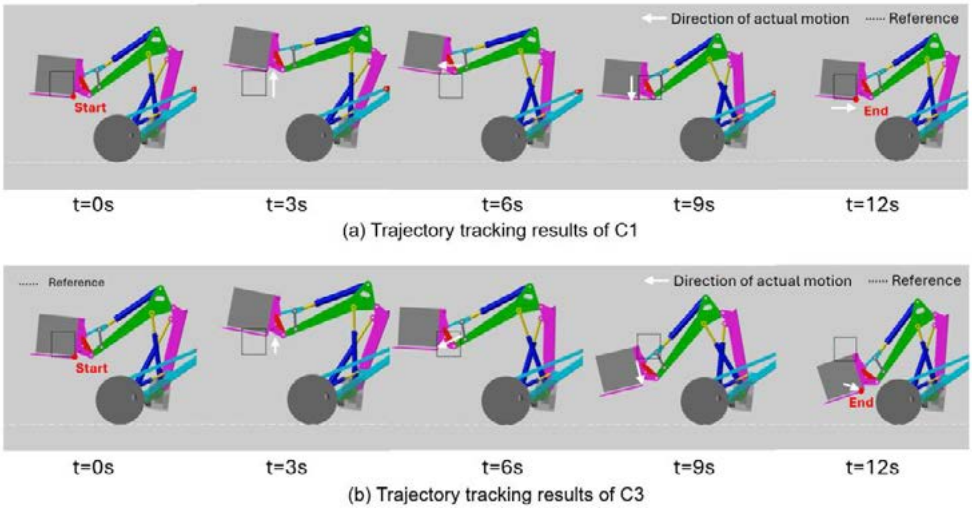


Figure 5.6. Comparative results shown on the machine

6. CONCLUSION AND FUTURE WORK

This paper presented a unified framework integrating constrained optimal trajectory planning and adaptive robust motion control for electrohydraulic mobile machines. The multi-layer architecture enables coordinated, constraint-satisfied, and robust motion execution. A nonlinear filter-based planner and an adaptive robust controller were combined to achieve precise, time-optimal tracking under nonlinearities and uncertainties. Experimental verification on an unmanned wheel loader demonstrated that the proposed approach outperformed baseline linear and open-loop controllers in tracking accuracy, robustness under uncertainty, and constraint compliance. A lightweight controller for pressure-compensated hydraulic systems was also introduced for practical industrial use. The proposed framework provides an effective and scalable basis for the autonomous operation of future electrohydraulic mobile machines.

In the future work, the integration of higher-level perception and task planning modules, such as environment sensing and adaptive task sequencing, will be explored to enable fully autonomous operation of mobile machines. Furthermore, future work will emphasize experimental validation across various types of machines under varying payloads and external disturbances. These extensions will further enhance the robustness, scalability, and industrial applicability of the proposed automation framework.

ACKNOWLEDGEMENT

This work was carried out as part of a research project funded by the Ministry of Economics, Labour and Tourism of Baden-Württemberg, Germany. Furthermore, we would like to thank the project partner QuiMo for designing the mechanical system of the unmanned wheel loader and providing its desired movement (i.e., Setpoints x_d in Section 5.3).

APPENDIX

The videos can be accessed via the link:
https://gitlab.kit.edu/bobo.helian/2RadLader_vedios.git .

7. REFERENCES

[1] M. Geimer, Mobile Working Machines, SAE International, Warrendale, PA, USA, 2020.

[2] B. Helian, Z. Chen, B. Yao, 'Constrained motion control of an electro-hydraulic actuator under multiple time-varying constraints', IEEE Trans. Ind. Inform., vol. 19, no. 12, pp. 11878–11888, 2023.

[3] A. Stentz, J. Bares, S. Singh, P. Rowe, 'A robotic excavator for autonomous truck loading', Auton. Robots, vol. 7, no. 2, pp. 175–186, 1999.

[4] Z. Xu, G. Qi, Q. Liu, J. Yao, 'ESO-based adaptive full-state constraint control of uncertain systems and its application to hydraulic servo systems', Mech. Syst. Signal Process., vol. 167, Art. no. 108560, 2022.

[5] Z. Lin, J. Yao, W. Deng, 'Input constraint control for hydraulic systems with asymptotic tracking', ISA Trans., vol. 129, pp. 616–627, 2022.

[6] W. Zang, Q. Zhang, J. Su, L. Feng, 'Robust nonlinear control scheme for electro-hydraulic force tracking with time-varying output constraint', Symmetry, vol. 13, no. 11, Art. no. 2074, 2021.

[7] M. Yuan, Z. Chen, B. Yao, X. Liu, 'Fast and accurate motion tracking of a linear motor system under kinematic and dynamic constraints: An integrated planning and control approach', IEEE Trans. Control Syst. Technol., vol. 29, no. 2, pp. 804–811, 2021.

[8] C. Guarino Lo Bianco, F. M. Wahl, 'A novel second-order filter for real-time trajectory scaling', Proc. IEEE Int. Conf. Robot. Autom. (ICRA), pp. 5813–5818, 2011.

[9] B. Helian, Z. Chen, B. Yao, et al., 'Accurate motion control of a direct-drive hydraulic system with an adaptive nonlinear pump-flow compensation', IEEE/ASME Trans. Mechatronics, vol. 26, no. 5, pp. 2593–2603, 2021.

[10] J. Yao, W. Deng, 'Active disturbance rejection adaptive control of hydraulic servo systems', IEEE Trans. Ind. Electron., vol. 64, no. 10, pp. 8023–8032, 2017.

[11] A. R. Plummer, N. D. Vaughan, 'Robust adaptive control for hydraulic servosystems', ASME J. Dyn. Syst., Meas., Control, vol. 118, no. 2, pp. 237–244, 1996.

[12] M. Rigotti-Thompson, M. Torres-Torriti, F. Auat Cheein, G. Troni, 'H∞-based terrain disturbance rejection for hydraulically actuated mobile manipulators with a non-rigid link', IEEE/ASME Trans. Mechatronics, vol. 25, no. 6, pp. 2686–2697, 2020.

[13] J. Yao, W. Deng, Z. Jiao, 'Adaptive control of hydraulic actuators with LuGre-model-based friction compensation', IEEE Trans. Ind. Electron., vol. 62, no. 10, pp. 6469–6477, 2015.

[14] J. Koivumäki, J. Mattila, 'Stability-guaranteed impedance control of hydraulic robotic manipulators', IEEE/ASME Trans. Mechatronics, vol. 22, no. 2, pp. 601–612, 2017.

[15] B. Yao, F. Bu, J. Reedy, G. T. C. Chiu, ‘Adaptive robust motion control of single-rod hydraulic actuators: Theory and experiments’, IEEE/ASME Trans. Mechatronics, vol. 5, no. 1, pp. 79–91, 2000.
[16] B. Helian, P. Mustalahti, J. Mattila, et al., ‘Adaptive robust pressure control of variable-displacement axial-piston pumps with a modified reduced-order dynamic model’, Mechatronics, vol. 87, Art. no. 102879, 2022.
[17] B. Helian, X. Liu, Z. Liu, M. Geimer, ‘Reinforcement learning-based autonomous control methodology of hydraulic excavators’, Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), pp. 4881–4888, 2025.
[18] Quimo GmbH, ‘Technical highlights’, https://www.quimo-gmbh.de/technology/, 2025.

Biographies



Bobo Helian received the Ph.D. degree in mechatronic engineering from Zhejiang University, Hangzhou, China, in 2021. Since 2021, he has been a Postdoctoral Researcher with the Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. His research interests include advanced controls of electrohydraulic systems (e.g., EHA motion control, trajectory planning, constrained control, and adaptive robust control) and intelligent perception of mobile working machines (e.g., machine learning and computer vision).



Christoph Keller received the M. Sc. degree in mechanical engineering from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2023. Since 2024, he has been a Ph.D. student with the institute of Mobile Machines (Mobima), KIT, Karlsruhe, Germany. His research interests include the simulation of mobile working machines (e.g. simulation with lumped parameters, co-simulation) and the product development of mobile working machines.



Marcus Geimer received the Diploma degree in mechanical engineering from the RWTH Aachen University, Aachen, Germany, in 1990, and the Ph.D. degree from the Institute of Hydraulics and Pneumatics, today named the Institute for Fluid Power Drives and Systems, RWTH Aachen University. Since 2005, he has been a Full Professor and the Director of the Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. He started his industrial career in 1995 in the field of construction at the company Krupp Berco Bautechnik GmbH, Essen, Germany, where he was the Leader of the research group for hydraulic breakers. In 2000, he was with Bucher Hydraulics GmbH, Klettgau, Germany, where he led the construction and customer development for mobile hydraulics.

Exploratory Analysis of Optimisation Results for the Multi-Pump System Using Data-Driven Methods

Artur Tozzi de Cantuaria Gama, Samuel Kärnell, Liselott Ericson

Linköping University, Department of Management and Engineering, Division of Fluid and Mechatronic Systems, Linköping, Sweden, artur.tozzi@liu.se

Abstract.

Reduced climate impact is a major topic in the mobile machinery industry, and electrification is one of the ways to achieve that. Electrified vehicles provide new opportunities to improve the hydraulic system design to achieve higher efficiency. A Multi-Pump System (MPS) approach has been proposed, that provides connection paths between any hydraulic machine port to any actuator chamber through a manifold of on/off valves. This flexibility provides multiple alternative ways to control the system that are challenging to describe intuitively, specially when multiple actuators are considered. This paper presents an exploration of different machine learning algorithms for data analyses applied to the output of control optimisation of the MPS. The results show that dimensionality reduction might not be enough to provide a visualisation approach with clear borders between operating modes for the design of control rules. A supervised classification approach called k-nearest neighbours (kNN) is then trained on the optimisation results to predict the hydraulic machines operating modes based on the actuators' states. Using part of the data for training and testing allows a measurement of the accuracy of this approach. The accuracy of the algorithm varies between 53 % and 96 %, depending on the parameters set for the kNN, greatly reducing the number of options to consider for the design of a control system.

Keywords. Multi-pump system, data-driven analysis, machine learning, dimensionality reduction, classification.

1. INTRODUCTION

The global trend of electrification and digitalisation of hydraulic systems brings opportunities for novel subsystems architectures. Electric vehicles provide new degrees of freedom for the design and control of these systems and also require more efficient solutions to minimise the effects of the overall lower energy density compared to fossil fuel alternatives.

Mobile machinery hydraulic systems have traditionally relied on centralised pump architectures, where multiple actuators share flow through proportional valves. When several actuators operate simultaneously, significant energy losses occur due to throttling and flow sharing [1]. Consequently, conventional valve-controlled systems exhibit low overall efficiency. In contrast, a comparative study between a conventional system and a self-contained electro-hydraulic cylinder demonstrated an improvement in efficiency from 22 % to 57 % [2].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

An electrified vehicle can adopt similar solutions to those applied to combustion engines, such as displacement control [3, 4]. The disadvantage is that the hydraulic machines share a single shaft, limiting their operation to the same angular speed, thus variable displacement is needed. Electro-hydraulic actuators (EHAs) aim to decentralise the hydraulic system and eliminating flow sharing between actuators can improve efficiency [2, 5, 6]. The downside is the increase in total installed power, as each EHA must be sized to provide the maximum corresponding actuator power.

Common- or multi-pressure rail solutions can also be employed [7–9], but they typically still required some level of throttling for finer control, although a more recent proposal aims to achieve varying pressure levels [10]. Finally, a novel solution proposes a network of variable-speed drives to control the hydraulic machines for multi-actuator operation and to remove the directional valves [11–13], although sizing components and defining the ideal short-circuits between actuators' chambers is non-trivial.

This paper focuses on another proposal called the Multi-Pump System (MPS) [14]. The system aims to minimise throttling losses by using only on/off valves and focuses the control system on the variable-speed, fixed-displacement, 4-quadrant operation pump/motors. Since the hydraulic machines can be shared between actuators and recuperation and regeneration are available, the total installed power can be reduced.

Controlling such a system is challenging, as there are many options to achieve the same behaviour. An optimisation strategy available in [14, 15] determined the operating conditions that minimise power consumption for a system with a single actuator. The results can be used to define the optimal operating conditions for the hydraulic machines based on the actuator's speed and force. Visualising the data and extracting control rules for this system is intuitive [16], but each additional actuator doubles the number of dimensions to consider.

1.1. *High-Dimensional Control Challenges in the Multi-Pump System*

Figure 1.1 shows an example of the MPS, referred to as 3P2A, with three pump/motors (3P), two actuators (2A), and a constant pressure source for the low-pressure line. The valve manifold allows any pump/motor port to connect to any actuator chamber. By sharing the hydraulic machines between the functions, the overall installed power can be reduced compared to alternatives such as individual electro-hydraulic actuators. The system design facilitates expansion, meaning that more actuators or hydraulic machines can be considered, which only requires adding a row or column of valves, respectively, for each new component [14].

One drawback of this solution is that the control strategy is tightly coupled to the operating state of each system component. The number of active hydraulic machines and their operating modes (pump or motor) depend on the combined operating quadrants of the actuators throughout the drive cycle. In theory, flow distribution among the hydraulic machines can be achieved in infinitely many ways, and should account for energy recuperation and regeneration, leading to a large number of possible configurations.

Previous research discussed how an optimisation approach can be used to explore the control variables design space [14, 15]. Based on the actuator(s) operating points, the algorithm can identify what are the optimal ways to distribute the flow throughout the system during steady-state operation to minimise power consumption. The goal is to identify unusual valve connections and define transition boundaries that determine when to activate the hydraulic machines and how to distribute the flow between them.

The design space is reduced, but remains high-dimensional and difficult to interpret. As a result, direct visualisation of the results might not be possible, and more structured analyses are required. This paper presents an exploratory application of established machine learning techniques for high-dimensional data analysis to use the optimisation results.

The characteristics and algorithms of the applied methods are not discussed in detail, as they are will established in the literature. Mathematical descriptions and examples are provided in [17, 18], while practical implementations in *Python* are presented in [19]. The methods used

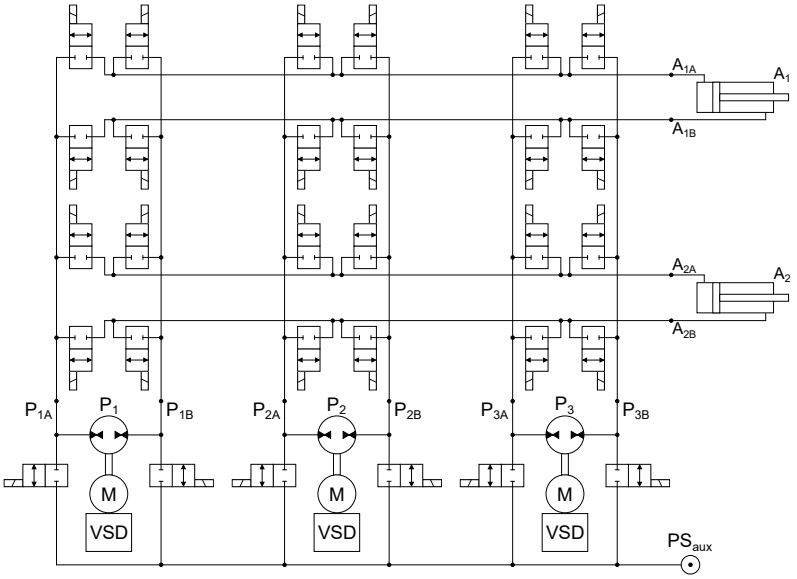


Figure 1.1: A 3P2A Multi-Pump System with three pump/motors, two actuators and a constant auxiliary pressure line.

in this work are available in the *Scikit-learn* library [20], and the following descriptions are based on these references unless stated otherwise.

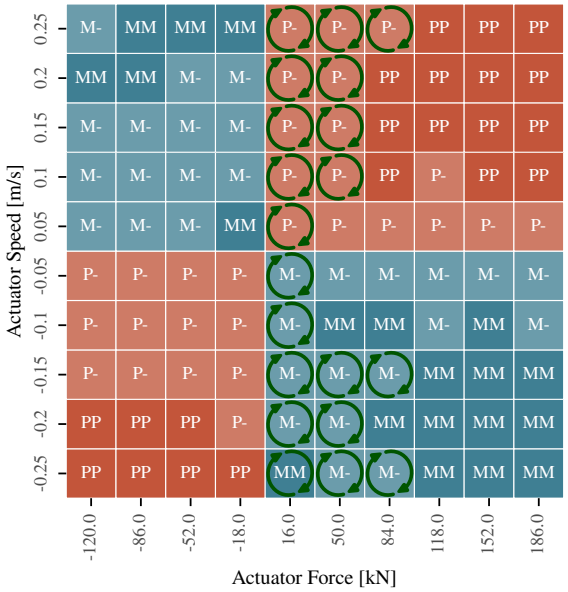
2. Problem Statement and Motivation for a Machine Learning Approach

For a system with a single actuator, the transition boundaries between operating modes are straightforward to define due to the low-dimensional operating space, as illustrated in fig. 2.1a for a 2P1A system. The axes represent the desired actuator speed and force–which can be determined from the chamber pressures and the corresponding actuator areas (piston- and rod-side).

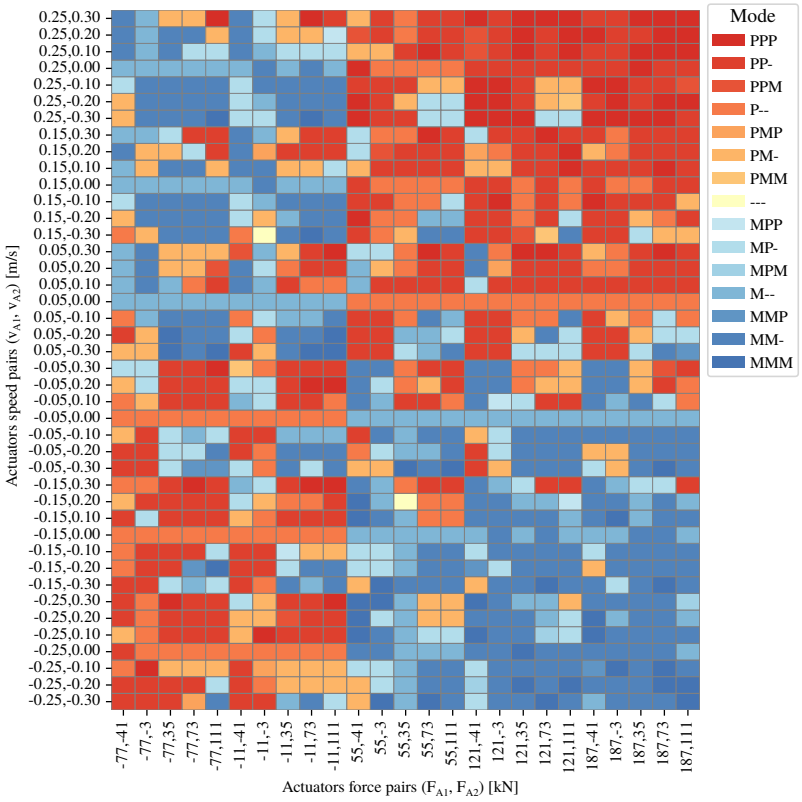
Each speed-force combination corresponds to an optimal operating mode for the two hydraulic machines, found by the optimisation algorithm. The two letters shown in each cell refer to pump/motors P_1 and P_2, respectively: "P" indicates pump mode, "M" is motor mode, and "-" an inactive machine. The green arrows indicate regenerative mode, when the actuator chambers are short-circuited to reduce the required pump/motor flow. The resulting division enables the design of a mode-based control strategy, where control decisions depend on the region in which the operating point lies.

Figure 1.1 illustrates a system with an additional actuator and one more pump/motor, a 3P2A configuration. To maintain a high degree of freedom for control, the pump/motors are sized so that two can provide the maximum combined actuators' flow. The third machine, of identical size, provides redundancy and allows the system to achieve high actuator velocities without requiring all three machines to operate at high angular speeds.

The additional actuator introduces two extra dimensions to the input space. To reduce the visualisation to two dimensions, a representation similar to Fig. 2.1b can be used. In this graph, the x-axis represents pairs of actuator speed values: the first actuator's speed is held constant while the second actuator's speed is varied across its discrete values. Once all values for the second actuator have been covered, the first actuator's speed is incremented to the next grid level, and the process repeats. The same approach is applied to the force values.



(a) 2P1A system [16].



(b) 3P2A system [14].

Figure 2.1: Optimal operating modes for two MPSs architectures: (a) 2P1A and (b) 3P2A systems.

The optimisation can assign any of the pump/motors to operate at each evaluated point, which may result in functionally redundant solutions—for example, the operating modes `P--`, `-P-`, and `--P` all represent the same system behaviour. To address this, the solutions are reorganised such that the first character always corresponds to the pump/motor with the highest absolute angular speed, which can be assumed to be P_1 in Fig. 1.1. For instance, a mode like "-PM" would be reordered as "MP-", indicating that the machine with the highest angular speed operates as a motor, the second as a pump, and the third is inactive.

The previous step reduces the number of combined operating modes to consider, but defining transition criteria from the resulting graph remains challenging. While the corner regions are relatively easy to interpret, the intermediate areas exhibit more complex and irregular boundaries. This lack of clear separation makes it difficult to extract consistent rules solely from the graph. Alternative arrangements of the axes values were tested, but they did not lead to significant improvements in interpretability.

Hydraulic systems for mobile machinery often have three or more actuators, increasing the challenge of defining the transition criteria. It is thus interesting to apply methods that can adapt to a varying number of input variables, or for the purposes of using machine learning terminology, the features or dimensions of the problem.

2.1. *Selected Machine Learning Methods*

Different machine learning algorithms could be employed to undertake this problem, which depends on what the aim is and what information one wants to employ for the analysis. The proposal for this paper can be summarised in the following questions:

- Which machine learning methods should be employed for the MPS problem based on the optimisation results?
- Assuming a system plant with limited sensors and a direct application to a control system using only the measurable variables such as actuator speed and pressure levels:
  - Is it possible to summarise the available information into a bi-dimensional representation that can be used as a guideline for the definition of rules and transition regions for a control system?
  - Can any method produce a model from the optimisation data available that can take the necessary control decisions with satisfactory precision?

The results presented in this paper therefore focus on two complementary approaches. The first considers dimensionality reduction approaches, which can be used for visualisation of patterns in the data in lower dimension or to produce a simplified model of the system. The second approach relies on supervised classification algorithms. As the operating modes are already defined by the optimisation procedure, they can be treated as class labels, enabling the algorithms to categorise the data based on the available input variables.

The following sections present results for the two approaches. The algorithms applied here were selected for their illustrative value and are not necessarily optimal, as the best choice depends on the specific problem context. This paper focuses on the dimensionality reduction techniques Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), as well as the supervised classification algorithm *k*-nearest neighbours (kNN).

2.2. *The Optimisation Results*

The Genetic Algorithm (GA) employed to produce the data used in this analysis is described in [15]. A similar process is done for the 3P2A system with varying actuator speeds and force levels. The goal of the optimisation is to minimise power consumption at steady-state, thus transitions between operating modes and valve switching are not accounted for.

The optimisation is performed six times for each combined operating point of the actuators in order to limit the computational cost. The reason is that the GA is used as a stochastic,

explorative method and does not guarantee global optimality. Instead, this optimisation problem can exhibit multiple local minima with similar performance. From a control perspective, having multiple feasible solutions is advantageous, as it provides flexibility when selecting operating modes and transitioning between configurations. The operating modes presented in fig. 2.1b refer to the best solution of the six runs.

The combined actuators' operating points, as shown in fig. 2.1b produced a total of 1 050 points, thus totalling 6 300 optimisation runs. The values of all system variables are available for analysis. Around 2,6 % of these points are discarded as they converged to undesired operation conditions, e.g. with too large error in the actuators' speeds. Thus, the following methods are applied to a data set with a total of 6 134 points.

The 3P2A has at least four features, corresponding to the force and speed of each actuator, but more could be employed. Assuming that in a real system only pressure levels can be measured, the following results use the actuators' reference speed values and pressures on each chamber as the features for analysis. The aim is to evaluate whether this feature set is sufficient to support control decisions at pump/motors level.

3. Dimensionality Reduction for Visualisation

The main idea with dimensionality reduction is to represent a high-dimensional data set into fewer dimensions. Some methods focus on identifying redundant aspects of the data set that have minimal effect on the distribution and thus could be ignored without high loss of information. Figure 3.1 summarises the process employed in this work.

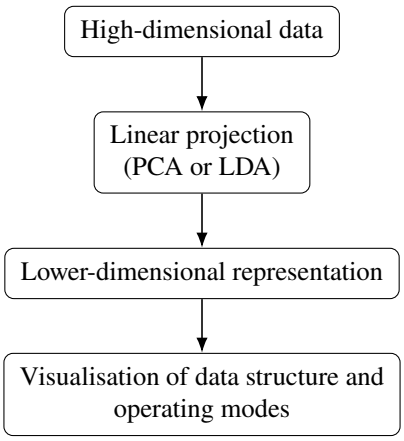
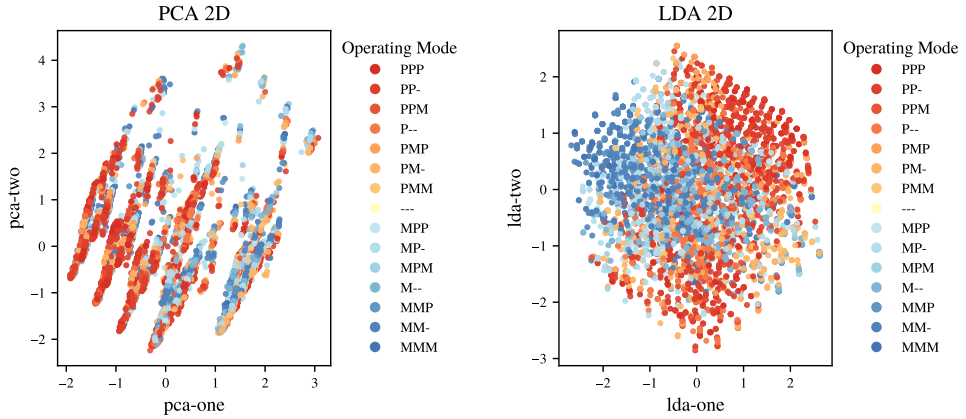


Figure 3.1: Overview of the dimensionality reduction workflow using PCA and LDA.

PCA and LDA are both dimensionality reduction techniques. They create new variables—linear combinations of the original features—that aim to represent the data in fewer dimensions. PCA is an unsupervised method that identifies directions of maximum variance in the data, thereby spreading the data points as widely as possible. LDA, on the other hand, is a supervised technique that uses known class labels to find directions that maximise the separation between classes. Although LDA can also be used for classification, it is often use for dimensionality reduction in a way that preserves the separation.

As these are linear methods, it is also possible to determine the features values from a data point in the lower dimension space. This could be used to determine the operating modes for a point that was not evaluated during the optimisation. Their downside is that the linear approach may not capture the non-linear characteristics of the system, thus underperforming regarding the data separation. Nevertheless, they can provide simple models from a data set with a large number of inputs.

Figure 3.2 shows the two-dimensional projections obtained by using the first two components as the x- and y-axes for both PCA and LDA. In both cases, the first components describe the dominant structure of the data, which results in the most distinct separation. Figures 3.2a and 3.2b show the resulting data clustering for PCA and LDA, respectively.



(a) Principal Component Analysis (PCA).

(b) Linear Discriminant Analysis (LDA).

Figure 3.2: Visualisation of the first two components for the dimensionality reduction algorithms.

From these results, limited information can be extracted from visualisation alone, although some initial clustering can be seen. Table 3.1 shows the explained variance (*PCA Var. Ratio*) and cumulative explained variance (*PCA Cum. Var.*) for each PCA component, as well as the equivalent discriminant ratios (*LDA Discr. Ratio*) and cumulative discriminant ratio (*LDA Cum. Discr.*) for LDA. There are a total of six features, the actuators' speeds and the pressure levels in each chamber. Accordingly, PCA can yield up to six components, while LDA produces at most one fewer discriminant component than the number of classes.

Table 3.1: Explained variance ratios for PCA and discriminant ratios for LDA, including cumulative values.

Component	*PCA Var. Ratio*	*PCA Cum. Var.*	*LDA Discr. Ratio*	*LDA Cum. Discr.*
1	0,22	0,22	0,77	0,77
2	0,19	0,42	0,13	0,90
3	0,19	0,60	0,06	0,96
4	0,17	0,77	0,02	0,98
5	0,17	0,93	0,02	1,00
6	0,07	1,00	-	-

The first two PCA components explain just 42 % of the total variance in the data, indicating that a two-dimensional projection provides a limited representation of the data, as in fig. 3.2a. LDA, on the other hand, captures 90 % of the cumulative discriminant ratio with the first two components; however, this does not translate into clear visual separation due to the large number of operating-mode classes and their possible overlap.

Overall, neither method yields a two-dimensional representation with clear separation between operating modes. This might be expected, as the analysis attempts to determine the pump/motor operating modes based only on actuators information, thus neglecting non-linear behaviour and discontinuous characteristics of the system.

4. Supervised Classification with K-Nearest Neighbours

An alternative approach is to use the operating modes obtained from the optimisation to train supervised classification methods. Once trained, these models can predict the pump/motor operating modes (classes) based only on the input variables provided during training.

The k-nearest neighbours (kNN) algorithm is selected due to its simplicity and intuitive interpretation. For a given input point, the algorithm identifies the k closest points in the training data and assigns a class based on a majority vote. Two weighting strategies are considered: uniform weighting, where all neighbours contribute equally, and distance-based weighting, where closer neighbours have a stronger influence. Distances are computed in the multi-dimensional feature space, allowing the method to incorporate multiple input features simultaneously. Figure 4.1 presents the steps employed in this paper, and section 4.1 expands on these steps.

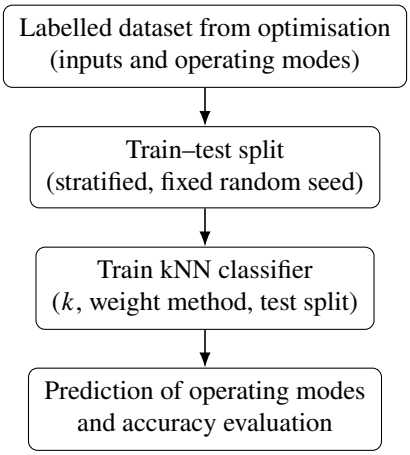


Figure 4.1: Overview of the supervised classification workflow using kNN.

4.1. *Accuracy Evaluation and Hyperparameters*

The classification accuracy is evaluated by dividing the dataset into separate training and test sets. The trained model is applied to the inputs of the test set, and the predicted operating modes are compared with those obtained originally from the optimisation.

To ensure that that all operating modes are accounted for during training, a stratification approach is used to ensure that both training and test sets contain examples of each class. Therefore, the data split is generated through a randomised procedure with stratification, and a fixed random seed is used to provide the same train-test separation when comparing results with different hyperparameters.

The parameters used and varied for the kNN classifier are:

- **Number of neighbours:** $k = 5$.
- **Weighting method:** uniform or distance.
- **Test set fraction:** 10 %, 20 %, 30 %.

The small number of neighbours was chosen to prioritise local information. Since several operating modes occur infrequently in the dataset, using a small k reduces the influence of frequent classes and distant samples, thus mitigating misclassification of rare operating modes.

4.2. *Classification Results*

Table 4.1 shows the accuracy for different evaluations with kNN using the unmodified optimisation data. One characteristic of the MPS is that it might have multiple viable solutions for each combined operating point of the actuators. In the table, the *Exact* column refers to the kNN accuracy in finding the exact operating mode determined by the optimisation runs in the test set.

The results are also presented for the training and test sets. Using the same training set to evaluate the kNN serves as a sanity check on the model memorisation. Generalisation is assessed from the test set.

Table 4.1: Comparison of kNN classification accuracies (%) for different test set sizes and weighting methods. Exact = strict accuracy, Accept. = acceptable accuracy (any valid mode for same operating point), Top-3 = accuracy within three most probable modes.

Split	**Uniform weights**			**Distance weights**		
	Exact	*Accept.*	*Top-3*	*Exact*	*Accept.*	*Top-3*
Training set						
10 %	68,3	95,2	99,8	100,0	100,0	100,0
20 %	67,6	94,3	99,8	100,0	100,0	100,0
30 %	67,5	93,2	99,7	100,0	100,0	100,0
Test set						
10 %	54,2	90,2	98,5	58,3	94,6	98,5
20 %	54,0	89,2	98,7	58,2	95,3	99,0
30 %	53,2	86,6	98,9	57,9	94,9	99,2

In this context, the distance-weighted approach may not be ideal. If two optimisation runs converge to the same mode for a given operating point of the actuators, the features used for training may be identical due to model simplifications. Consequently, the test point for the kNN algorithm may exactly match one of the training points, resulting in nearly zero distance, as the points coincide in the feature space. This explains the 100 % accuracy for the training set in tab. 4.1.

For the *Exact* columns, the accuracy varies between 53,2 %-68,3 %, which could be misleadingly low. Since multiple solutions are viable, a more reasonable approach would be to evaluate if the kNN determines any of the operating modes that the optimisation found for the corresponding operating point of the actuators. These are the values in the *Accept.* columns, increasing the accuracy to 86,6 %-95,3 %. This reflects the case where multiple operating modes are interchangeable from a control perspective.

The kNN output is also not a single value, but rather a probability for each of the possible classes. For a given actuator operating point, the highest probability mode from the kNN may not match the one found by the optimisation, but the second or third most likely class might. This is shown in the *Top-3* column, which verifies whether any of the most probable classes selected by the kNN corresponds to the optimisation result, moving the accuracy to 98,5 %-99,8 %.

5. DISCUSSION

The kNN classifier is a viable approach for predicting the operating modes of hydraulic machines. Using only reference speeds and chamber pressure values—which can be available in real applications—the algorithm can estimate the hydraulic machines' operating modes with acceptable accuracy when appropriate considerations are taken into account.

Using the ranked output of the algorithm, based on neighbour votes, can also be helpful for the control design. It greatly reduces the number of options to consider, making it possible to efficiently evaluate which of the options would lead to fewer transition losses.

A current limitation of this approach is that the hydraulic machines operating modes do not cover short circuiting actuator chambers. In a dynamic system, if the external force changes, the pressures will vary according to the current valve combination. The kNN would then provide the operating mode that best fits the current valve setting.

Future work could therefore perform a similar analysis but focused on the valve signals, define operating modes for the actuators, or determine short-circuited chambers from the optimisation results. A separate kNN, or other algorithm, could then be employed to find the viable system connections.

The dimensionality-reduction methods were also evaluated as potential preprocessing steps. While PCA maximises variance and LDA maximises class separability assuming linear projections, neither method improved the separability of operating modes in lower-dimensional spaces for the kNN. This suggests that non-linear relationships in the system play a significant role in determining the operating modes.

The next step is to apply the proposed classification to a dynamic model to evaluate how well the control system will perform. This includes setting criteria to select the operating mode based on neighbour vote confidence of the kNN and evaluating the model's ability to interpolate operating-mode decisions for actuator conditions not covered during training.

5.1. *Other Machine Learning Methods and Possible Applications*

Other machine learning methods exist but were not considered in this work. Unsupervised clustering and classification are not suitable in this case, as the classes-corresponding to operating modes- are already defined based on the torque and angular speed values obtained from the optimisation.

Neural networks and reinforcement learning are also commonly used approaches but were outside the scope of this study. These methods typically require large datasets for training, and parameter tuning is challenging as they are, in a way, complex optimisation problems. One of the objectives of this work was to identify a compromise between model capability and practical application, favouring methods that are easier to implement while still capable of supporting the control decision process.

Regression methods were also considered but not explored in this paper. No satisfactory compromise was found for the development of a simple yet reliable model based on limited inputs. A possible explanation is that information from the actuators is transmitted to the pump/motors through the valve manifold, and relevant details are lost when valve states are not included. This does not mean that the approaches are not viable, but rather that a more detailed analysis is required.

6. CONCLUSION

A Multi-Pump System poses a combinatorial control problem due to the large number of possible connections between hydraulic machines and actuators. While optimisation can identify power-efficient operating modes, the direct use of optimisation results in a control strategy is impractical.

This work showed that linear dimensionality-reduction techniques such as PCA and LDA are insufficient to achieve a clear separation of operating modes for a 3P2A MPS, due to the non-linear characteristics of the system. This motivated the use of a supervised classification approach. A kNN classifier was therefore applied to map operating conditions to operating modes. The achieved accuracy ranged from 53.2 % to 99.5 %, depending on whether exact operating mode prediction or the identification of a feasible alternative was required.

The results show that, although low-dimensional projections are insufficient for direct interpretation, supervised classification can be used as a tool to analyse and structure optimisation outcomes. Further work is required to assess the robustness of this approach for operating points outside the training data and to evaluate its relevance when applied in a

dynamic or control-oriented context. In addition, the current approach does not address other aspects of the control process, such as valve configurations or the conditions to short-circuit actuator chambers, which might be necessary for a effective design of the control system.

7. REFERENCES

[1] L. Lynch and B. T. Zigler, 'Estimating Energy Consumption of Mobile Fluid Power in the United States', en, Tech. Rep. NREL/TP–5400-70240, 1408087, Nov. 2017. DOI: 10.2172/1408087.

[2] D. Hagen, D. Padovani, and M. Choux, 'A Comparison Study of a Novel Self-Contained Electro-Hydraulic Cylinder versus a Conventional Valve-Controlled Actuator—Part 2: Energy Efficiency', en, Actuators, vol. 8, no. 4, p. 78, Dec. 2019, ISSN: 2076-0825. DOI: 10.3390/act8040078.

[3] T. Patel, L. D. S. Franquilino, A. Vacca, and C. Young, 'Comparison Study of Fully Individualized System Architectures for Electrified Mini-Excavators: Displacement Control (DC) vs. Electro-Hydraulic Actuation (EHA)', en, in 14th International Fluid Power Conference, Dresden, Germany: River Publishers, 2024, pp. 22–36, ISBN: 978-87-7004-222-2. DOI: 10.13052/rp-9788770042222C02.

[4] D. Fassbender, C. Brach, and T. Minav, 'Energy-Efficiency Comparison of Different Implement Powertrain Concepts to Each Other and Between Different Heavy-Duty Mobile Machines', in International Journal of Fluid Power, Jul. 2024, pp. 127–144. DOI: 10.13052/ijfp1439-9776.2521.

[5] F. Wang, J. Wu, Z. Lin, H. Zhang, and B. Xu, 'A power-sharing electro-hydraulic actuator system to downsize electric motors for electric mobile machines', Energy, vol. 284, p. 129 343, Dec. 2023, ISSN: 0360-5442. DOI: 10.1016/j.energy.2023.129343.

[6] P. Casoli, F. Scolari, C. M. Vescovini, and M. Rundo, 'Energy Comparison between a Load Sensing System and Electro-Hydraulic Solutions Applied to a 9-Ton Excavator', en, Energies, vol. 15, no. 7, p. 2583, Jan. 2022, Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1996-1073. DOI: 10.3390/en15072583.

[7] K. Heybroek, 'On Energy Efficient Mobile Hydraulic Systems : with Focus on Linear Actuation' (Linköping Studies in Science and Technology. Dissertations), en. Linköping: Linköping University Electronic Press, Nov. 2017, vol. 1857, ISBN: 978-91-7685-511-9. DOI: 10.3384/diss.diva-142326.

[8] Z. Xu, M. Bertolin, A. Vacca, and J. Nilsson, 'A Hydraulic Architecture Based on Multi-common Pressure Rail Principle Using Multi-chamber Cylinders for Excavators', en, in Advancements in Fluid Power Technology: Sustainability, Electrification, and Digitalization, L. Ericson and P. Krus, Eds., Series Title: Lecture Notes in Mechanical Engineering, Cham: Springer Nature Switzerland, 2025, pp. 339–356, ISBN: 978-3-031-84504-8 978-3-031-84505-5. DOI: 10.1007/978-3-031-84505-5_22.

[9] J. Wills and P. Y. Li, 'Comparison of Real-Time Pressure Rail Selection Algorithms for the Hybrid Hydraulic Electric Architecture: Case Study on a Track Loader', in BATH/ASME 2024 Symposium on Fluid Power and Motion Control, Bath, United Kingdom: American Society of Mechanical Engineers, Sep. 2024, V001T01A027, ISBN: 978-0-7918-8819-3. DOI: 10.1115/FPMC2024-140443.

[10] X. Guo, J. Lengacher, and A. Vacca, 'A Variable Pressure Multi-Pressure Rail System Design for Agricultural Applications', en, Energies, vol. 15, no. 17, p. 6173, Jan. 2022, Number: 17 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1996-1073. DOI: 10.3390/en15176173.

[11] L. Schmidt and K. V. Hansen, 'Electro-Hydraulic Variable-Speed Drive Networks—Idea, Perspectives, and Energy Saving Potentials', en, Energies, vol. 15, no. 3, p. 1228, Feb. 2022, ISSN: 1996-1073. DOI: 10.3390/en15031228.

[12] L. Schmidt, M. v. Binsbergen-Galán, R. Knöll, M. Riedmann, B. Schneider, and E. Heemskerk, 'Energy Efficient Excavator Implement by Electro-Hydraulic/Mechanical Drive Network', en, International Journal of Fluid Power, pp. 413–438, Dec. 2024, ISSN: 2332-1180. DOI: 10.13052/ijfp1439-9776.2541.

[13] M. van Binsbergen-Galán, B. Videbæk, K. Vorbøl Hansen, and L. Schmidt, 'Experimental Investigation of Hydraulic Power Sharing Potential in a Dual Cylinder Electro-Hydraulic Variable-Speed Drive Network', en, American Society of Mechanical Engineers Digital Collection, Oct. 2024. DOI: 10.1115/FPMC2024-140322.

[14] A. Tozzi De Cantuaria Gama, 'Multi-Pump Systems for Electrified Mobile Machinery : Addressing Combinatorial Control Complexity through Simulation-Based Optimisation' (Linköping Studies in Science and Technology. Licentiate Thesis). Linköping: Linköping University Electronic Press, Dec. 12, 2025, vol. 2022, 64 pp., ISBN: 978-91-8118-303-0. DOI: 10.3384/9789181183047.

[15] A. Tozzi de Cantuaria Gama, S. Kärnell, and L. Ericson, 'The Multi-Pump System Combinatorial Problem: A Filtering Approach Using Genetic Algorithms', en, in The 19th Scandinavian International Conference on Fluid Power, SICFP'25, Linköping University, Sweden: River Publishers, 2025, ISBN: 978-87-438-0825-1. DOI: 10.13052/rp-9788743808251A33.

[16] S. Kärnell, A. Tozzi De Cantuaria Gama, and L. Ericson, 'Control of Multi-Pump Systems', in Volume 6: 19th International Conference on Micro- and Nanosystems (MNS); 21st International Conference on Multibody Systems, Nonlinear Dynamics, and Control (MSNDC); 37th Conference on Mechanical Vibration and Sound (VIB); 38th Fluid Power and Motion Control Symposium (FPMC), Anaheim, California, USA: American Society of Mechanical Engineers, Aug. 2025, V006T12A017, ISBN: 978-0-7918-8926-8. DOI: 10.1115/DETC2025-164506.

[17] S. L. Brunton and J. N. Kutz, 'Data-driven science and engineering: machine learning, dynamical systems, and control', eng. Cambridge University Press, 2019, ISBN: 978-1-108-42209-3.

[18] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. E. Taylor, 'An introduction to statistical learning: with applications in Python' (Springer texts in statistics), eng. Cham: Springer, 2023, ISBN: 978-3-031-38747-0.

[19] A. Géron, 'Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems', eng, Second edition. O'Reilly, 2019, ISBN: 978-1-4920-3261-8.

[20] F. Pedregosa et al., 'Scikit-learn: Machine learning in Python', Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. DOI: 10.48550/arXiv.1201.0490.

BIOGRAPHIES



Artur Tozzi de Cantuaria Gama. He received his M.Sc. degree in Mechanical Engineering from the Federal University of Santa Catarina (UFSC), Brazil, in 2020. He is currently pursuing doctoral studies in Fluid and Mechatronic Systems at Linköping University (LiU), Sweden. His research interests include simulation and control of multi-pump systems for electrified mobile machinery, optimisation, and machine learning.



Samuel Kärnell. He received his Ph.D. in Hydraulics at Linköping University (LiU), Sweden, in 2022. He currently works as a Assistant Professor at Fluid and Mechatronic Systems at LiU. His research interests include pump and motor design, electro-hydraulic systems, and electrification of mobile machinery.



Liselott Ericson. She received her Ph.D. in Hydraulics at Linköping University (LiU), Sweden, in 2012. She currently works as a professor at Fluid and Mechatronic Systems at LiU. Her research interests include pump and motor design, electro-hydraulic systems, modelling and simulation, and optimisation.

Analysis of High-Pressure Pneumatic Stage Separation System for Reusable Launch Vehicles Based on Real Gas Equation of State

Ruiyao Tao[1], Yaobao Yin[1,2*], Naiping Gao[1], Pengfei He[3]

[1] *School of Mechanical Engineering, Tongji University,Shanghai,201804, P.R.China*
[2] *Shanghai Key Laboratory of Wearable Robotics and Human-Robot Interaction Technology, Shanghai 201804, P.R.China*
[3] *School of Aerospace Engineering and Applied Mechanics, Shanghai, 200082, P.R.China*

Corresponding author: y-yin@tongji.edu.cn

Abstract.

Interstage separation of launch vehicles is essential for reliable payload delivery into orbit. Helium, owing to its chemical inertness and low density, is widely used as the working medium in pneumatic separation systems. While high-pressure helium deviates significantly from ideal-gas behavior, and traditional cubic equations of state (EoS) introduce cumulative errors. To overcome these issues, a new corresponding state EoS is proposed, considering intermolecular interactions and the influence of temperature on molecular forces and volume. Based on the new EoS, a transient thermodynamic model is established, with strategies developed for both open and closed systems, and applied to the analysis of interstage separation. Results indicate that within –30 °C to 120 °C and 1–35 MPa, the proposed EoS achieves a maximum absolute error below 0.82%, significantly improving computational accuracy. The transient model reproduces the thermodynamic behavior of fixed-volume open, variable-volume open, and variable-volume closed systems, with average deviations from experiments below 5.2%, confirming its validity. Under heavy-load conditions of 90 t and 120 t, the minimum supply pressure of the gas tank is determined to be 12.5 MPa to ensure a separation velocity ≥ 5 m/s while keeping the separation duration under 1000 ms. A theoretical basis and key parameters are provided for the design and optimization of interstage separation systems.

Keywords. Real-gas equation of state, Helium, Interstage separation, Launch vehicles, Transient thermodynamic modelling.

1. INTRODUCTION

During the flight of a launch vehicle, multiple separation events are required, among which the interstage separation system plays a crucial role in reducing structural mass and improving payload efficiency, as shown in Figure. 1. The primary task of this system is to achieve non-interfering separation of the expended stage within a millisecond timescale [1]. This process involves precise mechanisms that must be triggered at the exact moment while minimizing disturbances to the attitude and trajectory of vehicle.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Currently, two main technical approaches are employed for interstage separation. Traditional expendable configurations predominantly rely on pyrotechnic devices, in which energetic materials drive the separation mechanism by breaking or releasing structural connections between stages [2]. In contrast, non-pyrotechnic separation technologies significantly reduce the adverse effects of high temperature and explosive shock, offering greater flexibility and enabling safe separation under low-thermal-load conditions [3]. Typical non-pyrotechnic approaches include mechanical separation [4], electromechanical separation [5], electromagnetic separation, smart-material–driven separation [6], and pneumatic separation [7]. Among these, pneumatic separation is the most widely applied owing to its advantages of reusability, versatility, and low cost [8]. This technique relies solely on compressed gas to generate thrust, thereby enabling the relative acceleration between adjacent stages.

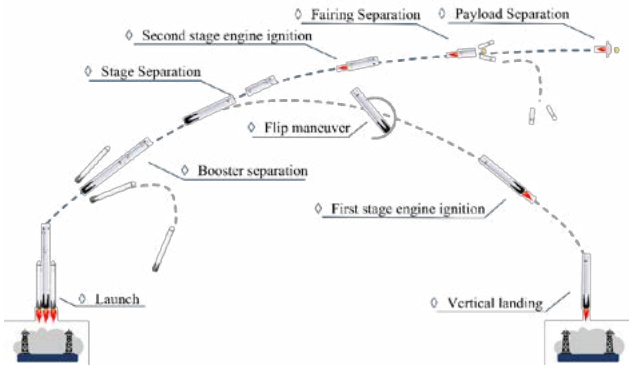


Figure 1.1 Multiple Separation Events of a Launch Vehicle

For safety considerations, helium is commonly selected as the working medium in pneumatic separation systems, primarily due to its chemical inertness and low specific gravity [9]. The molar mass of helium is only one-seventh that of air, which enables a significant reduction in the volume and weight of storage vessels while meeting system performance requirements. Consequently, the structural mass of the launch vehicle can be lowered, and the required amount of pressurant for a given structural configuration can be reduced, thereby further decreasing overall energy consumption. In high-pressure helium systems used for rocket interstage separation, the working fluid often exists under high-pressure or even supercritical conditions. Under such circumstances, the ideal gas law fails to accurately capture the thermodynamic behavior, making the development of an equation of state (EoS) capable of representing real-gas effects particularly essential [10].

Among existing formulations, cubic EoS, typified by the van der Waals (VDW) EoS [11], have attracted considerable attention in both engineering practice and scientific research. To improve accuracy in representing vapor-phase behavior, Redlich and Kwong [12] modified the temperature-dependent term of the VDW EoS. Building on this foundation, Soave [13] and Peng and Robinson [14] introduced further refinements to the temperature and volume functions, resulting in the SRK and PR EoSs, respectively, which significantly enhanced the predictive accuracy of pure substances over a wider temperature range and under vapor–liquid equilibrium conditions. Alternatively, virial-type EoS, expressed as a power series expansion, exhibit reliable predictive performance in the low-pressure region [15].

Extensions to this class of EoS have been achieved by introducing empirical parameters, as demonstrated by Benedict, Webb, and Rubin, and subsequently by Lee et al. [16], thereby broadening their applicability to multicomponent mixtures.

With the advancement of statistical mechanics and computational science, the focus of EoS research has gradually shifted toward theoretical formulations rooted in molecular interaction mechanisms. Based on perturbation theory, Chapman et al. [17] developed the Statistical Associating Fluid Theory (SAFT), which incorporates successive contributions of intermolecular interactions and provides accurate descriptions of complex fluid behavior. Building upon this framework, Alanazi et al. [18] introduced thermodynamic models that account for chain structures and molecular association phenomena, and compared them with cubic EoS for hydrogen-containing mixtures. Their results indicated that cubic EoS retain superior predictive performance for such systems. Despite their sound physical foundation, EoS derived from perturbation theory and virial coefficients are computationally intensive and exhibit limitations in predicting thermophysical properties under high-pressure conditions.

To address these issues, a new real-gas EoS is proposed by combining the advantages of cubic formulations and perturbation theory. Through the introduction of composite molecular correction factors and temperature-dependent coefficients, the accuracy of helium property prediction under high-pressure conditions is significantly improved. On this basis, a transient thermodynamic system model incorporating real-gas effects is established to capture the dynamic variations of pressure and temperature within the interstage separation chamber. Finally, the validity of the proposed model is verified through a case study of a high-pressure helium separation system for rocket interstages, providing critical theoretical insights and technical parameters for the design of aerospace separation systems.

2. Real-Gas Equation of State Applicable to High-Pressure Helium Systems

When analyzing high-pressure helium systems, the choice of an appropriate EoS is essential for accurately predicting the physical properties of the gas. Helium, an inert gas with a stable closed-shell electronic structure, is a monoatomic and nonpolar species characterized by a symmetric electron cloud. Moreover, due to its very low molecular weight, helium exhibits pronounced quantum effects and is regarded as a prototypical quantum gas.

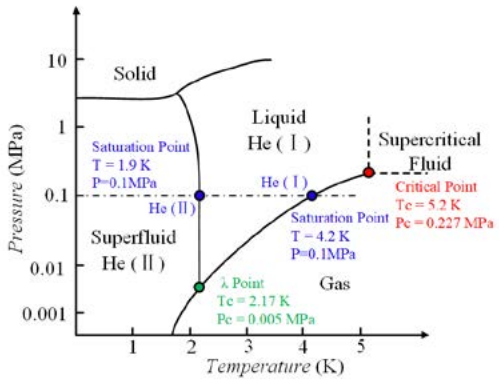


Figure 2.1. Phase Diagram of Helium

As shown in Figure. 2.1, the phase diagram of helium is unique in that it exhibits two liquid phases: the normal liquid phase (He I) and the superfluid phase (He II). Consequently, helium does not possess a conventional solid–liquid–vapor triple point. Instead, a λ-point appears at the intersection of the vapor, liquid, and superfluid phases.

The ideal-gas EoS assumes that gas molecules are point masses without volume and that no intermolecular forces exist other than elastic collisions. In reality, helium is a quantum system with intermolecular interactions, and its behavior deviates substantially from the classical ideal model of non-interacting point particles. This deviation becomes particularly significant under high-pressure conditions, where the weak but non-negligible dispersion forces arising from its nonpolarity and inertness become increasingly important, and the finite molecular volume is more pronounced. Under such circumstances, the ideal-gas EoS fails to provide accurate predictions, and the adoption of a real-gas EoS is therefore indispensable.

2.1. Real-Gas Equation of State

Common real-gas EoS can be classified from several perspectives, as summarized in Table 2.1. Virial-type EoS involve complicated mathematical expressions, multiparameter EoS are

Table 2.1. Types of Commonly Used Real Gas EoS

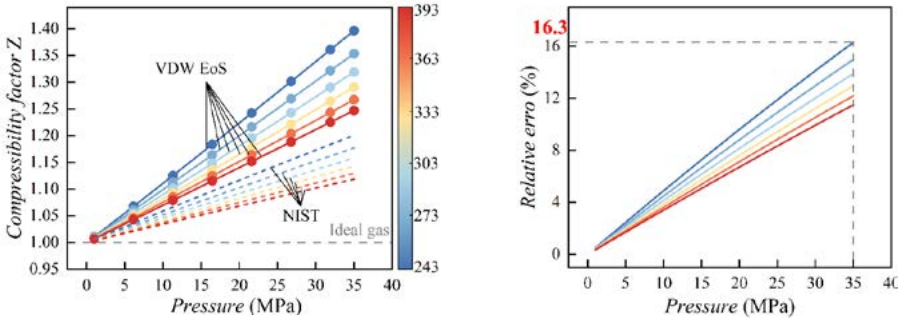
Type	Theoretical basis	Representative equations
Virial	Statistical mechanics	$p = RT\rho(1 + B\rho + C\rho^2 + D\rho^3 + \cdots)$
Cubic	Van der Waals concept	Van der Waals EoS: $p = \frac{RT}{V_m - b} - \frac{a}{V_m^2}$ Redlich-Kwong EoS: $p = \frac{RT}{V_m - b} - \frac{a}{T^{0.5}V_m(V_m + b)}$ Soave-Redlich-Kwong EoS: $p = \frac{RT}{V_m - b} - \frac{a(T)}{V_m(V_m + b)}$ Peng-Robinson EoS: $p = \frac{RT}{V_m - b} - \frac{a(T)}{V_m(V_m + b) + b(V_m - b)}$
Multiparameter	Empirical fitting	Benedict-Webb-Rubin EoS: $p = RT\rho + (B_0RT - A_0 - C_0/T^2)\rho^2 + (bRT - a)\rho^3 + a\alpha\rho^6 + C\rho^3/T^2\,(1 + \gamma\rho^2)e^{-\gamma\rho^2}$
Perturbed-chain	Statistical mechanics	Statistical associating fluid theory: $A^{total} = A^{idealgas} + A^{monomer} + A^{chain} + A^{association} + A^{other}$
Crossover	Scaling laws	Universal method [19]: $p = -\left(\frac{\partial A}{\partial v}\right)_T = \frac{RT}{v_{0c}}\left[-\frac{v_{0c}}{v_c}\left(\frac{\partial \Delta\bar{A}}{\partial \Delta\eta}\right)_T + \left(\frac{1}{b} - \frac{T_{0c}}{T}\frac{\Omega_a a_0}{Z_{0c} b_2 b_3}\right) + \frac{v_{0c}}{v_c}\left(\frac{\partial K}{\partial \Delta\eta}\right)_T\right]$

strongly dependent on experimental data, and perturbation-chain or crossover-type EoS are suitable for describing complex molecular systems. By contrast, cubic EoS require fewer parameters and involve simpler calculations, making them more applicable for analyzing high-pressure helium systems.

The compressibility factor Z is an important dimensionless parameter that characterizes the deviation of real gases from ideal-gas behavior, defined as

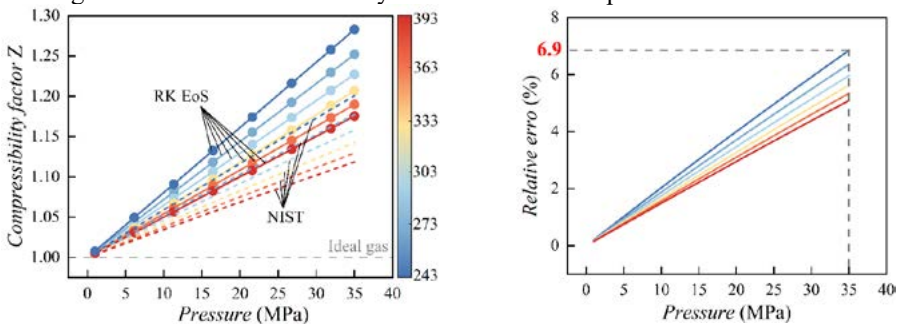
$$Z(p,T) = \frac{pV_m}{RT} \tag{2.1}$$

For an ideal gas, Z is always equal to 1. In the case of real gases, however, Z is a function of both pressure and temperature, reflecting the influence of intermolecular interactions. For helium, the compressibility factor varies with both pressure and temperature.



(a) Compressibility factor Z comparison (b) Absolute error in Z

Figure 2.2. Prediction Accuracy of VDW EoS: Comparison with NIST Data



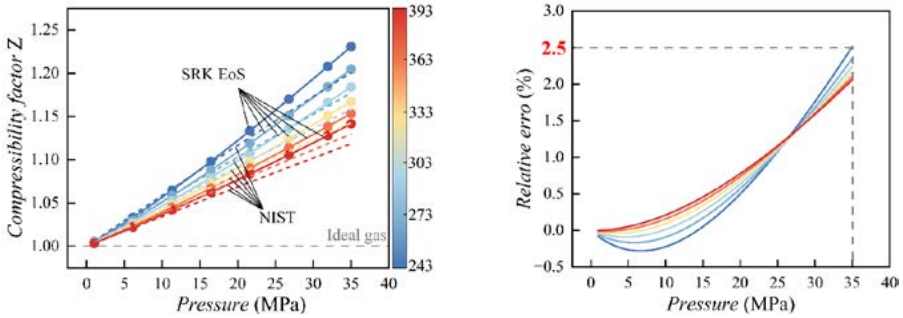
(a) Compressibility factor Z comparison (b) Absolute error in Z

Figure 2.3. Prediction Accuracy of RK EoS: Comparison with NIST Data

The variation of the compressibility factor of helium with pressure and temperature is investigated. To assess the predictive accuracy of classical cubic EoS, four representative EoSs are selected: van der Waals (VDW), Redlich–Kwong (RK), Soave–Redlich–Kwong (SRK), and Peng–Robinson (PR) EoSs. The compressibility factors of helium are calculated using these formulations and compared systematically with the reference data from the REFPROP database developed by the National Institute of Standards and Technology (NIST). The results are presented in Figures 2.2–2.5.

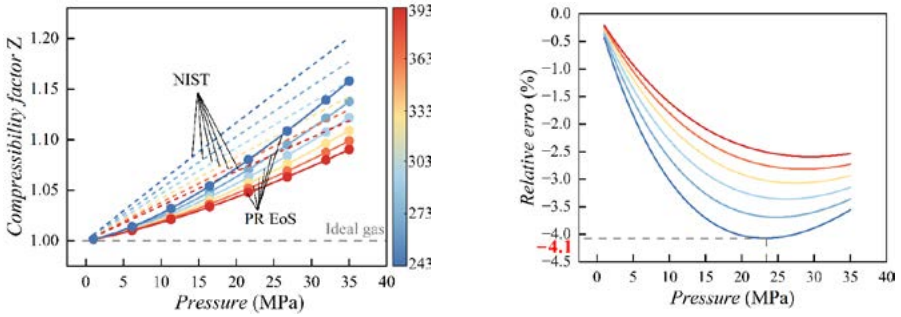
The comparison shows that within the typical operating range of high-pressure helium systems for interstage separation, with temperatures from 243 K to 393 K (−30 °C to 120

°C) and pressures from 1 MPa to 35 MPa, the VDW EoS exhibits the largest absolute error of 16.3%. The errors for the RK and PR EoSs are 6.9% and −4.1%, respectively, whereas the SRK EoS yields the smallest deviation of 2.5%. It should be noted that due to inherent assumptions and simplifications in the EoSs, systematic deviations arise in the prediction of high-pressure helium properties. Such deviations may further accumulate in practical engineering modeling. Therefore, in order to improve the accuracy of high-pressure helium system calculations, it is necessary to employ a real-gas EoS with smaller intrinsic prediction errors.



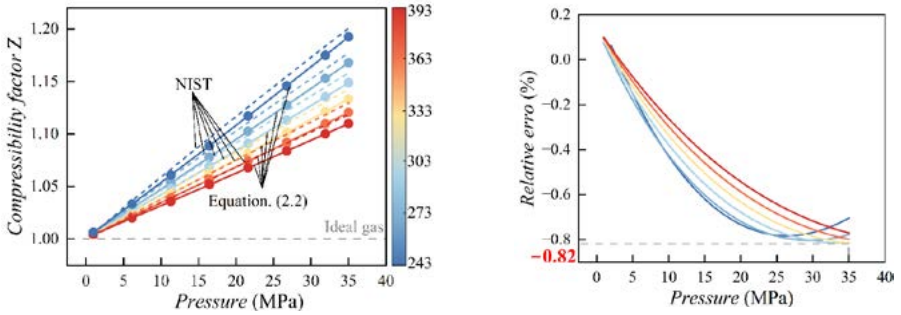
(a) Compressibility factor Z comparison (b) Absolute error in Z

Figure 2.4. Prediction Accuracy of SRK EoS: Comparison with NIST Data



(a) Compressibility factor Z comparison (b) Absolute error in Z

Figure 2.5. Prediction Accuracy of PR EoS: Comparison with NIST Data



(a) Compressibility factor Z comparison (b) Absolute error in Z

Figure 2.6. Prediction Accuracy of (2.2) EoS: Comparison with NIST Data

2.2. A New Real-Gas Corresponding State EoS

By combining the advantages of cubic EoS and perturbation theory, and considering the molecular association phenomena under high-pressure conditions, a new real-gas corresponding state EoS is developed based on statistical thermodynamics. The specific form of the equation is expressed as follows [20]:

$$p_r = \frac{8}{3}\left[\frac{T_r}{V_{mr}/\gamma - \xi/3} - \frac{9/8}{T_r^{\alpha}(V_{mr}/\gamma)^2}\right] \tag{2.2}$$

According to the principles of statistical thermodynamics, a composite molecular correction factor is introduced to represent the ratio between the actual number of molecules N_t contained in 1 kg of gas under arbitrary states and the number of molecules N_c under the critical state:

$$\gamma = \frac{N_t}{N_c} = \left(\frac{3}{8Z_c}\right) \cdot \left[1 - \left(1 - \frac{8}{3}Z_c\right) \cdot T_r^{-\Delta I/2} \cdot e^{(1/T_r - 1)} \cdot V_r^{-1}\right] \tag{2.3}$$

In addition, the effects of temperature on molecular volume and intermolecular forces are taken into account. The resulting molecular volume correction coefficient is given as:

$$\xi = \frac{b}{b_c} = \sqrt{\frac{0.5 + \sqrt{0.25 + 0.375}}{0.5 + \sqrt{0.25 + 0.375T_r}}} \tag{2.4}$$

And the intermolecular force correction coefficient is expressed as:

$$\alpha = \left[\ln\left(\frac{27}{8}\right) - \ln\sqrt{\frac{0.5 + \sqrt{0.25 + 0.375}}{0.5 + \sqrt{0.25 + 0.375\,T_B/T_c}}}\right] \Big/ \ln(T_B/T_c) - 1 \tag{2.5}$$

Within the same ranges of temperature and pressure, calculations are performed using Equation (2.2), and the predicted values are systematically compared with reference data from the NIST database. The absolute error distribution is also plotted, as shown in Figure 2.6. The results indicate that the maximum absolute error of the proposed EoS is only 0.82%, demonstrating high predictive accuracy. Furthermore, the proposed formulation is concise in form and clear in structure, which facilitates its practical application in engineering systems.

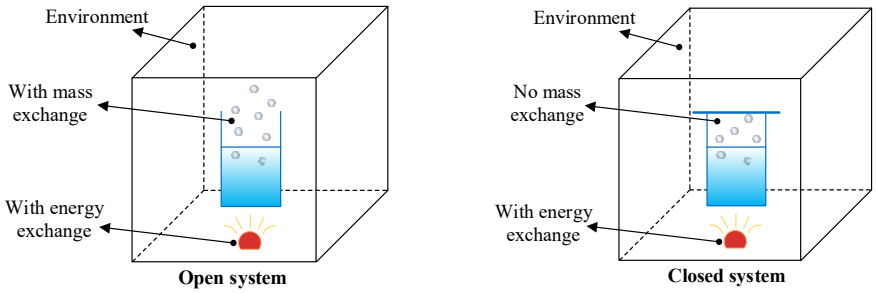
3. HIGH-PRESSURE GAS ANALYSIS METHOD BASED ON A REAL-GAS EOS

High-pressure helium thermodynamic systems can generally be categorized into open and closed systems, as illustrated in Figure 3.1. An open system allows both mass and energy exchange with the surroundings, whereas a closed system permits only energy exchange without mass transfer. When formulating the thermodynamic conservation equations for high-pressure helium systems, pressure and temperature are adopted as the primary variables to characterize mass and energy conservation. Based on the first law of thermodynamics and incorporating the real-gas equation of state, time-dependent governing equations are derived with pressure and temperature as state variables. This provides the foundation for transient analysis of the system.

3.1. Thermodynamic Model of Open System

The real-gas equation of state in Equation (2.2) can be expressed as:

$$f(m,p,V,T) = 0 \tag{3.1}$$



(a) Open system (b) Closed system

Figure 3.1. Mass and Energy Exchange in Open and Closed Thermodynamic Systems

Assuming gas pressure, mass, temperature, and volume are taken as variables, the partial differential form is obtained. Accordingly, the time-dependent variations of pressure, mass, temperature, and volume can be written as:

$$\frac{\partial f}{\partial p}\frac{\mathrm{d}p}{\mathrm{d}t} + \frac{\partial f}{\partial m}\frac{\mathrm{d}m}{\mathrm{d}t} + \frac{\partial f}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t} + \frac{\partial f}{\partial V}\frac{\mathrm{d}V}{\mathrm{d}t} = 0 \tag{3.2}$$

From this relation, the governing equation with pressure as the primary variable is derived as:

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial f}{\partial m}\frac{\mathrm{d}m}{\mathrm{d}t} - \frac{\partial f}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t} - \frac{\partial f}{\partial V}\frac{\mathrm{d}V}{\mathrm{d}t} \tag{3.3}$$



(a) Typical fixed-volume open system (b) Typical variable-volume open system

Figure 3.2. Typical Open Systems



Figure 3.3. Typical Closed Systems

The gas inside the high-pressure chamber follows the first law of thermodynamics, where the increment of system energy equals the difference between the input and output energy. This is expressed as:

$$\mathrm{d}U + \delta E_k + \delta E_p = h_{in}\mathrm{d}m_{in} - h_{out}\mathrm{d}m_{out} + \delta Q - \delta W \tag{3.4}$$

Combined with the second law of thermodynamics, the internal energy can be formulated as a function of any two independent variables among T, p, and v. Expressed in terms of T and v., the first internal energy equation is given by:

$$\mathrm{d}u = c_v \mathrm{d}T + \left[T\left(\frac{\partial p}{\partial T}\right) - p\right] dv \tag{3.5}$$

For fixed-volume open systems such as gas cylinders, shown in Figure 3.2(a), the system performs no external work, and only a single inlet or outlet exists for the gas. Neglecting changes in kinetic and potential energy, the first law of thermodynamics for a fixed-volume open system is expressed as:

$$u\mathrm{d}m + m\mathrm{d}u = (u + pv)\mathrm{d}m + \delta Q \tag{3.6}$$

By substituting Equation (3.5) into Equation (3.6), the time-dependent governing equation for temperature in a fixed-volume open system is obtained as:

$$\mathrm{d}T = \frac{1}{mc_v}\left[T\left(\frac{\partial p}{\partial T}\right)\right]\left(\frac{V}{m}\right)\mathrm{d}m + \frac{\delta Q}{mc_v} \tag{3.7}$$

where the partial derivative of pressure with respect to temperature is evaluated based on the real-gas EoS in Equation (2.2).

Due to the constant volume of the open system, the time-dependent pressure governing equation is given by Equation (3.3):

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial f}{\partial m}\frac{\mathrm{d}m}{\mathrm{d}t} - \frac{\partial f}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t} \tag{3.8}$$

For variable-volume open systems, such as the rodless chamber of an intake cylinder shown in Figure 3.2(b), and by neglecting changes in gas kinetic and potential energy, the first law of thermodynamics is expressed as:

$$u\mathrm{d}m + m\mathrm{d}u = (u_{in} + pv_{in})\mathrm{d}m + \delta Q - p\mathrm{d}V \tag{3.9}$$

By substituting Equation (3.5) into Equation (3.9), the time-dependent governing equation for temperature in a variable-volume open system is obtained as:

$$\mathrm{d}T = \frac{1}{mc_v}(u_{in} - u)\mathrm{d}m + \frac{1}{mc_v}\left[T\left(\frac{\partial p}{\partial T}\right)\right]\left(\frac{V}{m}\right)\mathrm{d}m - \frac{1}{mc_v}\left[T\left(\frac{\partial p}{\partial T}\right)\right]\mathrm{d}V + \frac{\delta Q}{mc_v} \tag{3.10}$$

where the variation of specific internal energy is derived by integrating Equation (3.5), which yields:

$$u_{in} - u = \int c_v \mathrm{d}T + \int \left[T\left(\frac{\partial p}{\partial T}\right) - p\right] dv \tag{3.11}$$

The time-dependent governing equation for pressure in a variable-volume open system is obtained as Equation (3.3).

3.2. Thermodynamic model of closed systems

For variable-volume closed systems, such as the rod chamber of a cylinder shown in Figure 3.3, the gas volume, temperature, and pressure vary with time, while the mass remains constant. By neglecting changes in gas kinetic and potential energy, the first law of thermodynamics is expressed as:

$$m\mathrm{d}u = -p\mathrm{d}V + \delta Q \tag{3.12}$$

By substituting Equation (3.5) into Equation (3.12), the time-dependent governing equation for temperature in a variable-volume closed system is obtained as:

$$\mathrm{d}T = -\frac{1}{mc_v}\left[T\left(\frac{\partial p}{\partial T}\right)\right]\mathrm{d}V + \frac{\delta Q}{mc_v} \tag{3.13}$$

Since the mass of a variable-volume closed system remains constant, its time-dependent governing equation for pressure is derived from Equation (3.3) as:

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial f}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t} - \frac{\partial f}{\partial V}\frac{\mathrm{d}V}{\mathrm{d}t} \tag{3.14}$$

By solving the governing equations of pressure and temperature for the three typical thermodynamic systems, a transient thermodynamic system model that incorporates real-gas effects is established.

4. THEORETICAL ANALYSIS AND EXPERIMENTAL VALIDATION OF THE INTERSTAGE PNEUMATIC SEPARATION SYSTEM

The investigated interstage pneumatic separation system consists of four high-pressure supply gas tanks connected in series, solenoid valve, and a separation cylinder, as shown in Figure 4.1. Before system operation, the gas tanks are pre-charged with high-pressure helium as the energy source. Upon receiving the activation command, allowing the high-pressure gas to pass through the valve and flow into the rodless chamber of the separation cylinder. The piston is then driven to move, generating the thrust required for separation and enabling the rapid unlocking and separation of the first and second stages, as illustrated in Figure 4.2.

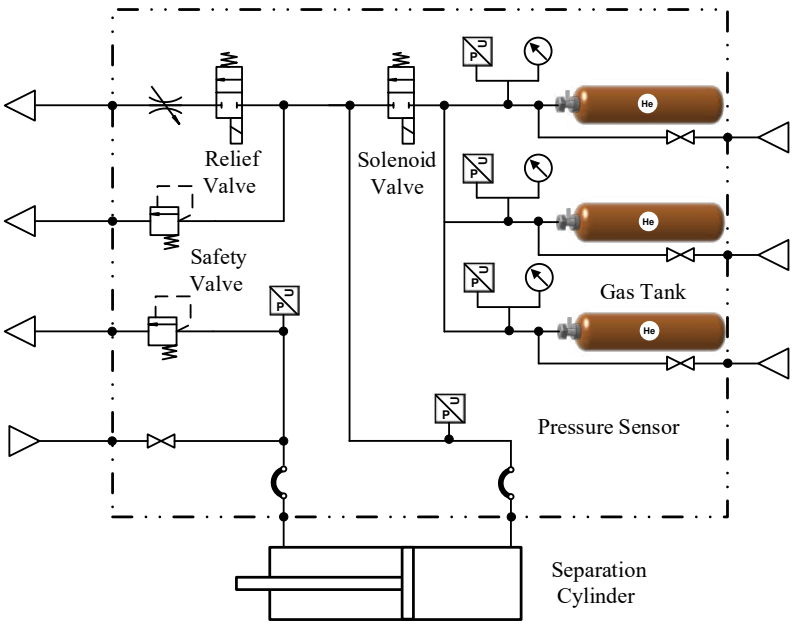


Figure 4.1 Schematic diagram of the interstage high-pressure helium separation system

For the first stage of the rocket, which is connected to the cylinder barrel, the force balance equation is expressed as:

$$p_1 A_1 - p_2 A_2 = m_1 \frac{\mathrm{d}^2 x_1}{\mathrm{d}t^2} + B_1 \frac{\mathrm{d}x_1}{\mathrm{d}t} + K_1 x_1 - \beta m_1 g \tag{4.1}$$

For the second stage, which is connected to the piston rod, the force balance equation is expressed as:

$$p_1 A_1 - p_2 A_2 = m_2 \frac{\mathrm{d}^2 x_2}{\mathrm{d}t^2} + B_2 \frac{\mathrm{d}x_2}{\mathrm{d}t} + K_2 x_2 - \beta m_2 g \tag{4.2}$$

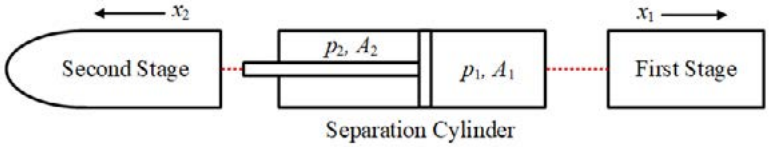


Figure 4.2 Connection relationship between the stages and the separation cylinder [21]

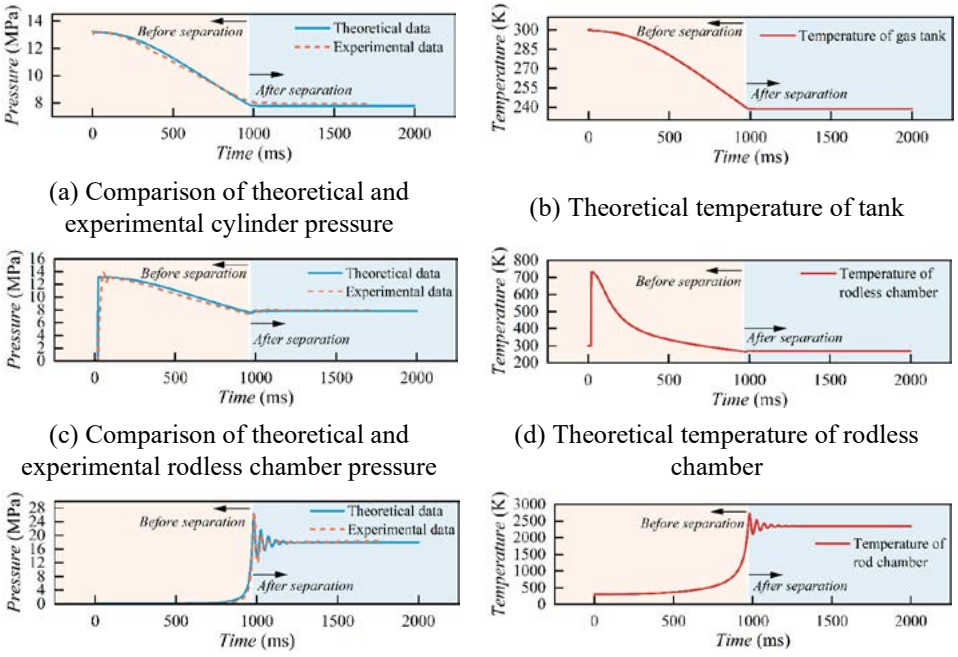
Table 4.1. Calculation Parameters and Values

Parameter	Symbol/Unit	Value
Critical pressure of helium	p_c/MPa	0.227
Critical temperature of helium	T_c/K	5.2
Critical volume of helium	V_{mc}/(m^3/mol)	57.3e-6
Molar mass of helium	M/(kg/mol)	4.003e-3
Gas cylinder volume	V/L	80
First-stage mass	m_1/t	90
Second-stage mass	m_2/t	120
Gravitational acceleration	g/(m/s^2)	9.8

To simulate the dynamic characteristics of the high-pressure chambers in this pneumatic system, a high-pressure gas analysis method based on a real-gas EoS is employed. According to the thermodynamic characteristics of the system components, the high-pressure gas tank is treated as a fixed-volume open system, with its pressure and temperature dynamics governed by Equations (3.7) and (3.8). The rodless chamber of the cylinder is treated as a variable-volume open system, described by Equations (3.10) and (3.3). The rod chamber is modeled as a variable-volume closed system, with pressure and temperature variations governed by Equations (3.13) and (3.14). Together, these equations establish a complete mathematical model for the system's dynamic response. The differential equations are solved using an eighth-order Dormand–Prince algorithm, with relevant constant parameters listed in Table 2, where load conditions are referenced to a representative heavy-lift rocket stage of approximately 100 t as reported by NASA [22]. Since the interstage separation process occurs on a millisecond time scale, which is much shorter than the characteristic timescale of heat transfer, the system is treated as adiabatic.

During the construction of the ground test bench for interstage separation, the first and second stages are simulated by coaxially arranged rail-mounted flat cars. Pressure sensors are installed in the gas supply system to monitor the pressures in the gas tank, the rodless chamber, and the rod chamber of the actuator cylinder. The supply pressure during the test

is 13.2 MPa. The separation instant is defined as the moment when the first and second stages just lose contact, which is measured to occur at 967 ms. A comparison between the theoretical model and the experimental results shows that the average percentage errors of the three chambers are 1.62%, 3.15%, and 5.14%, respectively, thereby validating the accuracy of the theoretical model. The comparison of pressure characteristics between theoretical predictions and experimental results, as well as the theoretical temperature variation curve, is presented in Figure 4.3.



(a) Comparison of theoretical and experimental cylinder pressure

(b) Theoretical temperature of tank

(c) Comparison of theoretical and experimental rodless chamber pressure

(d) Theoretical temperature of rodless chamber

(e) Comparison of theoretical and experimental rod chamber pressure

(f) Theoretical temperature of rod chamber

Figure 4.3. Comparison of Theoretical and Experimental Results for Interstage Separation Pressure Characteristics and Theoretical Temperature Profile

After the separation command is issued, the gas tanks begin supplying the rodless chamber through the solenoid valves. The pressure evolution during separation can be divided into three stages:

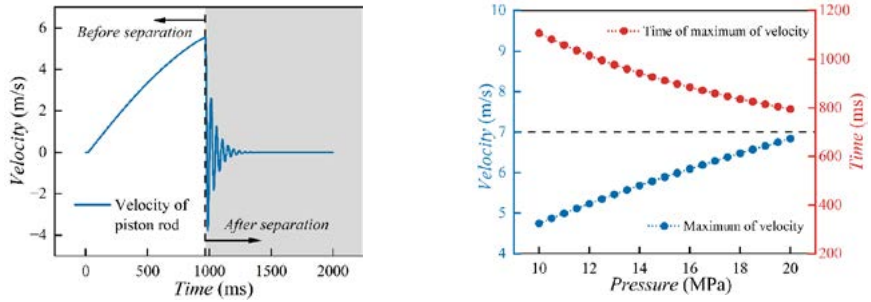
(1) Before separation: The pressure in the gas tanks decreases gradually (Figure 4.3(a)), while the rodless chamber pressure rises rapidly until it matches the tank pressure (Figure 4.3(c)). Once the rodless chamber pressure exceeds the piston start-up pressure, the piston rod begins to extend, leading to an increase in chamber volume and a gradual decrease in pressure.

(2) At the separation point: The piston rod reaches its maximum displacement, and the rodless chamber volume stabilizes. At this moment, the gas tanks and rodless chamber are fully connected, maintaining equal pressures. Meanwhile, the volume of the rod chamber reaches its minimum, and its internal pressure rises sharply to a peak value (Figure 4.3(e)).

(3) After separation: The cylinder system can be abstracted as a mass–spring system under inertial effects. Due to gas compressibility, the piston rod undergoes oscillatory motion, which gradually decays to an equilibrium state.

The gas temperature variations in each chamber follow a similar trend to the pressure changes, as shown in Figure 4.3(b), (d), (f), consistent with fundamental thermodynamic principles.

The fundamental task of the interstage separation system is to achieve reliable separation of the two stages within an extremely short time and over a limited stroke. The relative separation velocity serves as the most direct indicator of this function. As shown in Figure 4.4(a), at the moment of separation, the relative velocity reaches a maximum of 5.52 m/s. A higher separation velocity indicates a faster separation action. As illustrated in Figure 4.4(b), with increasing inlet pressure, the separation velocity increases accordingly, and the time required to reach the maximum velocity is shortened. To ensure that the separation velocity is not lower than 5 m/s and that the separation duration is less than 1000 ms, the minimum supply pressure of the gas tank must be no less than 12.5 MPa.



(a) Velocity of the separation cylinder piston rod

(b) Separation velocity and the time to achieve

Figure 4.4 Relative Velocity Curves during Stage Separation

5. CONCLUSIONS

To address the significant deviation of high-pressure helium from ideal gas behavior and the cumulative errors associated with conventional cubic EoS, a new corresponding state EoS is proposed based on the intermolecular interaction mechanism and the temperature dependence of molecular forces and molecular volume. On this basis, a transient thermodynamic model is developed using the proposed real-gas EoS, with dedicated modeling strategies formulated for open systems and closed systems. This model is further applied to the case of the interstage high-pressure helium separation process. The main conclusions are as follows:

(1) Within the typical operating temperature range (−30 °C to 120 °C) and pressure range (1–35 MPa) of the interstage separation high-pressure helium system, the proposed real-gas EoS achieves a maximum absolute error below 0.82%. This effectively reduces systematic deviations and significantly improves computational accuracy.

(2) The transient thermodynamic model constructed with the proposed EoS can accurately capture the thermodynamic behavior of fixed-volume open systems, variable-volume open systems, and variable-volume closed systems in the interstage separation process. The

average percentage error compared with experimental data is below 5.2%, thereby validating the effectiveness of the modeling approach in high-pressure gas analysis.

(3) Under heavy-load conditions, where the first and second stages carry 90 t and 120 t, respectively, to satisfy the technical requirements of a separation velocity not lower than 5 m/s and a separation duration of less than 1000 ms, the supply pressure of the gas tank must be no less than 12.5 MPa.

ACKNOWLEDGEMENTS

This research is supported by the National Natural Science Foundation of China (No. 52575071, 52175059), the National Key Research and Development Program of China (No. 2024YFF0505300), and the Shanghai Science and Technology Commission (No. 24YF2745200) through the Rising-Star Program (Yangfan Plan).

NOMENCLATURES

A_1	Effective area of rodless chamber	K	Spring stiffness
A_2	Effective area of rod chamber	m	Mass
a	Molecular interaction strength constant	p	Pressure
B	Damping coefficient	Q	Heat
b	Molecular volume correction constant	T	Temperature
β	Gravitational acceleration correction	T_B	Boyle temperature
c_v	Specific heat capacity	U	Internal energy
$\mathrm{d}m$	Mass flow rate	u	Specific internal energy
E_k	System kinetic energy	v	Specific volume
E_p	System potential energy	V_m	Molar volume
h	Specific enthalpy	W	Work done
I	Degrees of freedom	x	Displacement
Subscripts			
in	Into the system	r	Reduced
out	Out of the system	1	First stage of launch vehicle
c	Critical	2	Second stage of launch vehicle

REFERENCES

[1] S.J. Isakowitz, International reference guide to space launch systems, NASA STI/Recon Technical Report A, vol. 91, pp. 51425, 1991.

[2] J. Lee, Y.-S. Kim, S.-J. Baek, H.S. Jeong, Study of separation mechanism and characteristics of ridge-cut explosive bolts using numerical analysis, International Journal of Aeronautical and Space Sciences, vol. 22, pp. 84–93, 2021.

[3] M. Lucy, R. Hardy, E. Kist, J. Watson, S. Wise, Report on alternative devices to pyrotechnics on spacecraft, 1996.

[4] M. Hihoud, A. Pagès, C. Benoit, F. Claeyssen, S. Sanchez, S. Trémolières, P. Guay, BRUCE – Electromagnetic actuated pin puller, Proc. of the 13th International Conference on New Actuators & 7th International Exhibition on Smart Actuators and Drive Systems, Bremen, Germany, pp. 18–20, 2012.

[5] B. Little, Electrically powered separation nuts, Proc. of the 39th Aerospace Mechanisms Symposium, pp. 7–9, 2008.

[6] W. Tak, M. Lee, B. Kim, Ultimate load and release time controllable non-explosive separation device using a shape memory alloy actuator, Journal of Mechanical Science and Technology, vol. 25, pp. 1141–1147, 2011.
[7] SpaceX, Falcon 9 Launch Vehicle User's Guide, User Manual SpaceX-SPEC-001, Hawthorne, CA, USA, 2021.
[8] E. Hand, Space-science hopes rest on rocket test, Nature, vol. 465, pp. 276–277, 2010.
[9] J.M. McMahon, M.A. Morales, C. Pierleoni, D.M. Ceperley, The properties of hydrogen and helium under extreme conditions, Reviews of Modern Physics, vol. 84, pp. 1607–1653, 2012.
[10] R.J. Thomas, R. Dutta, P. Ghosh, K. Chowdhury, Applicability of equations of state for modeling helium systems, Cryogenics, vol. 52, pp. 375–381, 2012.
[11] J.O. Valderrama, The state of the cubic equations of state, Industrial & Engineering Chemistry Research, vol. 42, pp. 1603–1618, 2003.
[12] Redlich, J.N.S. Kwong, On the thermodynamics of solutions. V. An equation of state. Fugacities of gaseous solutions, Chemical Reviews, vol. 44, pp. 233–244, 1949.
[13] G. Soave, Equilibrium constants from a modified Redlich–Kwong equation of state, Chemical Engineering Science, vol. 27, pp. 1197–1203, 1972.
[14] D.-Y. Peng, D.B. Robinson, A new two-constant equation of state, Industrial & Engineering Chemistry Fundamentals, vol. 15, pp. 59–64, 1976.
[15] A.J. Schultz, D.A. Kofke, Virial equation of state as a new frontier for computational chemistry, The Journal of Chemical Physics, vol. 157, 2022.
[16] B.I. Lee, M.G. Kesler, A generalized thermodynamic correlation based on three-parameter corresponding states, AIChE Journal, vol. 21, pp. 510–527, 1975.
[17] W.G. Chapman, G. Jackson, K.E. Gubbins, Phase equilibria of associating fluids, Molecular Physics, vol. 65, pp. 1057–1079, 1988.
[18] A. Alanazi, M. Ali, S. Bawazeer, N. Yekeen, H. Hoteit, Evaluation of cubic, PC-SAFT, and GERG2008 equations of state for accurate calculations of thermophysical properties of hydrogen-blend mixtures, Energy Reports, vol. 8, pp. 13876–13899, 2022.
[19] S.B. Kiselev, D.G. Friend, Cubic crossover equation of state for mixtures, Fluid Phase Equilibria, vol. 162, pp. 51–82, 1999.
[20] R. Tao, Y. Yin, X. Liu, et al., Novel corresponding states principle equation for real hydrogen under ultra-high pressure considering molecular association, Journal of Mechanical Science and Technology, 2025, pp. 1–13.
[21] R. Tao, Y. Yin, S. Zhu, et al., Performance evaluation and experimental validation of high-pressure pneumatic interstage separation system for reusable launch vehicles considering real gas polytropic process, Aerospace Science and Technology, 2025, 111080.
[22] M. Clarke, et al., Project Columbiad: Mission to the Moon. Book 1: Executive Summary. Vol. 1: Mission trade studies and requirements. Vol. 2: Subsystem trade studies and selection, NASA Technical Report, NAS 1.26:192019, 1992.

Biographies



Ruiyao Tao is a PhD student in the School of Mechanical Engineering, Tongji University, Shanghai, China. Her research interests focus on ultra-high-pressure gas characteristics and flow control properties.



Yaobao Yin is currently a Professor in the School of Mechanical Engineering, Tongji University. He received his Ph.D. from Saitama University, Japan. His main research fields include high-speed aerodynamic control and fluid control in extreme environments, fundamental theories of hydraulics.



Naiping Gao is currently a Professor in the School of Mechanical Engineering, Tongji University. He received his Ph.D. from The Hong Kong Polytechnic University. His main research fields include ventilation and thermal issues in buildings and underground spaces, efficient utilization of low- and medium-temperature thermal energy, and clean combustion of natural gas.



Pengfei He is a Professor in the School of Aerospace Engineering and Applied Mechanics, Tongji University. His research focuses on composite mechanics and computational mechanics, applying molecular dynamics and multiscale modeling to study the mechanical and transport properties of advanced materials such as ionomer membranes, carbon nanotubes, graphene composites, and nanostructured alloys.

Radial flow force reduction in hydraulic spool valve using balanced through-hole design

Yuhang Zhang, Shipei Zheng, Jiaxi Zhang,

Jianshe Feng, Miao Yang, Guangwei Zhang, Beichen Ding*

School of Advanced Manufacturing & Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, People's Republic of China
**Corresponding author: E-mail address: dingbch@mail.sysu.edu.cn*

Abstract.

Hydraulic spool valve (HSV) is a key regulating component in fluid power systems owing to its capability for precise flow and pressure control. Unbalanced radial flow force (RFF) is a major contributor to spool instability and friction-induced failure, particularly in open-neutral configurations where flow field asymmetry is prominent. Although conventional design assumes that radial forces are self-balanced due to geometric symmetry, practical valve structures constrained by standardized port layouts frequently induce asymmetric pressure and velocity distributions. To address this issue, a novel balanced through-hole design is proposed and integrated into the mid-shoulder region of the spool. Computational fluid dynamics (CFD) simulations are employed to investigate the generation mechanism of RFF and to evaluate the effectiveness of the proposed structure. The Realizable k-ε turbulence model is adopted to capture the dominant turbulent features, and the individual contributions of pressure and viscous forces to the RFF are quantified. A parametric study is conducted to examine the influence of through-hole inclination angle on force reduction performance. The study results indicate that RFF is predominantly pressure-driven, accounting for more than 85% of the total force, with pronounced asymmetry concentrated in the mid-shoulder region. By introducing a dual through-hole structure inclined at 15°, the RFF acting on the mid-shoulder is reduced by 45.5%, while the total spool RFF is decreased by 29.9%. The proposed design effectively improves flow symmetry and suppresses local pressure gradients, providing a practical solution for RFF mitigation in compact hydraulic spool valves.

Keywords. Hydraulic spool valve, Radial flow force, Through-hole design, CFD simulation, Structural optimization.

1. INTRODUCTION

Hydraulic systems are extensively applied in critical engineering domains such as construction machinery, marine equipment, and aerospace, where they serve as core components in applications, e. g. electric loader bucket and boom drive control [1]. These

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

systems are valued for their rapid response, high power density, and precise controllability [2]. As a core regulating component within a hydraulic system, hydraulic valve plays a pivotal role in determining system responsiveness, stability, and energy efficiency. Among them, the hydraulic spool valve (HSV) employing thin-walled orifice structures exhibit reduced sensitivity to oil viscosity and temperature variations, making them widely used in advanced components such as proportional and servo valves [3].

The spool-to-sleeve clearance of HSVs typically ranges between 5 to 20μm, serving functions such as clearance sealing and motion guidance [4]. Ideally, axial movement of the spool should only induce minimal and stable friction. However, in practical applications, hydraulic forces, thermal deformation, and particulate contamination alter the lubrication state [5], leading to spool hysteresis, jamming, or stiction. Such phenomena can significantly deteriorate control performance and, in severe cases, induce valve failure or safety incidents [6]. Consequently, understanding and mitigating spool stagnation has become a critical issue in hydraulic valve research, traditionally attributed to unbalanced flow forces and lubricant viscosity effects.

Among the influencing factors, flow force is widely recognized as one of the primary contributors to control performance degradation [7]. HSVs regulate flow direction and magnitude through axial displacement of the spool. When the orifice formed between the spool shoulder and sleeve is partially opened, fluid acceleration occurs, which alters the flow momentum and generates force on the spool. Flow force is commonly categorized into steady-state and transient components. The transient component is usually considered negligible compared to the steady-state force. When fluid flows through the throttle port, it experiences changes in velocity and direction, resulting in hydraulic forces on the spool surface, both in axial and radial directions [8].

While research on axial flow force has matured, the radial flow force (RFF) has received relatively limited attention. Many studies have assumed that due to geometric symmetry, RFFs are self-balanced and can be neglected [9]. However, in actual engineering designs, the flow chamber geometry is often asymmetric. This asymmetry results from practical constraints related to valve block layout and standardized port configurations, especially for base-mounted HSVs that follow ISO mounting standards. These designs lead to uneven pressure and velocity distributions along the cylindrical surface of the spool and the annular gap between spool and sleeve [10]. The resulting unbalanced RFF may push the spool toward one side, causing increased friction and possible motion failure.

To mitigate RFFs, researchers have developed a variety of structural modifications. One common method involves machining circumferential pressure-equalizing grooves on the spool shoulder to improve pressure distribution. Hong et al. proposed the use of spiral grooves along the spool surface [11]. These grooves enable simultaneous axial and circumferential pressure release, which effectively reduces pressure peaks and lowers total RFF by over 60 percent. Han et al. introduced a guide groove connected to a central pressure-equalizing groove, which distributed pressure more uniformly and enhanced the response frequency of the valve [12]. Additionally, the shape of the spool edge, such as right-angle, conical, or curved forms, influences the throttle jet direction and thereby affects the RFF at the outlet [13]. However, experimental results show that in many cases, the inlet section's RFF remains largely unchanged despite structural modifications.

Most existing studies focus on negative-opening HSVs, whereas open-neutral configurations remain underexplored. Moreover, complex multi-step throttling structures, such as claw-shaped notches [14], limit the applicability of circumferential balancing grooves. Importantly, traditional groove-based methods act as passive compensation mechanisms, redistributing pressure after radial forces have already developed. In contrast, the present study aims to suppress RFF at its point of origin. By introducing pressure-relief through-holes directly within the spool shoulder, where jet impingement and pressure stagnation occur, the proposed design aims to suppress the formation of pressure asymmetry rather than redistributing it downstream.

2. PHYSICAL MODEL DESCRIPTION

The typical multi-way spool valve (MSV) working principle and structure of the loader hydraulic system are shown in Figure 2.1. Positioned within the main hydraulic circuit, the MSV dis-tributes flow from the main pump to the boom/bucket cylinders, enabling efficient operation of the work equipment. The MSV features a series-parallel integrated two-section design, structurally composed of the valve body, bucket spool valve section, boom spool valve section, and check valves, as shown in Figure 2.1(a). The inlet of two sections is arranged in series, granting priority to the bucket section from Figure 2.1(b). When the bucket section active, it prevents simultaneous operation of the boom section. The return oil passages adopt a parallel configuration, allowing both sections to return oil concurrently.

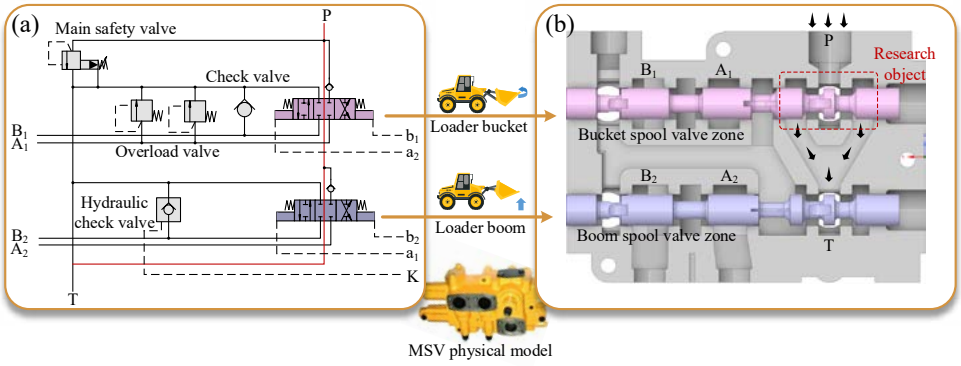


Figure 2.1. MSV working principle and structure of the loader hydraulic system: (a) Working principle; (b) Valve structure.

Due to the asymmetric and discontinuous internal flow passage geometry, the fluid develops an unbalanced turbulent flow. This generates uneven fluid pressure distribution on the spool surface, resulting in unbalanced RFFs. Furthermore, the unbalanced RFF tends to induce spool tilting, leading to asymmetric variation in the clearance between the spool and valve body. In severe cases, this may cause direct contact between the spool and valve body, significantly increasing frictional resistance during spool movement and potentially resulting in sticking or even complete seizure. Specially, when high-pressure fluid enters the valve chamber, it initially impacts the shoulder of the P-T flow path on the spool. Thus, this paper focus on the unbalanced RFF mechanism and reduction approach based on the P-T flow path of the bucket spool valve. The simplified HSV geometry is shown in Figure 2.2.

This study investigates a full circumferential opening spool valve, whose structural characteristics are illustrated in Figure 2.2(a). The HSV is composed of valve body and valve spool. Specially, as shown in Figure 2.2(b), the valve spool mainly consists of five components. Hydraulic fluid enters through the inlet, passes through four throttling orifices, and exits via both left and right outlets (as shown in Figure 2.2c). The valve body and spool exhibit strict bilateral symmetry. When the spool is in its neutral position, x_{01}=x_{04}=0.4 mm and x_{02}=x_{03}=0.35 mm.

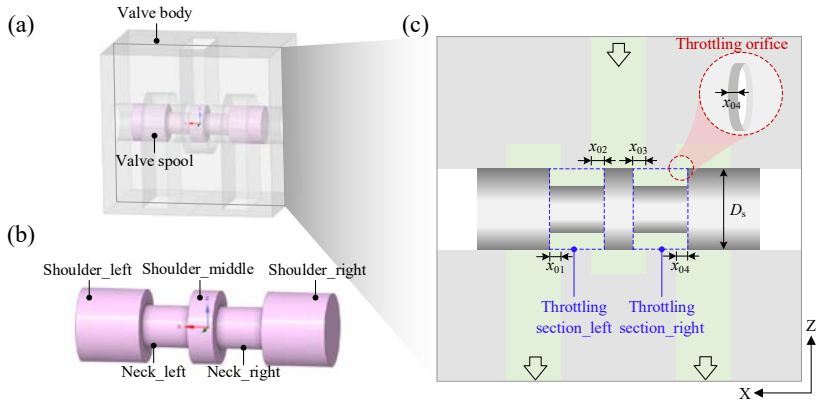


Figure 2.2. Object HSV structure: (a) HSV assembly drawing; (b) Valve spool structure; (c) Central cross-section view. In (c), x_{01} ~ x_{04} represent valve port openings between the valve body and spool shoulders, and D_s is the spool shoulder diameter.

3. NUMERICAL SIMULATION

3.1. Computational fluid domain

The main valve structure is simplified and the extracted fluid domain is shown in Figure 3.1(a). It is a symmetric structure and thus the fluid domain can be simplified as a half to decrease simulation time, shown in Figure 3.1(b). The boundary condition status is shown in Figure 3.1(c) to satisfy the realistic working condition and symmetry flow field.

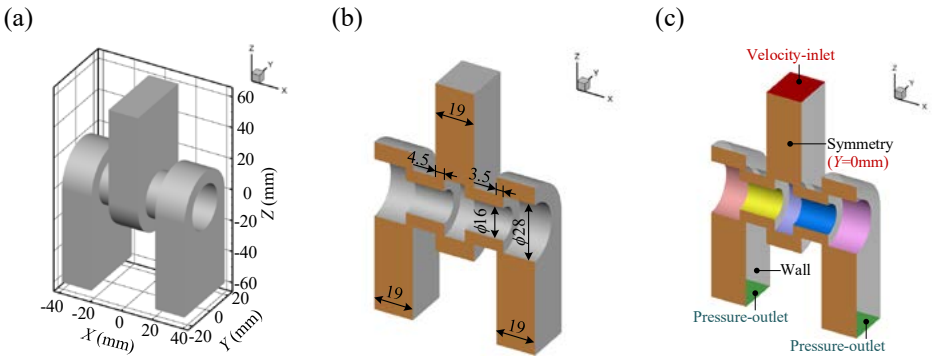


Figure 3.1. Fluid domain model and key structure size: (a) Complete fluid domain; (b) One-half fluid domain from *x*-axis; (c) Fluid domain boundary condition settings.

3.2. Meshing and independence verification

This investigation employs ANSYS Fluent® software to perform numerical simulations associated with HSVs. Compared to structured grids, unstructured grids are more efficient and easier to control in terms of grid numbers. Given the complex configuration of the HSV's orifice and its compact volume, the entire computational domain is divided into a three-dimensional unstructured mesh using Fluent Meshing® software to ensure grid quality and improve computational efficiency. To capture complex turbulence behaviour such as high-speed jet of valve ports and vortex, five 'Body of Influence' (Boi) regions are developed to refine the mesh, which are shown in Figure 3.2(a). Consequently, at least four mesh models of varying densities are produced in one fluid domain.

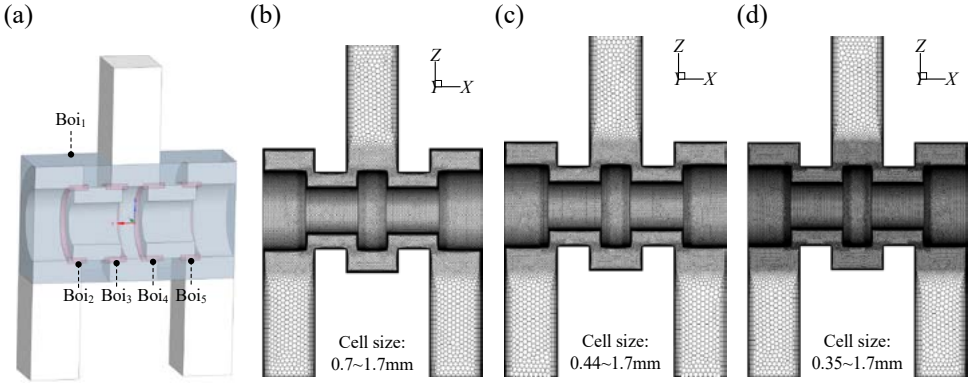


Figure 3.2. Boi regions setting for mesh refinement and mesh quantity comparison: (a) Boi regions setting for mesh refinement; (b) Mesh A with cell size 0.7~1.7mm; (c) Mesh B with cell size 0.44~1.7mm; (d) Mesh C with cell size 0.35~1.7mm.

Table 3.1. Mesh independence verification setting

Region	Mesh A	Mesh B	Mesh C
Overall size (mm)	1.70	1.70	1.70
Boi_1 size (mm)	1.20	0.75	0.60
Boi_2 size (mm)	0.70	0.44	0.35
Boi_3 size (mm)	0.80	0.50	0.40
Boi_4 size (mm)	0.80	0.50	0.40
Boi_5 size (mm)	0.70	0.44	0.35

Mesh division is a pivotal factor in numerical simulations, with the number of meshes directly impacting computational accuracy and determining the convergence and correctness of residual results. Consequently, to guarantee calculation precision and enhance efficiency by reducing computation time, it is imperative to assess the mesh density's relevance prior to the commencement of calculations. The mesh independence test is conducted for a steady simulation. Three mesh models of varying densities are produced, as depicted in Figure 3.2(b)-(d). The detailed mesh setting of different Boi regions is shown in Table 3.1. When the liquid flows at a velocity of 4.8733 m/s, the model with three different mesh numbers is calculated with the outlet pressure set to 0 MPa as the boundary condition. The pressure of inlet (P_{in}) and the radial fluid force (F_r) on the valve core are monitored for the mesh

independence. The final results of the mesh independent verification and the error of mesh independence index are presented in Table 3.2. As observed from Table 3.2, an increase in the number of grids leads to a closer alignment between the monitored P_{in} and F_r. When the mesh quantity is increased to 841 908, the error of P_{in} and F_r are at 0.33% and 0.14%, which has been maintained within 0.5%. For a better radial fluid force mechanism revealing and faster computation speed, Mesh C is chosen for the subsequent simulation.

Table 3.2. Mesh independence verification results

Mesh type	Mesh elements	Inlet pressure (Pa)	% Change	Spool RFF (N)	% Change
Mesh A	219 473	61764.3	--	1.436	--
Mesh B	524 611	61006.4	1.20%	1.405	2.20%
Mesh C	841 908	60803.2	0.33%	1.406	0.14%

The selection of the turbulence model and near-wall mesh strategy is guided by the specific flow characteristics of the HSV rather than an assumption of fully developed turbulence. Although hydraulic oil exhibits relatively high viscosity, the flow inside the valve is dominated by strong acceleration through narrow throttling orifices, high velocity gradients, and jet–wall interactions, which promote locally turbulent and separation-dominated flow structures even at moderate Reynolds numbers

In regards to the near-wall approach for Realizable k-ε, the particular complexity of the valve geometry leads to the separation and reattachment of boundary layer easily. For this reason, the 'enhanced wall treatment' (EWT) is employed in the present work, which solves viscous substrates directly for the large pressure gradient of the valve port [15]. The number of boundary layers is 10, and the growth ratio is 1.2. The critical value of y+ of the fluid domain wall, which signifies the non-dimensional wall distance, is currently maintained below 5. This configuration enables adequate resolution of near-wall pressure gradients and wall shear stresses without excessive mesh refinement. Mesh independence verification further confirms that the selected turbulence model and boundary layer settings yield stable predictions of inlet pressure and radial flow force, with variations below 0.5%, indicating that the numerical results are not unduly sensitive to the chosen near-wall discretization.

3.3. Physical properties and boundary conditions

In this investigation, the calculation is performed under steady-state conditions. The physical properties of liquids used in the simulation are shown in Table 3.3.

Table 3.3. Physical properties and boundary conditions for the CFD simulations

Targets			Values or settings
Physical properties	Hydraulic oil L-HM46	Temperature (°C)	40
		Density (kg/m^3)	870
		Viscosity (Pa·s)	0.04
Boundary conditions	Inlet	Velocity (m/s)	4.8733
	Outlet	Pressure (MPa)	0
	Symmetric face	Symmetry	--
	Valve body Valve spool	Wall	No-slip and adiabatic

Monitoring faces are established at the inlet/outlet positions of the throttle holes to monitor pressure pulsations and flow rate pulsations. These data points are utilized to assess the stability of the calculation results and will be compared and analysed with the experimental findings at a subsequent stage. The boundary condition status is shown in in Figure 3.1(c) to satisfy the realistic working condition and symmetry flow field. Given that the HSV possesses a stable structure and its internal flow dynamics are solely determined by the inlet and outlet boundary conditions, velocity-inlet and pressure-outlet conditions are employed for the calculations while assuming an adiabatic wall surface and no-slip condition.

3.4. Solution strategy

The pressure-based solver is used in the solution. Different discretisation formats are used for the control equation terms to accelerate convergence, 'Least Square Cell Based' discretisation for the gradient term, 'PRESTO!' for the pressure term, 'Second Order Windward' discretisation for the momentum term and each turbulence term. The final discretised control equations are solved by 'Coupled' algorithm. Especially, these parameter settings are verified stable in many HSV simulations. In the simulation process, convergence is deemed to have been achieved when the corresponding parameters within an identical grid cell remain invariant between two successive calculations. Nevertheless, this ideal condition is seldom encountered in practical computations. To affirm the convergence of computational results, the following criteria are typically satisfied: (1) the residual value of the continuous equation should be less than 10^{-5}, with a remaining residual value of less than 10^{-6}; (2) monitor changes such as spool RFF within the fluid domain are less than 0.5%.

4. RESULTS AND DISCUSSION

4.1. Spool radial unbalanced flow force analysis

Before analysing the radial unbalanced force characteristics within the HSV, the flow characteristics of the flow channel should be given. Therefore, the average velocity of the centre cross-section (Y=0 mm) just as Figure 3.1(c) of the valve is first obtained, as shown in Figure 4.1.

Since the hydraulic oil flow inside the valve chamber is restricted by the walls of the spool and the valve body, distinct flow characteristics can be observed. The flow zone division and flow characteristics are shown in Figure 4.1(a) and (b). As the oil enters through the inlet, firstly a part of the oil impinges on the middle shoulder (region A_1), secondly the convergence enters the inlet throttling port (region A_2) and forms a free jet, and then it collides with the wall of the valve neck (region A_3) to form a wall jet region. Next, the wall jet enters the outlet throttling port (region A_4), where the chamber area increases, allowing the jet to fully develop, which is called the outflow jet region. When the outflow jet contacts the wall, another wall jet begins to form, which has a lower velocity magnitude and a wider jet width than the previous wall jet region. At the same time, a portion of the oil is trapped in the narrow region formed by the spool shoulder and the valve body (region A_5), forming a strong vortex.

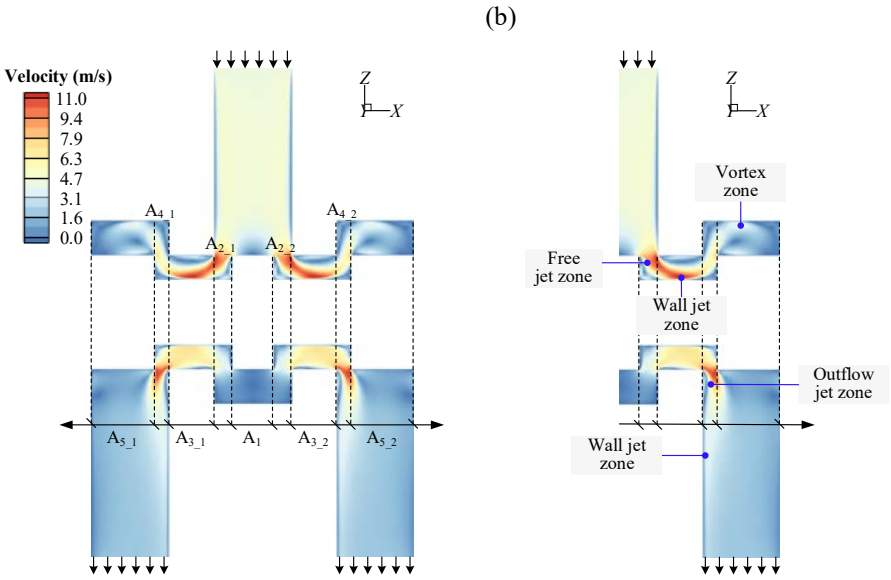


Figure 4.1. Average velocity and flow characteristics of the centre cross-section (Y=0 mm): (a) Flow zone division; (b) Flow characteristics. Especially, due to the symmetrical velocity contour distribution of the left and right parts, figure (b) only shows the right fluid domain.

The RFF is dominated by the pressure component and the viscous component, and Figure 4.2 shows the values and percentages of these two components in the Z-direction for different wall parts (W_1-W_5). Due to the symmetrical contour distribution of the spool wall left and right parts, pressure and WSS-Z contours are placed in a separate contour. From Figure 4.2(a) and (b), the RFF on each spool wall is mainly dominated by the pressure force component, and the viscous force only has a certain contribution (23%) in the middle shoulder (W_1), while the viscous force in the rest walls accounts for less than 10%. Among them, valve neck wall jet zones ($W_{3\text{-}1}$ and $W_{3\text{-}2}$) are the main positive RFF sources, showing significant positive pressure force component (3.34 N), and the viscous force contribution is small but in the same direction (+Z). Conversely, outlet shoulder zones ($W_{5\text{-}1}$ and $W_{5\text{-}2}$) show the most significant negative RFF (-4.28 N). Especially, the middle shoulder (W_1) has a smaller force value but possesses the opposite viscous force direction.

In short, the overall spool wall exhibits positive RFFs, with pressure still dominating (89% of the total). The above analyses indicate that there is a significant pressure asymmetry in these zones, which is the main source of unbalanced RFFs. In addition, since the spool is in the neutral position, the RFF on each region of the spool is symmetrically distributed, further indicating that the simulation is reliable by taking 1/2 flow domain as Figure 3.1.

It should be noted that the pressure values reported in this study are expressed in terms of gauge pressure, with the outlet pressure specified as 0 MPa. Under this formulation, negative pressure values may appear in regions of strong flow acceleration, jet impingement, or local flow separation. These negative values represent pressure levels lower than the outlet reference pressure and do not correspond to negative absolute pressures.

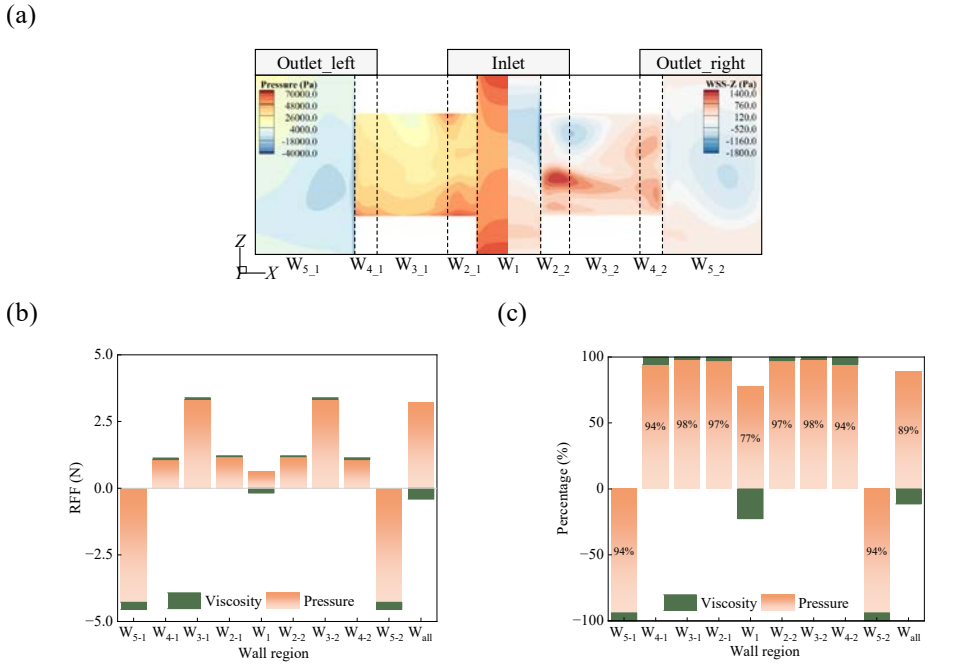


Figure 4.2. Pressure and viscous components of RFF on the spool wall and their proportion: (a) Spool walls division; (b) Pressure and viscous force value; (c) Pressure and viscous force value proportion. Especially, figure (a) exhibits the wall pressure and *Z*-direction wall shear stress (WSS-*Z*) in the *XOZ* view.

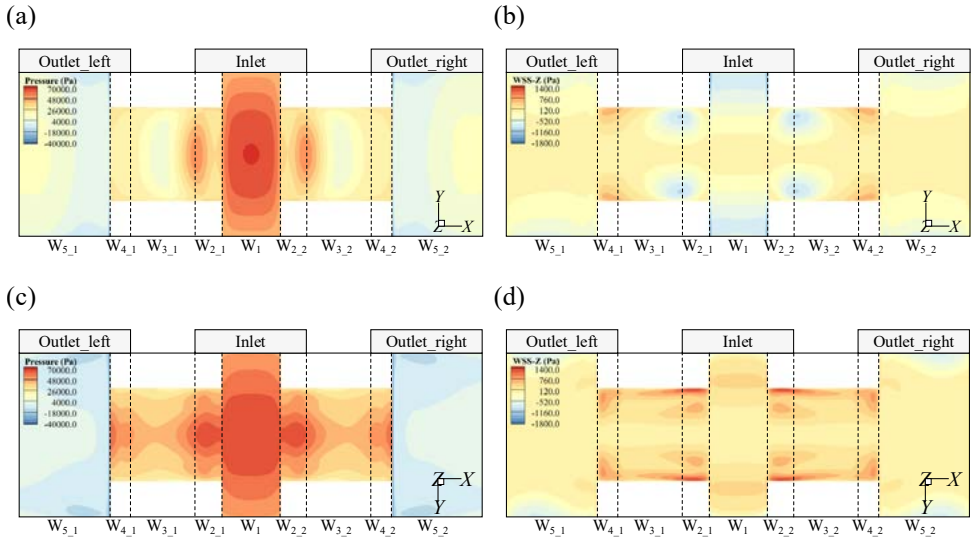


Figure 4.3. Pressure and WSS-*Z* distribution of the upper and lower spool walls: (a) and (b) Pressure and WSS-*Z* distribution of the upper spool wall; (c) and (d) Pressure and WSS-Z distribution of the lower spool wall.

The wall pressure and *Z*-direction wall shear stress (WSS-*Z*) of the upper and lower spool walls in Figure 4.3 further reveal the asymmetry in the spatial distribution of pressure and shear stress. From Figure 4.3 (a) and (c), a more pronounced pressure difference is visible at the upper and lower walls in W_3, corresponding to the larger radial positive force here in Figure 4.2. In W_5, the lower spool wall pressure decreases significantly, creating a significant negative RFF. In contrast, the high WSSs shown in Figure 4.3 (b) and (d) are mainly concentrated in the throttling regions of W_2 and W_4, and despite the large local gradients, their values are much lower than the pressure differences, verifying that the contribution of viscous forces is relatively small.

4.2. Spool radial unbalanced flow force reduction

Above analysis of pressure distribution relative to the middle shoulder clearly reveals that elevated pressure difference predominantly occurs within the central region of the upper and lower walls. To effectively reduce unbalanced RFF, optimization of these specific areas is required. Figure 4.4(a) shows the original valve spool (D_0). Figure 4.4(b) presents an optimized model (D_1) incorporating two perpendicular through-holes (diameter: 2.5 mm) positioned at the mid-section of the shoulder. This design aims to minimize the pressure differential between the upper and lower wall surfaces, thus reducing the unbalance RFF.

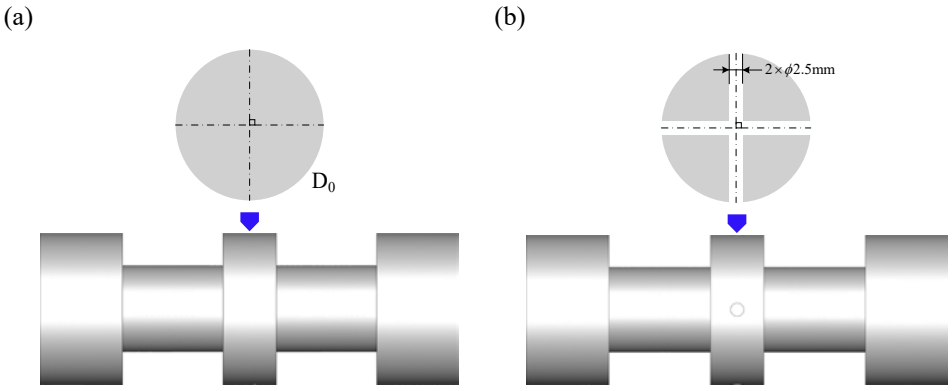


Figure 4.4. Original and optimized valve spool: (a) Original valve spool D_0; (b) Optimized valve spool D_1 with two perpendicular ϕ2.5mm through-holes.

4.2.1. Effect of balanced through-holes structure

Figure 4.5(a) and (b) shows the wall RFF components of two different shoulder structure. The optimized model shows a 20.3% decrease in pressure component, a 15.5% decrease in viscosity component and a 22.8% decrease in net RFF compared to the original model. It is shown that the balanced through-hole structure effectively guides the fluid diversion through the orifice and reduces the pressure gradient and shear effect of the shoulder upper and lower walls, thus weakening the unbalanced RFF. For the entire spool, the pressure component is reduced by 7.3%, the viscous component by 7.0%, and the combined force by 7.3%. Although the overall RFF change is relatively small, the local optimisation effect of the shoulder region is significant. This study is more concerned with an analysis of the effect of pressure on the middle shoulder, given the high pressure-RFF share.

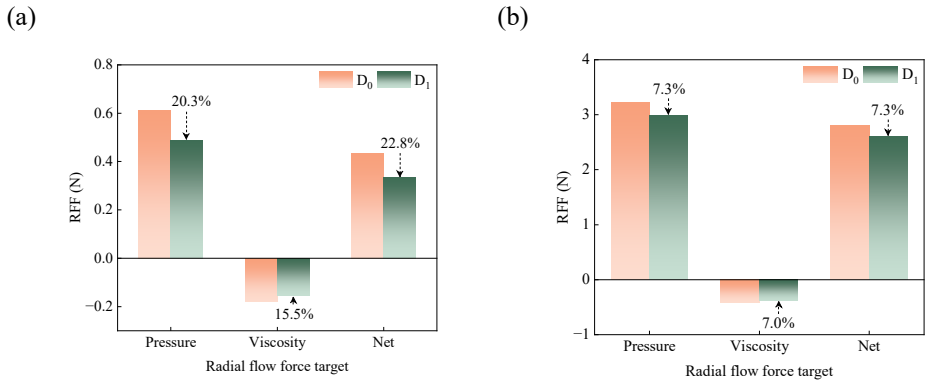


Figure 4.5. RFF on the middle shoulder wall (W_1) and the overall spool (W_{all}) of spools D_0 and D_1: (a) RFF components on the W_1; (b) RFF components on the W_{all}.

Pressure distribution on the upper and lower walls of two shoulder models are shown in Figure 4.6. It can be found that the centre area of the upper and lower walls shows an elliptical high-pressure region, and the boundary pressure decreases abruptly, indicating that the main flow passes around the shoulder resulting in a pressure build-up at the stagnation point. Meanwhile, the area of the high-pressure region on the lower wall is larger than that on the upper, especially at the edge position, resulting in a radial pushover trend (Figure 4.6a and b). With the interference of the through-holes structure, the pressure at the centre is effectively released as the pressure relief channel guides the fluid across the convex shoulder (Figure 4.6c and d). The reduction of the high-pressure areas improves the symmetry and balance of the upper and lower walls, which helps to reduce the RFF difference.

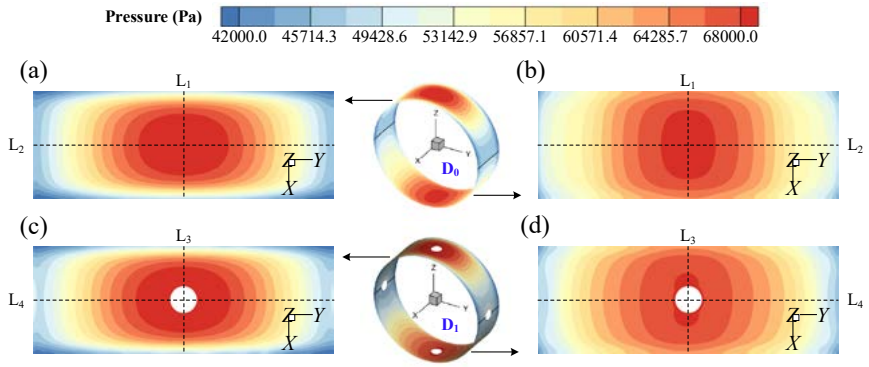


Figure 4.6. Pressure distribution on the upper and lower walls of two shoulder models: (a) and (b) Pressure distribution on the upper and lower walls of D_0; (c) and (d) Pressure distribution on the upper and lower walls of D_1.

The pressure values on the vertical bisectors (L_1~L_4) of the upper and lower walls are further extracted, as shown in Figure 4.7. From Figure 4.7(a), the optimisation resulted in an overall decrease in the lower wall (L_3) pressure and an overall increase (L_1) in the upper wall pressure. Meanwhile, the pressure difference of L_3-L_1 can be found in Figure 4.7(b). It can be found that the positive (+Z) pressure difference is reduced as a whole (near -5<X<-3mm

and 3<X<5mm), and the negative (-*Z*) pressure difference is increased (near -3<X<3 mm), which indicates the positive RFF decrease (+*Z*) as the negative RFF (-*Z*) increase. Then, the overall RFF decrease (+*Z*). The polar plots of L_2 and L_4 are shown in Figure 4.7(c) and (d). The inhomogeneity is defined to measure the pressure difference, which is shown as

$$\sigma_i = \frac{\left|A_{\text{lower}} - A_{\text{upper}}\right|}{A_{\text{lower}}} \times 100\% \tag{4.1}$$

where A_{lower} and A_{upper} are closed curve area enclosed by L_2 and L_4 and the coordinate axes, *i* represents the spool model.

Specially, $\sigma_0 = 1.39\%$ and the $\sigma_1 = 0.67\%$, which indicates that the overlapping cross area of the graph lines increases and the radial symmetry is enhanced after optimisation.

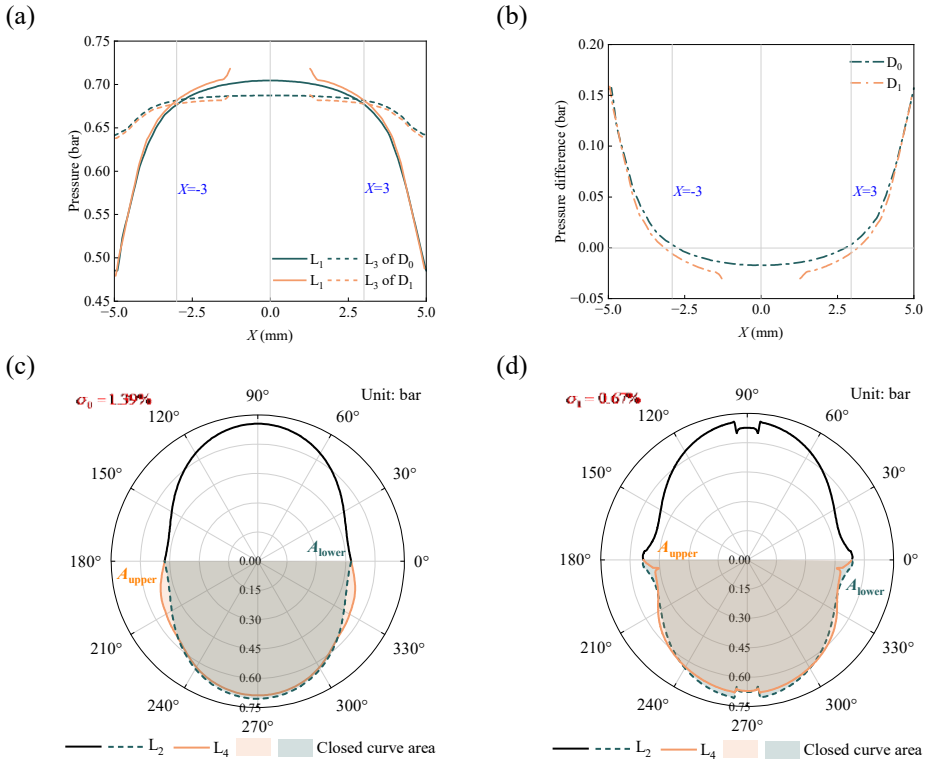


Figure 4.7. Pressure value extracted from L_1~L_4 of the shoulder D_0 and D_1 walls: (a) Pressure value extracted from L_1 and L_3; (b) Pressure difference of L_1 and L_3; (c) Polar plots from L_2 and L_4 on model D_1; (d) Polar plots from L_2 and L_4 on model D_2. Especially, in figure c and d, the cross areas are drawn for identifying the pressure difference between L_2 and L_4.

4.2.2. *Effect of balanced through-hole angle*

In this section, the effect of different through-hole angles on the suppression of radial unbalance force is comparatively analysed. The four through-hole angle models used in this study are shown in Figure 4.8. They are named as model D_0~D_4.

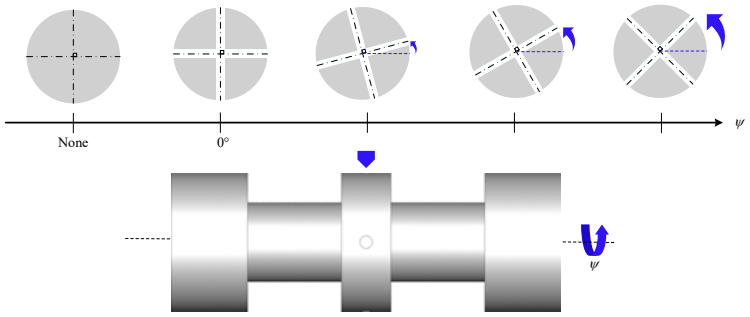


Figure 4.8. Original model and four different balanced through-holes angle (ψ) models.

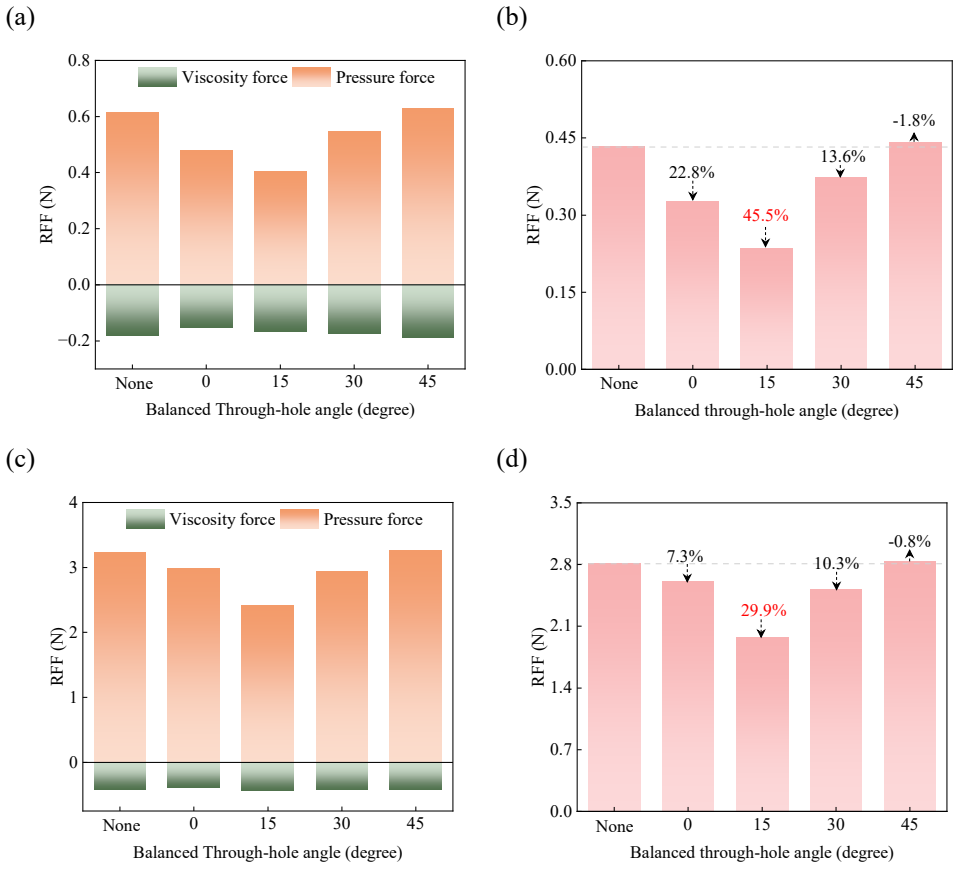


Figure 4.9. RFF (F_r) on the middle shoulder wall (W_1) and overall wall (W_{all}) of different balanced through-hole angle model: (a) Pressure and viscosity RFF components of W_1; (b) Overall RFF of W_1; (c) Pressure and viscosity RFF components of W_{all}; (d) Overall RFF of W_{all}.

The spool RFFs (comprising pressure force and viscous force) for different through-hole angles are demonstrated in Figure 4.9. As illustrated in Figure 4.9(a) and (c), the pressure force component remains predominant in the RFF, and the trend in the performance of the pressure component for each angle structure is consistent with the trend in the total RFF change. The viscous force component is less affected by the through-hole angle and remains stable, indicating that the viscous effect varies weakly within the range of structural parameters in this study. As the through-hole angle is adjusted, the total RFF undergoes a corresponding alteration from Figure 4.9(b) and (d), reaching its most significant decline (W_1 reduce by 45.5% and W_{all} reduce by 29.9%) of D_2. This observation signifies a substantial RFF balancing effect. Notably, For D_4, the RFF approaches the level observed in a structure devoid of holes.

5. CONCLUSION

This study presents a comprehensive investigation of the generation and mitigation for unbalanced RFFs in open-neutral HSVs. Through detailed CFD simulations, the research reveals that RFF is primarily caused by asymmetric pressure distribution across the spool shoulder surfaces, especially in regions with strong vortex formation and abrupt geometric transitions. The viscous component, while present, is comparatively minor. A novel structural strategy introducing through-holes in the mid-shoulder region was proposed to alleviate this imbalance. These through-holes create a preferential flow path that effectively redistributes pressure, promotes local flow symmetry, and reduces radial pressure differentials. Among the tested configurations, the 15° inclined through-hole model demonstrated the most effective performance, reducing local middle shoulder RFF by up to 45.5% and total spool RFF by 29.9%.

In contrast to traditional groove-type pressure balancing methods, the proposed design is more adaptable to compact valve architectures and complex multi-step throttling geometries. It also avoids secondary pressure peaks often introduced by overly steep hole angles, thus ensuring stable flow transition and minimal disturbance. Experimental validation of the proposed through-hole design will be conducted in a subsequent stage of this research to further assess its effectiveness under practical operating conditions.

6. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 52275008); and the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (Grant No. SML 2024SP007).

7. REFERENCES

[1] T. Lin, Y. Lin, H. Ren, H. Chen, Q. Chen, and Z. Li, ‘Development and key technologies of pure electric construction machinery’, Renew. Sustain. Energy Rev., vol. 132, pp. 110080, Oct. 2020.

[2] G. Zhang, K. Zhang, W. Shen, G. Xu, and Y. J. Lin, ‘Intelligent fault diagnosis of multi-way directional valves in hydraulic systems using digital twin and deep learning approaches’, Mech. Syst. Signal Process., vol. 230, pp. 112579, May 2025.
[3] L. Romagnuolo, R. De Rosa, E. Frosina, and A. Senatore, ‘Study of a proportional spool valve noise by means of functional data analysis: cavitation and intensity detection’, Mech. Syst. Signal Process., vol. 209, pp. 111100, Mar. 2024.
[4] W. Li, H. Jiang, C. Feng, Z. Wang, S. Ma, and X. Xu, ‘A hybrid intelligent diagnostic approach for spool jamming faults of hydraulic directional valves’, Measurement, vol. 241, pp. 115706, Feb. 2025.
[5] D. Li et al., ‘A prediction model of fluid–solid erosion wear in hydraulic spool valve orifice’, Wear, vol. 540–541, pp. 205235, Mar. 2024.
[6] M. Simic and N. Herakovic, ‘Reduction of the flow forces in a small hydraulic seat valve as alternative approach to improve the valve characteristics,’ Energy Conversion and Management, vol. 89, pp. 708–718, Jan. 2015.
[7] R. Amirante, E. Distaso, and P. Tamburrano, ‘Sliding spool design for reducing the actuation forces in direct operated proportional directional valves: Experimental validation’, Energy Conversion and Management, vol. 119, pp. 399–410, Jul. 2016.
[8] M. Yang, Y. Zhang, C. Ai, G. Yan, and W. Jiang, ‘Multi-objective optimisation of K-shape notch multi-way spool valve using CFD analysis, discharge area parameter model, and NSGA-II algorithm’, Eng. Appl. Comput. Fluid Mech., vol. 17, no. 1, pp. 2242721, Aug. 2023.
[9] Y. Zhang, S. Wang, J. Shi, and X. Wang, ‘Evaluation of thermal effects on temperature-sensitive operating force of flow servo valve for fuel metering unit’, Chin. J. Aeronaut., vol. 33, no. 6, pp. 1812–1823, 2020.
[10] W. Yang et al., ‘CFD-based axial steady-state hydrodynamic study and structural optimization of miniature switching valve’, Flow Measurement and Instrumentation, vol. 100, pp. 102744, Dec. 2024.
[11] S. H. Hong and K. W. Kim, ‘A new type groove for hydraulic spool valve’, Tribology International, vol. 103, pp. 629–640, Nov. 2016.
[12] T. A. Zaki, ‘Turbulence from an Observer Perspective,’ Annu. Rev. Fluid Mech., vol. 57, no. 1, pp. 311–334, Jan. 2025.
[13] S. Zhang, Z. Lin, S. Chen, Z. Jin, and J. Qian, ‘Unbalanced force analysis on valve core clamping in hydraulic spool valves’, Annals of Nuclear Energy, vol. 203, pp. 110505, Aug. 2024.
[14] Y. Zhang, B. Ding, and G. Yan, ‘Reduced-order estimation and optimisation of claw-shaped throttling notch via CFD analysis, surrogate models, and CCMO algorithm’, Eng. Appl. Comput. Fluid Mech., vol. 19, no. 1, p. 2444418, Dec. 2025.
[15] K. Zhang, D. Wu, J. Wang, and Y. Song, ‘Numerical and experimental investigation of cavitation characteristics of the electronic expansion valves’, Phys. Fluids, vol. 36, no. 4, pp. 45144, Apr. 2024.

<DOI>

Biographies



Yuhang Zhang received the B.Eng degree from Nanjing Institute of Technology, Nanjing, China in 2021 and the M.Eng degree from Yanshan University, Qinhuangdao, China in 2024. He is currently working toward the Ph.D. degree with Sun Yat-sen University, Shenzhen, China. His research interests include cavitation erosion of high-pressure flows in confined spaces, and high-performance hydraulic components.



Shipei Zheng received the B.Eng. degree in mechanical engineering from Wuhan University of Technology, Wuhan, China in 2020 and the M.Eng. degree in robotics engineering from King's College London, U.K., in 2021. He is currently working toward the Ph.D. degree with Sun Yat-sen University, Shenzhen, China. His research interests include electrical machines and drive, motion control and robot control.



Jiaxi Zhang received the B.Eng. degree in mechanical engineering from Sun Yat-sen University, China, in 2021. He is currently working toward the M.Eng. degree with Sun Yat-sen University, Shenzhen, China. His research interests include cavitation erosion of high-pressure flows in confined spaces, and high-performance hydraulic components.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International



Jianshe Feng received the Ph.D. degree from University of Cincinnati, USA, in 2020. He is currently an Associate Professor with the School of Advanced Manufacturing, Sun Yat-sen University, Shenzhen, China. His research interests focus on industrial artificial intelligence, intelligent monitoring and optimization of manufacturing processes, predictive maintenance, and industrial big data analytics.



Miao Yang received the Ph.D. degree in mechatronic engineering from Harbin Institute of Technology, Harbin, China, in 2019. He is currently an Associate Professor with the School of Advanced Manufacturing, Sun Yat-sen University, Shenzhen, China. His research interests include compliant mechanisms, precision motion systems, parallel and robotic manipulators, robot force control, and mechatronic system integration.



Guangwei Zhang received the Ph.D. degree in mechanical engineering from the University of Tokyo, Tokyo, Japan, in 2019. He is currently an Associate Professor with the School of Advanced Manufacturing, Sun Yat-sen University, Shenzhen, China. His research interests include electrostatic film actuators and motors, precision motion control, artificial muscles, and soft robotic systems.



Beichen Ding received the Ph.D. degree in mechanical engineering from the University of Bath, Bath, U.K., in 2019. Currently, he is an Associate Professor with the School of Advanced Manufacturing, Sun Yat-sen University, Shenzhen, China. He is a member of the Intelligent Fluid Power Control of CSME. His research interests include actuation and motion control of smart machine systems and robot-assisted machining and dynamic modeling of electrical drives.

Modelling and Analysis of the Performance for a Pressure Compensated Flow Control Valve

K. Bhaskar[1,2], N. Johnston[1], A. Plummer[1], S. Westwood[2]

1. University of Bath, 2. Hydraulic System Products Ltd

Abstract.

Pressure compensated flow control valves are an essential requirement in the areas where flow control is necessary to achieve constant speeds regardless of the change in pressure difference or the load on the system. However, maintaining a constant flow rate remains a challenge with increasing pressure drop which directly impacts the performance of the system. The aim of this study is to integrate numerical modelling into the design process for a hydraulic valve in order to predict and optimise its performance. The example of a pressure compensated flow control valve is presented in this paper. The effects of flow forces and flow coefficients on the pressure compensation and flow control are examined. Computational Fluid Dynamics (CFD) models were created and used to obtain the flow forces and flow coefficients for different designs and spool displacements under constant flow rate conditions, and a system model was created in order to examine the performance, incorporating coefficients predicted from the CFD study. Lastly, experimental studies were carried out with the proposed design changes and their impact on the flow control across different pressures were studied.

Keywords. Flow forces, CFD, mathematical model, Orifice, Pressure compensated flow control valve

1. INTRODUCTION

Pressure-compensated flow control valves play an important role in the speed control of hydraulic systems. The output flow rate remains nearly constant, regardless of variations in the inlet pressure. The constant rate of flow is ensured by the adjustment of orifice area depending on the fluctuations in the system pressure. However, it is quite challenging to design a simple, stable valve which maintains the flow within a specified tolerance. As the pressure increases, the compensation by reduction of the orifice area introduces flow forces that act on the displacing element such as a spool or a poppet and contribute to deviations in the flow rate. These flow forces arise due to low pressure regions around the exit of the flow and directly influence the performance characteristics of the valve. Accurate prediction of these flow forces requires combining mathematical models, CFD simulations and validation data from experimental tests. Earlier analytical studies on a system level laid the fundamental approach for the modelling of compensated valves. Weber, Johnston and Edge

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

[1][2] developed steady state CFD models that were used to extract the effective jet angles and discharge coefficients for use in valve dynamic models. The jet angle was identified as the crucial parameter that led to significant changes in the flow force acting on the spool. Faulkner, Johnston and Weber [3] applied this methodology to pressure cartridge valves showing that the geometric parameters such as chamfers and introduction of fillets in place of sharp corners altered the discharge coefficients. The research also emphasised the sensitivity of valve performance to manufacturing tolerances.

Cheng [4] proposed analytical models linking the damping orifice flow characteristics to the transient response time. Parameters such as the spring constant and the pre-compressed length of the spring directly affect the steady flow rate, demonstrating how the behaviour of the flow forces influences compensation accuracy. Borghi et al. [5] conducted CFD-based numerical analysis and experimental studies to determine axial flow forces on compensated spool valves. The dependency of the flow forces on the geometry was modelled and verified through measurements from experimental studies. The Coanda effect was noted to be a significant influence on the actual flow at small openings of the valve. CFD studies and experiments on spool valves by Heraković et al. [6] concluded that the implementation of simple changes to a combination of geometrical parameters of the spool reduced the axial flow forces in the valve and helped eliminate the non-linear digressive flow-force characteristics. The CFD results agreed well with the experimental results. Further studies using CFD and mathematical models implemented in Matlab/Simulink were carried out for a proportional flow control valve by Lisowski et al. [7]. The study explored a triangular shape for the spool opening which reduced the dynamic flow forces acting on the spool. The jet angle was also studied and found to vary with the throttle gap width of the opening area; it was concluded that the functional relationship between the two may be obtained with an approximation.

Mao et al. [8] analysed a bidirectional throttling slide valve using CFD and experiments. It was found that the asymmetric vortices formed caused unequal steady flow-force curves and a guiding shoulder with an annular cavity was proposed that effectively reduced the asymmetry in the flow forces. Altare, Rundo and Olivetti [9] coupled the 3D CFD studies with a 0D dynamic model to study the effects of flow force compensation on a conical poppet pressure relief valve. The research experimentally explored the use of a deflector geometry for reducing the flow force. The data extracted were supplied as an input to a lumped parameter model of the valve. The study highlights the important role of the relation between the geometry, the flow forces and the deflection of the flow jet. Amirante et al. [10] described a CFD analysis of flow forces in a hydraulic proportional valve with circumferential notches. The CFD predictions were validated with experimental flow rate data. A reduction in the flow forces was observed at large valve openings, but smaller benefits were obtained at small openings. There was a decrease in the discharge coefficient with the valve opening due to changes in the flow direction. The study further shows that alterations to the spool profile could help in further reductions of the flow forces.

Across the studies, flow forces are shown to have a direct effect on the flow performance of the valve, and the jet angle as well as the geometry emerge as the principal contributors. The present study adopts an approach for the estimation of the flow forces and the flow coefficients and combining them with a mathematical model using Matlab/Simulink to optimise the performance of the pressure compensated flow control valve.

2. DESCRIPTION OF THE VALVE MODEL

The pressure compensated flow control valve examined in this study was developed jointly by Hydraulic System Products Ltd and University of Bath. The valve is of cartridge type with a flow control range between 20 lpm to 30 lpm and a working pressure up to 250 bar. The valve consists of a main body with large radial holes as the outlet, a sliding spool with smaller radial holes which would act as the variable metering orifice, a needle and a sliding 'hat' that would form the fixed orifice. There is an adjustable spring to set the flow limit manually. Figure 1 shows the cross-section of the valve.

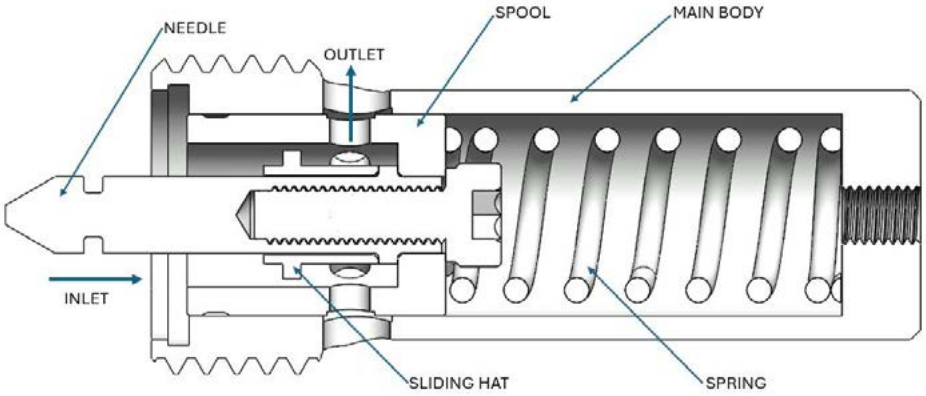


Figure 1: Cross-section of the pressure compensated flow control valve.

3. ANALYSIS OF FLOW USING CFD MODELLING

3.1. CFD model for the geometry and flow paths

A parametric CFD model was developed by tracing the flow path from the CAD model. The model consists of an inlet, radial outlets for the spool and the radial outlets for the main body. The interface between the radial outlets of the spool and the body are connected by a 0.2 mm deep groove. The spool displacements are set from 0 mm from the fully open, normal position as shown in Figure 1, to 3.25 mm which leaves a smaller orifice opening of 0.15 mm, with the displacement of the spool for closing the orifice completely being 3.4 mm. Separate CFD simulations are conducted for each opening of the valve based on the spool displacement under steady flow conditions to test the entire range of the valve opening. A quarter model was used for the CFD studies as the model is axisymmetric, and this saves computational resources and time. Figure 2 shows the quarter model used in the CFD studies.

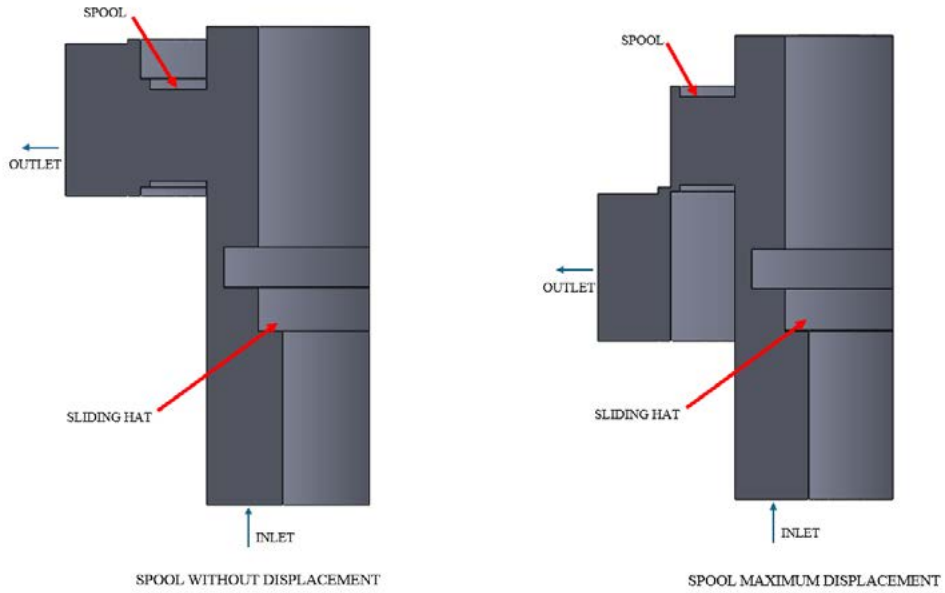


Figure 2: CFD quarter model

3.2. Mesh and mesh settings.

The meshing of the model and the CFD studies were carried out using the ANSYS FLUENT software which also enabled the application of local mesh refinements. The smallest element size was limited to 0.005mm with the largest size being 0.08mm. 5 boundary layers were added to the wall regions with an initial height of 0.005mm for better resolution near the boundary layers. Each model consisted of about 1.4 million poly-hexcore cells with a maximum skewness of 0.56 and a minimum orthogonal quality of 0.5. Figure 3 shows the mesh for the quarter model of the flow path.

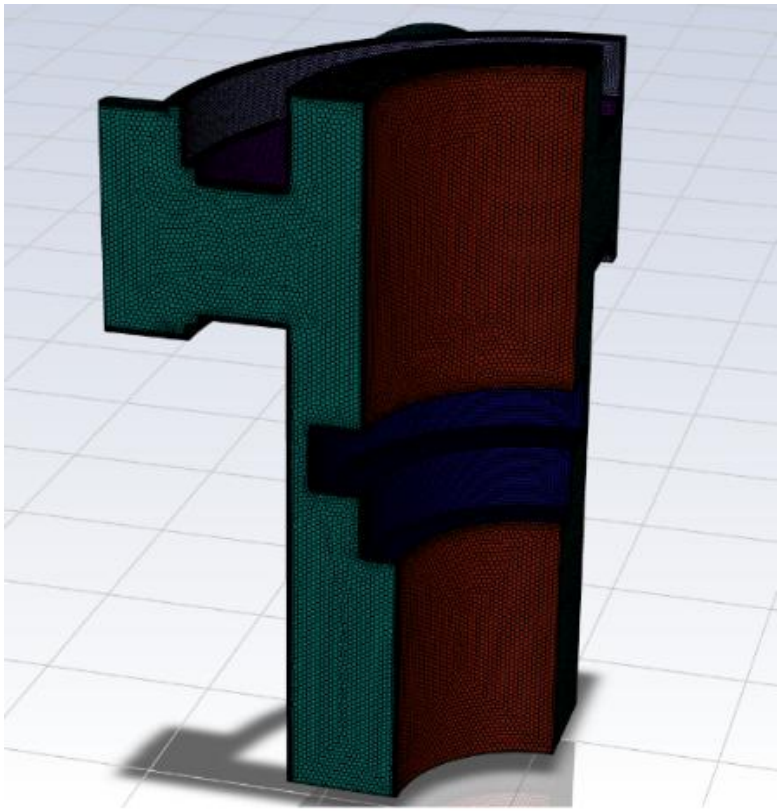


Figure 3: Mesh geometry for the CFD quarter model

3.3. CFD model parameters

A pressure-based model was used for the studies. There are many different turbulence models which can be used for the flow through the valve. The 'Realisable k-ε' model is equipped with enhanced wall treatments and considers the pressure gradient effects. Several CFD studies including [1], [7], and [10] support the use of this model for internal flow through valves, and so this model was chosen.

This turbulence model has been validated thoroughly for a wide range of flows and has been found to perform better than the standard k-ε model [11].

The fluid viscosity is set to 32 centistokes with a density of 860 kg/m^3. The inlet velocity is set at 11.7 m/s while the pressure at the outlet is set to 0 Pa gauge pressure. The convergence criteria were set to 10^{-4} for the momentum and the forces acting on the spool surface. It was found to be difficult to obtain valid, converged CFD results for openings less than 0.15 mm.

3.4. CFD results

The CFD results provided pressure and velocity distributions throughout the flow path. The pressures acting on the spool, the forces acting on the spool, the pressure drops across the orifices and the flow rate at the outlet were extracted from the CFD simulations. The pressure drops across each orifice were determined from the CFD pressure distribution plots.

The total force acting on the spool extracted from the CFD is the resultant force which includes the flow force. The flow force (F_f) is calculated by subtracting the pressure force from the resultant force (F_T) on the spool as shown in equation 1. Figure 5 shows a plot of flow force versus spool displacement, where the displacement of the spool increases as the pressure increases.

$$F_f = F_T - PA_s \qquad (1)$$

Where F_f is the total force acting on the spool, F_T is the total resultant force acting on the spool, P is the pressure acting on the spool at the inlet and A_s is the diametrical area of the spool. Whilst this flow force is related to the axial momentum flux entering as well as leaving a control volume within the spool, it is convenient to relate it to the downstream radial orifice. The effective jet angle is calculated using the flow force relation given by the equation:

$$F_f = \rho Q V \cos\theta \qquad (2)$$

where ρ is the density of the fluid, Q is the flow rate at the exit of the valve, θ is the jet angle and V is the velocity of the flow through the orifice which is calculated by the equation:

$$V = \sqrt{\frac{2\Delta P}{\rho}} \qquad (3)$$

Here, ΔP is the pressure drop across the orifice. The flow coefficient C_q is calculated using the orifice equation (Equation 4) with an approximation of the orifice area (Equation 5) which resembles the lateral surface area of a cylinder formed by the internal grooves on the spool and the main body as shown in figure 4.

$$Q = C_q A_{orifice} \sqrt{\frac{2\Delta P}{\rho}} \qquad (4)$$

$$A_{orifice} = \pi D_o x \qquad (5)$$

Where, $A_{orifice}$ is the orifice area, D_o is the outer diameter of the spool and x is the opening displacement of the orifice. This equation for the area is an approximation to the true orifice area to establish a functional relation between the flow and the orifice area. Any error in the approximation is compensated for in the flow coefficient C_q.

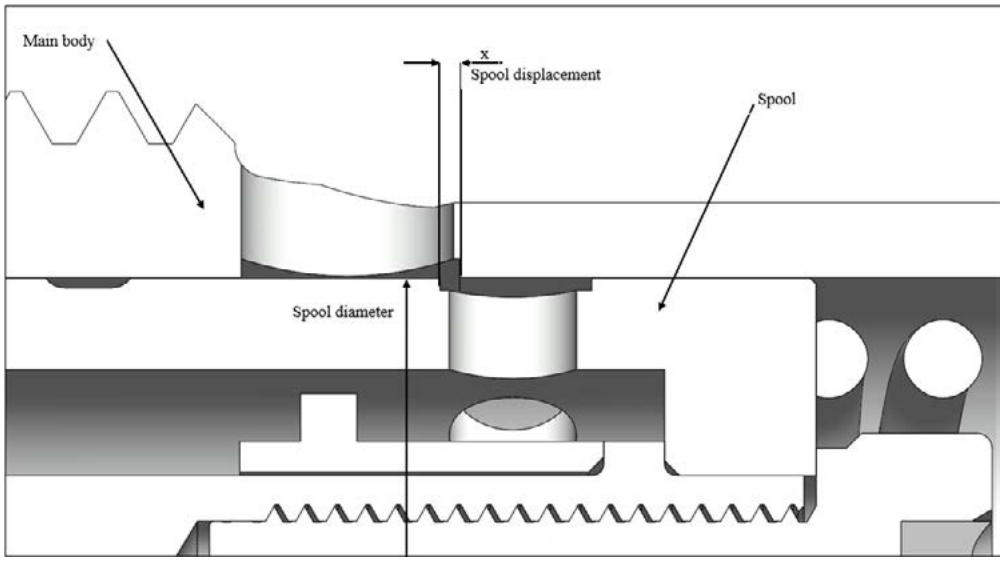


Figure 4: Flow path through the valve for the CFD model.

Figure 5 shows the plot of the flow force against spool stroke for different valve openings. Positive flow force indicates the force is acting from left to right in figure 4 helping to close the orifice, while negative flow force indicates the force is acting from right to left helping to open the orifice. It was observed that the flow force steadily increases with a decrease in the outlet orifice area, but for very small openings, the flow force dips to a negative value. This effect is largely due to the change in the jet angle (Figure 6) of the exit flow [1],[2],[7], except that for large openings the inlet momentum is significant.

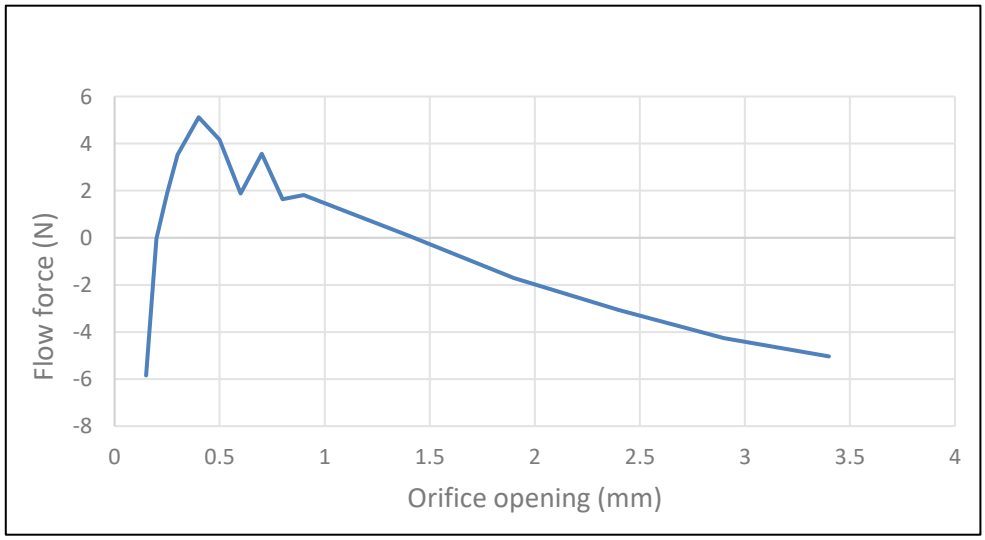


Figure 5: Spool v/s Orifice opening

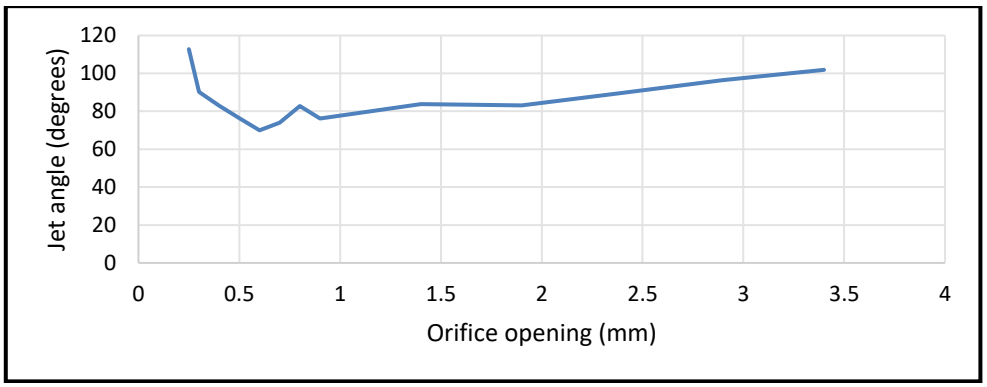
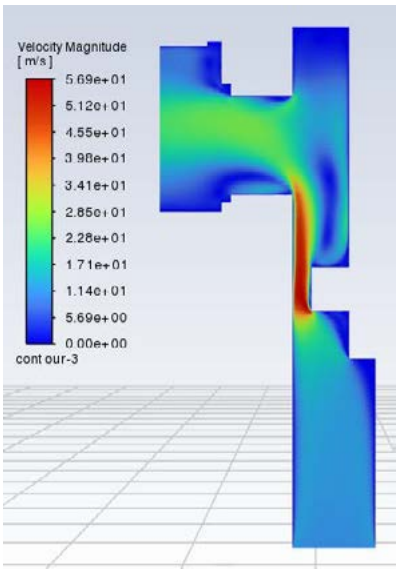
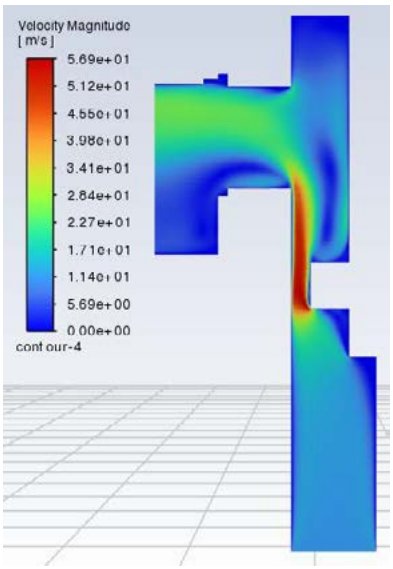


Figure 6: Jet angle v/s Orifice opening

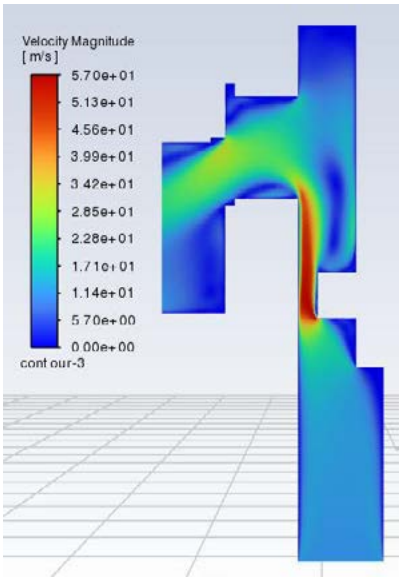
The figures from 7(a) through 7(f) show the velocity distribution of the flow through both the orifices. The angle of the jet is considered as the angle made from the inlet flow line to the line the exit jet follows at the exit of the second orifice.



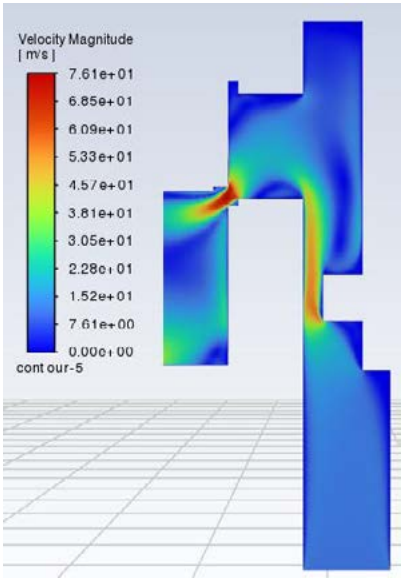
(a): Orifice fully open



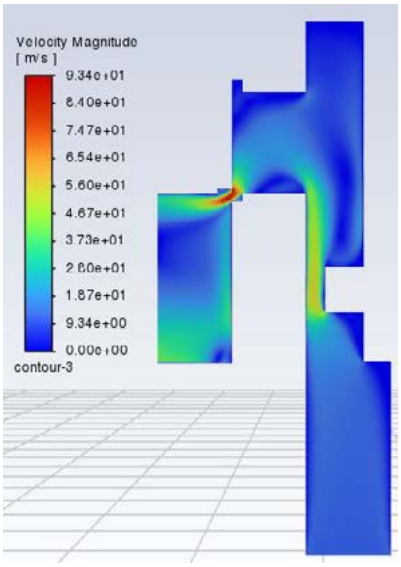
(b):Orifice 2.4mm open



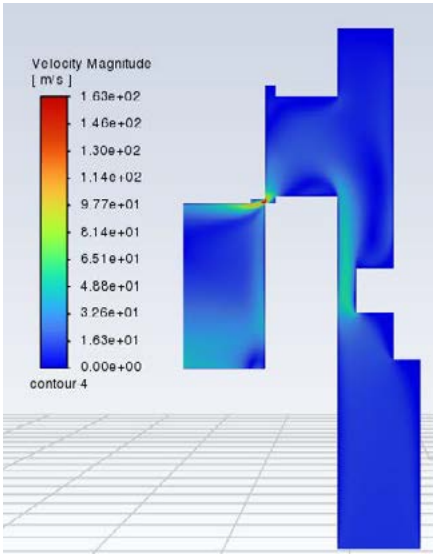
(c): Orifice 1.4mm open



(d): Orifice 0.4mm open



(e): Orifice 0.15mm open



(f): Orifice 0.1mm open

Figure 7: CFD flow patterns through a radial plane, for different openings.

It is observed from the velocity distribution figures that the jet angle starts off greater than 90° and reduces as the area of opening for the variable orifice reduces. The reduction in the jet angle increases the value of $\cos\theta$ in equation (2). This would indicate smaller quantity of flow forces for large opening of the valve. However, for openings less than 0.4 mm, the jet angle can be seen to increase towards the radial direction, and this correlates with the increase in angle in figure 6. It should be noted that the orifice geometry and jet patterns are

complex and three-dimensional because of the radial holes in the spool and housing (figure 3), so the 2D plots in figures 7(a) to 7(f) do not tell the full story.

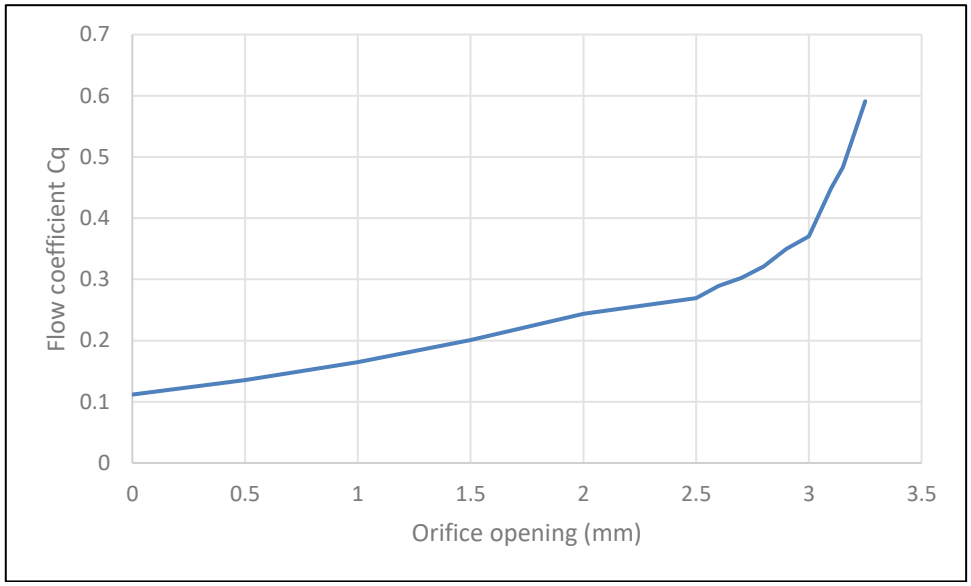


Figure 8: Flow coefficient vs orifice opening

Figure 8 shows the flow coefficients calculated from equation (4) plotted against the opening of the orifice which is a result of the displacement of the spool. The flow coefficient is calculated for flows across the variable orifice. The value of the flow coefficient C_q is observed to increase steadily as the area of opening for the decreases. This change observed is mainly due to the approximation in the calculation of the orifice area. The values of the jet angle and the flow coefficients were input as a table into the mathematical model.

4. MATHEMATICAL MODEL

A mathematical model was developed and implemented in Matlab® Simulink. The model consists of a pressure ramp at the inlet as the input, pressure drop across the input (sliding hat) orifice as well as the variable orifice, a spring-mass system to derive the displacement of the spool, a leakage model and the flow at the outlet of the second orifice. Figure 9 shows the block diagram for the mathematical model.

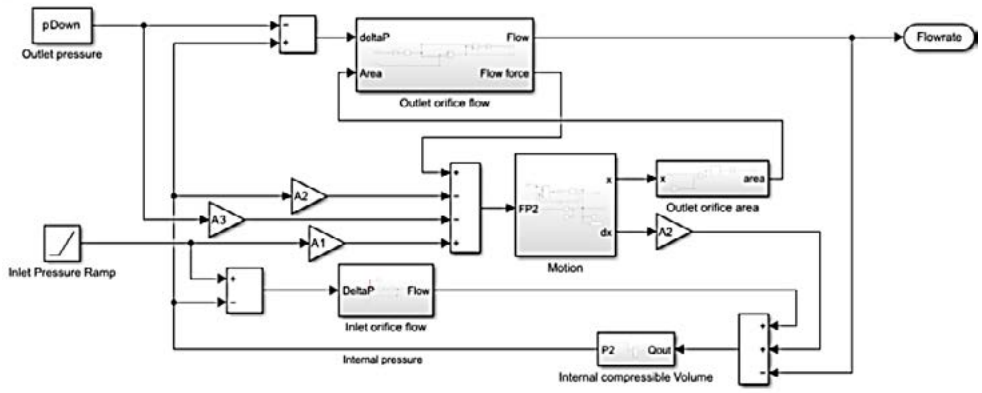


Figure 9: Matlab® Simulink block diagram of valve model

The pressure ramp values are converted to force acting on the sliding hat and the spool which pushes the spring. The displacement of the spool is derived with the equation

$$F = m\ddot{x} + c\dot{x} + kx \qquad (6)$$

Where, m is the mass of the spool and the sliding hat, c is the damping coefficient, k is the spring stiffness, x is the displacement of the spool and $\ddot{x}$ and $\dot{x}$ represent the acceleration and the velocity of the spool. The inputs for the flow coefficient and the jet angle were taken from the CFD results and compiled into a 1-D table for different points of displacement and approximate area value for the variable orifice. The flow force is calculated using the equation (1) and is subtracted from the pressure force acting from equation (6). The flow across the orifice is calculated using the equation (4). A curve for the flow vs the pressure drop is obtained from the Simulink model shown in figure 10, for upward and downward pressure ramps.

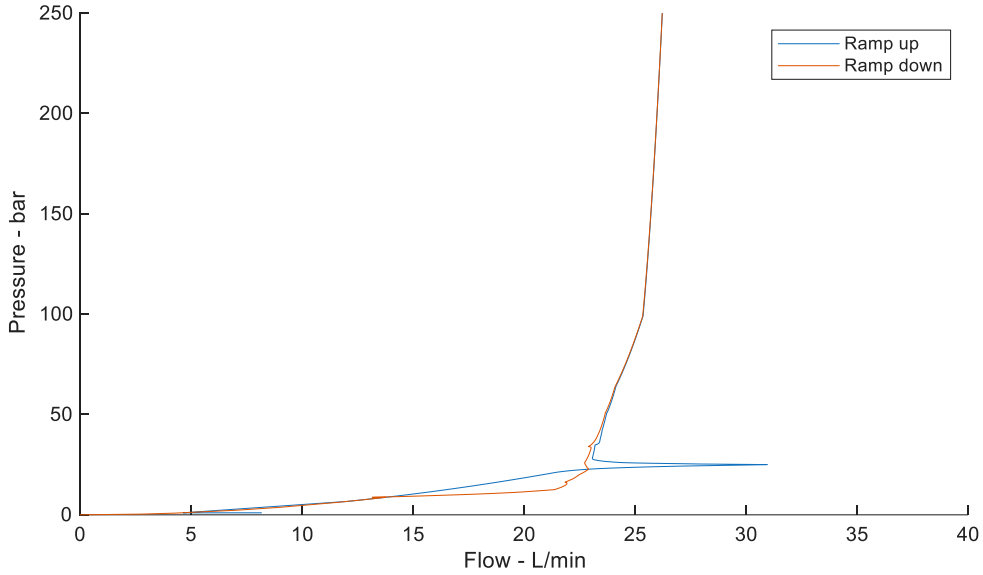


Figure 10 – predicted pressure-flow characteristics

From the figure, it is observed that for pressures above about 30 bar the valve meters the flow close to the setpoint of 25 L/min as desired, with a slight increase in flow with pressure. It was found that this variation was quite sensitive to the flow force. For pressures below about 30 bar there is insufficient pressure force to compress the spring, so the valve is fully open, and the flow is not controlled. There is some hysteresis apparent between about 10 bar and 30 bar, and for increasing pressure ("ramp up") a large spike is observed in the flow. This is because the predicted flow coefficient (figure 7) decreases with opening for large openings, such that as the spool starts to move in the direction to close the orifice, the flow resistance decreases resulting in a spike in flow for increasing pressure. This spike is not observed for decreasing pressure.

5. EXPERIMENTAL RESULTS

The experimental rig was setup at the premises of Hydraulic System Products Ltd. The rig was driven by an axial piston pump capable of achieving a maximum flow of 300 lpm and a maximum pressure of 350 bar. The measurements for the flow and the pressure were recorded using a Webtec HPM4000 data logger with a sampling rate of 60 Hz. The schematic of the test rig and assembly of the valve is shown in figure (11).

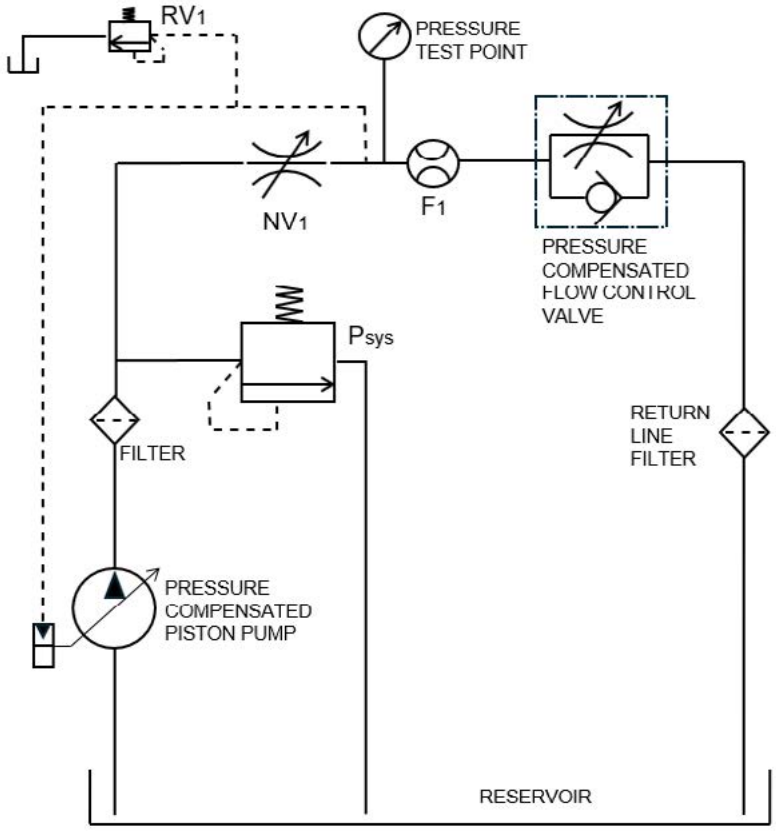


Figure 11: Schematic diagram of the hydraulic test rig

The oil is pumped from the reservoir by the piston pump through the system relief valve P_{sys} and a flowmeter F_1 into the valve. The flow is controlled by the needle valve NV_1 and the pressure through a restrictor valve RV_1. The outlet of the valve is connected back through the tank. The image of the valve on the test rig is shown in figure (12).

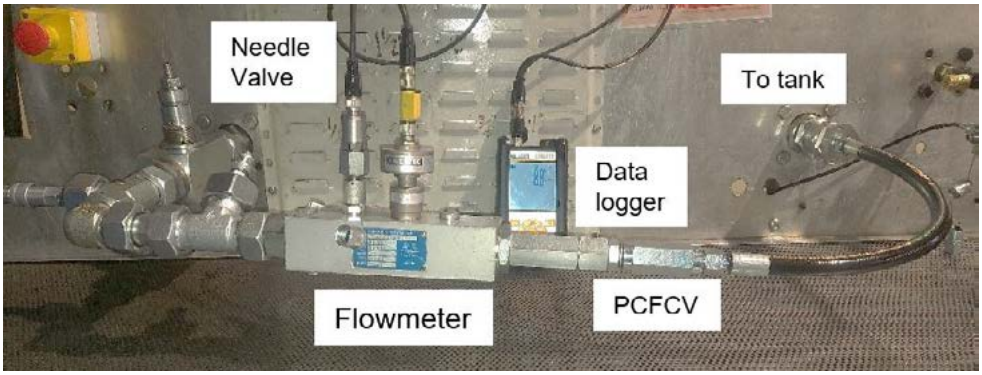


Figure 12: Hydraulic test rig at Hydraulic System Products LTD

The pressure was increased in small increments from 0 to 250 bar while recording the values of the pressure and flow. The recorded data form the data logger was used to plot the graph of flow versus the pressure as seen in figure (13).

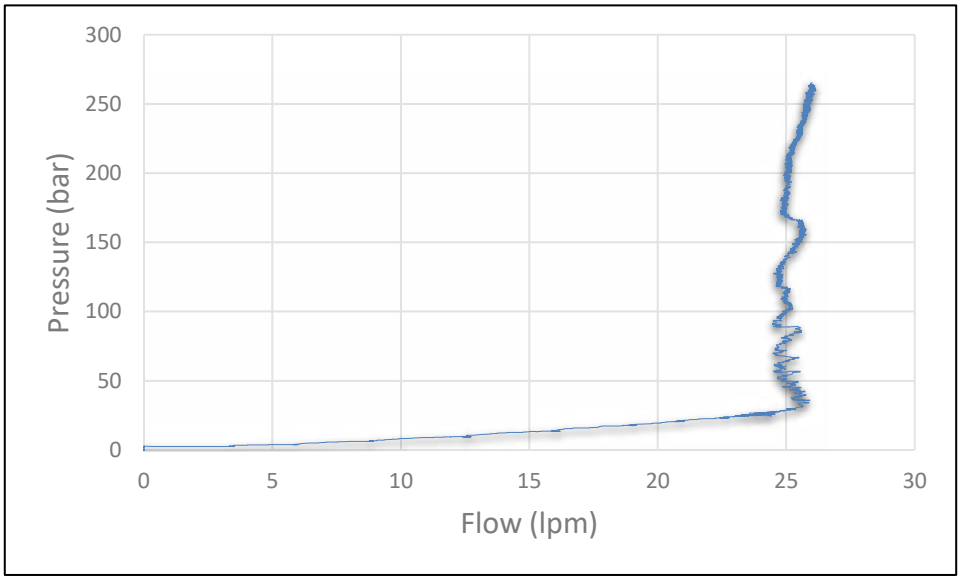


Figure 33: Experimental results - flow v/s pressure.

From the experimental results, it is observed that the behaviour of the valve follows closely with the results of the mathematical model. The flow is found to increase as the pressure increases and a slight overshoot at the set flow. The slight increase in the flow until the maximum pressure of 250 bar is observed. The spike in flow and hysteresis in the 10-30 bar region were not observed in the experimental results. The deviation from the set constant flow suggests the flow forces impacting the performance of the valve at higher inlet pressures and lower area of the exit orifice.

6. CONCLUSION

A method to integrate Computational Fluid dynamics (CFD) and mathematical modelling approach for analysing and optimising the performance for a pressure compensated flow control valve is presented. The study demonstrates that combining fluid parameters form the CFD simulations with a mathematical model provides an accurate and reliable method of predicting the valve behaviour. The CFD analysis quantified the forces acting on the spool which allowed for the calculations of the flow force and the jet angle. The results show that the flow force increases as the orifice area reduces but may reverse the direction at very small openings due to an increase in the jet angle. The calculated jet angles and the flow coefficients were integrated into the Simulink based mathematical model which produced the compensation behaviour of the valve, predicting the gradual increase in flow with an increase in the pressure. Experimental testing confirmed the accuracy of the model with the measured characteristics of the flow v/s pressure curve closely matching the results of the mathematical model. The findings show that the jet angles and the flow forces were the core contributors to the deviations from the ideal constant flows at higher pressures, and the predicted valve performance was very sensitive to the flow force.

The mathematical modelling was a key part of the design process. It highlighted the importance of flow force on the valve performance and helped to produce a design in which the flow force was reduced resulting in a good pressure-flow characteristic. Overall, this methodology offers a design framework for prediction of flow performance with the jet angle, flow force and the flow coefficient that would aid in optimising the valve geometry for future designs of the pressure compensated flow control valve. Future work will expand on the current methodology to parametrise the design and optimise the orifice geometry.

7. REFERENCES

[1] Weber, S.T., Johnston, D.N., and Edge, K.A., Mathematical modelling of a load control valve, *3 Iternationales Fluidtechnisches Kolloquium 2002*, Aachen, March 2002

[2] Weber S.T., Johnston D.N., Edge K.A. Recent Research into Cartridge-Style Hydraulic Valves. Bath Workshop on Power Transmission and Motion Control, PTMC 2000, Bath, Sept 2000

[3] Faulkner, J., Johnston, D.N., Weber, S.T., Computer Modeling of a Pressure Relief Cartridge Valve, Paper presented at *International Fluid Power Exposition*, Las Vegas, March 2011

[4] Borghi M., Milani M., Paoluzzi R. (2000). "Stationary axial flow-force analysis on compensated spool valves." *International Journal of Fluid Power*, 1(1), 17–25. DOI: 10.1080/14399776.2000.10781079.

[5] Cheng M. (2005). "Modeling and Analysis of a Pressure Compensated Flow Control Valve." *ASME FEDSM 2005 Proceedings.* DOI: 10.1115/FEDSM2005-77057.

[6] Heraković N., Dular M., et al. (2019). “Experimental and CFD analysis of flow in hydraulic valves.” *Energy Conversion and Management*, 196, 16–27. DOI: 10.1016/j.enconman.2019.05.010.

[7] Lisowski E., Filo G. (2016). “CFD analysis of the characteristics of a proportional flow control valve with an innovative opening shape.” *Energy Conversion and Management*, 123, 15–28. DOI: 10.1016/j.enconman.2016.06.025.

[8] Mao Q., Jia X., Liu Z., Li G., Cao Y., Yang Q. (2024). “Study on Steady Flow Force of a Bidirectional Throttling Slide Valve and Its Compensation Optimization.” *Applied Sciences*, 14, 11037. DOI: 10.3390/app142311037.

[9] Altare G., Rundo M., Olivetti M. (2016). “3D Dynamic Simulation of a Flow Force Compensated Pressure Relief Valve.” *ASME IMECE 2016 Proceedings.* DOI: 10.1115/IMECE2016-65624.

[10] Amirante R., Moscatelli P.G., Catalano L.A. (2007). “Evaluation of the flow forces on a direct (single-stage) proportional valve by means of CFD analysis.” *Energy Conversion and Management*, 48, 942–953. DOI: 10.1016/j.enconman.2006.08.024.

[11] ANSYS/Fluent 2025 R2: Users guide

Thermal-fluid coupling analysis of the Deflector Jet Servo Valve: effects of internal air region, oil viscosity, and coil heat source

Junhao Chen, Junru Hui, Mingjie Li*, Xiaoping Ouyang*

State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310030, China

**Corresponding author: E-mail address: mingjielee@zju.edu.cn, ouyangxp@zju.edu.cn*

Abstract.

The thermal behavior of the deflector jet servo valve (DJSV) is influenced by complex interactions among electromagnetic, fluid, and thermal fields, making accurate temperature prediction challenging under wide operating conditions. However, most existing simulations rely on simplified assumptions (such as omitting the internal air region, assuming constant oil viscosity, or ignoring the coil heat source) without quantitatively assessing their effects on the predicted thermal field. In this study, a three-dimensional steady-state thermal-fluid coupled model of a DJSV is developed to evaluate the influence of these modeling simplifications. Six simulation schemes are designed by varying the treatment of the internal air domain, oil viscosity-temperature dependence, and coil heat generation, and are assessed under six representative operating conditions covering combinations of inlet oil temperature, ambient temperature, and pressure. Results show that omitting the internal air region mainly affects electromagnetic components, while neglecting temperature-dependent oil viscosity and coil heating causes distributed deviations across multiple parts. The coil and armature exhibit the highest thermal sensitivity. These findings clarify the thermal implications of modeling simplifications and provide reference for selecting appropriate model fidelity in engineering-oriented DJSV analyses.

Keywords. Deflector jet servo valve, thermal-fluid coupling, temperature field prediction, temperature dependence

1. INTRODUCTION

As a core component of electro-hydraulic servo control systems, the deflector jet servo valve (DJSV) critically determines system precision and reliability. In aerospace applications, hydraulic oil temperatures can vary widely, from below -50 °C to above 120 °C, and may even exceed 200 °C in high-performance systems. Excessive temperature rise can lead to severe consequences, such as spool jamming and degradation of electromagnetic components. Therefore, accurate prediction of the temperature field is essential for effective thermal management and reliable system design.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Numerous studies have investigated thermal effects in servo valves. Li et al. [1] developed a thermodynamic model that takes throttling losses and convective heat exchange into account and has been experimentally validated. Zhang et al. [2] applied a heat-fluid-solid FEM and found that thermal deformation of the spool-sleeve pair reduced the clearance and caused clamping above 80 °C. The zero-drift model in [3] attributed over 15 % control error (20-270 °C) to variations in flow coefficient and magnetic reluctance. An AMESim-based model [4] demonstrated that temperature reached equilibrium within 2.5 hours, with pressure having a strong influence. Recent efforts have focused on DJSVs. A nonlinear ninth-order model [5] demonstrated that temperature-induced flow-coefficient variation and pilot-stage deformation jointly degrade accuracy by over 30 %. Temperature-dependent flow-force evolution was quantified in [6], and wide-temperature modeling [7] linked magnetic-flux loss and air-gap deflection to temperature drift. Other studies [8-12] linked thermal friction and clearance reduction to jamming, demonstrated that reduced viscosity enhances self-excited oscillations, and clarified the leakage-clamping trade-offs at elevated temperatures. Collectively, these works confirmed that oil temperature significantly affects valve dynamics and stability.

However, most existing thermal analyses emphasize the hydraulic stage while neglecting the torque motor and its internal air region. The internal air region is often omitted or solidified, overlooking its role in convective heat transfer. Moreover, many models are confined to zero-opening conditions, where oil flow and heat convection are minimal, and oil viscosity is treated as constant rather than temperature-dependent. Such simplifications limit the accuracy of temperature-field prediction, particularly for electromagnetic components under realistic operating states. To address these limitations, this study establishes a three-dimensional steady-state thermal-fluid coupled model of a DJSV. Six simulation schemes with different simplifications, regarding the internal air region, oil-viscosity-temperature dependence, and coil heat generation, are evaluated under representative conditions to quantify their respective effects and clarify component-level thermal sensitivities.

2. THERMAL GENERATION AND HEAT TRANSFER IN DJSV

2.1. Operating Principle

The DJSV mainly consists of three functional subsystems: a torque motor, a jet prestage, and a power spool stage. The feedback rod passes through a pair of high-pressure jets and receiver ports on the jet pan, with V-shaped grooves machined on its surface. When an electric current flows through the coils, electromagnetic torque is generated in the magnetic circuit, producing a small angular deflection of the armature assembly. This deflection causes a displacement of the V-groove, which changes the pressure difference between the two receiver ports and consequently drives the spool motion in the power stage. During this energy conversion process, multiple heat generation and dissipation phenomena occur simultaneously. These include Joule heating within the coils, iron losses in the magnetic conductors, and throttling-induced viscous dissipation within the valve body. The generated heat is dissipated through several mechanisms: heat conduction, forced convection through hydraulic oil flow, and natural convection from the external surfaces of the valve to the surrounding air.

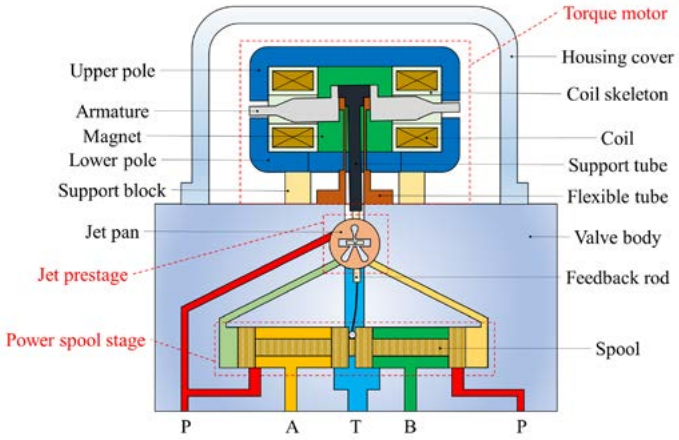


Figure 2.1. Structure of the deflector jet servo valve.

2.2. Heat Generation

(a) Electromagnetic Losses

When an electric current I passes through the coil, Joule heating occurs, with the instantaneous power expressed as:

$$P_{cu}=I^2R \tag{1}$$

The resistance of the coil varies with temperature according to:

$$R=R_0[1+\alpha(T-T_0)] \tag{2}$$

where R_0 is the resistance at the reference temperature T_0, and α is the temperature coefficient of resistivity. Assuming uniform current density and homogeneous heating, the winding region is treated as a uniformly distributed volumetric heat source. The volumetric heat generation rate is then expressed as:

$$q_{\mathrm{cu}}=\frac{I^2R}{V} \tag{3}$$

where V is the coil volume.

The present model neglects dynamic magnetic losses (iron losses) but incorporates the temperature-dependent copper loss to account for the resistivity-temperature coupling in the windings.

(b) Throttling-Induced Heat Generation

Viscous dissipation during throttling arises from viscosity-induced energy losses. As hydraulic oil flows through the orifice, the abrupt reduction in flow area accelerates the fluid and produces a high shear rate. Viscous stresses resist the velocity gradient, converting flow work and turbulent kinetic energy into internal energy and heat. The resulting temperature rise inside the valve due to viscous dissipation is given by:

$$\Delta T=\frac{\Phi}{\rho c_{\mathrm{p}}} \tag{4}$$

where ρ is the oil density, c_p is the specific heat at constant pressure, and Φ is the viscous dissipation function:

$$\Phi=\mu\{2[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial w}{\partial z}\right)^2]+\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}\right)^2+\left(\frac{\partial v}{\partial z}+\frac{\partial w}{\partial y}\right)^2\} \quad (5)$$

where μ is the fluid dynamic viscosity, u, v, w are the velocity components of the fluid in the x, y, z directions. The dynamic viscosity is treated as temperature-dependent, according to the fitted viscosity-temperature relation in Section 3.2.

2.3. Heat Transfer

(a) Forced Convection

Within the valve body, hydraulic oil flow driven by the pressure differential transfers heat away from internal components. The convective heat flux can be expressed by Newton's law of cooling:

$$q=h(T_w-T_f) \quad (6)$$

where h is the convective heat transfer coefficient, T_w is the wall temperature, and T_f is the fluid bulk temperature. The coefficient h can be estimated using a Nusselt number correlation of the form:

$$N_u=\frac{hL_c}{\lambda}=CR_e^mP_r^n \quad (7)$$

where L_c is the characteristic length, λ is the thermal conductivity, R_e is the Reynolds number, P_r is the Prandtl number, C, m, n are the empirical coefficients. The dynamic viscosity of the hydraulic oil varies exponentially with temperature and is described by:

$$\mu(T)=\mu_0e^{-b(T-T_0)} \quad (8)$$

where μ_0 is the viscosity at the reference temperature T_0, and b is the empirical constant.

(b) Natural Convection

Heat dissipation from the external surfaces of the valve to ambient air occurs primarily via natural convection, which can be described as:

$$q=h_{nat}(T_s-T_\infty) \quad (9)$$

where h_{nat} is the natural convection coefficient, T_s is the surface temperature, and T_∞ is the ambient air temperature.

(c) Heat Conduction and Internal Air Domain

Heat conduction follows the Fourier's law:

$$q=-k\nabla T \quad (10)$$

where k is the thermal conductivity, ∇T is the temperature gradient vector.

Within the enclosed housing, natural convection is limited, allowing the air to be approximated as a solid with low thermal conductivity.

3. NUMERICAL SIMULATION MODEL AND METHODOLOGY

3.1. Modeling Overview and Meshing strategy

The model preprocessing of the DJSV is shown in **Figure 3.1**, ignoring non functional details such as chamfers, screw holes, and fillets. Assuming an input current of 40 mA, the corresponding displacement of the spool valve is set to -0.5 mm, and the displacement of the jet prestage deflection is set to +0.025 mm. Consequently, the computed temperature field reflects a quasi-steady-state thermal balance at a representative operating point.

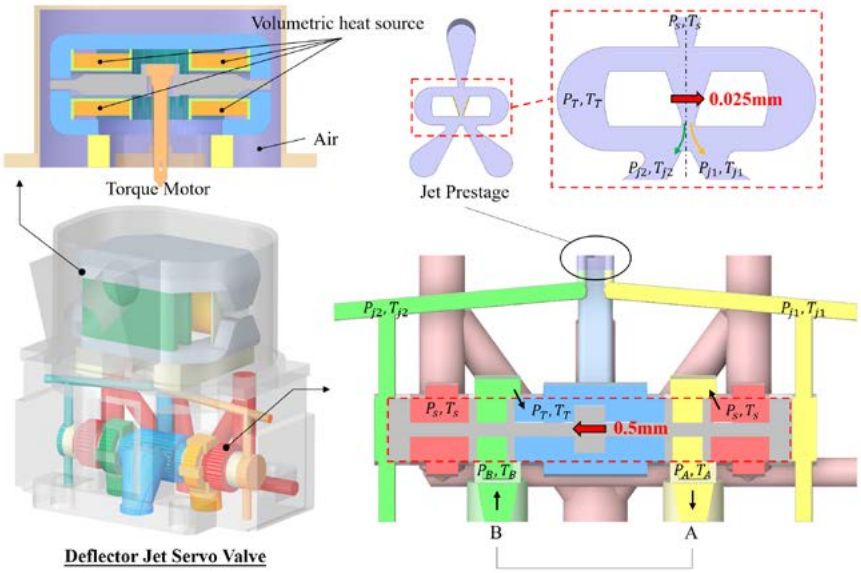


Figure 3.1. Model preprocessing of the DJSV

The mesh generation result for the entire valve is shown in **Figure 3.2**. Local refinement was applied to the orifices of the jet pilot stage and the spool, and the total number of grid cells reached 2 million.

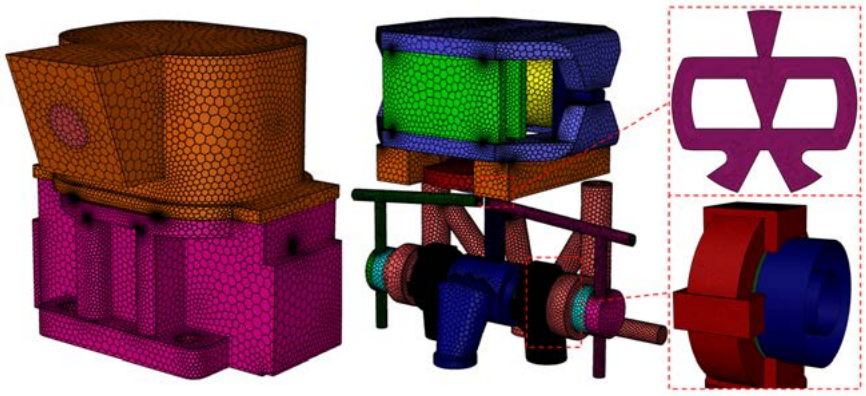
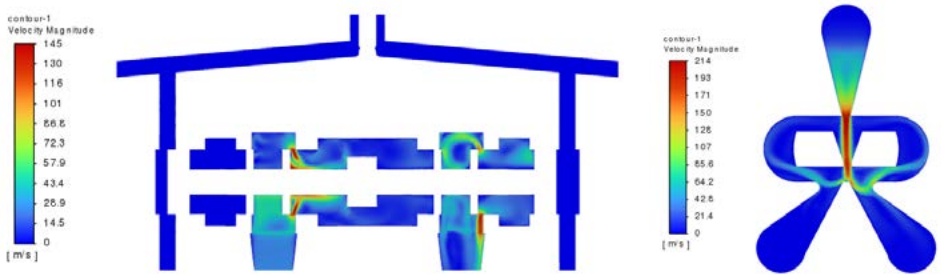


Figure 3.2. Mesh generation and local refinement

The velocity contour results in **Figure 3.3** confirm that the mesh adequately resolves the flow structures and velocity gradients within the critical regions, including the jet prestage and throttling orifice.



(a)valve body section (b) jet prestage section

Figure 3.3. Representative velocity contour results

3.2. Material Properties

The thermal physical properties are summarized in **Table 1**. The component "Solid-air" is the equivalent solid region.

Table 1: Materials settings

Component	Material	Density (kg/m^3)	Specific heat capacity (J/(kg·K))	Thermal conductivity (W/(m·K))
Housing cover Support block	0Cr18Ni9	7930	500	21.5
Upper pole Lower pole Armature	1J50	8100	460	15
Magnet	LNG32	7500	465	32.7
Support tube Feedback rod	3J1	8100	460	30
Flexible tube	Qbe2	8300	380	100
Spool	9Cr18	7700	470	29.3
Valve body	440C	7800	480	24.2
Coil	Copper	8930	385	390
Coil skeleton	Polyimide	1400	1150	0.25
Solid-air	-	1.205	1012	0.026
Oil	YH-15 aviation hydraulic oil	850	1880	0.12

The hydraulic medium is YH-15 aviation hydraulic oil, the temperature dependence of the oil viscosity was obtained from experimental data (see **Table 2**). A double-exponential correlation derived by least-squares fitting (R^2=0.99996) is expressed as:

$$\mu(\mathrm{T})=0.0017e^{-0.1182T}+0.0389e^{-0.0312T} \tag{11}$$

where $\mu(\mathrm{T})$ is the viscosity of the oil (kg/(m·s)), T is the temperature of the oil (℃).

Table 2: Viscosity test results of YH-15 aviation hydraulic oil [13]

Temperature/℃	Viscosity/(mm^2/s)	Temperature/℃	Viscosity/(mm^2/s)
-55	1620.993	60	9.683
-40	390.924	80	7.088
-20	111.064	100	5.579
0	42.734	120	4.460
20	22.482	135	4.027
40	13.928		

3.3. Boundary Conditions and Simulation Settings

The energy equation with viscous dissipation was activated, and the Realizable k-epsilon model was employed as the turbulence model. The boundary condition settings are as follows:

Table 3: Boundary condition settings

Boundary Location	Type	Parameter	Value
Port P	Pressure inlet	pressure temperature	**inlet_pressure** **inlet_temp**
Port A	Pressure outlet	pressure	(**inlet_pressure**+0.5MPa)/2
Port B	Pressure inlet	pressure temperature	(**inlet_pressure**+0.5MPa)/2 Outlet A temperature
Port T	Pressure outlet	pressure	0.5 MPa
Outer surfaces	Convection	temperature coefficient	**ambient_temp** 10 W/(m²·K)

All solid-fluid and solid-solid interfaces employed shared topology to ensure continuous heat flux. The temperature-dependent volumetric heat source of the coil was expressed as:

$$q_v(T)=(I^2\cdot\frac{234+T}{254}\cdot R_{20})/V=160000\cdot\frac{234+T}{254} \quad (12)$$

where $q_v(T)$ is the volumetric heat power (W/m^3), T is the temperature of coil (℃), I is the current (A), R_{20} is the resistance of the coil at 20℃ (Ω), V is the coil volume (m^3).

Table 4: Multiple operating conditions

Condition	**ambient_temp**/℃	**inlet_pressure**/MPa	**inlet_temp**/℃
C1	25	21	25
C2	25	35	25
C3	300	21	25
C4	25	21	135
C5	25	35	135
C6	300	21	135

As shown in **Table 4**, the boundary conditions cover six representative operating environments (C1-C6) to evaluate the modeling assumptions under normal, high-pressure, and high-temperature conditions. Case C3 (300 °C ambient, 25 °C inlet oil) reflects hot external exposure with a relatively low inlet-oil temperature, creating a high-ambient and low-supply contrast to quantify the oil-cooling effect and the resulting temperature gradients. To systematically assess the influence of modeling assumptions, six simulation

schemes (S1-S6) were constructed as shown in **Table 5**. Each scheme isolates one or more simplifications related to the internal air region, oil viscosity–temperature dependence, and coil heat-source modeling. Scheme S1 serves as the reference model, while S2-S6 progressively omit selected coupling effects.

Table 5: Summary of simulation schemes

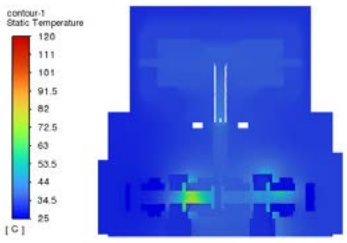
Scheme	Air region treatment	Viscosity model	Coil heat source
S1	Equivalent solid	Variable $\mu(T)$	Variable $q_v(T)$
S2	Equivalent solid	Constant μ	Constant q_v
S3	Equivalent solid	Constant μ	Ignored
S4	Ignored	Variable $\mu(T)$	Variable $q_v(T)$
S5	Ignored	Constant μ	Constant q_v
S6	Ignored	Constant μ	Ignored

4. SIMULATION RESULTS AND DISCUSSION

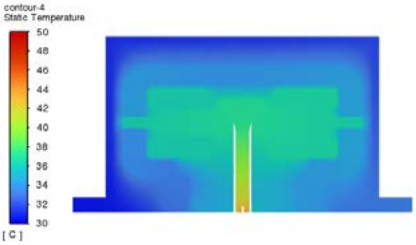
4.1. Overall Temperature Distribution under Typical Condition

To intuitively demonstrate the influence of model simplification on thermal field prediction, the temperature distributions obtained from the most detailed scheme (S1) and the most simplified scheme (S6) were compared under two representative operating conditions: C1 and C6. These two cases correspond to the normal and high-temperature extremes, enabling a direct comparison of the valve's thermal behavior under ordinary and harsh environments.

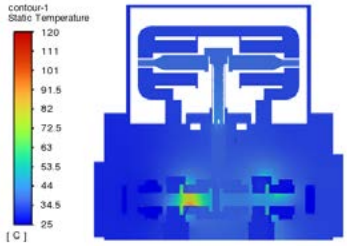
As shown in **Figure 4.1**, under the normal temperature condition, the overall temperature gradient across the valve is relatively small. The high temperature zones are mainly concentrated at the spool throttling orifice and the jet prestage, where viscous dissipation is strongest. A temperature gradient exists along the feedback rod within the torque motor, whereas the temperature distribution in the remaining components is relatively uniform.



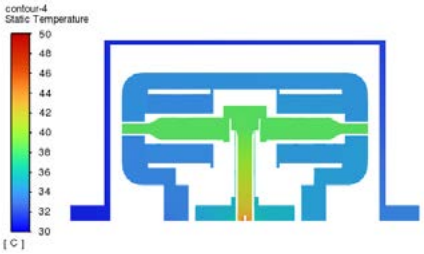
(a) C1-S1-whole valve section



(b) C1-S1-torque motor section



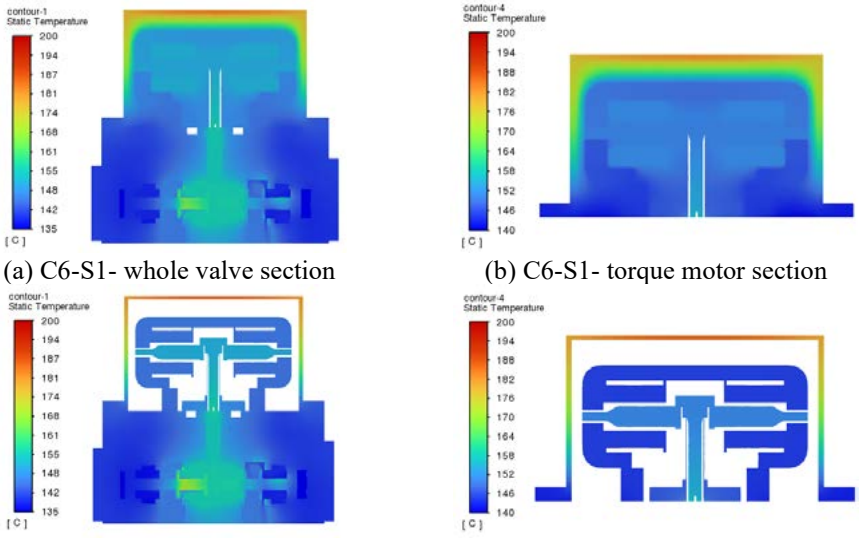
(c) C1-S6-whole valve section



(d) C1-S6-torque motor section

Figure 4.1. Temperature contours under C1 between S1 and S6

As shown in **Figure 4.2**, under the high temperature condition, however, the thermal gradient becomes much more pronounced. The highest temperature occurs at the top of the housing cover, while the return oil region in the valve body is also relatively hot.



(a) C6-S1- whole valve section (b) C6-S1- torque motor section

(c) C6-S6- whole valve section (d) C6-S6- torque motor section

Figure 4.2. Temperature contours under C6 between S1 and S6

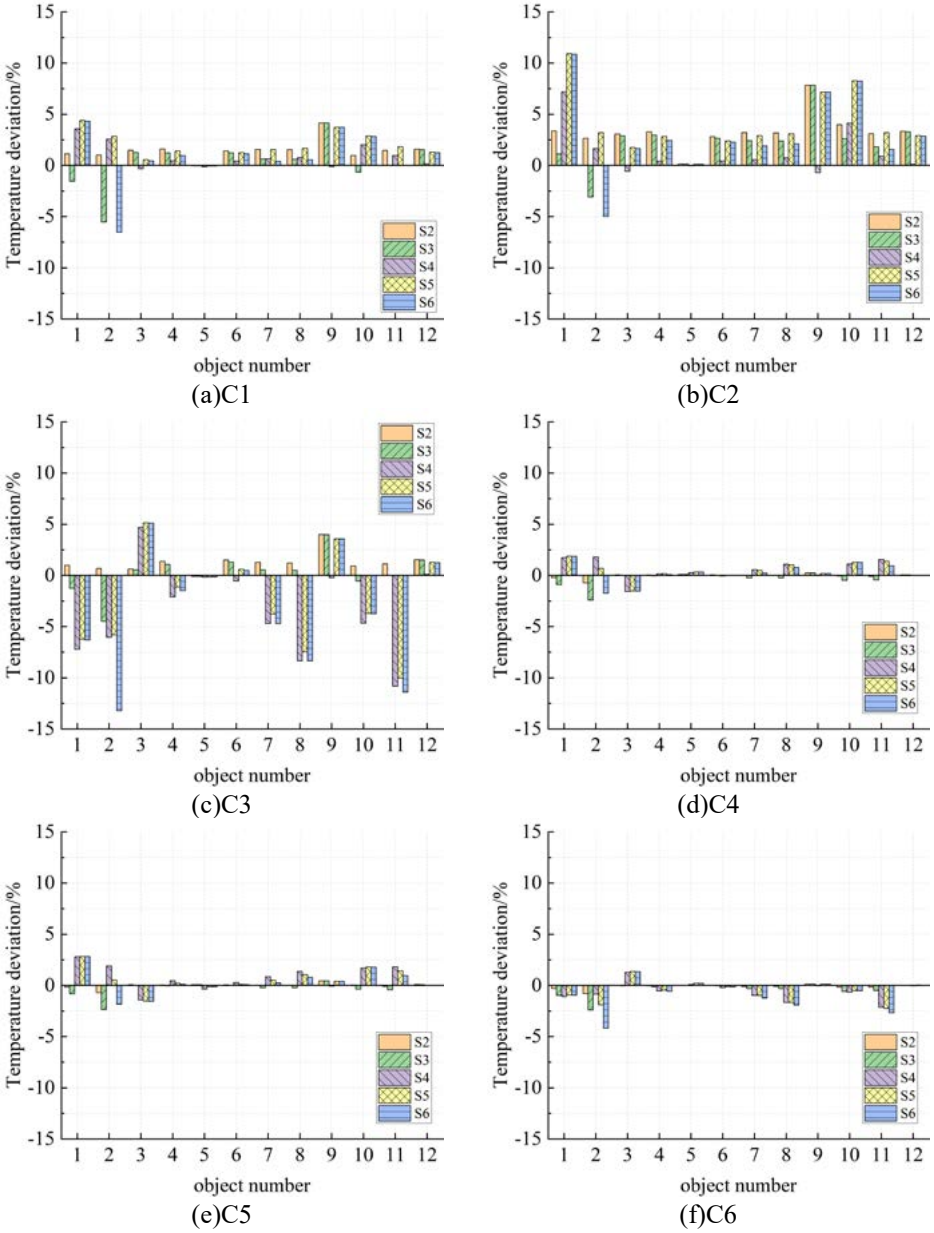
When comparing S1 and S6, the global temperature distributions of the valve appear similar under both C1 and C6, suggesting that at a macroscopic level the main heat-transfer paths and primary hot zones are preserved in the simplified model. Nevertheless, the torque motor cross section reveals localized discrepancies between the two schemes. To assess these deviations, a quantitative comparison of component-level temperatures is presented in the next section.

4.2. Deviation Analysis of Different Modeling Schemes

Twelve structural components of the DJSV were defined as monitoring bodies for thermal evaluation, as listed below: (1) armature, (2) coil, (3) housing cover, (4) support block, (5) feedback rod, (6) flexible tube, (7) lower pole, (8) magnet, (9) spool, (10) support tube, (11) upper pole, and (12) valve body. The volume averaged steady-state temperature of each component was extracted under all modeling schemes (S1-S6) and operating conditions (C1-C6) for quantitative comparison.

Figure 4.3 presents the relative temperature deviations of twelve components under six operating conditions (C1-C6). Under the conventional condition C1, temperature deviations mainly occurred in the armature, coil, spool, and support tube. When the pressure increased (C1→C2), The deviations of the armature and support tube became larger, while other components remained nearly unchanged. During ambient temperature elevation (C1→C3), a wider range of components showed pronounced changes. The deviations of the armature, coil, magnet, and both magnetic poles shifted significantly, with several instances of sign reversal and magnitudes up to 10-12%. However, under high-temperature oil conditions, the transitions C4→C5 and C4→C6 produced negligible changes. When comparing oil

temperature increase pairs (C1→C4, C2→C5, C3→C6), several components demonstrated evident deviations. For example, the armature and support tube showed large variations in C2→C5 (-8%), while in C3→C6 the coil, magnet, and upper pole exhibited positive deviations up to 9%.



1-armature; 2-coil; 3-housing cover; 4-support block; 5-feedback rod; 6-flexible tube; 7-lower pole; 8-magnet; 9-spool; 10-support tube; 11-upper pole; 12-valve body

Figure 4.3. Bar charts of temperature deviations relative to S1 under condition C1-C6

4.3. Average Absolute Deviation and Sensitivity Evaluation

To further quantify the global influence of model simplifications, the average absolute deviation (AAD) of the twelve components was calculated for each scheme under all six operating conditions, as shown in **Figure 4.4**. The AAD provides a single comprehensive metric representing the overall discrepancy introduced by each modeling assumption.

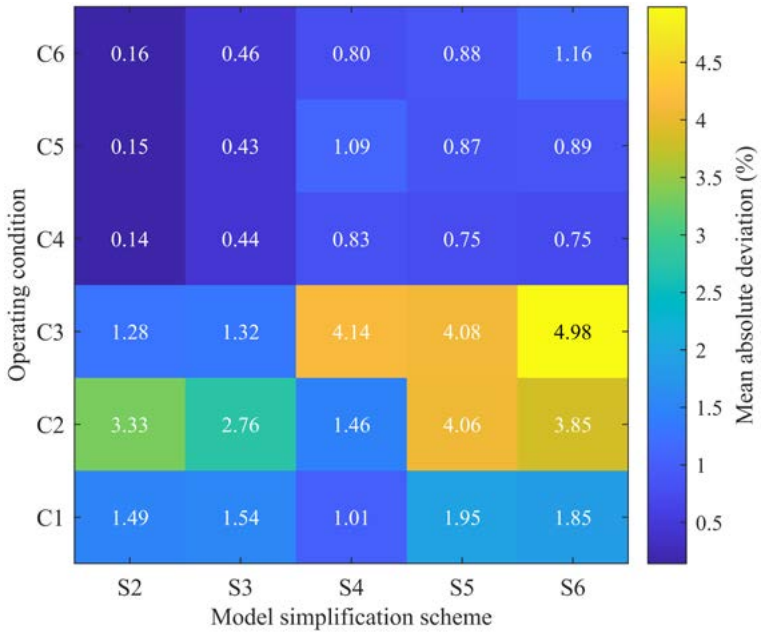


Figure 4.4. Heatmap of average absolute deviation under C1-C6

Under high oil inlet temperatures (C4-C6), the AAD values of all simplified schemes remain relatively low, suggesting that the convective heat transfer of the hot oil plays a primary role in shaping the valve's overall thermal response. In this regime, the temperature viscosity dependence of the oil tends to stabilize, diminishing the impact of the assumed viscosity model on thermal dissipation. Therefore, neglecting the internal air domain and the temperature dependence of viscosity and coil resistance has a negligible impact on the overall error. In contrast, under moderate thermal conditions (C2 and C3), the deviation increases notably, particularly for the more simplified schemes (S4-S6), reaching up to about 5%. This indicates that, when the inlet oil temperature remains at room temperature while the ambient temperature intensifies, neglecting the internal air region removes an important thermal resistance path, thus introducing a discernible discrepancy in the predicted temperature field. Moreover, under high pressure conditions (C2 and C5), the intensified viscous dissipation in the throttling orifices elevates local temperatures, which may be inadequately represented if the temperature dependence of oil viscosity is ignored.

To isolate the influence of each modeling factor, including the omission of the internal air region, the neglect of temperature-dependent effects (for both viscosity and coil heat generation), and the complete omission of the coil heat source, pairwise comparisons were conducted such that only the factor of interest varied while the other two remained fixed.

The scheme pairings used for each isolated factor are summarized in **Table 6**. This approach ensures that the extracted deviations reflect the individual contribution of each simplification rather than the combined influence of multiple assumptions.

Table 6: Scheme pairs used for single-factor sensitivity analysis

Factor	Scheme Pair for Comparison
Omit air region	S4-S1, S5-S2, S6-S3
Temperature dependence(Viscosity/Coil heat source)	S2-S1, S3-S1, S5-S4, S6-S4
Ignore coil heat source	S3-S2, S6-S5

As shown in **Figure 4.5**, the coil consistently exhibits the highest temperature sensitivity, with deviations exceeding 1.6% for all factors and reaching up to 2.9% when the coil heat source is completely neglected. The armature demonstrates pronounced sensitivity primarily to the air-domain omission, with deviations up to 2.75%. When the temperature-dependent effects of viscosity and coil heating are ignored, deviations are moderate but widely distributed across multiple components, particularly the spool, support tube, and magnet, reflecting the system-wide influence of assuming constant material properties. In contrast, the valve body, flexible tube, and feedback rod exhibit minimal changes across all factors. Therefore, accurate inclusion of both the air domain and temperature dependent effects is essential for reliable prediction of local thermal gradients within the torque motor assembly.

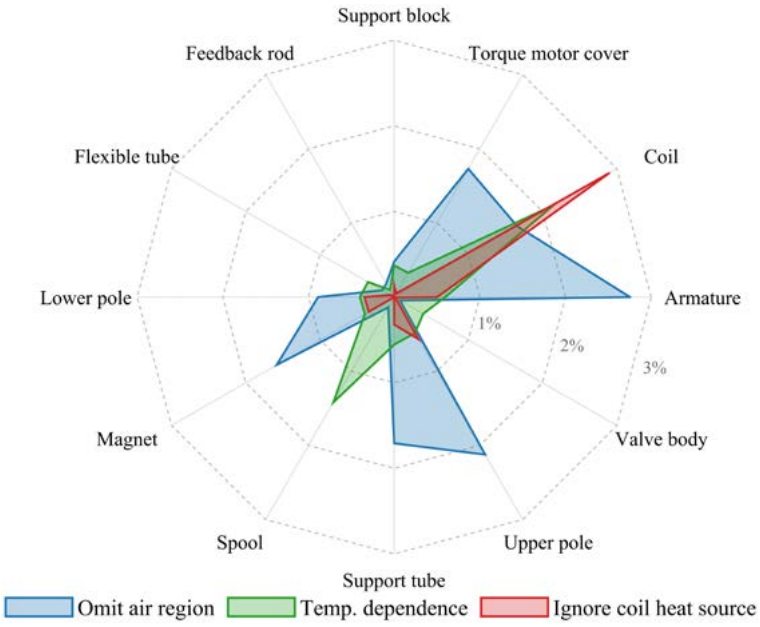


Figure 4.5. Sensitivity of component temperature deviation to modeling simplifications

5. CONCLUSIONS

A steady-state three-dimensional thermal-fluid coupled analysis of a deflector jet servo valve (DJSV) was performed. The analysis quantified the effects of common modeling simplifications, including the omission of the internal air domain, neglect of temperature-

dependent oil viscosity and coil heat generation, and complete exclusion of the coil heat source. Six simulation schemes were examined under six representative operating conditions to evaluate the resulting thermal deviations. The results indicate that the simplified models generally preserve the primary heat-transfer paths and dominant hot zones. However, neglecting the internal air region primarily affects electromagnetic components, particularly under elevated ambient temperatures. Ignoring the temperature dependence of oil viscosity and coil heating introduces distributed temperature deviations across multiple components. Among all factors, the coil exhibits the highest sensitivity, the armature is mainly affected by the omission of the air region, while the spool, support tube, and magnet show moderate deviations when temperature-dependent effects are disregarded. In contrast, the valve body, flexible tube, and feedback rod remain largely insensitive to these simplifications.

Acknowledgement

This work was supported by the National Key Research and Development Program of China under Grant 2024YFF0505300, National Natural Science Foundation of China under Grant 52405078.

6. REFERENCES

[1] Lei L, Gen W, Libo Q, Cunkun C. Research on thermodynamic modeling and temperature prediction of double flapper-nozzle servo valves. *Case Studies in Thermal Engineering*. 2024; 60: 104664. doi: 10.1016/j.csite.2024.104664.

[2] Zhao J, Zhou S, Lu X, et al. Numerical simulation and experimental study of heat-fluid-solid coupling of double flapper-nozzle servo valve. *Chinese Journal of Mechanical Engineering*. 2015; 28: 1030-1038. doi: 10.3901/CJME.2015.0417.045.

[3] Ma L, Yan H, Cai C, Ren Y. Mechanism of temperature-induced zero drift on electro-hydraulic servo valve. *AIP Advances*. 2021; 11: 065210. doi: 10.1063/5.0049045.

[4] Zhu Y, Zhang Q, Tao J, Tan D, Wang X. Heat characteristics analysis of electro-hydraulic servo valve. *Journal of Thermal Science and Engineering Applications*. 2019; 11(3): 031008. doi: 10.1115/1.4041880.

[5] Ma L, Yan H, Cai C, Ren Y. Temperature drift mechanism of deflector jet servo valve under temperature shock. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*. 2022; 237(3): 462-478. doi: 10.1177/09596518221132457.

[6] Ma L, Liu Y, Wan L, Wang J, Yan H. Temperature characteristics of flow force in the pre-stage of deflector jet servo valve across a wide temperature range. *Case Studies in Thermal Engineering*. 2025; 74: 106967. doi: 10.1016/j.csite.2025.106967.

[7] Xiaoxue L, Yaobao Y, Shuwei Z, Tao J, Qiang X. The modeling and analysis of operating characteristics and temperature drift of deflector jet servo valve under the wide temperature range. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*. 2024; 238(12): 5921-5936. doi: 10.1177/09544062231219313.

[8] Zhang Y, Wang S, Shi J, Wang X. Evaluation of thermal effects on temperature-sensitive operating force of flow servo valve for fuel metering unit. *Chinese Journal of Aeronautics*. 2020; 33(6): 1812-1823. doi: 10.1016/j.cja.2019.09.011.

[9] Ma L, Yan H, Ren Y, Li L, Cai C. Numerical investigation of flow force and cavitation phenomenon in the pilot stage of electrical-hydraulic servo valve under temperature shock. *Machines*. 2022; 10(6): 423. doi: 10.3390/machines10060423.
[10] Chen Z, Wang W. Analysis of temperature characteristics of electro-hydraulic servo valve based on AMESim. In: *Proceedings of the 2018 International Conference on Mechanical, Electrical, Electronic Engineering and Science (MEEES 2018)*. Atlantis Press; 2018: 47-53. doi: 10.2991/meees-18.2018.10.
[11] Zhang X, Yin Y, Huang W. Influence of temperature on null position pressure characteristics of flapper-nozzle servo valve. In: *2010 International Conference on Computer Design and Applications (ICCDA)*. Qinhuangdao, China; 2010: V3-257-V3-261. doi: 10.1109/ICCDA.2010.5540780.
[12] Xu L, Ma H, Ren D. Numerical simulation for multi-way valves and fit clearance research based on heat-fluid-solid coupling. *The Journal of Engineering*. 2019; 2019(13): 247-252. doi: 10.1049/joe.2018.9036.
[13] Huang H, Mao Yang, Sun Yongbin, et al. Test and analysis of YH-15 aviation hydraulic oil viscosity-temperature characteristics. *Ordnance Industry Automation*. 2015, 34(05), 42-43. doi: 10.7690/bgzdh.2015.05.012.

Biographies



Junhao Chen received the bachelor's degree in mechanical engineering from Zhejiang University of Technology in 2023. He is currently working toward the Ph.D. degree in mechanical engineering with the State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou, China. His research areas include multi-physics coupled modeling and reliability enhancement of servo valves.



Junru Hui received the bachelor's degree in energy and power engineering from Lanzhou University of Technology in 2019 and master's degree in mechanical engineering from The University of Sheffield in 2021. He is currently working toward the Ph.D. degree in mechanical engineering with the State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou, China. His research areas include the internal performance changes of electro-hydraulic servo valves and the design of new configurations.



Mingjie Li received the bachelor's degree and Ph.D. degree in mechatronic engineering from Zhejiang University, Hangzhou, China, in 2013 and 2019. He is currently an assistant researcher in school of mechanical engineering, Zhejiang University. His research interests include control of electro-hydraulic systems and robotic systems.



Xiaoping Ouyang received the bachelor's degree and master's degree from Northeast Petroleum University Daqing, China, in 1997 and 2000, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2005, all in mechanical engineering. He is currently a Professor with the State Key Laboratory of Fluid Power and Mechatronic Systems, and the Director with the Institute of Fluid Power and Intelligent Control, Zhejiang University. His research interests include electrohydraulic components and systems, and exoskeleton and humanoid robots.

Model-Based Valve Design: Potential of CFD Simulations as a Database for Prediction Models

Timo Rothhardt, Jakob Wirtz, Katharina Schmitz

RWTH Aachen University – Institute for Fluid Power Drives and Systems (ifas),
timo.rothhardt@ifas.rwth-aachen.de, sc@ifas.rwth-aachen.de

Abstract.

Conventional valve development includes measurements and computational fluid dynamics (CFD) simulations. Both are used in an iterative process to achieve the desired characteristic curve. To accelerate the valve development process considerably, data-based expert systems should be introduced. These models are trained with CFD simulation data of numerous geometries. Usually, CFD simulation parameters are tuned for one specific geometry.

This paper presents CFD simulation parameters applicable for a wide range of 2/2 directional poppet valve geometries. The parameters are validated with measurement data from both normally open (NO) and normally closed (NC) valves. Based on previous research, the theoretical background in CFD simulation and preliminary studies, the simulation parameters used in this study are carefully selected. For the validation, 164 points are generated from continuous measurements of nine industrial poppet valves. The comparison between the simulations and measurements show a very high accuracy in the prediction of the pressure difference with a mean deviation of 7.9 %. The accuracy in the simulation of flow forces is lower but the trends in the simulations are correct and high flow forces can be predicted accurately. The results prove that it is possible to use one set of CFD simulation parameters for different geometries. This makes CFD simulations suitable as a database for prediction models.

Keywords. Model-based valve design, poppet valves, flow forces, CFD simulation, validation.

1. INTRODUCTION

Oil hydraulic valves are used in a wide range of industrial and mobile hydraulic applications for simple switching functions as well as for control tasks. The variants within a valve series are often as diverse as the applications themselves. Nevertheless, there are always customer requirements that call for a further, adapted valve geometry. Being able to implement these requirements quickly and cost-effectively plays a decisive role in global competition.

Currently, developing valves for specific applications is time-consuming and costly. This is because valve development is an iterative process involving trials and CFD simulations,

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

which requires expert knowledge and is usually based on the experience of individual company employees. Demographic change and a changing work environment with higher job fluctuation among employees requires knowledge to be documented and formalized [1]. In a recent study by PwC, almost three quarters of all respondents expect personnel costs to rise by 4.7 % in 2026 and see the shortage of skilled workers and increasing cost pressure as obstacles for corporate growth [2].

These challenges can be addressed with an innovation of the development process. So called expert systems can represent complex relationships in an easy-to-use single model. Such a model can calculate the same correlations in a few seconds with the same accuracy. An expert system for valve development could allow the user to define a customer-specific characteristic flow curve, which the system uses to calculate the optimum valve geometry.

To make this step, there is a lack of fast calculating models that can be inverted. Measurements take long but are still the most exact way to determine pressure losses and flow forces. CFD simulations are faster but a little less accurate [3]. Analytical equations are fast to calculate and invertible but not as exact, especially in the determination of flow forces [4].

The solution is a data-based approach, where data is collected and metamodels are trained. As such, CFD simulations should be used as the database, since geometries are easy to manipulate and data is easily gained across a wide range of geometries. When simulating hundreds of different geometries, it is not possible to tune the simulation parameters for each geometry. Thus, generally applicable CFD simulation settings are needed.

This paper presents the selection and validation of generally applicable CFD simulation parameters for 2/2 directional poppet valves with nominal flow rates from 20 l/min to 40 l/min. To determine optimal parameters, at first, previous research about CFD simulations of poppet valves is elucidated and a detailed review of CFD simulation parameters as well as their influence is given. The next chapter outlines the conducted simulations. The selection of points for simulation from continuous measurement data is explained, the preliminary studies on CFD simulation parameters presented and the methodology for the simulation of numerous operating points described. In the last chapter, the simulation results are compared to measurement data, and the accuracy of the simulations is discussed and improved.

2. FUNDAMENTALS

This chapter presents previous research about the CFD simulation of flow forces in poppet valves and gives an overview about the most important CFD simulation parameters and their influence on the simulation model. The focus of the reviewed literature lies on the flow forces and not the pressure difference since it is more challenging to achieve a high accuracy in the CFD simulation of flow forces [3]. The aim is to identify commonly used CFD simulation parameters as a basis for this study. For the flow forces, the convention used is that a positive flow force acts in the closing direction. The detailed definition and an example cross-sectional view of a valve with labelled ports are provided in [5].

2.1. Simulation-based Analysis of Flow Forces in Poppet Valves

The numerical simulation of flows in poppet valves started in the early 90s. Especially in the beginning, the focus was to show that simulated flow patterns equal the flow patterns

from measurements [6]. Recent research focusses on improving the accuracy of CFD simulations by implementing new models (e.g. turbulence or cavitation models) or optimizing components by varying selected geometrical parameters. In the optimization of valves, the pressure drop and the flow induced forces on the poppet are the most relevant quantities. Investigations show that the simulation of the flow forces is more inaccurate than the simulation of the pressure drop [7]. Particularly for the numerical simulation of flow forces, there is no generally valid combination of simulation parameters (e.g. mesh type, mesh size and turbulence model) for accurate results [3]. Zong et al. [8] show that the turbulence model has a strong influence on the results of the simulated flow force while the pressure drop was nearly unchanged. Therefore, the focus of this chapter lies on previous research that simulated flow forces in poppet valves and the employed simulation parameters. Due to large steps in CFD modelling and computing power in the last years, only recent research from 2010 and later is considered. Not all CFD simulation parameters are mentioned in the following brief content description of the papers but summarized afterwards in a table.

Domagala and Fabis-Domagala [9] published a detailed review of research about the CFD simulation of flow forces in both spool and poppet valves. The paper covers the theoretical background of flow forces in valves and of CFD methods as well as numerous publications about steady-state and transient flow simulations. The review results in commonly used assumptions in the CFD simulation of flow forces:

- The fluid is homogeneous, single-phase, incompressible and has constant properties.
- All valve geometries are ideal.
- The roughness of components is neglected, and walls are modelled as smooth.
- Heat transfer and phase changed are neglected.
- The turbulence models used are a standard k-ε or RNG k-ε

Bernad and Susan-Resiga [10] use a 2D CFD simulation to investigate cavitation in a highly simplified seat valve. The fluid is modelled as multi-phase and the main dynamics of the cavitating flow can be mapped. Hao et al. [11] examine three simple cone geometries in a water valve with the aim of reducing flow forces and cavitation in the valve. Steady-state 2D CFD simulations are performed with a two-phase (water and steam) mixture. On the third valve, two housing parameters at the valve outlet are varied individually. By combining the optimal parameters, the flow force is reduced by 44.2 %. Simic and Herakovic [12], [13] optimize the geometry of a normally closed (NC) valve geometry with the aim of reducing flow forces in static operation. The influence of five geometric parameters on the flow forces is investigated by using 3D CFD simulations. For one parameter, additional experiments confirm the accuracy of the CFD simulations with a maximum deviation of 7 %. Xie et al. [14] and Liu et al. [15] investigate basic cone geometries and vary different geometrical parameters in 3D CFD simulations. The method one-factor-at-a-time (OFAT) is used by both authors and the influence of the parameter changes on the flow forces are evaluated.

Rundo and Altare [16] compare two analytical flow force models (Bernoulli and momentum conservation) with stationary CFD simulations. Both flow directions and three poppet angles are considered. In flow direction 1-2, the analytical calculation according to Bernoulli achieves similar results to the CFD simulation, while in direction 2-1, Bernoulli calculates a flow force that is seven times greater. This shows that physical models are not well suited for the analytical calculation of flow forces. Roemer et al. [17] develop their own analytical

model for calculating the forces on the poppet of a fast-acting valve. This model considers flow-induced and motion-induced forces. The parameters for the analytical model are determined using a series of transient CFD simulations, which means that a one-time high simulation effort is required to obtain the analytical model.

Transient CFD simulations are used by Gomez et al. [18] to investigate the dynamics of a commercially available valve. In the simulations, the switching frequencies, inlet pressures and poppet angles of the valve are varied, and the pressure loss and flow force are evaluated. The poppet angle has little effect on pressure loss and mainly affects the flow force. Finesso and Rundo [19] develop a transient simulation of a pressure relief valve to predict the valve characteristics for different geometries and spring settings. The CFD simulation is validated with measurements and shows good agreement. Based on few CFD simulations, a 0D model is tuned and can predict the flow characteristics of the valve. Hirose et al. [20] use transient 3D CFD simulation to investigate the force hysteresis induced by flow patterns on the valve poppet. In preliminary steady-state simulations, the flow patterns described by Johnston et al. [21] are confirmed by simulation. The transient simulations show that the change in flow pattern occurs at different positions when the valve is opened and closed. All CFD simulation parameters mentioned in the presented papers, are summarized in **Table 2.1**. The properties not mentioned are marked with [-].

Author	2D or 3D Steady-state or transient	Turbulence model	Fluid model (Fluid specification / fluid phases / cavitation)	Discretization scheme (pressure / momentum / k / ε or ω)
[10]	2D Steady-state	RNG k-ε	[-], two-phase, cavitation	SIMPLE / second-order / first-order / first-order
[11]	2D Steady-state	Realizable k-ε	Water, two-phase, cavitation	PRESTO! / second-order / second-order / [-]
[12]	3D Steady-state	k-ω SST	Oil VG 32, single-phase, [-]	[-]
[13]	3D Steady-state	k-ω SST	Oil VG 32, single-phase, [-]	[-]
[14]	3D Steady-state	k-ε	[-]	[-]
[15]	3D Steady-state	RNG k-ε	Density and viscosity, [-], [-]	[-]
[16]	3D Steady-State	k-ε	Oil, two-phase, cavitation	First-order upwind
[17]	3D Steady-state and transient	Laminar flow	Oil, single-phase, [-]	[-]
[18]	3D Transient	k-ε	Oil VG 32, single-phase, [-]	SIMPLE / second-order / second-order / second-order
[19]	3D Transient	k-ε	Oil DTE 25, two-phase, cavitation	PRESTO! / QUICK / first-order / first-order
[20]	3D Steady-state and transient	k-ε	Density and viscosity, [-], [-]	[-]

Table 2.1: Overview of CFD simulation parameters used in previous investigations

Based on the overview in the table it can be said that most of the simulations of poppet valves in the last years were 3D CFD simulations. Depending on the research objective, both steady-state and transient simulations were set up. The most used turbulence model was, as described by [9], some kind of k-ε model. Some publications name the fluid with the operating temperature and the resulting density and viscosity; others name only the density and viscosity or do not name any fluid properties. The discretization schemes used differ between the investigations, but for the momentum the second-order upwind scheme is the most common. The most commonly used simulation environment is Ansys® Fluent® and the mesh type an unstructured tetrahedron-mesh.

2.2. Simulation Parameters and Their Influence

The construction process of a CFD simulation entails a wide range of choices and parameters that require careful consideration and are essential for its success. This section will provide an overview of crucial decisions during the setup process of a typical CFD simulation in Ansys® Fluent®. The general workflow for a CFD simulation first involves the conception of a geometry which is then discretized into a mesh and finally examined by a CFD solver. The geometry is usually created with a CAD software and then imported into the CFD environment. Therefore, the focus of the simulation parameters lies on the mesh and solution settings. Mesh creation and numerical simulation in general is always a balancing act between accuracy and computational cost.

2.2.1. Mesh

Meshes can be categorized into two main mesh types – structured and unstructured meshes. In structured meshes the cell coordinates align with the cartesian coordinates and a cell always has a predefined number of neighbours. The structured grid is limited to simple geometries because both complex walls and tight spaces, that require the use of a finer mesh, would introduce unnecessarily small spacings in other parts of the solution domain in order to maintain the structured nature [22]. Structured meshes also minimize the effects of numerical diffusion when using the first-order upwind scheme for cell discretization, because the second-order derivatives from the Taylor series expansion, that cause numerical diffusion, are ideally one-dimensional when the flow is aligned with the grid [23].

Unstructured meshes are able to fit an arbitrary solution domain and are thus the most flexible of all grid types when it comes to meshing irregular geometries. They are not limited by the number of cell neighbours and allow for all types of cell shapes, including tetrahedrons or polyhedrons in three dimensions. This type of mesh involves a computational overhead during construction due to its arbitrary nature and lack of simple indexing system [24]. It also requires more complex methods for gradient computation as it is not able to rely on cartesian directions and is additionally very vulnerable to numerical diffusion when using first-order upwind schemes for cell discretization [25]. Despite these challenges, unstructured meshes have become increasingly popular and there are several methods to alleviate the drawbacks, as will be elaborated upon in chapter 3.2. Comparing tetrahedron- and polyhedron-meshes, polyhedron-meshes require more computation time during the mesh generation. This overhead is compensated, however, in the total computation time of the solution due to a lower overall cell count [24].

In all cases, when flow becomes oblique to the grid cells or is characterized by high gradients, the solution is prone to discretization errors [22]. Said errors are proportional to the truncation errors from the Taylor series expansion of the used discretization schemes. In the case of this first-order upwind scheme, the leading error term is a second derivative which corresponds to the mathematical description of physical diffusion. The truncation error thus adds a diffusion term to the solution making it overly diffusive. This effect is commonly referred to as numerical or false diffusion. A reduction of the overall spatial discretization error can be achieved by reducing the cell size or by the choice of the discretization scheme. Important differences are between the first- and second-order upwind scheme, because the first-order scheme introduces numerical diffusion effects, as previously explained, whilst the latter alleviates these effects by eliminating the second derivative in the truncation error [25].

Furthermore, it is important to appreciate that over-refining the mesh across the entire solution domain does not necessarily improve the overall results. Convergence of the solution depends on the residuals of the discretized conservation equations throughout the entire domain. When determining the appropriate mesh size and refinement regions, it is thus important to attempt a balance in the discretization error across the entire domain, namely a similar residual in all domain parts [22].

Finally, the design of a meaningful mesh requires the consideration of near-wall effects, in particular the adequate resolution of the viscous sublayer and buffer layer. To achieve this, meshes are typically either equipped with boundary-layer elements or with wall functions. Boundary layers essentially provide fine resolution in the wall-normal direction. In this way, near-wall effects can be represented accurately with only a modest increase in computational cost. Wall functions are semi-empirical formulas that interpolate the viscous sublayer and buffer layer, omitting the necessity of a complete resolution of the near-wall region [25].

2.2.2. Solution

This section focusses on the interpretation of the mesh in terms of discretized conservation equations and also endeavour to elucidate relevant decisions during turbulence modelling. Before beginning any type of solution, it should first be examined what kind of characteristics the flow is expected to have; strictly laminar or turbulent. If the flow is strictly laminar, the underlying conservation-equations are usually solved using any type of numerical solver. If the flow is presumed to assume turbulent characteristics at some point in the domain, the solution complexity increases drastically. The numerical examination of turbulent flow can generally be categorized into three types of simulation: the direct numerical simulation (DNS), the large-eddy simulation (LES) and the Reynolds-averaged Navier-Stokes (RANS)-approach [22].

The DNS is the most exact approach for turbulent simulations but entails significant computational cost. LES is a compromise between computational cost and accuracy. It is usually used in cases where the DNS is too computationally expensive like in high-Reynolds flows or for complex geometries. The RANS-approach is the least computationally expensive and entails the replacement of flow variables in the conventional Navier-Stokes equations by quantities that underwent Reynolds-averaging, namely the separation of flow variables into an averaged and a fluctuating quantity. During this averaging process an additional unknown quantity emerges called the Reynolds stress tensor (RST). This tensor can be approximated with the help of the turbulent kinetic viscosity which, in return, is then modelled with additional equations [22].

Said modelling can be implemented in the form of zero-, one- or even six-equation models, the most common, however, being the two-equation model. In the realm of two-equation-models, the most prominent are the k-ε and the k-ω models. Both cases institute two transport equations, one for the turbulent kinetic energy k and one for either the dissipation rate of k, namely ε, or the specific dissipation rate of k, namely $\omega = \varepsilon/k$. The k-ε is very popular in general industrial CFD applications because it provides reasonable accuracy for a wide range of simulation problems. However, it is not integrable through the viscous sublayer without the addition of further equations, limiting its accuracy in the prediction of near-wall phenomena. The k-ω model addresses this limitation and is conceived in a way that allows it to adequately predict the viscous sublayer, making it the better choice for simulations involving extensive wall-modelling [25].

3. CFD SIMULATIONS

The CFD simulation parameters should be validated based on the measurements of 9 poppet valves. Due to the measurement procedure with a moving poppet, required to capture the flow forces, the measurement data is continuous [5]. At first, this chapter shows how discrete points for CFD simulation are extracted from the continuous measurement data. Since generally applicable CFD simulation settings are needed, preliminary studies are conducted to determine most of the CFD simulation parameters. Finally, the methodology and procedure for the batch simulation of different valve geometries with different opening widths and boundary conditions are described briefly.

3.1. Determining the Point of Operation

The simulated points of operation should be as close as possible to the real point of operation. How the point of operation is determined, differs between NO und NC valves. For NO valves, the actual operating point can be easily read from the measurement data. It is the point at which the rod of the valve actuation and force measurement system lifts off the poppet. At this position, there is a balance of forces at the poppet between the flow force acting in the closing direction and the spring force acting in the opening direction. The measurements in [5] were conducted with different tank pressures and pressure differences over the valve.

Figure 3.1 shows the poppet position x, the flow rate Q and the measured force F plotted over time t for one cycle of a NO valve. The actual operating point is located at the position where these variables begin to change with the poppet position, since this is the point where the rod of the through-drive starts touching the poppet. From this point on, the measured position also corresponds to the poppet position.

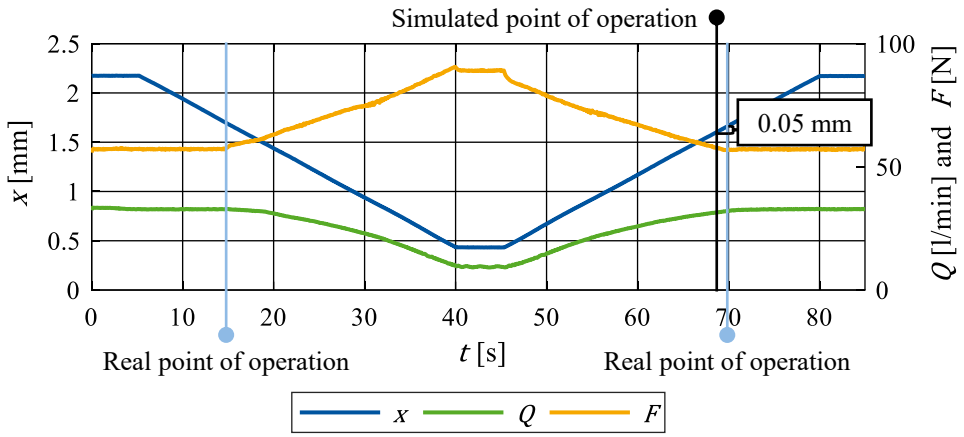


Figure 3.1: Position, force and flow rate curve from a NO valve, flow direction 1-2, tank pressure 30 bar, pressure difference 15 bar

The flow forces should be evaluated during movement, as there is no static friction in this case and the flow force can be calculated from the measurements with and without flow. For this reason, the point 0.05 mm away from the actual operating point is used for the CFD simulation. The actual operating point occurs once during both the closing and opening of the valve. For the simulation, the point at which the actuator extends is used, as there is less possibility of the rod tilting during the extension movement.

For NC valves, it is not possible to determine the actual operating point without further effort. The real operating point is established through a balance of forces between the flow force, spring force and actuator force. The force of the actual solenoid is position-dependent and unknown. For this reason, the actual operating point cannot be read from the measurements and must be assumed. The measurements of all NC valves show a buckle in the flow rate and force curve. This buckle occurs at the point where the valve is opened so far that the smallest cross-section is no longer at the sealing edge. It is assumed that the valves are designed so that point of operation is close to this point where the flow is stable and the flow rate reaches its maximum. This buckle lies at approximately 90 % of the maximum flow rate for all valves, which is why this point is used for simulation.

For validation, all cycles with a tank pressure of 30 bar are used. For nine valves, this results in 164 points that have to be simulated. Once the poppet positions of the operating points to be simulated have been determined, all other required values can be evaluated at these points. These are the flow rate, the flow force as well as the pressures and temperatures upstream and downstream of the test valve. For these variables, an average is calculated for each value using the 50 surrounding values.

3.2. Preliminary Studies

This section details the process and the choices that were made to establish the settings for the geometry, the mesh and the solution which are used for the simulations carried out in this paper. For all simulations Ansys® Fluent® and Fluent® Meshing were used.

3.2.1. Geometry

The simulated geometries belong to industrial poppet valves and were used in their original shape. During the preliminary studies, one geometric variation was tested for their impact on the simulation results. The variation was an extension of the in- and outlet of the manifold by an 80 mm pipe to ensure fully developed flow at the in- and outlet. The pipe influences both the pressure drop and the flow force on the valve. None of the results were implausible so that this variation is to be further investigated based on measurement data.

3.2.2. Mesh

Considering the vast number of simulations that should be conducted, the difference between polyhedron- and tetrahedron-meshes in terms of cell count is significant, facilitating the decision to use polyhedron-meshes. As mentioned in section 2.2.1, unstructured polyhedron-meshes are excessively vulnerable to numerical diffusion for first-order upwind schemes. However, the problem of numerical diffusion is a phenomenon that is limited to first-order upwind discretization schemes. Fluent® uses the second-order upwind scheme as the default discretization for convection terms. Higher order schemes were also tested but did not yield any noticeable effect in the simulation results. In summary, the unstructured polyhedron-mesh with second-order upwind discretization schemes proved to be the best fit for the underlying simulation task. This was also validated by measurements.

The choice of mesh size was an iterative process to find an optimum between desired accuracy and simulation time. A refinement region was placed around a reasonable radius of the poppet, separating the fluid domains with pipe flow and those with, partially turbulent, narrow gap flow. In the next step, both the reference cell size across the entire domain and the refinement factor of the refinement region were iteratively decreased, and increased respectively, until the accuracy of the simulation did not further improve. Said investigations

culminated in the establishment of a reference cell size of maximum 0.3 mm and minimum 0.005 mm. The chosen refinement was the factor two. This setup resulted in an average of around 10 million or less cells per simulation.

Another important manipulation of the mesh includes the definition of boundary layers. As mentioned in section 2.2.1, boundary layers are especially crucial for turbulent flows. Ten boundary layers with a growth rate of 1.2 and an aspect ratio of 15 are selected. The first simulations with and without boundary layers showed differences in pressure difference and flow force. In addition, boundary layers increase the number of cells, so the use of boundary layers is to be further investigated based on measurement data.

3.2.3. Solution

The main decision that was to be made in the solution process is the choice of turbulence modeling. An extensive analysis of the Reynolds number across the solution domain yielded the insight that turbulent flow was only present in narrow gaps at high flow rates. Due to the lowest computational costs, RANS simulations were chosen. As mentioned in section 2.2.2 for the turbulence model, the k-ω model is dominant in including near-wall effects. Especially in light of the fact, that boundary layers would not be included in every mesh, it was of vital importance that the turbulence model is able to accurately model near-wall effects in the narrow regions of the fluid domain. For this reason, the widely used k-ω SST (shear stress transport) turbulence model is employed. The alternative of wall functions was also tested by activating the tabulated choice in the near-wall treatment section of Fluent® Setup, however this did not affect the results in a meaningful way.

In the final step of the preliminary investigations, the boundary conditions were examined. Both a flow rate and a pressure boundary condition can be used at the inlet. The investigations showed that the simulations with the flow rate boundary condition converge more quickly, while the pressure boundary condition leads to slight oscillations in the solution. For this reason, the flow rate boundary condition is selected at the inlet. A pressure boundary condition is used at the outlet. A pressure of 500 bar is specified so that positive pressures are always present throughout the solution area during the calculation, and the solution can be calculated more stably. The solver is pressure-based, and all calculations are caried out as steady-state simulations. As convergence criteria, a residual of 0.0005 was used for the outlet flow rate, the pressure difference between inlet and outlet, and the axial flow force acting on the poppet. All CFD simulation parameters specified in the preliminary investigations are summarised in **Table 3.1**. The parameters that are examined in detail in chapter 4 are shown in italics.

Parameter	**Selection**
Simulation environment	Ansys® Fluent®
Mesh type and form	Unstructured polyhedron-mesh
Discretization scheme	Second-order upwind scheme (pressure / momentum / turbulent kinetic energy / specific dissipation rate)
Mesh size	Maximum 0.3 mm; Minimum 0.005 mm
Mesh refinement	In the area of turbulent flow around the poppet with a factor of 2
Type of simulation	RANS
Turbulence model	k-ω SST
Boundary conditions	Inlet: flow rate; Outlet: Pressure 500 bar
Solver	Pressure-based, steady-state
Convergence criteria	Residual of 0.0005 for outlet flow rate, pressure difference, poppet flow force

Boundary layers	*With and without boundary layers, count 10, growth rate 1.2, aspect ratio 15*
Inlet and outlet	*With and without 80 mm pipe*

Table 3.1: Preselected simulation parameters for all simulations

3.3. Methodology and Procedure

This chapter gives a brief overview of the workflow used for the CFD simulations of nine different valve geometries with varying opening widths and boundary conditions. The goal of the simulations is the computation of the pressure loss over the valve and the flow force in axial direction on the poppet. These two quantities are used to validate the CFD simulations with measurement data.

The CAD files for all valves were either provided by the manufacturer or self-created, based on precise measurements. Acute angles in the geometry are to be avoided for a good mesh. Therefore, the geometries were adapted if required. To make the simulations as comparable as possible, the valves are screwed in the manifold with the angle recorded during the measurements [5]. This CAD model is imported to Ansys® with the opening width of the poppet set as a variable. Based on this model, the volume fill is created. The filled area is meshed with the parameters named in Table 3.1.

The CFD simulation uses the poppet position, flow rate and temperatures from the measurements. Before the meshing, the poppet is positioned according to the measured value, and the flow rate is used as an input boundary condition. An ideal Fuchs RENOLIN B 15 VG 46 oil with constant density and viscosity is used in the simulations. The two fluid parameters are determined in advance by a calculation depending on the temperature, and the fluid is modelled as single-phase without cavitation. The temperature used for the calculation is the average of the temperature measured before and after the test valve. After setting the remaining simulation parameters, a so called case-file is exported from Fluent®. The case-file contains the information about the geometry, mesh and all simulation parameters. This file was used to execute the simulations on the High-Performance Computing (HPC) cluster of RWTH Aachen University.

4. RESULTS AND DISCUSSION

A total of 164 points from the measurements of all valves are available for validating the CFD simulations. 86 of these points belong to flow direction 1-2 and 78 to the opposite direction. The 164 points are simulated with four different meshes, but otherwise with the same settings. In the first variant, the geometries are meshed between the male stud connectors using the parameters specified in the preliminary investigations (Table 3.1). In the second variant, an 80 mm inlet and outlet pipe is attached to each of the geometries at the male stud connectors. In variants three and four, the first two variants are each meshed with boundary layers. The parameters listed in Table 3.1 are used for the boundary layers.

In the following, the accuracy of the four different simulation setups will be compared to the measurements. The comparison is carried out individually for the pressure difference Δp and the flow force F_{flow}. Starting with the pressure difference, for a first overview, the simulated values were plotted over the measured values. For both flow directions, the values were evenly distributed around the ideal line and deviate more strongly from the ideal line as the pressure difference increases. A few points in flow direction 2-1 were an exception and had a larger deviation than the rest. These values all came from simulations with boundary layers. A closer look at the points revealed that they all belong to one NC valve. When

examining the mesh at this valve, it was found that it has a height of 0.55 mm at the smallest cross-section and that there is only one cell layer between the boundary layers. The lack of an adequate volume mesh between the boundary layers of opposite walls leads to increased deviations. Consequently, there is a systematic error in these simulations. However, all other points indicate that the simulations of the pressure differences are correct and that there is no fundamental error.

To identify which mesh provides the most accurate results, the relative deviation is calculated for each point with **equation 4.1**.

$$relative\ deviation = \left|\frac{simulation - measurement}{measurement}\right| \tag{4.1}$$

The results are illustrated in boxplots in **Figure 4.1**. The boxes mark the range from the lower to the upper quartile. 25 % of all values are below the lower quartile and 25 % of all values are above the upper quartile. The solid line within the box marks the median and the cross marks the mean. All values that are more than 1.5 times of the interquartile range (distance from the upper to the lower quartile) away from the upper or lower quartile are considered outliers. Within the remaining values, the whiskers mark the maximum and minimum values. In flow direction 2-1, the models without boundary layers are the most accurate.

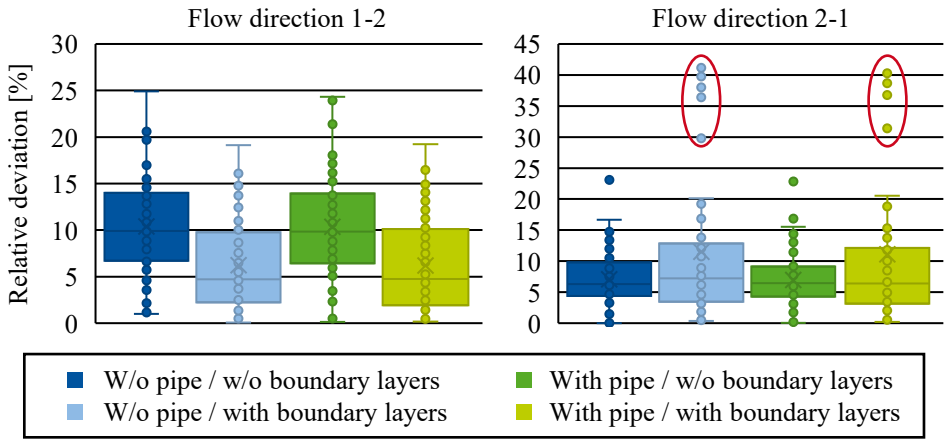


Figure 4.1: Relative deviation of simulated to measured pressure differences for four different meshes in flow direction 1-2 (left) and flow direction 2-1 (right)

The upper quartiles are each below 10 % deviation and the medians are around 6.3 %. The outliers in the models with boundary layers are the same points from the one NC valve already mentioned before. In flow direction 1-2, the simulations with boundary layers are the most accurate. Here, the upper quartiles are approximately 10 % and the medians 4.7 %. In both flow directions, the meshes with and without pipes achieve similar accuracies. In this case, the simulations without pipes are preferable because the mesh is smaller and therefore shorter computing times are possible. The biggest differences are in the simulations with and without boundary layers.

Regarding the accuracy in flow force simulation, **Figure 4.2** shows the simulated flow force plotted over the measured flow force. In contrast to the pressure difference, also negative

values are possible for the flow force. The measured values deviate from the ideal line by a similar amount across the entire range and do not converge in a funnel shape. In addition, there is a particularly large number of points at which a flow force close to 0 N was measured, while the simulation yields a larger or smaller value. This occurs in both flow directions, and the areas close to the measured 0 N are marked in red in both diagrams. When calculating the relative values, high percentage deviations would occur at these points, which would reduce the significance of the percentage deviation. For this reason, all points with a measured absolute force of less than 3 N are excluded from the calculation of the relative deviations. One notable feature is the points in flow direction 1-2, which deviate more strongly from the ideal line than the rest. These are also outlined in red and, after verification, could be assigned to the NC valve.

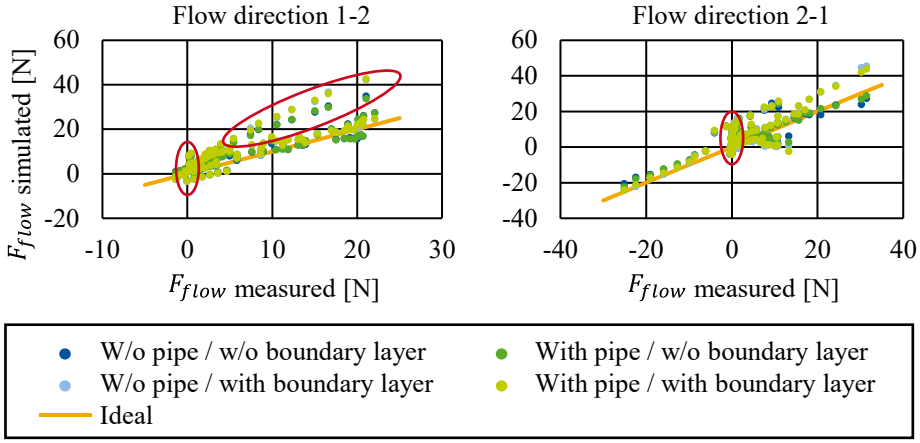


Figure 4.2: Simulated flow force over measured flow force for four different meshes in flow direction 1-2 (left) and flow direction 2-1 (right)

The relative deviation of the flow forces is shown in boxplots in **Figure 4.3**. As already described, all values for which the measured force was less than 3 N are excluded.

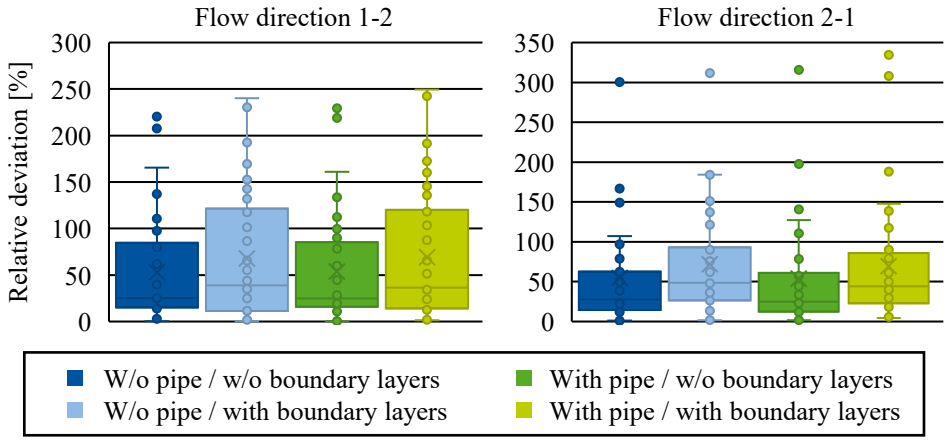


Figure 4.3: Relative deviation of simulated to measured flow forces for four different meshes in flow direction 1-2 (left) and flow direction 2-1 (right)

In flow direction 1-2, 50 of the initial 86 points are still used for the evaluation, and in flow direction 2-1, 44 of the 78 points are still used. The figure shows that the relative deviations in the flow force simulation are significantly greater than in the pressure differences. In both flow directions, the simulations without boundary layers are the most accurate. For these meshes, the upper quartile in flow direction 1-2 is approximately 85 % and the median is approximately 25 %. In the opposite direction, the simulations are more accurate, with the upper quartile at approximately 60 % and the median also at approximately 25 %. This raises the question whether the accuracy of the flow force simulation is sufficient. To clarify this, it is examined which relative deviations occur at which flow forces. It was found that from a measured flow force of 15 N, the relative deviation at each point is less than 26.6 %. High flow forces can therefore be reliably predicted by the simulations.

In summary, in three out of four cases, the mesh without boundary layers was the most accurate. In the case of the pressure difference in direction 1-2, the meshes with boundary layers are more accurate, but the meshes without boundary layers still deliver very good results with an upper quartile of approximately 13 %. For these reasons, a mesh without boundary layers is selected. There are marginal differences between the meshes with and without pipes. Since the mesh without pipes is smaller and the simulations therefore have a shorter calculation time, the mesh without pipes and without boundary layers is ultimately selected for the following simulations which create the database for the prediction models.

Deviations of around 10 % between a CFD simulation and measurements of the real system are currently state of the art in the CFD simulation of pressure differences. This is achieved by the conducted simulations. The accomplished deviations are greater for the flow force, which is why the reason for the deviations should be investigated and discussed at this point.

The first suspicion was a numerical error in the flow force calculation due to different mesh sizes on the poppet surface. For this reason, the poppet geometry and the zone of the mesh refinement were adjusted. Before the adjustment, small shoulders on the poppet led to a locally finer mesh on the surface of the poppet in the pressure compensation area. In addition, the zone of the mesh refinement ended in the pressure compensation bore. During the adjustment, the shoulders were removed in the CAD model and the refinement zone extended, resulting in a mesh with equal size around the poppet.

These changes lead to better matching flow forces, but it is noticeable that the simulations calculate higher values, especially around the range of 0 N measured flow force (Figure 4.2). For this reason, simulations were specifically carried out in which the expected flow force is close to 0 N. This is the case with a low flow rate, which is why all valves were simulated with flow rates from 1 l/min to 5 l/min in 1 l/min increments. The poppet position used was the position that was also used in the simulation with the lowest pressure difference. For two valves, the results show constant flow forces between 6 N and 10 N for these flow rates. These results are implausible and a systematic error in the simulation must be assumed.

Upon closer inspection of the CAD models of the valves, it was noticed that in both models, the manufacturer had also depicted the fits, resulting in clearance between the poppet and the sleeve and thus a gap. This gap was removed for the CFD simulation. To do this, the diameter of the poppet was increased by 0.1 mm so that the poppet and the sleeve overlap. The overlap means that the pressure compensation no longer works, and the poppet is no longer free of forces under static pressure. This problem was solved by matching the poppet diameter with the diameter of the hole in the sleeve exactly.

Another possible reason for the larger percentage deviations in the simulation of the flow forces may be the hysteresis in the flow force [5] in some valves. Previously, the value at the simulated operating point was used as the measured value for comparison. The force hysteresis was not considered. In the following, force hysteresis is taken into account by calculating the mean value of the flow forces between the closing and opening movements and using this as the comparison value.

Overall, the mean average error (MAE) and the mean average percentage error (MAPE) are used to present the final accuracy of the CFD simulation with all the discussed improvements. The two characteristic values are listed for each valve type and flow direction for the pressure difference and flow force in **Table 4.1**. The MAE in the simulation of the pressure difference is between 0.7 bar and 1.7 bar, depending on the valve type and flow direction. This corresponds to average percentage errors of 5.5 % to 11.1 %. The accuracies achieved are mostly below this value and, in the case of NC valves in flow direction 1-2, just above it. When simulating the flow forces, there are greater deviations from the measured values. The average deviation is between 2.8 N and 3.6 N. In percentage terms, this results in a deviation of between 36.6 % and 67.3 %. Bordovsky [7] obtains a $MAPE$ of 32 % with a simulation optimized for one valve. So, the $MAPE$ achieved with general simulation settings applied on nine different geometries can be rated as good. Despite the higher deviations, the trends in the flow force simulation are correct. This means that values in this accuracy range are also suitable for setting up a prediction model.

Δp	NO 1-2	NO 2-1	NC 1-2	NC 2-1	F_{flow}	NO 1-2	NO 2-1	NC 1-2	NC 2-1
MAE [bar]	1.1	0.7	1.7	0.7	MAE [N]	3.5	3.4	2.8	3.6
$MAPE$ [%]	8.8	6	11.1	5.5	$MAPE$ [%]	38.4	58.3	67.3	36.6

Table 4.1: MAE and $MAPE$ of simulated to measured values of pressure difference and flow force divided according to valve type and flow direction

In summary, it can be said that CFD simulations have potential as a database for prediction models. The achieved accuracies are satisfactory. For an individual valve, the simulation parameters could be further improved. Since the results originate from nine different valves, the CFD simulation parameters for the various geometries of this type and size of 2/2 directional poppet valves are validated. It can be assumed that simulations of other geometries using the same CFD simulation parameters will achieve similar accuracies.

5. CONCLUSION AND OUTLOOK

In this paper generally applicable CFD simulation parameters for 2/2 directional poppet valves were acquired. In extensive preliminary investigations, several simulation parameters, such as a polyhedron-mesh with a maximum size of 0.3 mm, were selected. An overall number of 164 points, generated from measurement data, were simulated with four different meshes. The mesh without an additional pipe and without boundary layers turned out to be the best compromise between accuracy and computing efficiency. The simulation of the pressure difference has a mean deviation of 7.9 % to measured data. The flow force simulation has a mean deviation of 50.2 %, but the trends in the flow force simulation are correct. This makes CFD simulations with the presented parameters suitable as a database for pressure drop and flow force prediction models of 2/2 directional poppet valves. The CFD simulations are particularly reliable for the pressure drop and for flow forces above 15 N. In future work, the determined simulation parameters can be applied to a variety of

geometries and thus create a database for prediction models of 2/2 directional poppet valves. Before using the determined parameters in CFD simulations of other valves, the parameters must be validated separately.

6. ACKNOWLEDGEMENTS

The IGF research project 01IF22852N of the research association Forschungskuratorium Maschinenbau e. V. – FKM, Lyoner Straße 18, 60528 Frankfurt am Main was supported from the budget of the Federal Ministry of Economic Affairs and Climate Action through the DLR within the scope of a program to support industrial community research and development (IGF) based on a decision of the German Bundestag.

Computations were performed with computing resources granted by RWTH Aachen University under project rwth1768.

7. REFERENCES

[1] Deloitte: 'Wachstumsmotor Maschinenbau - Vier Szenarien für eine erfolgreiche Zukunft in 2030', 2021.

[2] PwC: 'Maschinenbau-Barometer - Ausblick 2026', 2025.

[3] Yuan, Q., Singh, S. et al. 'Flow Forces Investigation through Computational Fluid Dynamics and Experimental Study', 9th International Fluid Power Conference (9th IFK), pp. 100–109, 2014.

[4] Bordovský, P. 'Evaluation of Steady-State Flow Forces in Spool Valves', Dissertation, Aachen, 2018.

[5] Rothhardt, T., Schmitz, K. 'First Steps Towards Model-Based Valve Design: Measuring Pressure Drop and Flow Forces of Industrial 2/2 Directional Poppet Valves', The 19th Scandinavian International Conference on Fluid Power, SICFP'25, pp. 424–435, 2025. DOI: 10.13052/rp-9788743808251A31.

[6] Vaughan, N. D., Johnston, D. N., Edge, K. A. 'Numerical Simulation of Fluid Flow in Poppet Valves', Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, Vol. 206, Nr. 2, pp. 119–127, 1992. DOI: 10.1243/PIME_PROC_1992_206_105_02.

[7] Bordovský, P., Schmitz, K., Murrenhoff, H. 'CFD Simulation and Measurement of Flow Forces Acting on a Spool Valve', 10th International Fluid Power Conference, pp. 473–484, 2016.

[8] Zong, C., Zheng, F. et al. 'Computational Fluid Dynamics Analysis of the Flow Force Exerted on the Disk of a Direct-Operated Pressure Safety Valve in Energy System', Journal of Pressure Vessel Technology, Vol. 142, Nr. 1, 2020. DOI: 10.1115/1.4045131.

[9] Domagala, M., Fabis-Domagala, J. 'A Review of the CFD Method in the Modeling of Flow Forces', Energies, Vol. 16, Nr. 16, 2023. DOI: 10.3390/en16166059.

[10] Bernad, S. I., Susan-Resiga, R. 'Numerical Model for Cavitational Flow in Hydraulic Poppet Valves', Modelling and Simulation in Engineering, Vol. 2012, pp. 1–10, 2012. DOI: 10.1155/2012/742162.

[11] Hao, Q., Wu, W., Tian, G. 'Study on reducing both flow force and cavitation in poppet valves', Proceedings of the Institution of Mechanical Engineers, Part C:

Journal of Mechanical Engineering Science, Vol. 236, Nr. 23, pp. 11160–11179, 2022. DOI: 10.1177/09544062221110729.

[12] Simic, M., Herakovic, N. 'CFD Simulation and Optimization of Hydraulic ON/OFF Valve for Small Volume Flow', 8th International Fluid Power Conference, 2012.

[13] Simic, M., Herakovic, N. 'Reduction of the flow forces in a small hydraulic seat valve as alternative approach to improve the valve characteristics', Energy Conversion and Management, Vol. 89, pp. 708–718, 2015. DOI: 10.1016/j.enconman.2014.10.037.

[14] Xie, H., Tan, L. et al. 'Research on Influence of Different types of Orifice on Axial Steady-state Flow Force in Cartridge Proportional Valve', International Journal of Fluid Power, 2019. DOI: 10.13052/ijfp1439-9776.2021.

[15] Liu, J., Li, R. et al. 'Flow force research and structure improvement of cartridge valve core based on CFD method', Heliyon, Vol. 8, Nr. 11, e11700, 2022. DOI: 10.1016/j.heliyon.2022.e11700.

[16] Rundo, M., Altare, G. 'Comparison of Analytical and Numerical Methods for the Evaluation of the Flow Forces in Conical Poppet Valves with Direct and Reverse Flow', Energy Procedia, Vol. 126, pp. 1107–1114, 2017. DOI: 10.1016/j.egypro.2017.08.261.

[17] Roemer, D. B., Johansen, P. et al. 'Modeling of movement-induced and flow-induced fluid forces in fast switching valves', 2015 International Conference on Fluid Power and Mechatronics (FPM), pp. 978–983, 2015. DOI: 10.1109/FPM.2015.7337257.

[18] Gomez, I., Gonzalez-Mancera, A. et al. 'Analysis of the Design of a Poppet Valve by Transitory Simulation', Energies, Vol. 12, Nr. 5, pp. 889, 2019. DOI: 10.3390/en12050889.

[19] Finesso, R., Rundo, M. 'Numerical and Experimental Investigation on a Conical Poppet Relief Valve with Flow Force Compensation', International Journal of Fluid Power, Vol. 18, Nr. 2, pp. 111–122, 2017. DOI: 10.1080/14399776.2017.1296740.

[20] Hirose, N., Masuda, S. et al. 'Flow Patterns and Hysteresis Characteristic of a Poppet Valve', JFPS International Journal of Fluid Power System, Vol. 15, Nr. 2, pp. 53–61, 2022. DOI: 10.5739/jfpsij.15.53.

[21] Johnston, D. N., Edge, K. A., Vaughan, N. D. 'Experimental investigation of flow and force characteristics of hydraulic poppet and disc valves', Proceedings of the Institution of Mechanical Engineers, Nr. 205, pp. 161–171, 1991.

[22] Ferziger, J. H., Perić, M., Street, R. L. 'Computational Methods for Fluid Dynamics', Springer International Publishing, Cham, 2020. DOI: 10.1007/978-3-319-99693-6.

[23] Vinchurkar, S., Longest, P. W. 'Evaluation of hexahedral, prismatic and hybrid mesh styles for simulating respiratory aerosol dynamics', Computers & Fluids, Vol. 37, Nr. 3, pp. 317–331, 2008. DOI: 10.1016/j.compfluid.2007.05.001.

[24] Lintermann, A. 'Computational Meshing for CFD Simulations', 2021. DOI: 10.1007/978-981-15-6716-2_6.

[25] Ansys, Inc. 'Ansys Fluent Theory Guide - 2025R2', https://ansyshelp.ansys.com, 2025.

Biographies



Timo Rothhardt received the bachelor's degree in mechanical engineering from RWTH Aachen University in 2020 and the master's degree in product development in 2023. He is currently working at RWTH Aachen as scientific staff in the Institute for Fluid Power Drives and Systems (ifas). As part of the team high-performance components, his research areas include valves, CFD simulation, design of experiments and machine learning.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz’s awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Experimental Comparison of Polymer and Non-Ferrous Metal Slippers for Oil-Hydraulic Axial Piston Machines

Felix Schlegel*,[1], Katharina Schmitz[1]

*[1] institute for fluid power drives and systems, RWTH Aachen, Campus Boulevard 30, 52074 Aachen; *corresponding author: felix.schlegel@ifas.rwth-aachen.de,*

Abstract.

In many hydraulic machines, particularly axial piston units, tribological contacts commonly rely on leaded non-ferrous metal alloys. Increasing legislative restrictions, accelerated oil-aging and copper-corrosion are motivating the search for alternatives. This study demonstrates, using single-piston test rig experiments, that carbon-fiber-reinforced PEEK can replace copper-alloys in the slipper–swashplate contact. It is shown, that polymer slippers show only about 1.3 % lower efficiency than a non-ferrous metal reference but exhibit reduced friction in the mixed lubrication regime. Wear is slightly higher yet remains within the same order of magnitude ($\approx 10^{-10}$ mm³/Nm), indicating suitability for long-term operation. Wear mechanisms were analyzed using scanning electron microscopy and energy-dispersive X-ray spectroscopy. The results also reveal that polymer deformation significantly influences the effective hydrostatic relief. Variations of several percent with pressure and speed must be considered in polymer slipper design to ensure functional stability.

Keywords. Slipper, Polymer, Friction, Wear, Efficency, lead-free, Axial Piston Machine

1. INTRODUCTION AND STATE OF THE ART

Hydraulic drive systems are widely employed in mobile, industrial, and aerospace applications due to their high power density, controllability, and efficiency. They operate by converting mechanical into hydraulic energy and vice versa, enabling the transmission of large forces and torques within compact systems. Among hydraulic units, axial piston machines (APMs) are the most prevalent design, capable of functioning as both pumps and motors.

The structure of the most common APM type, the swashplate machine is shown in Fig. 1. In pump mode, the drive motor rotates the shaft and cylinder block, moving the pistons, which are pushed outward by the chamber pressure. The inclined swashplate converts this rotation into a reciprocating motion so that fluid is drawn in during the outward stroke and discharged during the return stroke. The stationary valve plate directs the flow to the high-pressure (HP) line during delivery and to the low-pressure (LP) line during suction. In motor mode, the principle reverses when the swashplate is tilted beyond zero in the opposite direction, causing the pistons to be driven outward by pressurized fluid and retracted as the fluid expands. The high hydrostatic forces on the pistons are supported by slippers. Together, the

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

piston–bushing, cylinder block–valve plate and slipper–swashplate interface form the three main tribological contacts determining machine efficiency and lifetime. [X]

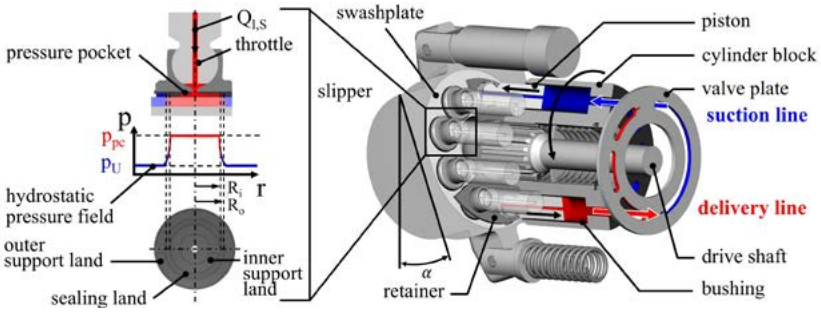


Figure 1.1: Structure of a swashplate type axial piston machine

To enable slippers to follow the swashplate with six degrees of freedom, they are connected to the pistons via spherical joints. This results in complex micro-motions influenced by frictional, centrifugal, and flow forces. This joint may be located on the piston (conventional) or on the slipper (inverse), with the latter allowing larger swashplate angles but leading to less favourable deformation. Because of the high normal loads and the need for operation at low speeds, the slipper–swashplate contact cannot rely solely on hydrodynamic lubrication. A small leakage volume flow from the displacement chambers is routed through bores in piston and slipper to form a hydrostatic pressure field beneath the running surface. There, the un-slitted sealing land defines the hydrostatic pressure drop, achieving 92–105 % load compensation [1, 2]. This hydrostatic relief represents a balance between leakage and friction reduction. Excessive relief may cause lift-off, resulting in leakage, high tilting induced wear or malfunction, which can be prevented by a retainer or, where feasible, an integrated limiting-orifice restricting flow. However, the latter cannot be implemented in all slipper geometries due to manufacturing constraints. Usually, slotted support lands increase slipper stiffness, promote hydrodynamic lift, and reduce local contact pressure. Mechanical slipper failures include joint or neck fracture, whereas tribological failure occurs when wear degrades sealing or alters hydrostatic balance, leading to loss of function.

Afterwards, Sec. 1.1 outlines the calculation of hydrostatic relief, efficiency, and wear of slippers, followed by material selection and the state of research on polymer slippers in Sec. 1.2. The experimental setup and test specimens are described in Sec. 2.1 and 2.2, and the experimental results are presented in Sec 3. Sec 4 provides a critical discussion, followed by Sec 5 concluding with a summary and outlook.

1.1. CALCULATION OF SLIPPERS

This section introduces only the equations required to interpret the results in Sec. 3, details in slipper dynamics are provided in [1]. In general, the hydrostatic pressure field $p(r)$ of a slipper can be described as a potential field along the sealing land radius r with the boundary pressures (pocket p_i, ambient p_a) and the gap height profile $h(r)$, as shown in Eq. (1.1).

$$p(r) = p_i - (p_i - p_a) \cdot \int_{R_i}^{r} \frac{1}{r \cdot h(r)^3} dr \cdot \left(\int_{R_i}^{R_o} \frac{1}{r \cdot h(r)^3} dr \right)^{-1} \tag{1.1}$$

This yields the leakage flow through the annular gap Q as a function of viscosity η and the

sealing land geometry defined by inner radius R_i and outer radius R_o (Eq. (1.2)).

$$Q = \frac{\pi}{12\,\eta}\left(\int_{R_i}^{R_o} \frac{1}{r \cdot h(r)^3}\,dr\right)^{-1} (p_i - p_a) \tag{1.2}$$

The hydrostatic force F_S reducing the slipper normal force F_N is obtained by integrating the pocket pressure profile and the sealing land pressure profile, as shown in Eq. (1.3).

$$F_S = \int_0^{Ro} 2\,\pi\,r\,p(r)dr = \underbrace{\int_0^{Ri} 2\,\pi\,r\,p_i\,dr}_{\text{pocket pressure}} + \underbrace{\int_{Ri}^{Ro} 2\,\pi\,r\,p(r)\,dr}_{\text{sealing land}} \tag{1.3}$$

According to Manring [1], this yields Eq. (1.4), and Eq. (1.5) for an even parallel gap.

$$F_{S,even} = p_i\,\frac{\pi(R_o^2 - R_i^2)}{2\ln(R_o/R_i)} \tag{1.4}$$

$$Q_{even} = \frac{\pi}{6\,\eta}\,\frac{h_G^3}{ln(R_o/R_i)}\,p_i \tag{1.5}$$

Based on this, the static degree of hydrostatic relief B_S can be defined with Eq. (1.6) [1] using the piston force loading the slipper F_P and the swashplate angle α.

$$B_S = F_S \cdot F_P^{-1} \cdot \cos(\alpha) \tag{1.6}$$

Beyond this simple approximation, a logarithmic (Eq. (1.7)) and two polynomial gap profiles of third (Eq. (1.8)) and fifth order (Eq. (1.9)) were studied, representing simulated deformed gaps [3]. The minimum gap height h_o is defined at R_o, and the maximum height h_i at R_i results from deformation (Eq. (1.10)). The scaling factor k (Eq. (1.11)) describes the latter as a function of socket height $H_{S,1}$, layer height $H_{S,2}$, and the young's moduli of socket E_1 and layer E_2. Eqs. 1 and 2 were solved analytically with Eq. (1.7), while Eqs. (1.8) and (1.9) were integrated numerically.

$$h(r) = h_a + \frac{p_i \cdot H_S}{E} \cdot \frac{ln(R_a/r)}{ln(R_a/R_i)} = h_a + k \cdot \frac{ln(R_a/r)}{ln(R_a/R_i)} \tag{1.7}$$

$$h(r) = h_a + k \cdot (8.75\,x^3 - 15.3\,x^2 + 7.12\,x) \tag{1.8}$$

$$h(r) = h_a + k \cdot (3.77\,x^5 - 18.23\,x^4 + 30.32\,x^3 - 20.24\,x^2 + 3.83\,x + 0.78) \tag{1.9}$$

$$h_i = h_a + k \tag{1.10}$$

$$k = p_i \cdot \left(\frac{H_{S,1}}{E_1} + \frac{H_{S,2}}{E_2}\right) \tag{1.11}$$

$$x = \frac{r - R_i}{R_o - R_i} \tag{1.12}$$

Due to hydrostatic relief, the frictional coefficient μ of a slipper (CoF) cannot be determined from the ratio of friction F_R to normal force F_N (Eq. (1.13)). Instead, the degree of relief must be considered (Eq. (1.14), [1]), making the calculated friction coefficient dependent on the gap-height profile. All CoFs shown below refer to the definition of Eq. 14.

$$\mu = \frac{F_R}{F_N} \tag{1.13}$$

$$\mu_S = \frac{F_R}{F_N \cdot (1 - B_S)} \tag{1.14}$$

As slipper design involves a trade-off between friction, leakage, and wear due to hydrostatic relief, it is not useful to compare slippers solely by CoF. For consistent comparison, friction and leakage are evaluated jointly by introducing a slipper efficiency η_{total} (Eq. (1.15), [1]), analogous to APMs, combining hydromechanical η_{hm} and volumetric efficiency η_{vol}.

$$\eta_{total} = \eta_{hm} \cdot \eta_{vol} \tag{1.15}$$

The efficiency definitions applied here differ from those of APMs, as only a single tribological contact is considered. In multi-stage tribosystems, efficiency is defined in series ($\eta_{serial,i}$, Eq. (1.16)), requiring knowledge of both stage losses $P_{loss,i}$ and stage output power $P_{out,i}$. The latter differs from the total machine output power P_{out}, which is reduced by the losses of all stages. As the slipper is isolated here, neither the total input power nor the losses of other stages (e.g., valve plate or piston) are known, rendering this definition inapplicable. Instead, a relative efficiency $\eta_{rel,i}$ is introduced (Eq. (1.17)), referencing stage losses to the total machine output power and thus slightly overestimating efficiency.

$$\eta_{total} = \prod_{i=1}^{n} \eta_{serial,i} = \prod_{i=1}^{n} \frac{P_{out,i}}{P_{in,i}} = \prod_{i=1}^{n} \frac{P_{out,i}}{P_{out,i} + P_{loss,i}} \tag{1.16}$$

$$with: P_{out,i+1} = P_{out,i} - P_{loss,i+1}$$

$$\eta_{rel,i} = \frac{P_{out}}{P_{out} + P_{loss,i}} \quad with: \eta_{total} \leq \prod_{i=1}^{n} \eta_{rel,i} \tag{1.17}$$

The difference between $\eta_{serial,i}$ and $\eta_{rel,i}$ varies with the stage under consideration. For n identical partial losses, the maximum absolute deviation can be expressed by Eq. (1.18).

$$\Delta\eta_{max}(\eta_{total}) = \eta_{total} - \left(1 + \frac{1}{n}\left(\frac{1}{\eta_{total}} - 1\right)\right)^{-n} \tag{1.18}$$

Pump losses are assumed to be evenly distributed among four tribo-contacts here: slipper–swashplate, piston–cylinder, valve plate–cylinder block, and auxiliary components such as bearings. For total pump efficiencies of 95 %, 90 %, and 85 %, the maximum deviation between relative and serial slipper efficiency amounts to +0.00094 %, +0.0038 %, and +0.0086 %, respectively. This difference is therefore regarded as negligible.

The calculation of individual efficiencies in an APM differs depending on whether it operates in pump or motor mode. In pump mode, η_{hm} is derived from the theoretical drive torque $M_{th,p}$ (Eq. (1.19), [1]), while in motor mode it is based on the effective output torque $M_{th,m}$ (Eq. (1.20), [1]). In pump mode η_{vol} is determined from the effective discharge flow $Q_{th,p}$ (Eq. (1.21), [1]) and in motor operation from the theoretical intake flow $Q_{th,m}$ (Eq. (1.22), [1]). Here, M_l is the friction loss torque and Q_l the hydrostatic leakage flow.

$$\eta_{hm,p} = \frac{M_{th,p}}{M_{th,p} + M_l} \tag{1.19}$$

$$\eta_{hm,m} = \frac{M_{th,m} - M_l}{M_{th,m}} \tag{1.20}$$

$$\eta_{Q,p} = \frac{Q_{th,p} - Q_l}{Q_{th,p}} \tag{1.21}$$

$$\eta_{Q,m} = \frac{Q_{th,m}}{Q_{th,m} + Q_l} \tag{1.22}$$

In both operating modes, M_{th} and Q_{th} are expressed by Eqs. (1.23) [1] and (1.24) [1], based on the pressure difference Δp, displacement volume V, and rotational speed n.

$$M_{th} = \frac{\Delta p \cdot V}{2\,\pi} \tag{1.23}$$

$$Q_{th} = n \cdot V \tag{1.24}$$

In an APM, slipper losses vary with the rotational angle due to kinematics and commutation. In this study, they are derived from measurement data using Eqs. (1.25) and (1.26), assuming that half of the j slippers operate at low pressure (p_{LP}) and the other half at high pressure (p_{HP}) along the path radius R. The low-pressure level is set to 50 bar, as measurements below this value show reduced quality with the existing test setup.

$$M_l(n,p) = \sum_{j=1}^{k} R \cdot F_{R,j}(n,p_j) = \frac{j}{2} \cdot R \cdot F_R(n,p_{LP}) + \frac{j}{2} \cdot R \cdot F_R(n,p_{HP}) \tag{1.25}$$

$$Q_l(n,p) = \sum_{j=1}^{k} Q_{l,S,j} = \frac{j}{2} \cdot R \cdot Q_{l,S}(n,p_{LP}) + \frac{j}{2} \cdot R \cdot Q_{l,S}(n,p_{HP}) \tag{1.26}$$

In addition to efficiency, slipper wear is evaluated in this study. Therefore, the wear coefficient K_W is calculated according to Eq. (1.27) from the weight loss m_W, with v_r denoting the sliding speed, t the test duration, and ρ the density of the running surface.

$$K_W = \frac{m_W}{F_N \cdot (1 - B_{S,even}) \cdot v_r \cdot t \cdot \rho} \tag{1.27}$$

1.2. STATE OF THE ART FOR POLYMER SLIPPERS

Slippers in axial piston machines are typically designed as hard–soft material pairs, combining hardened steel with non-ferrous metals (NFM) such as brass or bronze to ensure run-in capability, particle embedding, and reduced adhesion [4]. However, many conventional alloys contain toxic lead and promote lubricant degradation or corrosion due to copper, motivating research alternative soft materials. High-performance polymers, particularly PEEK-based compounds, are being investigated to address these limitations.

Research on polymer slippers for oil-hydraulic systems remains scarce. Stryczek [5] demonstrated the general feasibility of plastic components in fluid power applications, with performance strongly dependent on design. More comprehensive work exists for water-hydraulic pumps. Rokala [6], Schoemacker [7], Guan et al. [8], and Ni et al. [9] examined polymer slippers in such systems, showing that their lower friction can offset water's poor lubricity. Among tested materials, carbon-fiber-reinforced (CF) PEEK exhibited the most favorable performance. Li et al. [10] compared CF-PEEK against various steels in both "pin-on-disc" and water-hydraulic APM tests, reporting wear coefficients on the order of 10^{-7} mm³/Nm and improved behavior under hydrostatic pressure.

Building on these findings, the authors introduced a "slipper-on-disc" test [11] under oil-hydraulic conditions, revealing substantially lower friction and wear than under dry and water lubricated conditions. The present study extends this work by examining functional sample polymer slippers on a component-level single-piston test bench, to evaluate polymer slipper performance and, if possible, to provide proof of functionality.

2. METHODOLOGY AND EXPERIMENTAL SETUP

The implementation of polymer slippers in this study was preceded by parametric simulations [3], which showed that fully polymer designs deform excessively, preventing a planar load-carrying surface and a defined hydrostatic relief. Therefore, hybrid slippers with a steel

socket and a thin polymer running surface were developed. As the ball joint connection could not be manufactured, commercially available NFM-slippers were modified by adapting the steel socket and removing the running layer. For high-resolution optical and gravimetric wear assessment and proper sensor integration, comparatively large slippers were designed with a piston diameter of 25 mm and a base diameter of 34 mm. The pistons originate from a 165 cm³ pump for mobile applications, rated at 450 bar, 3000 rpm, and 326 kW. The test rig for these slippers is described in Section 2.1, the design and manufacturing of the slippers in Section 2.2.

2.1. SINGLE-PISTON TEST RIG

The test rig (Fig. 2.1) represents a single slipper-swashplate contact in inverse kinematics to simplify measurements: the slipper (1) remains stationary while an interchangeable flat disk (2, lapped, 100Cr6) is driven up to 3000 rpm (sliding speed 16.6 m/s) without piston stroke.

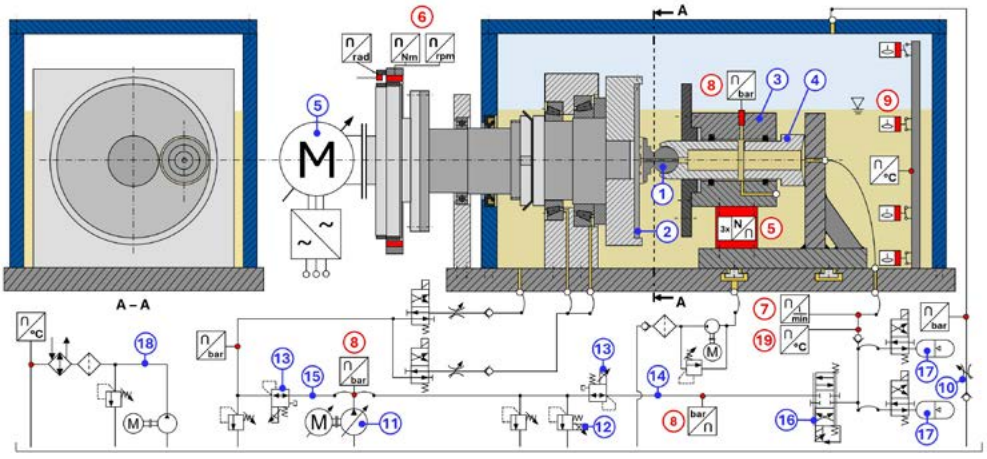


Figure 2.1: Structure of the single-piston polymer-slipper test rig

The slipper is mounted in a piston holder (3) that is axially force-balanced by a compensation piston (4) which is supported on a connecting plate. To enable the measurement of slipper friction with a three-axis force sensor (5, Kistler 9327C, ±8 kN, 0.5 % Full Scale Error (FSE)). A torque measurement shaft (6, Atesteo F1iS, 0-1 kNm, 0-20 krpm,0.03 % FSE, 60 ppr) provides an additional friction reference to validate the measurement principle and enables rotational speed measurement. As the compensating piston setup is inherently leaking and requires the piston holder's center of gravity to coincide with the force sensor centre, a second, sealed holder without force sensor is used to measure hydrostatic leakage flow by a volume flow sensor (7, VSE VS 0.1, 0.01-10 l/min, 0.3 %FSE), whereby the piston pressure is measured in both cases with a pressure sensor (8, Hydac HDA4744, 0-400 bar, 0.25 % FSE). Both holders are aligned by stops and pins for reproducibility. By float switches (9) and a variable throttle (10) in the case drain, the tests can be conducted under different fluid levels and with a controlled case pressure of up to 10 bar.

The hydraulic supply is provided by a pump (11) delivering up to 400 bar at 30 l/min. System pressure is controlled via an electro-proportional pressure valve (12, max. 350 bar), while two pressure-reducing valves (13) allow separate control of the slipper supply (14)

and external bearing lubrication (15). A fast servo valve (16) with a step response of 8 ms enables simulation of pump commutation and pulsing of the slipper. Optionally, pump pulsation can be damped with two accumulators depending on pressure (17). The fluid (HLP 46) is conditioned by a circuit (18) including 5 µm filtration, a 30 kW heater, and 40 kW cooling, controlling the slipper inflow temperature at 40 °C (19, Danfoss MBT3270, -50-300 °C, ±1 % FSE). Tests for friction and leakage are fully automated, measuring 30 sec speed ramps up to 3000 rpm at pressures from 0–300 bar in 50 bar increments

For long-term durability tests, the test rig is equipped with superimposed state controllers and enclosed in sound insulation. Surface changes and wear are assessed optically before and after testing using a focus variation microscope (lateral ±0.44 µm, vertical 10 nm) and a confocal profilometer (lateral ±0.5 µm, vertical ±1 µm). Mass loss is determined after ultrasonic cleaning in a class 7/8 DIN EN ISO 14644-1 cleanroom using a class I DIN EN 45501 analytical balance (σ = 0.016 mg).

2.2. POLYMER SLIPPER DESIGN AND MANUFACTURING

The running surfaces of the polymer slippers shown in Fig. 2.2 were machined from rod-extruded PEEK with a melt viscosity of 150 Pa·s at 400 °C (grade 150G). Higher-viscosity grades such as 450G, which offer improved wear resistance, are typically not extruded but could be suitable for mass production by injection molding. All polymer slippers share the same geometry as the NFM reference to ensure direct comparability. The hydrostatic relief degree of the functional samples amounts 92.75 %.

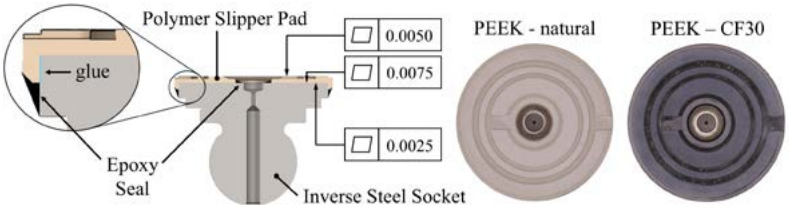


Figure 2.2: Structure of the polymer-slipper functional samples

Besides unfilled 150G PEEK, reinforced variants are tested: 30% carbon fiber (PEEK-CF30) and a composite with 10% carbon fiber, 10% graphite, and 10% PTFE (PEEK-CF10-GR10-PTFE10, HPV). After machining, the slippers were lapped, as surface finish significantly affects wear [11]. A critical challenge for the design investigated here is the tolerance chain involving flatness of the steel base and both pad surfaces. The achieved flatness, shown for the worst case in Fig. 2.2, remained below that of industrially grinded slippers. To avoid changing running surface flatness, the pads were glued only between the pad overhang and the socket outer diameter. The gap between both is sealed with epoxy resin to prevent detachment under pocket and case drain pressure. These steps were required only for the functional samples here, in industrial implementation, overmolding the steel base would enable higher flatness without the need of bonding or sealing.

3. EXPERIMENTAL RESULTS

This section first presents experimental results friction, leakage and efficiency of the polymer slippers in Sec. 3.1, followed by the evaluation of wear in Sec. 3.2.

3.1. FRICTION, LEAKAGE and EFFICIENCY

To evaluate friction and efficiency, the compensation-piston measurement principle introduced in Sec. 2.1 was first validated. Therefore Fig. 3.1 shows friction results of the commercial NFM reference slipper for acceleration (up, 0–3000 rpm) and deceleration (down, 3000–0 rpm) ramps, measured with both the force and the torque sensor. All shown data of this NFM slipper serve only as a reference within the observed conditions. Since material and surface finish of the disc differ, the presented results do not represent the commercial tribosystem. Tests were performed with the housing drained to eliminate flow-induced forces. The bearing friction of the test rig was characterized as a function of pressure and speed and subtracted from the torque data. The good qualitative and quantitative agreement between both methods confirms that the friction of the compensating piston is negligible.

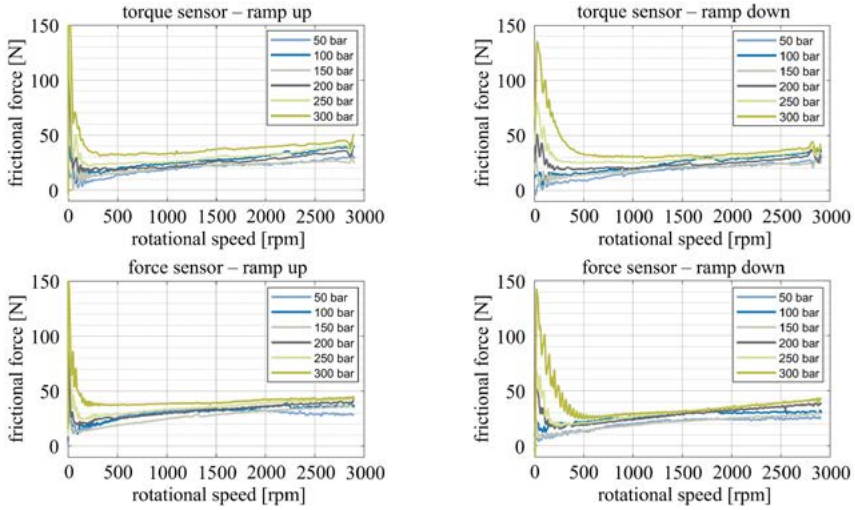


Figure 3.1: Comparison of friction measurements for non-ferrous metal

All curves in Fig. 3.1 show typical Stribeck behavior. During acceleration, the transition from mixed to full-film lubrication (release point) shifts with pressure from about 75 rpm at 50 bar to 250 rpm at 300 bar, consistent with previous findings [12]. A sensor-independent hysteresis is observed, where the difference between acceleration and deceleration at standstill represents static friction. This increases proportional with pressure up to 57 N at 250 bar and decreases afterwards to 16 N at 300 bar. In the mixed lubrication regime, the curves deviate due to hysteresis, leading to the release point occurring at higher speeds (≈450 rpm at 300 bar) during deceleration. As the NFM exhibits no significant viscoelasticity, the hysteresis is attributed to micro-motions of the slipper rather than material damping.

Fig. 3.2 compares the friction behavior of the NFM reference slipper with that of the polymer slipper functional samples during acceleration ramps. The natural PEEK slipper exhibits a distinctly different behavior compared to the reinforced variants. One reason is its poor machinability, which results in inferior surface flatness and sealing, leading to a significant increase in fluid friction. Additional contributing factors are discussed in Sec. 4. All polymer slippers show a pronounced pressure-dependent friction behavior, with a release point occurring below 100 rpm. At low speeds, the PEEK-CF30 slipper exhibits up to 43 % lower

friction compared to the NFM reference. In the full-film lubrication regime, however, its friction partially exceeds that of the NFM reference by up to 48 %.

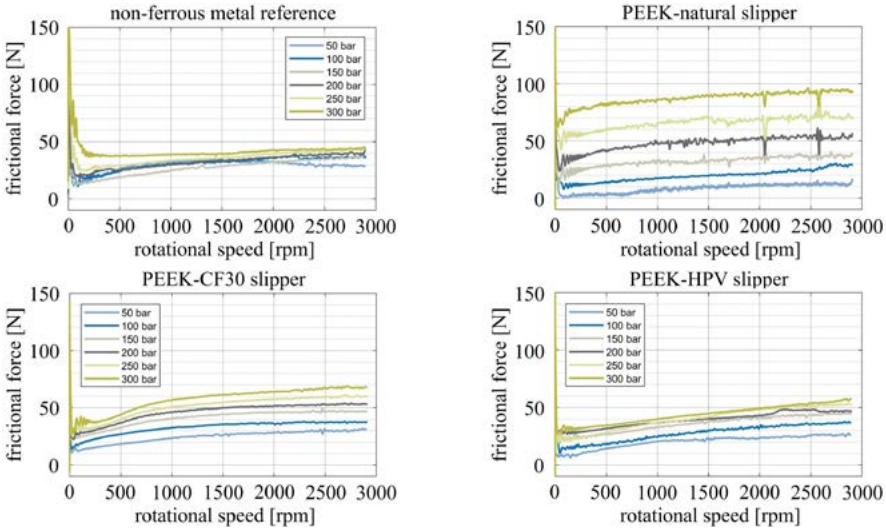


Figure 3.2: Friction comparison of the polymer slipper functional samples

Fig. 3.3 shows a comparison of slipper leakages. The measured leakage represents a superposition of the slipper–swashplate and ball-joint contact leakages. As all slippers were manufactured using pistons from the same production batch, the data is considered comparable.

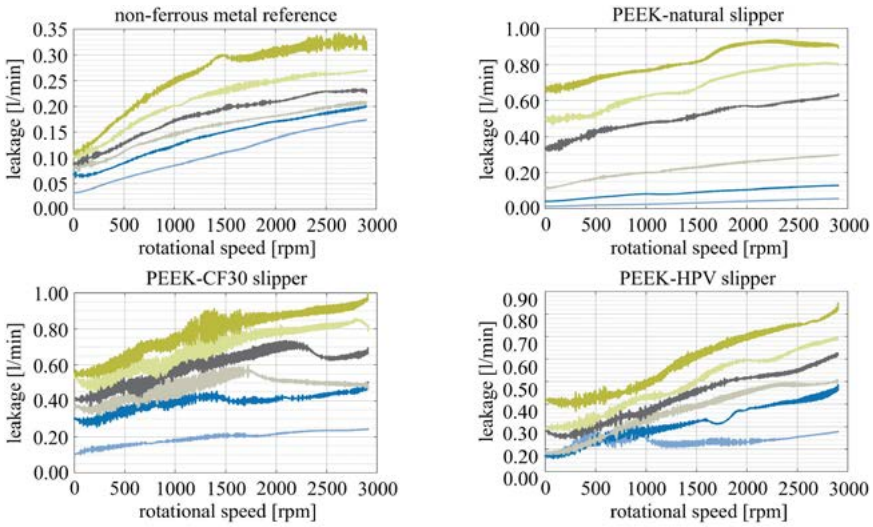


Figure 3.3: Leakage comparison of the polymer-slipper functional samples

The data shown is only lightly filtered, as leakage noise increased with pressure. A minor hysteresis between acceleration and deceleration was observed at low speeds too, with leakage being up to 10 % higher during deceleration but returning to the same value at standstill. Repeated tests revealed a reproducibility band of about ± 8 % for all slippers, including the

NFM reference. It is visible that at lower pressures, leakage increases almost linearly, while at higher pressures it tends to saturate. As expected from the limited flatness described in Sec. 2.2, the NFM reference exhibited the lowest leakage. However, the stronger pressure dependence of the polymer slippers suggests additional effects, discussed in Sec. 4. For the PEEK-CF30 slipper, leakage exceeded the NFM reference by about 26 % at 50 bar and 185 % at 300 bar, yet absolute values remain small compared to the pump's delivery flow (max. 495 l/min). That's why the resulting relative efficiencies are discussed in Fig. 3.4 using the equations from Sec. 1.1.

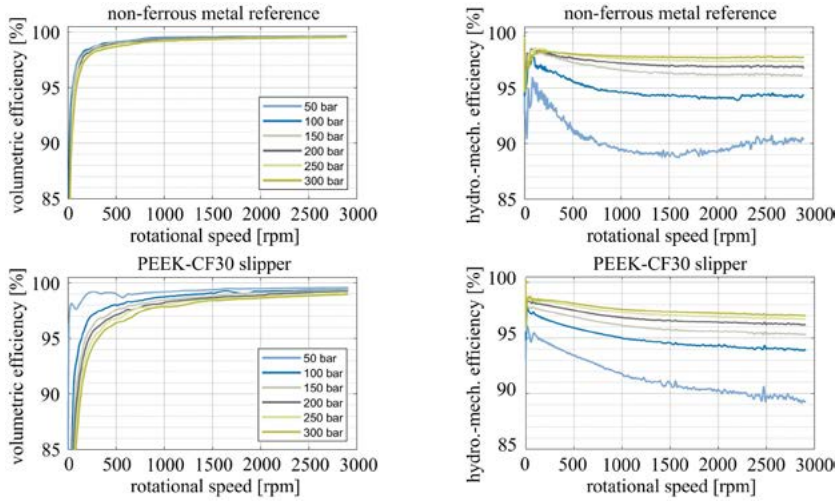


Figure 3.4: Comparison of volumetric and hydromechanical efficiencies

In addition to the partial efficiencies shown in Fig. 3.4, Fig. 3.5 presents a comparison of overall efficiency maps for all tested slippers.

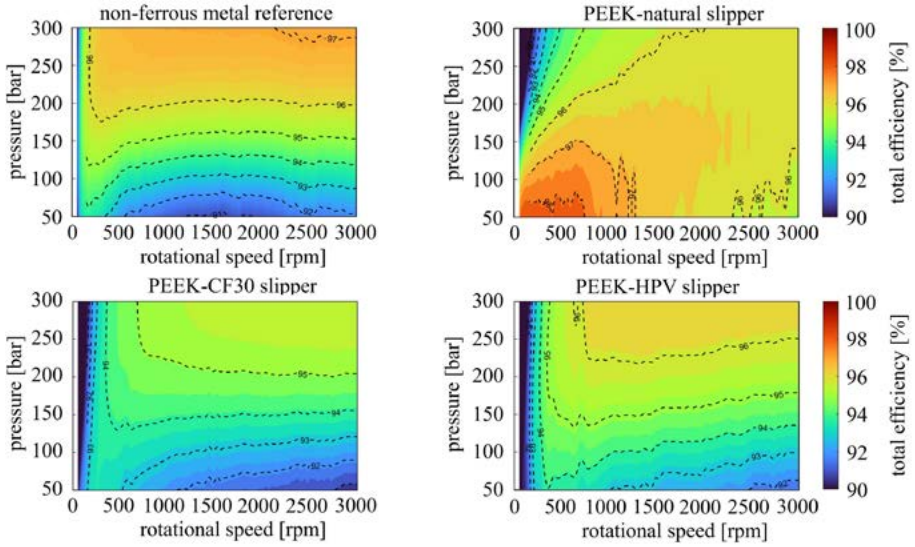


Figure 3.5: Comparison of total efficiencies

The volumetric and hydromechanical efficiencies exhibit the typical trends reported in the literature [1, 2]. Due to lower friction at low speeds, η_{hm} is higher for the polymer slippers, decreasing at higher speeds. However, η_{vol} reveals the stronger pressure dependence and increased leakage, resulting in lower overall efficiency.

The commercial pump is specified with a minimum operating speed of 400 rpm, consistent with the mixed lubrication regime and reduced total efficiency observed in this range. The recommended operating range is 1400–2200 rpm and 230–320 bar, where the overall efficiencies of the PEEK-CF30, PEEK-HPV, and natural PEEK slippers reach approximately 95.4 %, 96.1 %, and 96.2 %, respectively, being only slightly below NFM with 96.7 %.

3.2. WEAR

To assess wear behavior, the mass loss was first determined after 100 h of testing at 1450 rpm and 200 bar to evaluate the general material suitability. Due to the low density of polymers (PEEK-CF30 1.4 g/cm³ vs. NFM 8.1 g/cm³), shorter tests would not allow gravimetric resolution. Fig. 3.6 presents a boxplot with the resulting wear coefficients (Eq. 1.27) together with the friction coefficients over the entire speed range.

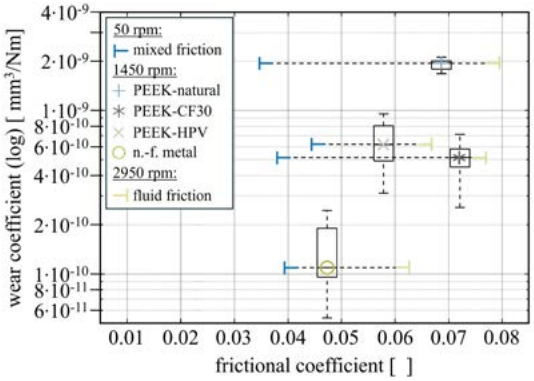


Figure 3.6: Gravimetric wear comparison after 100 h at 200 bar and 1450 rpm

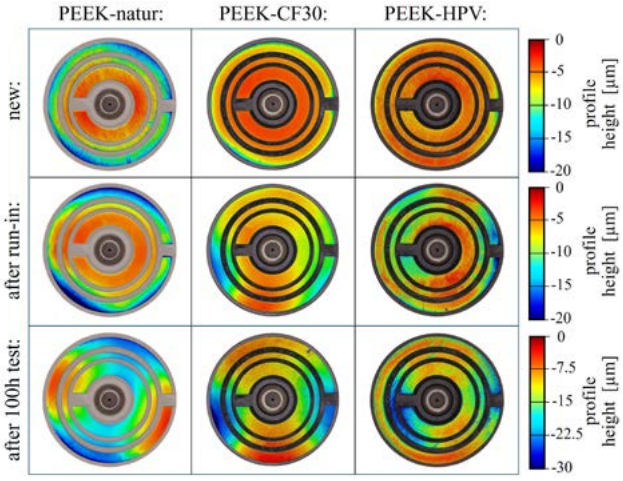


Figure 3.7: Optical wear comparison after 100 h at 200 bar and 1450 rpm

PEEK-CF30 and PEEK-HPV show wear coefficients in the same order of magnitude as NFM (< 10^{-9} mm³/Nm), typical for fully fluid-lubrication and suitable for long-lifetimes. On average, PEEK-CF30 exhibits roughly five-times higher wear, though the scatter indicates that gravimetrical evaluation mainly provides order-of-magnitude estimate in this wear range. In contrast, natural PEEK displays substantially higher wear and is deemed unsuitable. Accordingly, reinforced PEEK is generally suitable, though local wear at the sealing land may still limit durability. Therefore Fig. 3.7 presents optical wear localization.

The top of Fig. 3.7 shows the pronounced non-flatness of new slippers, especially for natural PEEK. Below it is evident that the local wear pattern forms mainly during the 30 min run-in procedure covering the entire operational range, with only quantitative changes thereafter. Wear concentrates near the pressure-relief grooves due to local deformation and stiffness loss, causing groove depression and partial wear transfer onto the sealing land. This may increase leakage or collapse hydrostatic relief over time. Therefore, polymer deformation necessitates geometric compensation as demonstrated in [3], meaning the NFM geometry cannot be directly adopted.

To identify the dominant wear mechanisms, scanning electron microscopy (SEM) and energy-dispersive X-ray spectroscopy (EDX) were performed before and after testing. The aim was to exclude thermal or chemical degradation and to determine whether the polymers form the thin transfer film, typically being responsible for the low wear of tribologically optimized polymers, even under the prevalent high gaps, hydrostatic relief, and smooth surfaces. Figs. 3.9–3.11 show the SEM images, and Figs. 3.12–3.14 the EDX spectra.

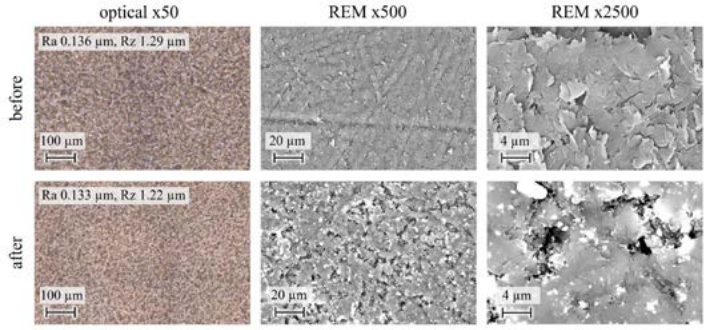


Figure 3.8: PEEK-natural surface before and after wear test

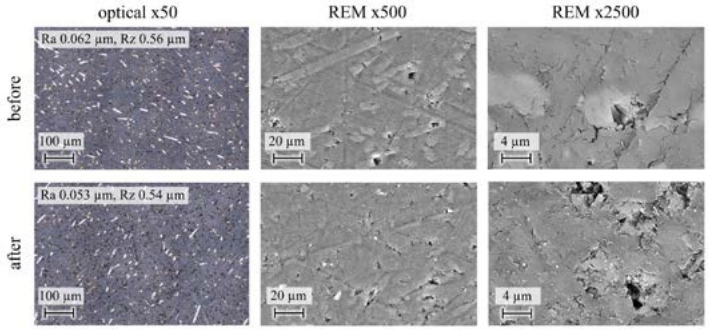


Figure 3.9: PEEK-CF30 surface before and after wear test

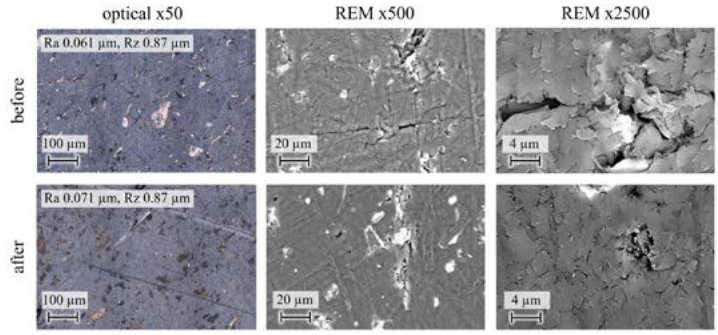


Figure 3.10: PEEK-HPV surface before and after wear test

The SEM images at 2500× magnification reveal surface smoothing for all polymers, with abrasion as the dominant wear mechanism and occasional fiber pull-out in reinforced grades. No signs of melting, structural damage, or chemical attack were detected.

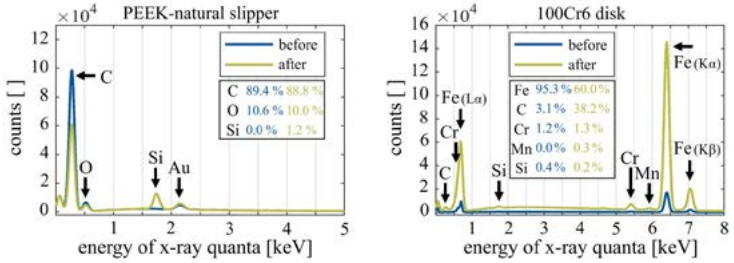


Figure 3.11: EDX-spectra of the PEEK-natural slipper and counter disc

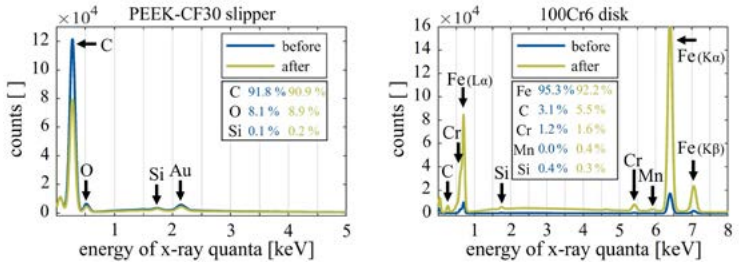


Figure 3.12: EDX-spectra of the PEEK-CF30 slipper and counter disk

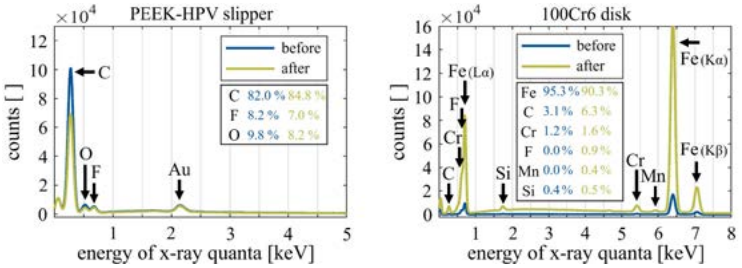


Figure 3.13: EDX-spectra of the PEEK-HPV slipper and counter disk

The EDX spectra peaks are semiquantitative, showing elemental presence. Calculated weight percentages are shown besides. Gold peaks originate from sputter coating for conductivity. All polymer slippers contain carbon and oxygen as main PEEK elements (hydrogen cannot be detected), with negligible compositional change during testing, as a thin graphite/PTFE layer forms already during lapping. Minor silicon traces may indicate reactions with silicone-based antifoam additives. On the 100Cr6 counter surfaces, increased carbon content after testing shows the formation of a transfer film. By locally dissolved measurements a continuous graphite/PTFE-based film could be confirmed, providing a stable, low-wear behavior in the slipper contact.

4. DISCUSSION

To explain the pressure-dependent behavior and differences between NFM and polymer slippers, leakage and hydrostatic relief were numerically evaluated using Eqs. (1.1)–(1.6) and (1.10)–(1.12) for deformed gaps with logarithmic (Eq. (1.7)) and polynomial (3rd- and 5th-order, Eqs. (1.8) (poly3), (1.9) (poly5) profiles based on simulated deformations from [3]. Fig. 4.1 shows the NFM and PEEK-CF30 gap geometries (left), the pressure dependence of the poly5 gap (center), and the resulting pressure profiles (right).

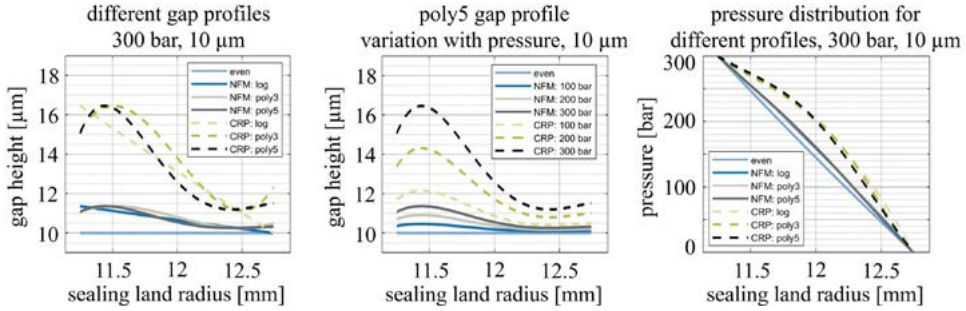


Figure 4.1: Comparison of pressure distributions for different gap profiles

The lower stiffness of polymers (NFM 80 GPa; natural PEEK 4.3 GPa; PEEK-HPV 5.9 GPa; PEEK-CF30 7.7 GPa) leads to significant pressure-dependent deformation, which enlarges the high-pressure area and shifts the hydrostatic pressure drop outward. This reduces flow conductance and leakage while increasing the hydrostatic relief (Fig. 4.2).

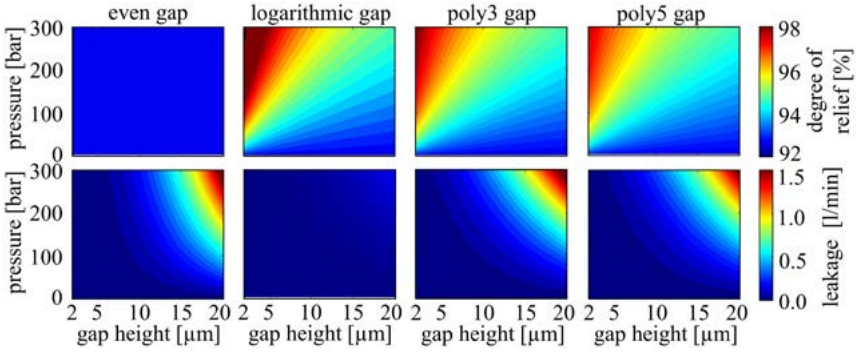


Figure 4.2: Influence of gap profile for PEEK-natural

Fig. 4.2 shows that the logarithmic gap poorly fits the measured leakage (Fig. 3.3), while the poly5 profile agrees reasonably. Moreover, for natural PEEK, the calculated hydrostatic relief deviates by up to 4.9 % at 300 bar and a 2 µm gap. Fig. 4.3 summarizes deviations for all materials: ≤ 2.2 % for NFM and up to 4.5 % for PEEK-CF30.

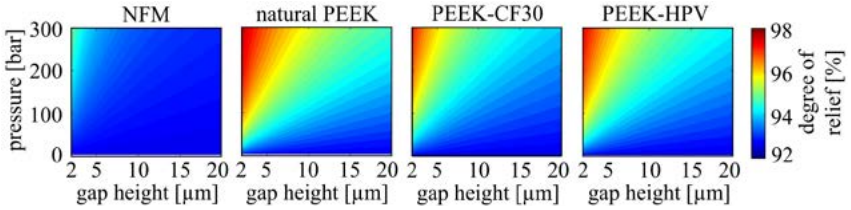


Figure 4.3: Comparison of relief change for different materials

Although approximate, these results highlight the deformation effects relevant for interpreting Sec. 3. Accurate relief design requires precise elastohydrodynamic (EHL) analysis. At 10 µm and 200 bar, PEEK-CF30 shows a 1.3 % higher relief than NFM, leading to 19 % lower residual load. Consequently, wear and friction coefficients become less comparable. Due to higher gaps and lower residual load, the transition to full-film lubrication shifts to lower speeds. The calculated leakage of PEEK-CF30 is 36 % higher than for NFM, indicating that surface non-flatness contributes at least similarly to leakage deviation. In summary, deformation effects can explain the measured results. While moderate deformation can be beneficial by enhancing pressure-dependent relief and reducing friction and wear, excessive elasticity, as seen for natural PEEK, renders the material unsuitable. Therefore, reinforced polymers are required, and uneven deformation must be prevented by geometric design adjustments to avoid non-uniform wear or early sealing land failure. Moreover, a precise lapping process is necessary to achieve even running surfaces.

5. CONCLUSION AND OUTLOOK

This study experimentally evaluated the potential of high-performance PEEK-based polymers to replace non-ferrous metals (NFM) in the slipper–swashplate contact of axial piston machines using a single-piston test rig. Polymer slippers showed a stronger pressure dependence due to deformation, making stiff grades such as PEEK-CF30 most suitable. PEEK-CF30 exhibited up to 43 % lower friction at low speeds and up to 48 % higher friction at higher speeds, the latter caused by increased fluid friction from higher manufacturing-related leakage. Leakage exceeded the NFM reference pressure dependent by about 26–185 %, yet remained small relative to the pump flow, yielding total efficiencies above 95 %, only about 1 % lower than NFM. The wear coefficient of PEEK-CF30 ($5.9 \cdot 10^{-10}$ mm³/Nm) is slightly higher, but in the same order of magnitude as NFM. SEM analysis revealed two-body abrasion as the main wear mechanism, with no thermal or chemical degradation, while EDX confirmed a thin transfer film, ensuring stable low friction and wear. Optical wear analysis indicated that the tested geometry is unsuitable for polymers due to excessive, asymmetric deformation. It can therefore be concluded that polymer slippers are feasible for use in axial piston machines, provided that geometric adaptations compensate deformation.

Future work will investigate PEEK-CF30 slippers with optimized geometry on the same test rig to quantify deformation compensation and achieve uniform wear for long-term operation.

Subsequently, full-machine tests will evaluate the efficiency and durability of pumps equipped entirely with polymer slippers.

6. ACKNOWLEDGEMENT

This work was created within the project “Improvement of oil-hydraulic polymer slippers using a physically based transient design method” (funding code 531424256) funded by the German Research Foundation DFG. The authors want to thank the DFG for this support.

7. REFERENCES

[1] N. D. Manring, ‘Fluid Power Pumps & Motors: Analysis, Design, and Control’, 1st ed. New York, USA: McGraw-Hill Education, 2013, ISBN: 978-0-07-181220-7.

[2] M. Ivantysynova and J. Ivantysyn, ‘Hydrostatic Pumps and Motors: Principles, Design, Performance, Modeling, Analysis, Control and Testing’, 1st ed. New Delhi, India: Tech Books International, 2003, ISBN: 81-88305-08-1.

[3] F. Schlegel, K. Schmitz, ‘Geometric design of axial piston machine slipper-bearings made of high-performance plastics’, Engineering Research, vol. 89, no. 26, 2025, https://doi.org/10.1007/s10010-025-00813-2.

[4] B. Reetz and T. Münch, ‘Challenges of novel lead-free alloys in hydraulics’, in Proc. 12th Int. Fluid Power Conf., vol. 1, pp. 17–25, Technische Universität Dresden, Dresden, Germany, 2020, https://doi.org/10.25368/2020.9

[5] J. Stryczek, M. Banaś, and J. Krawczyk, ‘The fluid power elements and systems made of plastics’, Procedia Engineering, vol. 176, pp. 600–609, 2017, https://doi.org/10.1016/j.proeng.2017.02.303.

[6] M. Rokala, ‘Analysis of Slipper Structures in Water Hydraulic Axial Piston Pumps’, Doctoral thesis, Tampere Univ. of Technology, Publ. 1055, Tampere, Finland, 2012, https://urn.fi/URN:ISBN:978-952-15-2872-9.

[7] F. Schoemaker, H. Murrenhoff, ‘Piston Slippers for Robust Water Hydraulic Pumps’, Proceedings of the 11th Int. Fluid Power Conf. Vol1, pp.236-247, RWTH Aachen University, Aachen, 2018, https://doi.org/10.18154/RWTH-2018-224460.

[8] Z. Guan et al., ‘Friction and wear characteristics of CF/PEEK against 431 stainless steel under high hydrostatic pressure water lubrication’, Materials & Design, vol. 196, art. no. 109057, 2020, https://doi.org/10.1016/j.matdes.2020.109057.

[9] S. Nie, F. Lou, and F. Yin, ‘Tribological performance of CF-PEEK sliding against 17-4PH stainless steel with various cermet coatings for water hydraulic piston pump application’, Coatings, vol. 9, no. 7, p. 436, 2019, https://doi.org/10.3390/coatings9070436.

[10] D. Li et al., ‘Difference in tribological characteristics between CFRPEEK and stainless steel under water lubrication in friction testing machine and axial piston pump’, *Lubricants*, vol. 11, no. 4, p. 158, 2023, https://doi.org/10.3390/lubricants11040158.

[11] F. Schlegel et al., ‘Swelling, wear and property changes of high-performance polymers in oil-hydraulic tribological contacts’, TuS, vol. 72, no. 1, 2025, https://doi.org/10.24053/TuS-2025-0002.

[12] F. Schlegel, A. Merkel, K. Schmitz, ‘Experimental and simulative Investigation of a partially hydrostatic relived contact in variable speed axial piston machines’, TuS, vol. 70. No. 4+5, 2025, https://doi.org/10.24053/Tus-2023-0023

Biographies



Felix Schlegel received his bachelor's degree in 2020 and his masters degree in 2022 from RWTH Aachen University. For his master's thesis on slippers in high-speed axial piston pumps, he was awarded by the German Society for Tribology. Since 2023, he has been a research associate at the Institute for Fluid Power Drives and Systems (ifas), focusing on tribology of piston pums, and assumed leadership of the tribology group in May 2025. His current doctoral research investigates polymer slippers using several experimental test rigs and electrohydrodynamic simulations.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Modeling and Simulation of Oil Film Influence on Hydraulic Piston Pump Vibrations

Avae Berger, Jiming Ma, Xiangxiang Tian, Yunrui Zhang

avae.berger@centralepekin.cn, jiming.ma@buaa.edu.cn, sy2424140@buaa.edu.cn, violetmoon@buaa.edu.cn

Abstract.

This paper presents a comprehensive numerical and analytical study of the oil film behaviour within the distribution plate of a hydraulic piston pump, focusing on its critical influence on vibration characteristics. A dedicated finite difference solver was developed to solve the Reynolds equation under realistic operating conditions. For the baseline case (μ=7.7×10^{-4} Pa·s, ω=4000 rpm, p_p=21.5 MPa, γ=14°), the simulation yields a mean pressure (p_{mean}) of 4.72 MPa and a stable mean oil film thickness (h_{mean}) of 5.70×10^{-4} m. These results were used to estimate the equivalent stiffness and damping coefficients of the lubricating film through analytical relations derived from squeeze-film theory. After calibration, the model yields an equivalent stiffness of 1.97×10^{7} N/m and a damping coefficient of 3.21×10^{-3} N·s/m. These coefficients were integrated into a single-degree-of-freedom vibration model to evaluate the natural frequency and damping ratio of the pump's distribution plate-piston assembly. The system exhibits a high natural frequency of approximately 5.77 kHz and a nearly zero damping ratio, indicating strong elastic support but severely limited viscous attenuation. The study conclusively demonstrates that while the oil film provides significant stiffness, its intrinsic damping capacity is insufficient to suppress oscillations, underscoring the necessity of incorporating additional structural and fluid-borne damping effects into the design of stable, low-vibration hydraulic pumps.

Keywords. Oil Film, Piston Pump, Dynamic Coefficients, Vibrations.

1. INTRODUCTION

1.1. Context

Rotating machines are at the heart of modern industry and energy systems, where operational reliability and efficiency often depend on the control of friction, wear, and vibration phenomena. Within this category, fluid power pumps and motors are essential components in aerospace, automotive, and heavy machinery applications, ensuring power transmission through pressurized fluids [1,2]. Among these, hydraulic piston pumps are particularly valued for their high-pressure capability. Compared with gear or vane pumps, axial piston pumps offer superior performance and are therefore widely adopted in demanding environments [1,2]. Their reliability is crucial because unexpected vibrations can lead to wear, energy losses, and catastrophic failure[3,4,5]. The tribological interface between

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

piston and cylinder, separated by a thin lubricating oil film, plays a central role [6,7,8]. This oil film simultaneously supports loads, damps vibrations, and prevents metal-to-metal contact. Lubrication plays a critical role in mechanical systems. Beyond reducing friction and wear, a proper lubricant film dissipates the heat generated at the contact interface, protects surfaces against corrosion, and prevents seizure during operation. The performance of this film directly influences energy efficiency, vibration stability, and the overall service life of the pump.

The lubrication behavior and vibration response of hydraulic piston pumps have been extensively studied using classical fluid power theory and hydrodynamic lubrication models [1,2,9]. More recent studies have further highlighted the importance of oil-film stiffness and damping in governing the dynamic response and stability of lubricated interfaces, particularly under oscillatory loading conditions [10,11]. These works confirm that lubrication-induced dynamic coefficients play a decisive role in the vibration behaviour of piston pumps and motivate the development of coupled lubrication–vibration modeling approaches.

1.2. State of the Art and Research Gap

Modeling of lubrication phenomena in hydraulic components is commonly based on the Reynolds equation, which describes the pressure distribution within thin fluid films and forms the theoretical foundation of hydrodynamic lubrication analysis. Various numerical strategies, including finite difference and finite element methods, have been developed to solve this equation for complex geometries and transient operating conditions [12]. These studies primarily focus on pressure fields, load-carrying capacity, and film thickness evolution under different pump configurations.

In parallel, extensive research has examined the dynamic and vibration behaviour of axial piston pumps under coupled mechanical-hydraulic excitation. Several investigations have analysed how operating conditions, pressure pulsations, and fluid-structure interactions affect the overall vibration response and noise generation of these machines [3,4,6,13,14,15]. Multibody dynamic formulations and numerical simulations have been applied to predict excitation mechanisms related to the swash-plate motion, piston reciprocation, and hydrodynamic pressure variations within the film [1,2,8]. Beyond pressure and kinematic analyses, lubrication theory also provides a dynamic description of the oil film response under oscillatory loading. Classical lubrication theory provides well-established analytical formulations for oil-film pressure, stiffness, and damping, which have been widely applied to hydraulic piston pumps and lubricated interfaces [2,9,11]. The tribodynamic coupling between the lubricating film and structural vibration has also been studied through elastohydrodynamic lubrication theory. Foundational models have described the elastic and viscous response of the oil film within hydraulic pumps, defining the equivalent stiffness and damping behaviour of the lubricated interfaces [1,2]. More recent analytical and experimental investigations have quantified the oil-film stiffness and damping in line or point contacts, linking film properties to vibration amplitude and frequency response [10,11]. These works confirm that oil-film stiffness and damping play a decisive role in the stability and dynamic response of lubricated systems. Despite these advances, most existing studies treat pressure as the primary simulation output [6,12,14],

while in vibration modeling, the time-varying film thickness governs the dynamic response and stability of the system. Only a few investigations explicitly incorporate lubrication-induced stiffness and damping into vibration analysis frameworks [10,11,13]. Therefore, a simplified yet physically consistent model coupling the Reynolds equation with a vibration analysis module remains necessary to predict the dynamic response of the oil film-piston system under realistic operating conditions.

1.3. Contribution of This Work

In this work, a finite difference solver is developed to compute the oil film thickness in a hydraulic piston pump, including contact correction effects to prevent non-physical film collapse. Based on the resulting hydrodynamic and contact forces, equivalent stiffness and damping coefficients of the lubricating film are estimated using analytical relations derived from classical squeeze-film lubrication theory. These coefficients are then integrated into a simplified single-degree-of-freedom vibration model to analyze the influence of lubrication-induced dynamic properties on the vibration response of the piston–distribution plate system. In this framework, pressure is not treated as a primary result, but rather as an intermediate variable required to update the oil film thickness and associated forces. The contribution of this work lies in providing a physically consistent coupling between lubrication modeling and vibration analysis, enabling a qualitative investigation of how oil-film stiffness and damping affect vibration regimes in hydraulic piston pumps. The proposed approach is intended as a preliminary modeling framework, highlighting key trends and limitations, rather than as a fully validated predictive tool.

2. THEORETICAL MODEL

2.1. Governing Reynolds Equation and Oil Film Thickness

The main computational basis for the distribution of the pressure field of the oil film gap is the Reynolds equation for fluid lubrication. The Reynolds equation is derived from the Navier-Stokes equation and the continuity equation under certain assumptions.

In this work, the form of the Reynolds equation in cylindrical coordinates is adopted:

$$\frac{1}{\mathrm{r}}\frac{\partial}{\partial \mathrm{r}}(\mathrm{rh}^3)+\frac{1}{\mathrm{r}^2}\frac{\partial}{\partial \theta}\left(\mathrm{h}^3\frac{\partial \mathrm{p}}{\partial \theta}\right)=6\mu\left(\mathrm{U}\frac{\partial \mathrm{h}}{\partial \mathrm{x}}+\frac{\partial \mathrm{h}}{\partial \mathrm{t}}\right) \tag{2.1}$$

where p is the hydrodynamic pressure, h is the film thickness, μ the viscosity, and U the sliding velocity. This formulation results from the standard thin-film lubrication assumptions, where the lubricant is considered Newtonian and incompressible, the film thickness h is much smaller than the piston radius, and inertial effects are negligible compared to viscous effects. Only the sliding velocity U between piston and cylinder and the squeeze effect $\partial h/\partial t$ are retained as sources of hydrodynamic pressure generation. A more general form of the Reynolds equation exists, including additional terms, as well as more complex film deformation contributions. While this extended formulation is appropriate for journal bearings under rotation or systems with significant multidirectional flows, it introduces additional unknowns and computational complexity. Since the hydraulic piston pump considered here is primarily subjected to axial translation, without imposed

radial or tangential velocities, these higher-order contributions can be neglected. Therefore, the simplified Reynolds equation is not only consistent with the physical operating conditions of the piston–cylinder system, but also well-suited for numerical solution with the finite difference method. It provides the necessary accuracy to evaluate the pressure distribution and to capture the coupled evolution of oil film thickness and vibration response, while avoiding unnecessary mathematical complexity.

The local oil film thickness between the piston and the cylinder is defined as:

$$\mathrm{h}(\mathrm{r},\theta) = \mathrm{H_c} - \mathrm{icos}(\theta)\tan(\alpha) - \mathrm{rsin}(\theta)\tan(\beta) \qquad (2.2)$$

where h is the instantaneous film thickness, H_c is the initial oil film thickness, and α and β are the tilt angles of the distribution plate with respect to the x and y axes, respectively.

When $\mathrm{h}(\mathrm{r},\theta)$ locally approaches a critical threshold h_{crit}, a corrective contact force is introduced to prevent non-physical negative film thickness:

$$\mathrm{h}(\mathrm{r},\theta) \geq \mathrm{h_{crit}} \qquad (2.3)$$

with h_{crit} representing contact limit. This mechanism captures the transition from hydrodynamic to mixed lubrication regimes, where partial solid occurs.

2.2. Finite Difference Method

For the numerical solution of the Reynolds equation, the finite difference method (FDM) is adopted because of its simplicity, stability, and suitability for structured cylindrical domains. The lubrication gap between piston and cylinder is naturally expressed in radial and circumferential coordinates, allowing the film to be discretized on a uniform two-dimensional grid. The FDM scheme ensures both accuracy and numerical stability through central-difference approximations of the pressure gradients and a consistent treatment of the film thickness derivatives. The discretized Reynolds Equation forms a sparse linear system of the form:

$$\mathrm{A} \times \mathrm{p} = \mathrm{b} \qquad (2.4)$$

where the matrix **A** depends on local film geometry $\mathrm{h}(\mathrm{r},\theta)$ and lubricant viscosity μ. An iterative relaxation procedure is applied at each time step until the residual between successive pressure fields falls below a predefined convergence tolerance (typically 10^{-6}). The choice of FDM is therefore motivated by its compatibility with the cylindrical geometry of the hydraulic piston pump, the reduced computational cost for time-dependent simulations, and its ability to capture pressure gradients and oil film thickness variations with sufficient accuracy for vibration analysis. It offers sufficient resolution to capture local pressure gradients and film thickness variations that directly affect stiffness, damping, and vibration response.

2.3. Dynamic Coefficients: Analytical Estimation of Stiffness K and Damping C

In a hydraulic piston pump, the oil film separating the rotating and stationary components behaves as a dynamic elastic–viscous layer. Its mechanical behaviour can be idealized as a

spring–damper system, where the stiffness K quantifies the resistance of the film to deformation, and the damping coefficient C represents the viscous energy dissipation during oscillatory motion. These coefficients govern the transmission of forces and the stability of the lubrication interface. In this work, these coefficients are estimated analytically from the mean pressure and film thickness obtained from the Reynolds solver. Assuming a circular equivalent contact region of effective radius R_{eff}, the hydrodynamic stiffness and damping coefficients are derived from the following closed-form relations:

$$K = \frac{3\pi R_{eff}^4 p_{mean}}{2h_{mean}^3} \qquad (2.5)$$

$$C = \frac{3\pi R_{eff}^4 \mu}{2h_{mean}^3} \qquad (2.6)$$

Where p_{mean} is the mean pressure, h_{mean} is the oil film mean thickness, and R_{eff} represents the effective film radius between the outer and inner sealing bands. These equations are derived from the analytical solution of squeeze-film theory, assuming laminar flow, negligible inertia, and small harmonic perturbations of the film thickness. The analytical expressions for stiffness and damping are derived from classical squeeze-film lubrication theory [9]. They reveal the cubic inverse dependence of both stiffness and damping on the film thickness, a fundamental characteristic of hydrodynamic lubrication behaviour. However, these idealized relations tend to over-predict real stiffness and damping, since they assume a fully loaded bearing area, uniform pressure over the contact surface, and negligible compliance of the structural components. It is well documented that such idealized formulations tend to overestimate effective dynamic coefficients due to assumptions of uniform pressure distribution and rigid surfaces [2]. A calibration factor of 10^{-4} is applied to align the theoretical predictions with experimental values reported in the literature for hydraulic piston pumps [4]. This scaling factor aligns the analytical predictions with experimental and numerical results reported in the literature, where typical oil-film stiffness and damping values for piston pumps range approximately from 10^6 to 10^8 N/m and 10^{-3} to 10^{-1} N·s/m respectively [10,11]. The analytical formulation offers a simple yet physically meaningful way to capture the viscoelastic behaviour of the oil film, and investigating how the operating parameters —viscosity, pressure, speed, and inclination angle — affect the equivalent stiffness and damping coefficients.

2.4. Vibration Model of the Oil Film System

Once the equivalent stiffness K and damping coefficient C of the oil film have been obtained from the analytical model (Section 2.3), a simplified single-degree-of-freedom (SDOF) dynamic system was constructed to predict the vibration behaviour of the distribution plate–piston assembly. The motion of the equivalent mass M is governed by the following differential equation:

$$M\ddot{x}(t) + C\dot{x}(t) + Kx(t) = F(t) \qquad (2.7)$$

where $x(t)$ represents the instantaneous displacement of the distribution plate, and $F(t)$ denotes an external excitation or imbalance force acting on the system. For a harmonic excitation $F(t) = F_0 \sin(\omega t)$, the steady-state vibration amplitude is obtained analytically as:

$$|X(\omega)| = \frac{1}{\sqrt{(K-M\omega^2)^2+(C\omega)^2}} \quad (2.8)$$

This expression defines the frequency response function (FRF) of the lubricated interface, where the amplitude $|X(\omega)|$ represents the steady-state displacement magnitude as a function of excitation angular frequency ω. The resonance peak occurs near the natural angular frequency, indicating the dynamic stiffness of the system. For free vibration, $F(t) = 0$, the displacement response is expressed as:

$$x(t) = X_0 e^{-\zeta\omega_n t}\cos\omega_d t \quad (2.9)$$

where the damped natural angular frequency is given by:

$$\omega_d = \omega_n\sqrt{1-\zeta^2} \quad (2.10)$$

and the parameters are:

$$\omega_n = \sqrt{\frac{K}{M}} \quad (2.11)$$

$$f_n = \frac{\omega_n}{2\pi} \quad (2.12)$$

$$\zeta = \frac{C}{2\sqrt{KM}} \quad (2.13)$$

Here, ω_n and f_n are the undamped natural angular and cyclic frequencies, respectively, and ζ is the dimensionless damping ratio. Since the damping coefficient C is relatively small, the system exhibits a nearly undamped oscillatory motion, characteristic of high stiffness and low energy dissipation in the lubricating film. The analytical vibration model provides direct access to two key quantities — the natural frequency f_n and the damping ratio ζ — which are essential indicators of the system's stability.

3. NUMERICAL SOLVER AND SIMULATION RESULTS

3.1. Computational Setup

To numerically evaluate the oil film behaviour in the distribution plate of a hydraulic piston pump, a dedicated solver is developed. This solver integrates the Reynolds Equation with oil film thickness correction and pressure boundary conditions under various operating parameters. Table 3.1. presents the list of parameters and variables used in the simulation:

Table 3.1. Parameters and Variables Description

Category	*Parameter Description*	*Symbol*	*Unit*
Geometry	Radial coordinate	*r*	m
	Circumferential coordinate	*θ*	rad

	Outer and Inner sealing band radius	R_1, R_2	m
	Outer and Inner inner-seal radius	R_3, R_4	m
	Piston and piston distribution radius	R_p, R_s	m
	Lenght Distribution plate	L_{dis}	m
	Number of pistons	z	-
	Angular pitch	$a = 2\pi/z$	rad
	Orifice opening angle	a_0	rad
Geometry Inclination	Initial central film	H_c	m
	Swash-plate inclination angle	γ	rad
Fluid & Contact	Dynamic viscosity	μ	Pa.s
	Density	ρ	$kg \cdot m^{-3}$
	Ambient pressure, suction and discharge pressure	P_{oil}, p_x, p_p	Pa
	Critical film thickness	h_{crit}	m
Kinematics	Rotational speed	ω	$rad \cdot s^{-1}$
	Piston force	F_{press}	N
Numerical Solver	Number of circumferential and radial nodes	N_θ, N_r	
	Mesh spacing (circumferential and radial)	$\Delta\theta$, Δr	rad, m
	Pressure solver tolerance	ε_p	
Analytical Dynamic Coefficients	Mean pressure and oil film thickness	p_{mean}, h_{mean}	Pa, m
	Effective equivalent radius	R_{eff}	m
	Analytical stiffness	K	N/m
	Analytical damping	C	N·s/m
Vibration Model	Moving mass	M	kg
	Natural angular frequency	ω_n	$rad \cdot s^{-1}$
	Natural frequency	f_n	Hz
	Damping ratio	ζ	
	External excitation	F(t)	N
Outputs	Pressure field	$p(r,\theta)$	Pa
	Oil film thickness field	$h(r,\theta)$	m
	Oil-film force and moment	F_{oil}, M_{oil}	N, N·m
	Contact force	F_c	N
	Stiffness and Damping	K, C	N/m, N·s/m
	Vibration response	X(t)	m

A schematic showing the computational process from initialization to convergence is also depicted in Figure 3.1:

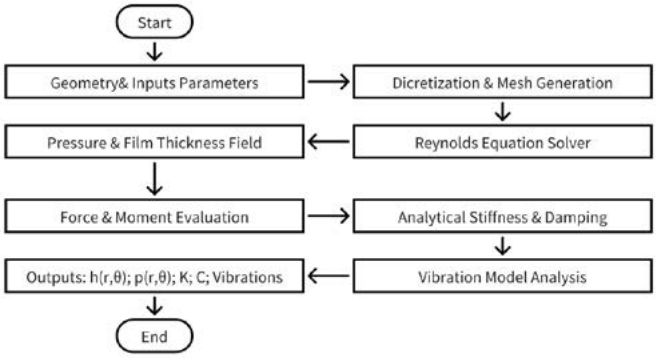


Figure 3.1. Flow Chart of the Numerical Model

The coupled Reynolds and elasticity equations were solved using FDM on a structured mesh with $N_r = 100$ radial and $N_\theta = 360$ circumferential nodes.

Convergence was achieved when both the total force and moment residuals were below the defined tolerance.

3.2. Oil Film Thickness and Pressure Distribution

This subsection analyzes the hydrodynamic field obtained from the Reynolds solver, focusing on the oil film thickness and pressure distribution across the sealing bands.

Figure 3.2 and Figure 3.3 illustrate the reference case corresponding to a discharge pressure of $p_p = 21.5$ MPa, rotation speed $\omega = 4000$ rpm, inclination angle $\gamma = 14°$, and oil viscosity $\mu = 7.7\times10^{-4}$ Pa·s. Under these conditions, the maximum hydrodynamic pressure reaches 21.5 MPa near the high-pressure crescent region, while the minimum pressure stabilizes at 0.2 MPa at the suction side. The pressure field exhibits a typical hydrodynamic wedge pattern, with a gradual pressure rise along the converging region and a sharp drop across the cavitation boundary. The corresponding oil film thickness varies between $h_{min} = 5.66\times10^{-4}$m and $h_{max} = 5.74\times10^{-4}$ m, with a mean value $h_{mean} = 5.70\times10^{-4}$ m, confirming that the interface operates entirely within the full-film lubrication regime.

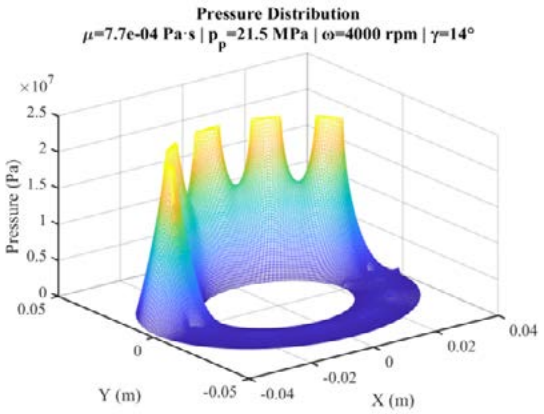


Figure 3.2. Pressure Distribution 1

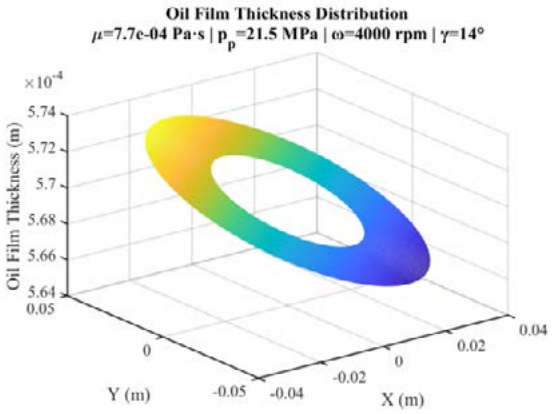


Figure 3.3. Oil Film Thickness Distribution 1

To assess the impact of operating parameters, additional simulations were simulated by varying discharge pressure, viscosity, rotational speed, and swash-plate inclination angle. Results are shown in Figures 3.4, 3.5, 3.6 and 3.7.

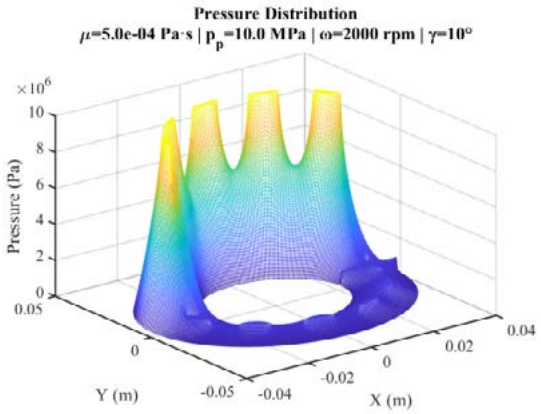


Figure 3.4. Pressure Distribution 2

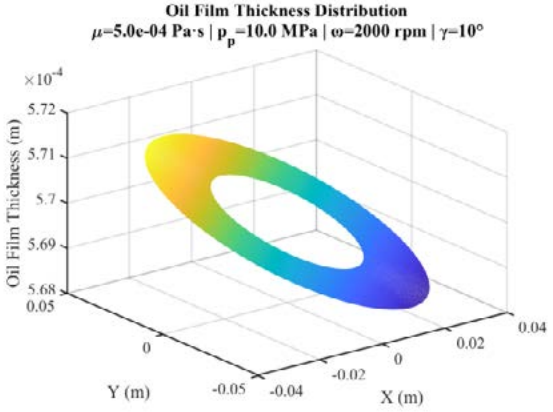


Figure 3.5. Oil Film Thickness Distribution 2

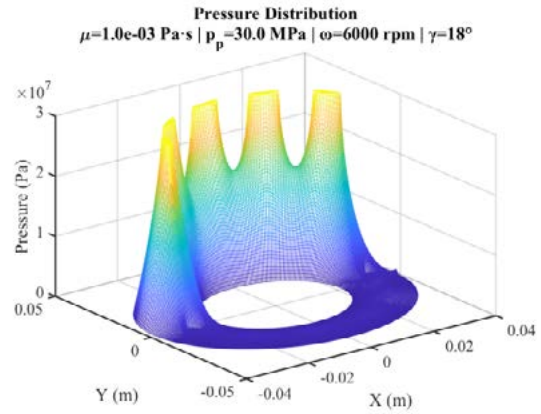


Figure 3.6. Pressure Distribution 3

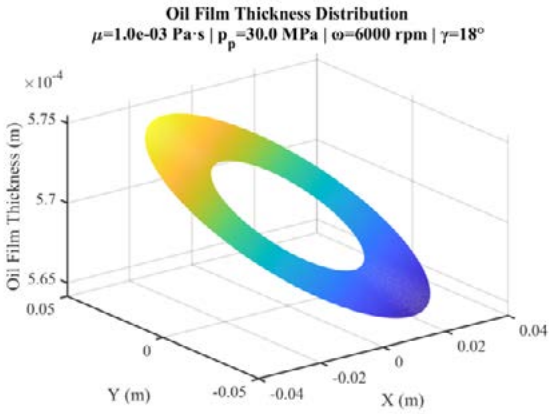


Figure 3.7. Oil Film Thickness Distribution 3

Table 3.2. Key Results for the Oil Film Thickness and the Pressure Distributions

Simulation	μ (Pa•s)	p_p (MPa)	ω (rpm)	γ (∘)	p_{mean} (MPa)	h_{min} (m)	h_{max} (m)
1	7.7×10^{-4}	21.5	4000	14	4.72	5.66×10^{-4}	5.74×10^{-4}
2	5.0×10^{-4}	10	2000	10	2.39	5.68×10^{-4}	5.71×10^{-4}
3	1.0×10^{-3}	30	6000	18	6.44	5.64×10^{-4}	5.75×10^{-4}

Figures 3.4, 3.5, 3.6, 3.7 and Table 3.2 show that the mean pressure p_{mean} increases proportionally with the discharge pressure p_p, validating the hydraulic consistency of the model. The mean film thickness h_{mean} remains constant across all simulations, $h_{mean} \approx 5.70 \times 10^{-4}$. These results indicate that, within the limited range of operating conditions investigated, the mean film thickness appears to be primarily influenced by geometric configuration and contact correction mechanisms.

From a physical standpoint, this weak dependence of h_{mean} on operating conditions can be explained by several factors. First, the contact correction mechanism imposes a minimum admissible gap h_{crit} that prevents non-physical interpenetration; the solver thus converges toward an equilibrium thickness constrained by this limit. Second, the inclination angle γ and initial central film height H_c geometrically define a baseline profile that dominates the final solution. Third, the model operates in a quasi-static regime, where transient squeeze effects are negligible compared with steady hydrodynamic pressure generation.

As a result, while the local pressure distribution adapts sensitively to load and viscosity, the overall mean film thickness remains locked near its geometric equilibrium value.

Overall, the film thickness variation across all cases confirms that the solver maintains a stable full-film lubrication regime, ensuring the reliability of subsequent stiffness and damping coefficient estimations.

3.3. Extraction of Dynamic Coefficients: Stiffness K and Damping C

This subsection evaluates the dynamic characteristics of the lubricating oil film by estimating its equivalent stiffness and damping coefficients through both analytical and force-based approaches. These coefficients characterize the elastic and viscous behaviour of

the film under small perturbations, directly linking lubrication dynamics to the vibration response of the piston–plate interface.

The force-based approach computes coefficients directly from the pressure induced force amplitude and film thickness variation.

Figures 3.8 and 3.9 illustrate the variation of the hydrodynamic stiffness K and damping coefficient C as functions of the mean film thickness h_{mean}. The analytical relations (blue curves) show the characteristic cubic inverse dependency (K, $C \propto h^{-3}$) [9], while the calibrated results (red curves) and force-based estimations (magenta markers) reveal the limitation of purely theoretical models.

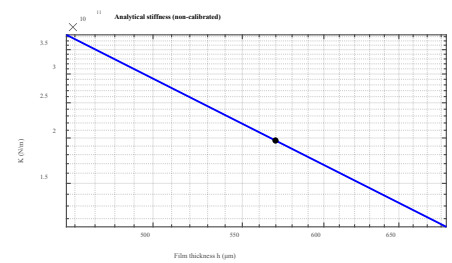


Figure 3.8. Hydrodynamic Stiffness K

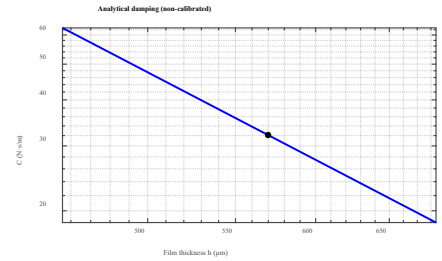


Figure 3.9. Damping coefficient C

Physically, this difference arises from several factors. First, the analytical expressions assume a fully developed hydrodynamic film with uniform pressure over the entire contact area, while the force-based estimation inherently accounts for non-uniform pressure distribution and localized deformation. Then, the contact correction mechanism used in the numerical solver restricts the minimum admissible film thickness, effectively increasing the apparent stiffness of the system. Additionally, the pressure integration in the force-based method tends to overestimate the effective load-bearing area when expressed as **A×p=b**, leading to higher values of K and C.

Table 3.3. Dynamic coefficients extracted for the three simulated operating conditions

Simulation	K_{cal} (N/m)	C_{cal} (N·s/m)	K_{force} (N/m)	C_{force} (N·s/m)
1	1.97×10^{7}	3.21×10^{-3}	4.36×10^{9}	1.13×10^{1}
2	9.97×10^{6}	2.09×10^{-3}	4.22×10^{9}	1.10×10^{1}
3	2.69×10^{7}	4.17×10^{-3}	4.43×10^{9}	1.14×10^{1}

As shown in Table 3.3, the analytical stiffness K_{cal} increases proportionally with discharge pressure, from 9.97×10^6 at 10 MPa to 2.69×10^7 at 30 MPa, following the hydrodynamic load capacity trend. The force-based method yields stiffness values approximately 200 to 400 times higher than the calibrated analytical results, highlighting its sensitivity to local pressure fluctuations and the strong non-linearity of the hydrodynamic field.

The damping coefficients show better consistency, with force-based values clustering around 11 N·s/m across all operating conditions, while analytical values range from 2.09×10^{-3}N/m to 4.17×10^{-3}N/m.

3.4. Vibration Analysis

The dynamic response of the piston-distribution plate system was evaluated using the equivalent stiffness and damping coefficients.

The film behaviour is modeled as a single degree-of-freedom oscillator with equivalent mass M_{eq}=0.0116 kg.

Table 3.4. Vibration parameters for the three different simulations

Simulation	f_n (kHz) *analytical*	f_n (kHz) *force-based*	ζ *analytical*	ζ *force-based*
1	6.56	97.5	0	0.001
2	4.67	96.0	0	0.001
3	7.66	98.4	0	0.001

The Table 3.4. demonstrate physically consistent behaviour: natural frequencies increase with operating pressure ranging from 4.67 kHz to 7.66 kHz. The consistent near-zero damping ratios ($\zeta \approx 0$). across all conditions confirm that the oil film's limited damping capacity is an inherent characteristic rather than an operating-condition dependent phenomenon.

The calibrated analytical approach exhibits distinct resonance peaks that shift with operating parameters as depicted in Figure 3.10.

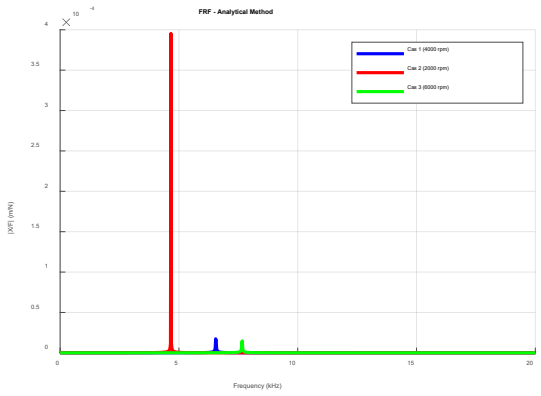


Figure 3.10. Frequency Response Function

The force-based parameters yield fundamentally different behaviour, with no discernible resonance peaks within the 0-20 kHz range. This flat response profile indicates that the excessively high stiffness values dominate the system dynamics, effectively suppressing vibrational motion.

The Free vibration analysis is depicted in Figure 3.11, and reveals critically different damping characteristics between the two methods.

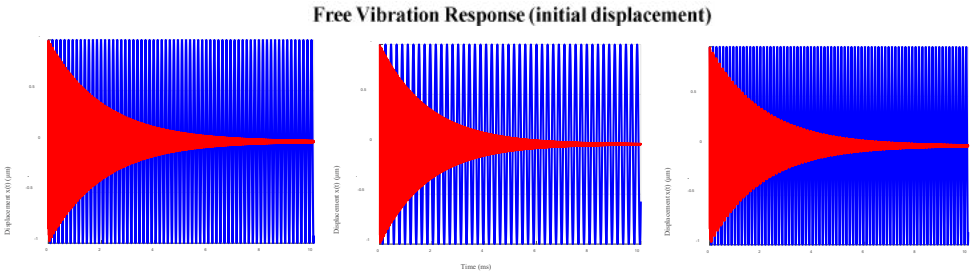


Figure 3.11. Frequency Response Function case 1, 2 and 3 respectively

The analytical solution shows persistent oscillations with negligible amplitude decay over 10 ms, consistent with the near-zero damping ratios calculated for the three cases. The force-base parameters produce immediate decay without oscillations. This quick dissipation contradicts practical observations of piston pump vibrations and underscores the physical inconsistency of the force-based stiffness estimation.

3.5. Physical Interpretation

The dramatic divergence between the two estimation methods provides critical insights for pump design. First, the oil film stiffness dominates dynamics. The analytical results confirm that the oil film provides substantial elastic support, with natural frequencies from 4.67 kHz to 7.66 kHz within typical ranges for piston pump vibrations. This stiffness is essential for maintaining structural integrity but necessitates complementary damping mechanisms.

Then, the persistent oscillations in free vibration response and sharp resonance peaks in forced response demonstrate that the oil film's viscous dissipation is insufficient for vibration control. The calculated damping ratios ($\zeta \leq 0.001$) are orders of magnitude below the critical damping threshold ($\zeta = 1$).

Finally, force-Based Method Limitations: While the force-based damping coefficients ($C \approx 11\text{N}\cdot\text{s/m}$) fall within plausible ranges.

The vibration analysis conclusively demonstrates that while the oil film provides excellent elastic support for the piston-distribution plate interface, its intrinsic damping capacity is fundamentally inadequate for vibration suppression. This finding emphasizes the necessity of integrated damping strategies in the design of high-performance, low-vibration hydraulic piston pumps.

In practical hydraulic piston pumps, measured vibration spectra typically exhibit distinct resonance peaks and sustained oscillations, rather than the flat response predicted by excessively high stiffness values [2,5,7].

4. CONCLUSION

This research successfully developed and validated a integrated model that couples fluid dynamics with structural vibration to analyse the impact of the lubricating oil film on the dynamic behaviour of a hydraulic piston pump. The key findings are summarized as follows:

1) Successful Quantification of Oil Film Dynamics: The Reynolds equation, solved via the FDM, effectively predicted the oil film pressure and thickness distributions under varying operational parameters. The results indicated that the mean film thickness remained stable across simulations, governed primarily by geometric configuration and the contact correction mechanism, confirming operation within a stable full-film lubrication regime

2) Comparative Analysis of Stiffness and Damping: The dynamic coefficients were extracted using both analytical and force-based approaches. The calibrated analytical method provided physically consistent stiffness values, which increased proportionally with discharge pressure. In contrast, the force-based method yielded stiffness estimates orders of magnitude higher, leading to unrealistic vibration response predictions, as shown by the absence of resonance peaks in the frequency response function.

3) The Dual Role of the Oil Film in Vibrations: The vibration analysis clearly delineates the dual role of the oil film. It provides substantial elastic support, resulting in high natural frequencies (4.67 - 7.66 kHz) essential for structural integrity. However, its damping capacity is negligible (damping ratio $\zeta \approx 0$), as evidenced by persistent oscillations in free vibration response and sharp resonance peaks in forced response, proving its inadequacy for vibration suppression.

This study could be further pursued by implementing a more sophisticated time-domain integration scheme, such as the Newmark method, for the vibration analysis. This would allow for the investigation of transient dynamic responses and non-linear phenomena under complex loading conditions. Furthermore, experimental validation of the simulated dynamic coefficients and vibration response is crucial for further model refinement.

The present study does not provide experimental validation. The results are intended to highlight qualitative trends and physical consistency with established lubrication and vibration literature. Experimental validation and advanced coupled fluid–structure models will be addressed in future work.

5. REFERENCES

[1] M. Ivantysynova, 'Hydrostatic Pumps and Motors: Principles, Design, Performance, Modelling, Analysis, Control and Testing', Fluid Power Net Publications, 2001.

[2] N. D. Manring, 'Fluid Power Pumps and Motors: Analysis, Design, and Control', McGraw-Hill Education, 2013.

[3] S. Kim, J. Lee, 'A study on the dynamics and vibration behaviour of an axial piston pump', Procedia Engineering, vol. 144, pp. 1105-1112, 2016. doi: 10.1016/j.proeng.2016.05.155.

[4] T. Zhang, P. Liu, 'Vibration model and characteristic analysis of hydraulic piston pump', Chemical Engineering Transactions, vol. 59, pp. 61-66, 2017. doi: 10.3303/CET1759011.

[5] W. Wang, 'Vibration model and characteristic analysis of hydraulic piston pump', Chem Eng Trans, vol. 59, pp. 61-66, 2017. doi: 10.3303/CET1759011.

[6] S. Ye, J. Zhang, X. Wang, H. Tang, Y. Zhang, 'A theoretical dynamic model to study the vibration response characteristics of an axial piston pump', Mechanical Systems and Signal Processing, vol. 150, 107237, 2021. doi: 10.1016/j.ymssp.2020.107237.

[7] S. Ye, J. Zhang, B. Xu, S. Zhu, J. Xiang, and H. Tang, 'Theoretical investigation of the contributions of the excitation forces to the vibration of an axial piston pump,' Mech Syst Signal Process, vol. 129, pp. 201-217, Aug. 2019. doi: 10.1016/j.ymssp.2019.04.032.

[8] Y. Sun, F. Gao, 'Dynamic modeling and vibration characteristics of multibody system in axial piston pump', Mechanical Systems and Signal Processing, vol. 162, 108034, 2022. doi: 10.1016/j.aej.2022.07.046.

[9] B.J. Hamrock, S.R. Schmind, and B.O. Jacobson, Fundamentals of fluid film lubrication. Marcel Dekker, 2004.

[10] X. Zhou, J. Zhang, 'Stiffness and damping models for the oil film in line contact elastohydrodynamic lubrication and applications in the gear drive', Applied Mathematical Modelling, vol. 62, pp. 695-708, 2018. doi: 10.1016/j.apm.2018.05.012.

[11] Y. Zhang, D. Zhu, 'Oil film stiffness and damping in an elastohydrodynamic lubrication line contact–vibration', Tribology International, vol. 95, pp. 121-130, 2016. doi: 10.1007/s12206-016-0611-x.

[12] Y. Zhang, X. Wang, 'Analysis of lubricating characteristics of valve plate pair of a piston pump', Tribology International, vol. 123, pp. 234-243, 2018. doi: 10.1016/j.triboint.2018.05.008.

[13] X. Li, Y. Wang, 'Vibration analysis of a centrifugal pump with healthy and defective impellers and fault detection using multi-layer perceptron', Mechanical Systems and Signal Processing, vol. 154, 107594, 2021. doi: 10.3390/eng5040131.

[14] H. Li, J. Zhang, 'Analysis of pressure pulsation mechanism and dynamic characteristics in axial piston pumps', Journal of Mechanical Engineering, vol. 55, no. 4, pp. 67-75, 2019. doi: 10.3969/j.issn.1674-8042.2022.01.003.

[15] P. Gao, T. Yu, Y. Zhang, J. Wang, J. Zhai, 'Vibration analysis and control technologies of hydraulic pipeline system in aircraft: A review', Chinese Journal of Aeronautics, vol. 34, no. 4, pp. 83-114, 2021. doi: 10.1016/j.cja.2020.07.007.

AUTHORS



Avae BERGER received the Bachelor's degree in Mechanical Engineering from Beihang University in 2023. She is currently pursuing a Master's degree in Systems Engineering at the Sino-French Engineering School, at Beihang University. Her research interests include hydraulic system dynamics, lubrication modeling, and vibration analysis in mechanical systems. She is dedicated to exploring advanced modeling and simulation techniques to improve the performance and reliability of hydraulic machinery.



Jiming MA, Ph. D, Associate Professor in Sino-French Engineer School (ECPkn, École Centrale de Pékin) of Beihang University(BUAA) (http://ecpkn.buaa.edu.cn/english/). Majored: Fluid transmission and control; Industrial science and technology; Reliability of hydraulic components; Multi-domain modeling.



Xiangxiang TIAN received the bachelor's degree in applied physics from Beijing University of Aeronautics and Astronautics in 2024. He is currently pursuing the master's degree in systems engineering at Beijing University of Aeronautics and Astronautics. His research areas include piston pump moving pair oil film calculation, high-speed piston pump dynamic balance, aircraft hydraulic system simulation, etc.



Yunrui ZHANG obtained a Bachelor's Degree in Applied Physics from Beihang University in 2024, possessing a solid foundation in physical theories and interdisciplinary insights into engineering disciplines. Currently, he is pursuing a Master's Degree at Beihang University, focusing on applied research in the field of engineering. His core research interests include structural fault diagnosis, vibration and noise analysis, among others. He places emphasis on exploring the practical application of relevant technologies in fields such as aviation and machinery equipment, and is committed to providing theoretical support for enhancing equipment reliability and optimizing operational performance.

Grain Microstructural Evolution of Friction Layers in Piston Pump Friction Pairs

Song Yuhan[1]* Angerhausen Julian[1] Schmitz Katharina[2]

[1] *Weichai Power European Science & Technology Innovation Center*

[2] *RWTH Aachen University – Institute for Fluid Power Drives and Systems (ifas)*

**Corresponding author: E-mail address: yuhan.song@weichaidestic.com*

Abstract.

In the frictional contact of piston pump friction pairs, the formation and evolution of the friction layer are accompanied by significant changes in the grain size, orientation, and distribution within the subsurface region. These grain-level structural characteristics critically influence the friction and wear behaviour of the contact interface. To clarify this process, cross-sectional EBSD (Electron Backscatter Diffraction) is carried out on specimens of port plate material after reciprocating sliding for different times under the same conditions. Changes in grain size, grain boundary misorientation angles, and crystal orientation reveal the rapid formation process of the friction layer. During the first hour of testing, the friction layer changes significantly. Grain size decreases, grain orientation rotation appears, and grain boundaries become pronounced with low-angle grain boundary (LAGB) to dominate. As the sliding continued, the layer thickness kept increasing until it stabilized after 2 h. From 2 h to 16 h, a stable hierarchical structure is established, consisting of an ultrafine-grained surface with persistent {111} alignment, an LAGB-rich transition zone, and the unaffected base material. These findings demonstrate that friction causes time-dependent stratification of microstructure and that the coupled evolution of refinement, boundary character, and orientation selection underpins the stability of the tribolayer, providing a quantitative basis for gradient design of wear-resistant hydraulic components.

Keywords. piston pump, friction layer, grain microstructure, EBSD.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

As a key element of hydraulic systems, axial piston pumps directly influence overall efficiency and reliability through their operational stability and durability. Under long-term high pressure and high-speed sliding, the friction surfaces gradually undergo wear , resulting in increased friction, lubrication deterioration, and leakage. This will reduce efficiency and durability, as well as cause damage to components. Therefore, an in-depth study of the wear mechanisms of key friction parts of piston pumps and their microstructural evolution is of great significance for extending service life and optimising lubrication design. Extensive research has been devoted to improving the tribological behaviour of piston pump components through surface texturing [1], operating-condition control [2], structural optimisation [3], material selection [4], coating design [5].

However, even with such optimisations, friction pairs inevitably experience a run-in stage. During this process, engineered systems transition from the initial state to a stable operating state by removing microscopic asperities via wear or plastic deformation, thereby improving surface conformity and friction stability [6]. R. Ivantysyn, et al analysed the running-in period of axial piston pumps, focusing on clarifying the effects of operating parameters and contact geometry on surface behaviour [7]. Tappeiner, et al. studied the optimisation of the surface roughness of the friction part of piston pumps, emphasising low friction and low wear achieved by matching the appropriate running-in mechanism with roughness parameters [8]. W. Jiao investigated the effects of initial surface roughness, speed and pressure on the tribological properties of axial piston pump port plate pairs under low-pressure conditions with respect to the run-in effect. The results show that higher speed and initial roughness shorten the run-in time, while higher pressure leads to an increase in the fluctuation amplitude of the friction coefficient. The wear mechanism is mainly characterised by abrasive wear [9]. A. Paulus studied the evolution of the friction layer when lead-free copper was used for the piston pair and found that under lubricated sliding contact conditions, both bronze and brass could form a stable boundary lubrication layer, and the formation mechanism was mainly attributed to the friction chemical reaction of tin oxide (SnO/SnO_2) and zinc sulphide (ZnS), which effectively reduced the coefficient of friction and improved the wear resistance [10]. At the contact interface, structural changes in the material can cover a depth range from a few nanometres (e.g., amorphization) to a few microns (e.g., phase transitions, recrystallization), which can lead to significant differences in the properties of the surface material compared to the substrate, including strength, corrosion resistance, lubrication properties, friction behaviour, and wear resistance [11].

Existing studies mainly focus on the macroscopic wear behaviour of the piston pump friction pair during the run-in stage, while there is a lack of systematic research on the microstructure evolution of the material during the run-in process and its influence on the tribological properties, as shown in Figure **1.1**. Especially, the evolution law of the microstructure of the piston pump friction pair still needs to be explored.

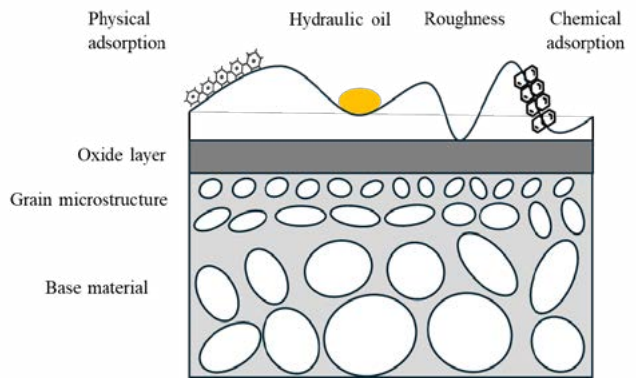


Figure 1.1. Schematic diagram of the microstructure of the friction layer

This study aims to systematically investigate the evolution of microstructure of the friction pairs from the running-in period to the steady state by using microanalysis techniques EBSD, and to analyse the characteristics of the friction layer's thickness, grain refinement, and orientation change in the process of friction, as well as their contribution to the wear resistance, to provide scientific basis for the optimization of the mating piston pump friction pair design.

2. MATERIAL AND METHODS

This section outlines the sample preparation, friction testing, and procedure for analysing the characterisation of friction layers.

2.1. Sample Preparation

Sample Preparation for friction test: the test specimens are the CuZn28Al4Ni3Si alloy, a widely used material for axial piston pump port plates due to its excellent wear resistance and mechanical properties. The specimens were machined into 24 mm × 6.9 mm discs, and surface roughness was controlled to Ra 0.4 μm, ensuring the same as the port plate surface. Before testing, all specimens were cleaned according to DIN 51834 standards, using ethanol and ultrasonic waves for cleaning, followed by drying with compressed air to remove residual contaminants. The other material in the friction pair consists of a QT600 steel pin with a diameter of 10 mm, which was manufactured and heat-treated in the same process as the cylinder block and installed as the upper test specimen.

2.2. Methods

The friction testing is performed on a sliding tribometer (SRV®5) in accordance with DIN 51834 standards. The test configuration consisted of a Pin-on-disc contact under lubricated conditions. The tribological experiments were conducted under the following conditions: frequency: 10 Hz、stroke length: 2 mm, normal load: 400 N, temperature: 50°C (preheating for 5 minutes before testing to ensure thermal stability), lubrication: mineral-based hydraulic oil (ISO VG 46 HLP46), refreshed before each test. Operating conditions

remain constant, only time is varying. Six durations were set: 0.5 h, 1 h, 2 h, 4 h, 8, and 16 h, with three independent specimens per duration (N = 18).

To investigate microstructural evolution and crystallographic orientation, micrographs and orientation maps were obtained using a field-emission SEM (Zeiss Gemini SEM 500) equipped with an EBSD detector (Oxford Instruments Symmetry2). EBSD mapping was performed at an accelerating voltage of 20 kV, with a 120 μm aperture and a step size of 50 nm. After acquisition, the EBSD datasets were indexed and cleaned in Oxford Instruments, and subsequently processed to extract grain size statistics, grain boundary distributions, and grain orientation (inverse pole figure, IPF) by software.

Phase analysis was conducted using X-ray diffraction (XRD) with a Bruker D8 Advance diffractometer, equipped with a Cu-Kα radiation source (λ = 1.5406 Å). measurement conditions, scanning range: 5°–90° (2θ), step size: 0.02°, scan speed: 1°/min. The main purpose is to obtain an unknown phase of the material, which is passed to the EBSD test for phase determination.

2.3. EBSD Sample Preparation

Obtaining a complete subsurface grain-structure is challenging; sectioning and polishing must minimise the removal of tribolayer. Edge rounding caused by cutting, inlaying, or polishing can remove part of the tribolayer, cutting off the layer and making the actual thickness no longer measurable. To preserve the as-formed structure, cross-sections were prepared as follows: The wear-track surface was first sputter-coated with a thin Pt film to protect against local tearing during sectioning. At the centre of the friction zone, vertical to the sliding direction, a low-deformation cutting process is employed. To avoid edge rounding and prevent subsequent disassembly, specimens were inserted into hard conductive resin using a hot-dipping process, enabling direct analysis of the cross-section after insertion. The polishing process involved using SiC sandpaper (grit sizes ranging from 800 to 4000) until scratches were eliminated, followed by final chemical-mechanical polishing: colloidal silica (OPS, approximately 0.04 μm SiO_2) was applied at 50 rpm under low load (10 N) for 5 minutes to remove residual deformation layers. After polishing, the wear zone was re-inspected to confirm that the tribolayer remained continuous and was not polished, and the target region was marked with conductive adhesive. The specimen was then cleaned, dried, and examined by EBSD.

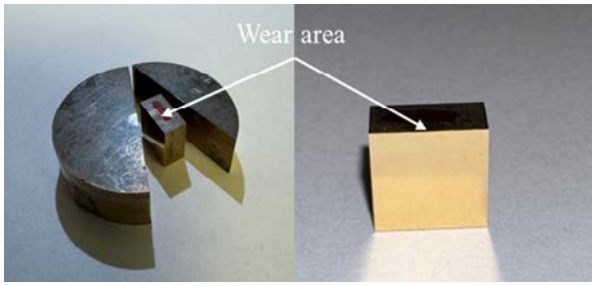


Figure 2.1. Sputter-Coated & Sectioned Specimen and Polished EBSD Cross-Section

3. RESULTS AND DISCUSSION

This section presents research findings spanning from tribological responses to subsurface mechanisms. Section 3.1 analyses friction-time behaviour, wear cross-sections, and the run-in and steady-state phases. Section 3.2, through X-ray diffraction combined with ICDD-PDF (International Centre for Diffraction Data – Powder Diffraction File) validation, analyses the phase composition of the distributor plate and provides critical background information for EBSD analysis. Section 3.3 employs EBSD techniques to investigate the depth-time distribution of equivalent grain size, grain boundary angle, and grain orientation across cross-sections at different time points.

3.1. Friction Behaviour and Surface Topography

In the study of hydraulic pump friction pairs, the coefficient of friction (COF) and wear are key parameters that characterise tribological performance. Both indicators are influenced by two concurrent mechanisms: material detaches from the surface, and on the other hand, the surface layer undergoes microstructural evolution under high stress and friction. These two mechanisms exist together under sliding lubrication conditions and directly influence the friction response and endurance of the friction pair.[12]

The evolution of the coefficient of friction over time is shown in Figure **3.1**. The results show the trend of the friction coefficient over time, presented as mean ± 1 standard deviation. The shaded areas indicate the standard deviation range with three specimens measured at each time point. During the initial 10 minutes (left figure), all groups showed low and quasi-constant friction coefficients. On a long-term scale (right figure, logarithmic time scale), the variability of the friction coefficient becomes more significant. The coefficient of friction increases between 0.5 h and 2 h. Between 2 h and 5 h, it remained at a high level with moderate fluctuations, before reducing again. Finally, after about 6 h, the coefficient of friction stabilized, with the range of the mean ± 1 standard deviation gradually narrowing, indicating that the system was approaching steady-state. This indicates that time dominates the evolution of friction behaviour. Regardless of the total test duration for individual specimens, results obtained at identical absolute time points remain consistent.

The curves for all specimens exhibit typical running-in characteristics: a rapid increase followed by a peak in the early stage, then a gradual decline toward stability. This closely aligns with Blau's summarised friction transition law, wherein the friction pair initially shows a rapid rise in the coefficient of friction before entering a stable wear phase [13].

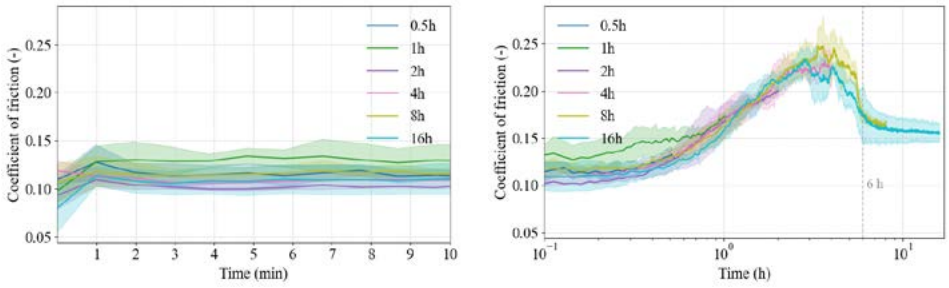


Figure 3.1. Time Evolution of the Coefficient of Friction

Within the first 6 h of measurement, the coefficient of friction changes significantly, indicating that the interface is still undergoing intense adjustment and the surface structure has not yet stabilised. As time progresses, the peak gradually decreases and its decay becomes more gradual, ultimately maintaining a low value and a stable range. This trend can be attributed to friction-induced grain refinement, changes in grain boundary angles and grain orientation [14–16]. The sustained friction process promotes the stabilisation of the surface microstructure, thereby effectively suppressing friction fluctuations.

The cross-sectional profiles of the wear scars (Figure **3.2**) illustrate the temporal evolution of wear depth. At the early stage (0.5 h and 1 h), the wear scars are shallow, and the cross-sections appear relatively smooth, indicating that the material removal is primarily governed by surface plastic deformation accompanied by slight delamination. By 2 h and 4 h, the wear depth increases significantly, and the cross-sections evolve into sharp "V-shaped" grooves. When the sliding time is further extended to 8 h and 16 h, the wear depth reaches its maximum. The good correlation between wear depth and friction coefficient indicates that the system is transitioning from run-in phase to steady-state friction.

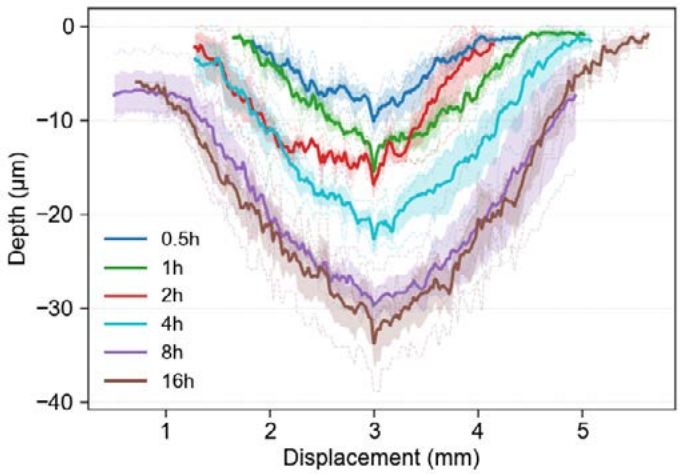


Figure 3.2. Central Groove Profile vs. Time

From a contact mechanics perspective, in the early stages of contact, the interaction is point-to-surface, resulting in extremely high localised stresses that easily lead to strong plastic deformation and ploughing, which increases the friction coefficient. As sliding progresses, the actual contact area increases and becomes more conforming, with the point-to-surface contact gradually transitioning to surface-to-surface contact. The bottom of the wear scar changes from flat to curved, and the contact state tends toward stability [17–19].

However, the above observations primarily reflected friction responses, revealing limited insights into the of the friction layer mechanisms. To uncover this, cross-sectional EBSD was used to construct depth–time maps, tracking grain refinement, orientation selection, and boundary evolution and relating these to friction and wear. This microstructure-to-property linkage clarifies how the tribolayer forms, evolves, and reaches a steady state.

3.2. Phase Identification

The XRD pattern of the port plate material is shown in Figure **3.3**. It displays the XRD diffraction pattern of the end plate material, analysed in conjunction with reference PDF cards for potential constituent phases. The diffraction peaks correspond to three phases: Co_3Ni, CuZn, and $Cu_{0.66}Zn_{0.34}$. Among these, only a faint Co_3Ni diffraction peak was detected, indicating an extremely low content of this phase.

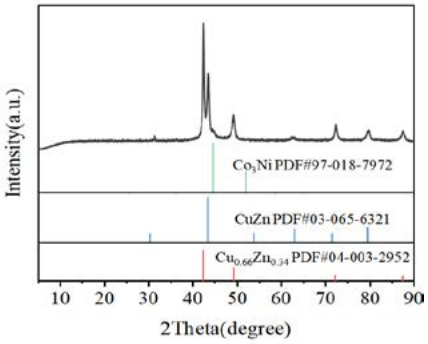


Figure 3.3. The XRD pattern of the port plate material

The results show that this phase belongs to the FM3M space group and has a face-centred cubic (FCC) structure. By analysing the diffraction peaks, the crystal structures of each phase were determined, providing fundamental data for the following EBSD testing. This can help deeply understand the microstructural characteristics of the port plate and their impact on material properties.

3.3. Time-dependent evolution of the microstructure of the friction layer

Figure **3.4** shows the SEM cross-section at 8 h: the interface is generally smooth, with no through-peeling observed, nor significant edge rounding, chamfering, or polishing marks, indicating that the sample preparation process preserved the true form of the friction layer.

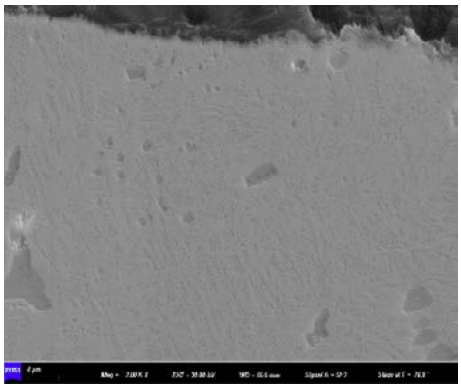
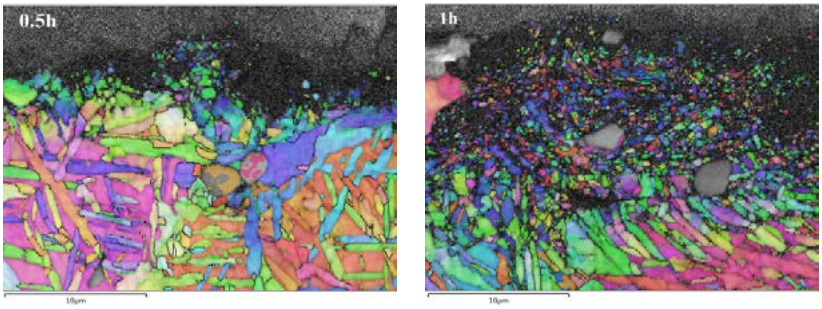


Figure 3.4. Cross-Sectional SEM at 8 h (Wear-Track Centre)

Figure **3.5** shows the EBSD results, with three results for each time point, only one of them is displayed here. All maps are displayed as IPF-ND (normal direction) colourings, with orientations defined by the accompanying inverse pole-figure triangle (red is {001}, green is {110}, blue is {111}). High-angle grain boundaries (HAGB, ⩾15°) are shown in black line; low-angle grain boundaries (LAGB, 2–15°) in green line. Regions affected by severe plastic deformation or delamination, as well as exogenous large particles (third-body debris), were masked and excluded from quantitative statistics. Near-surface nanograins approach the EBSD step-size and pattern-quality limits; values in this range should be regarded as lower-bound estimates.



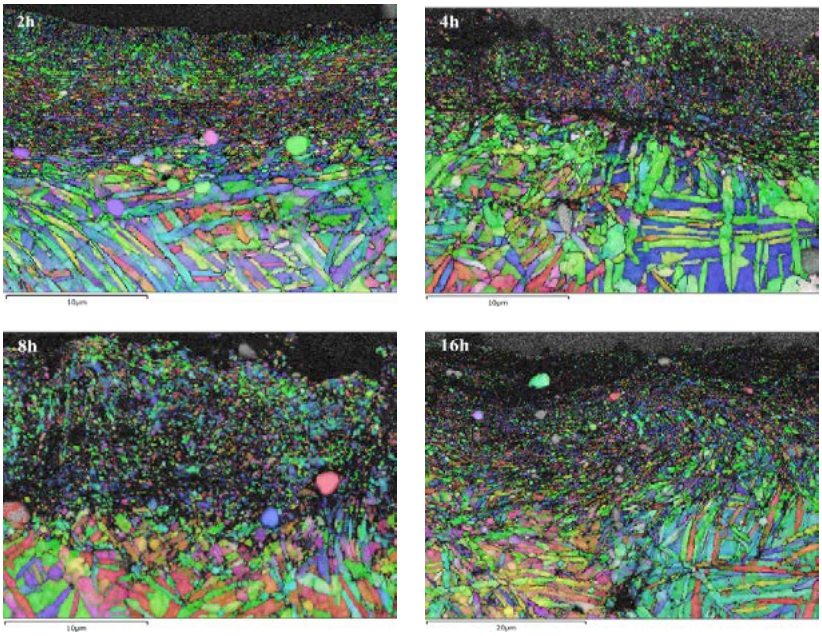


Figure 3.5. Time evolution of the friction layer cross-section at the centre of the grinding mark (0.5 h–16 h)

Grain Size Analysis. EBSD results show that the friction layer at different friction times shows a significant delamination and refinement: surface grains are clearly refined, while base material grains remain large and almost unaffected [20]. At 0.5 h, the friction layer was thinnest, with surface grains appearing refined, indicates that in the early stage, the local high stress and friction were already enough to cause a significant plastic deformation [21]. By 1 h, the fine-grain layer had significantly thickened and expanded, though its grain size remained larger than that at 2 h. Concurrently, EBSD images revealed external grains from the counter-friction material embedded within the copper alloy friction layer, indicative of third-body inclusion during friction [22]. At 2 h, the layer rapidly increases in thickness and reaches a local peak, indicating the intense surface refinement and rapid evolution during the early friction stage. From 2 h to 16 h, the friction layer thickness no longer increased significantly but stabilized at a higher thickness., as shown in Figure **3.6**. The layer thickness is defined as the combined thickness of the ultrafine-grained surface layer and the transition layer, excluding the unaffected base material. The gray dots represent individual sample measurements, while the gray shaded areas denote the mean ± 1 standard deviation for all samples at each time point, the black line as a trend reference.

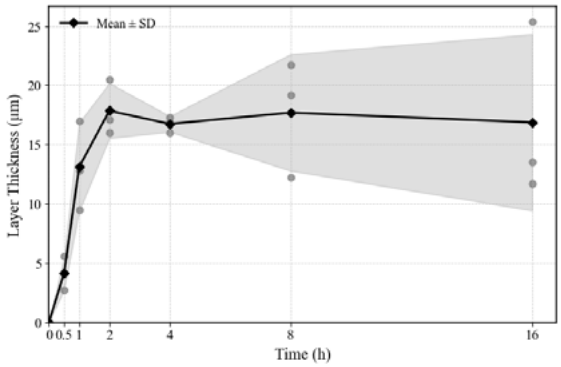


Figure 3.6. Thickness of the Friction-Induced Layer vs. Time

With increasing depth, the grain size increases gradually, with a medium-grain transition layer in the middle and the substrate grains forming the bottom layer. This gradient indicates that friction energy dissipates primarily at the surface and subsurface regions, decreasing with depth [23]. This evolution stems from the competition between grain refinement and surface spalling: on one hand, continuous friction and energy input promote the formation of new fine grains, while on the other, repeated sliding induces localized exfoliation and material removal [24,25].The early rise in COF matches the wear phase dominated by spalling; its decrease and subsequent stabilization match the phase where friction layer refinement/regeneration prevails. As the layer structure stabilizes, COF converges towards a steady state.

Grain Boundary Angle Analysis. Figure **3.7** shows the proportion of LAGB at different time points. At 0.5 h, low-angle grain boundaries appear in the outermost layer. At this stage, grains exhibit slight refinement accompanied by a large number of low-angle grain boundaries. The proportion of surface LAGB significantly increased, far exceeding that of the base phase. However, by 2 h, a distinct further stratification becomes evident: the surface layer consists of fine grains and a small proportion of high-angle grains (27.6%); the subsurface layer contained larger grains than the surface layer but exhibited a high proportion of low-angle grain boundaries (68.4%); the deepest layer comprised the base material (26.8%). During the long-term stage (8 h–16 h), as shown in Figure 3.5, the surface layer kept a low proportion of LAGB, while the subsurface layer maintained a significant proportion.

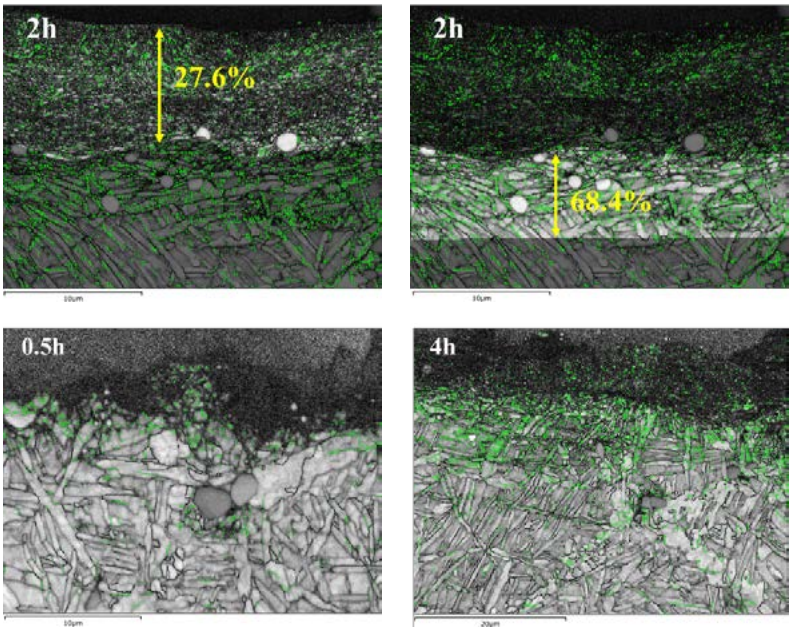


Figure 3.7. Distribution of Low-Angle Grain Boundaries

This indicates that during friction, grains first undergo dislocation or slip, leading to an increase in low-angle grain boundaries and the initiation of grain refinement. As friction persists and energy is continuously input, grain size further refines, and low-angle grain boundaries begin to transform into high-angle grain boundaries.

Before 2 h, the transition zone within the friction layer was not clearly defined. At this stage, the proportion of small-angle grain boundaries in the surface layer initially increased and then decreased. After 2 h, a distinct transition region emerged within the friction layer: grains of moderate size containing a significant number of small-angle grain boundaries. The surface layer and its evolution over time are shown in Figure **3.8**.

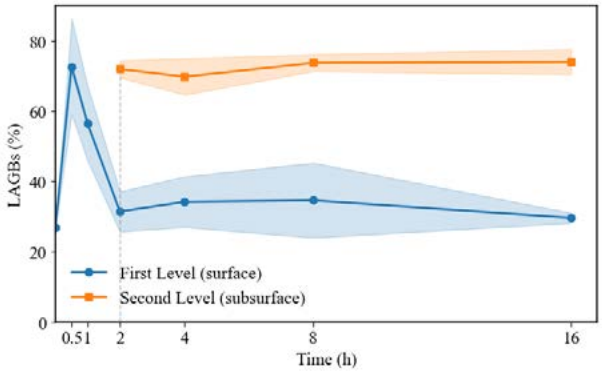


Figure 3.8. Time Evolution of Grain-Boundary Distribution in the Tribolayer

The friction process progressively establishes and maintains a stable delamination structure: fine-grained layers with a high proportion of HAGB surface layers, a large number of LAGB transition zones, and base grain boundaries.

The formation of this stratified structure can be attributed to two mechanisms. First, part of the LAGB in the surface layer gradually undergoes orientation rotation under continuous shear and evolves into HAGB, leading to structural stabilisation. Second, the surface experiences cyclic delamination while former subsurface layer transforms to surface layer. Strain is primarily accommodated in the subsurface, thereby maintaining a high fraction of LAGB in that region [26]. The evolution of microstructure is primarily confined to the friction layer and transition area, with the deeper base material remaining largely unaffected by this process and keeping its original grain boundary characteristics.

Grain Orientation Analysis. To quantify the temporal evolution of crystallographic orientation, the surface normal direction (ND) was taken as the reference for EBSD-based IPF analysis. In the base material, due to the properties inherent to casting and processing conditions, grain orientation remains essentially random. After sliding, however, the surface layer of most specimens consistently exhibits a pronounced {111} texture, while the subsurface retains a more random orientation distribution. This trend is evident in both short-term (0.5 h) and long-term (16 h) specimens, as shown in Figure **3.9**.

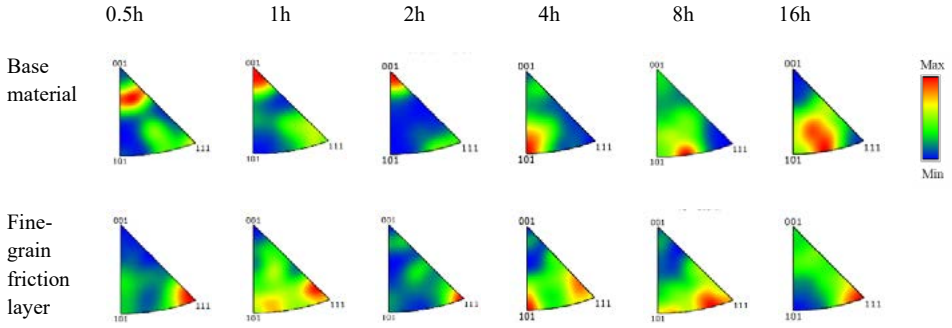


Figure 3.9. Time-Dependent Evolution of Grain Orientation

The evolution of grain orientation toward {111} can be explained by two mechanisms. First, during repeated sliding, surface grains are subjected simultaneously to tangential shear stress and normal contact stress, which induce progressive grain rotation. The rotation path is mainly governed by the sliding direction (SD), transverse direction (TD), and the initial orientation of each grain [16,27]. For FCC copper alloys, the combined shear and normal stresses during friction drive grains to rotate until their orientations align with the most easily activated and mechanically stable slip system. In this case, the favourable orientation is {111}, resulting in the predominance of {111}-oriented grains at the surface. The second mechanism is orientation selection: grains with less wear-resistant orientations are more prone to removal during sliding, leaving behind grains with the more stable and wear-resistant {111} orientation.

4. CONCLUSIONS

1) Under boundary lubrication, a friction layer forms on the planar distributor plate within around 0.5 h; surface grains are rapidly refined, with a significant {111} orientation and a higher proportion of LAGB grains.

2) The thickness of the tribological layer increases over time; after 2 hours, it shows a fluctuation pattern with a cycle of "formation-detachment-reformation," and the peak thickness no longer increases monotonically.

3) After 2 h, the LAGB proportion decreases and the friction layer enters a stable delamination state: the surface layer consists of ultrafine grains of {111} with less LAGB; the subsurface layer beneath is oriented and dispersed, enriched with LAGB, and has slightly larger grains; the base layer extends further downward.

4) The friction layer stabilisation and friction coefficient decrease occur concurrently, showing that grain-scale micro-evolution directly influences macro-scale tribological properties.

5. SUMMARY AND OUTLOOK

5.1. Summary

This study demonstrates that the friction layer of the distributor undergoes systematic evolution with increasing friction duration. Within a short timeframe, surface grain refinement occurs, LAGB increases, and preferred orientation develops. As friction time increases, ultrafine grains form in the surface layer, exhibiting preferred orientation and an increased proportion of HAGB; the subsurface layer contains numerous LAGB with larger grain sizes than the surface layer and random grain orientation. The stable emergence of this friction layer coincides with the oscillatory decrease in the coefficient of friction, indicating that the evolution of the friction layer governs the friction and wear of the distributor plate. This provides a basis for designing wear-resistant port plates.

5.2. Outlook

Although this study has clarified the time series evolution and formation mechanism of the friction layer, two questions still require answers. The specific effects of variable load pressures and shear stresses on grain-level microstructural evolution, along with their direct correlation to wear resistance, have not been quantitatively established. Furthermore, the structural evolution processes at finer scales before observable changes in grain size and orientation require further clarification. On this basis, effective stabilisation mechanisms for the friction-induced layer can be developed, providing a route to significantly enhance the wear resistance and operating efficiency of hydraulic pump friction pairs.

6. REFERENCES

[1] X. Ma, Q.J. Wang, X. Lu, V.S. Mehta, 'Piston surface design to improve the lubrication performance of a swash plate pump', Tribology International, pp. 275–285, Apr., 2019.

[2] M. Ivantysynova, R. Lasaar, ‘An Investigation into Micro- and Macrogeometric Design of Piston/Cylinder Assembly of Swash Plate Machines’, International Journal of Fluid Power, pp. 23–36, Jan., 2004.
[3] G. Schuhler, A. Jourani, S. Bouvier, J.-M. Perrochat, ‘Wear mechanisms in contacts involving slippers in axial piston pumps: A multi-technical analysis’, Journal of Materials Engineering and Performance, pp. 5395–5405, Oct., 2018.
[4] H. Wu, L. Zhao, S. Ni, Y. He, ‘Study on friction performance and mechanism of slipper pair under different paired materials in high-pressure axial piston pump’, Friction, pp. 957–969, Oct., 2020.
[5] A. Holzer, S. Koss, J. Rosefort, J.H. Schleifenbaum, K. Schmitz, ‘EHLA-Coated Carbide-Reinforced Control Plates in Axial Piston Pumps First Results from Real-Life Application’, Advances in Hydraulic and Pneumatic Drives and Control 2023, pp. 238–248, 2024.
[6] J. Jamari, ‘Running-in as an engineering optimization’, J@ Ti Undip: Jurnal Teknik Industri, pp. 1–10, 2007.
[7] R. Ivantysyn, A. Shorbagy, J. Weber, ‘Analysis of the run-in behavior of axial piston pumps’, 2018 Global Fluid Power Society PhD Symposium (GFPS), pp. 1–9, Samara, Jul., 2018.
[8] Z. Tappeiner, A. Holzer, K. Schmitz, ‘Experimental development and validation of tribological run-in strategies to reduce friction in hydraulic applications’, Tribologie und Schmierungstechnik, May, 2024.
[9] W. Jiao, Y. Xia, H. Ma, Z. Zhu, S. Xia, ‘Effect of running-in on the low-pressure tribological performance of valve plate pair in axial piston pumps’, Wear, p. 205113, Nov., 2023.
[10] A. Paulus, ‘Reaktionsschichtbildung auf bleifreien bronze- und messingwerkstoffen im kontakt von zylinder und steuerscheibe einer axialkolbenpumpe’, Doctoral Dissertation, RWTH Aachen University, Aachen, 2017.
[11] H. Czichos, K.-H. Habig, ‘Tribologie-handbuch: tribometrie, tribomaterialien, tribotechnik’, Springer-Verlag, 2010, ISBN: 3-8348-9660-8.
[12] A. Kapoor, J.A. Williams, K.L. Johnson, ‘The steady state sliding of rough surfaces’, Wear, pp. 81–92, 1994.
[13] P.J. Blau, ‘On the nature of running-in’, Tribology International, pp. 1007–1012, Nov., 2005.
[14] D.A. Rigney, ‘Large strains associated with sliding contact of metals’, Materials Research Innovations, pp. 231–234, Mar., 1998.
[15] H.A. Padilla, B.L. Boyce, C.C. Battaile, S.V. Prasad, ‘Frictional performance and near-surface evolution of nanocrystalline ni–fe as governed by contact stress and sliding velocity’, Wear, pp. 860–871, Jan., 2013.
[16] C. Haug, D. Molodov, P. Gumbsch, C. Greiner, ‘Tribologically induced crystal rotation kinematics revealed by electron backscatter diffraction’, Acta Materialia, p. 117566, Feb., 2022.
[17] S. Becker, K. Schulz, D. Scherhaufer, P. Gumbsch, C. Greiner, ‘Variations in strain affect friction and microstructure evolution in copper under a reciprocating tribological load’, Journal of Materials Research, pp. 970–981, Feb., 2021.
[18] J.R. Barber, ‘Contact mechanics’, Vol. 20, Springer, 2018, ISBN: 978-3-319-70938-3.

[19] X. Chen, Z. Han, K. Lu, ‘Friction and Wear Reduction in Copper with a Gradient Nano-grained Surface Layer’, ACS Applied Materials & Interfaces, pp. 13829–13838, Apr., 2018.
[20] C. Greiner, Z. Liu, L. Strassberger, P. Gumbsch, ‘Sequence of Stages in the Microstructure Evolution in Copper under Mild Reciprocating Tribological Loading’, ACS Applied Materials & Interfaces, pp. 15809–15819, Jun., 2016.
[21] F. Ruebeling, Y. Xu, G. Richter, D. Dini, P. Gumbsch, C. Greiner, ‘Normal load and counter body size influence the initiation of microstructural discontinuities in copper during sliding’, ACS Applied Materials & Interfaces, pp. 4750–4760, Jan., 2021.
[22] K.C. Ludema, ‘Third bodies: perspectives on modeling in lubricated contacts, in close fitting contacts, etc: following on the concepts of dr. Maurice godet’, pp. 3–19, Jan., 1996.
[23] M. Chandross, N. Argibay, ‘Friction of metals: A review of microstructural evolution and nanoscale phenomena in shearing contacts’, Tribology Letters, p. 119, Dec., 2021.
[24] N. P. Suh, ‘The delamination theory of wear’, Wear, pp. 111–124, Jul., 1973.
[25] M.M. Khonsari, S. Ghatrehsamani, S. Akbarzadeh, ‘On the running-in nature of metallic tribo-components: A review’, Wear, p. 203871, Jun., 2021.
[26] D.A. Hughes, N. Hansen, ‘High angle boundaries formed by grain subdivision mechanisms’, Acta Materialia, pp. 3871–3886, Sep., 1997.
[27] C. Haug, F. Ruebeling, A. Kashiwar, P. Gumbsch, C. Kübel, C. Greiner, ‘Early deformation mechanisms in the shear affected region underneath a copper sliding contact’, Nature Communications, p. 839, Feb., 2020.

Biographies



Yuhan Song received bachelor's degree in Mechanical Design, Manufacturing and Automation from Taiyuan University of Science and Technology in 2015, the master's degree in Engineering Mechanics from Taiyuan University of Science and Technology in 2018. His professional experience includes serving as a Hydraulic Component Development Engineer at Linde Hydraulic China (August 2018-June 2021), a Component Simulation Engineer at Linde Hydraulic Germany (July 2021-June 2023), and currently as a Hydraulic Component Development Engineer at Weichai Power European Science & Technology Innovation Center (July 2023-present). His primary research focus encompasses the development and simulation of hydraulic components, including pumps, motors, and valves.



Dr. Julian Angerhausen received his bachelor's degree (2012), master's degree (2014), and PhD (2020) in mechanical engineering from RWTH Aachen University. Since 2020, he has been with the Weichai Power European Science & Technology Innovation Center in Aachen, Germany, where he currently serves as Chief Engineer and Head of Hydraulic Component Technology. His research interests include tribology, hydraulic components (valves, pumps, motors), and hydraulic systems for off-road machinery.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Physics-Informed Neural Network-Based Prediction of Mixed Lubrication Characteristics on Cylinder Block/Valve Plate Interfaces via Mechanism-Data Integration

Dingchong LYU [1], Jian FU [2], Shoujun ZHAO [3]

[1]*Beihang University, by2207123@buaa.edu.cn*

[2]*Beihang University, fujian@buaa.edu.cn*

[3]*Beijing Institute of Precision Mechatronics and Controls, shoujunzhao@vip.sina.com*

Abstract.

Aerospace electro-hydrostatic modules (EHMs) play a critical role in precision actuation for flight control systems. The cylinder block/valve plate (CB/VP) interfaces of EHMs are subject to extreme lubrication failure under high-frequency alternating conditions. Typical failures include fluid starvation caused by the gate effect at ultra-high speeds, as well as accelerated wear resulting from boundary friction spikes during low-speed reversals. These issues pose great challenge to conventional analysis, as the high computational cost and inherent instability of high-fidelity simulations present a critical barrier to developing reliable EHM systems. To address these challenges, this study developed a physics-informed neural network (PINN) framework that integrates physical principles with data-driven deep learning. By conducting experiments on the CB/VP interface thermal characteristics under specified operational conditions, the acquired dataset served as the basis for developing a mapping model that correlates operational parameters (rotational speed and pressure) with interface temperature. This established relationship was subsequently integrated into a mixed lubrication model, yielding a thermo-elasto-hydrodynamic (TEHD) coupling simulation framework incorporated into the PINN architecture. Experimental validation demonstrated that the PINN-based model achieved significantly better agreement with measured outputs compared to conventional theoretical predictions, The implemented deep learning methodology enabled precise prediction of mixed lubrication characteristics, maintaining physical consistency while achieving data-driven accuracy. The proposed PINN framework will be extended to investigate rotor dynamics and vibrational behavior in future studies. This work provides enhanced insights into interfacial lubrication mechanisms and establishes an innovative paradigm for analyzing high-performance aerospace servo systems.

Keywords. electro-hydrostatic actuator, deep learning, thermo-elasto-hydrodynamic coupling, tribo-pair.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Aerospace Electro-Hydrostatic Module (EHM) work as the critical power components for precision applications such as thrust-vector control and flight-surface actuation, operate under extreme high-frequency quadrant-altering conditions [1–3]. The integrated EHM architecture induces significant rotor vibrations that critically destabilize the lubrication regime at the cylinder block/valve plate (CB/VP) interface [4,5]. This dynamic environment generates two distinct failure modes unobserved in conventional hydraulic systems. The first is gate effect-induced self-priming suppression: At ultra-high speeds [6], transient flow barriers cause lubricant starvation. While near zero-speed zones, frequent reversals provoke friction spikes that intensify adhesive wear and can lead to catastrophic seizure [7,8]. These phenomena constitute fundamental limitations for achieving reliable lubrication in aerospace EHMs.

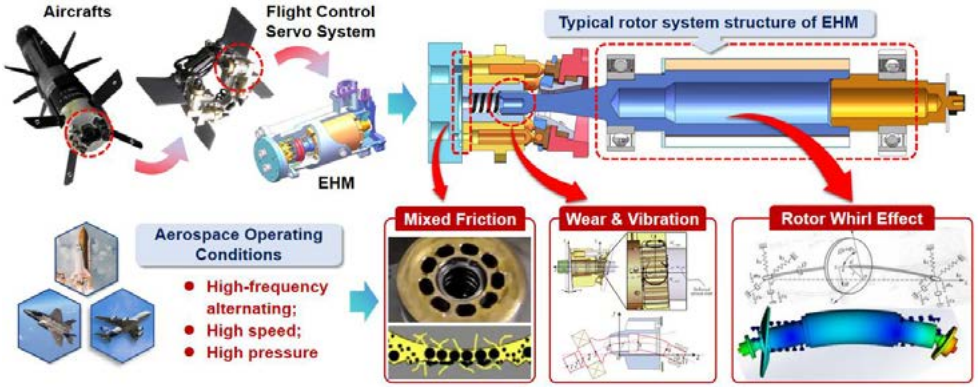


Figure 1.1. Challenges of aerospace highly-compact EHM

Traditional theoretical and numerical methods exhibit significant limitations: high-fidelity models suffer from convergence difficulties [9–12] while predictions consistently deviate from experimental friction torque measurements. System instability has been identified as the root cause [2], and rotor dynamics modelling has revealed critical speed behavior [13]. Although DLC-CrN coatings demonstrate wear resistance [6–8], these methods fail to accurately characterize friction transitions under transient conditions, reflect dynamic oil film rupture/reconstruction processes, or establish wear-vibration coupling frameworks. While artificial neural networks (ANN) show strong function approximation capabilities, purely data-driven approaches face substantial limitations.

To address these challenges, Physics-Informed Neural Networks (PINNs) embed governing equations into neural architectures, achieving synergistic integration of physical principles and data features [14,15]. Faras et al. [16] validated PINNs as effective solvers for elastohydrodynamic lubrication problems, and further demonstrated their versatility in handling both hydrodynamic and EHD regimes [17]. Addressing computational cost, they achieved significant acceleration in contact analysis compared to traditional methods [18]. Moreover, hard-constraint formulations [19] were introduced to strictly enforce physical boundaries, while Ahmed et al. [20] successfully integrated these surrogate models into digital twins for real-time monitoring of lubricated systems. And Yin et al. [21] established a coupled framework resolving cross-scale challenges in seawater-lubricated friction pairs.

This study integrates PINNs into a thermo-elasto-hydrodynamic (TEHD) coupling framework, enabling precise prediction of mixed lubrication characteristics and providing innovative research directions for characterizing complex dynamic behaviors of servo motor pump cylinder block/valve plate pairs under typical power spectra.

To overcome the limitations of conventional methods in analyzing the lubrication performance, this study developed a mechanism-data fused PINN model for the CB/VP interface. The research established a mixed lubrication numerical model, constructed the PINN framework, and validated its predictive accuracy against experimental data. The results demonstrate the efficacy of the proposed data-mechanism integration approach.

2. MIXED LUBRICATION NUMERICAL MODEL OF CB/VP INTERFACE

To analyze the tribological interface between the CB/VP tribo-pair, mathematical and numerical models are employed to simulate its mixed lubrication regime, characterized by partial fluid film and partial asperity contact. These models serve as traditional tools for performance prediction, design optimization, and service life extension [13,22].

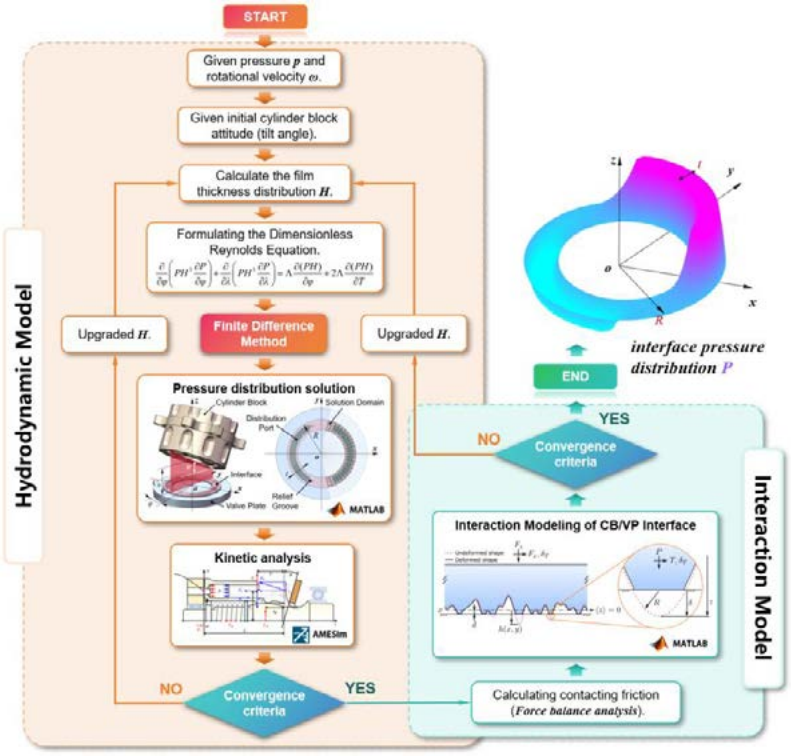


Figure 2.1. Elasto-hydrodynamic numerical modelling framework

2.1. EHD Coupling Modelling Framework

Based on the elasto-hydrodynamic (EHD) model that integrates geometric, lubrication, and contact performance analyses, the procedure for determining the optimal tolerance of the CB/VP interface is structured as follows: Initially, geometric tolerances, including

cylindricity error and dimensional deviations of the piston and cylinder, are characterized [9,10,23]. Subsequently, the fluid pressure distribution at the interface is solved by coupling fluid dynamics with the mechanical equilibrium of the cylinder assembly. This pressure distribution is then input into a contact mechanics model (the Greenwood-Williamson contact model, G-W model [24]) to achieve fluid-solid domain coupling. Finally, the resulting film thickness and pressure distributions are used to compute key performance parameters such as friction losses and leakage rates.

a) *Hydrodynamic Model*

The CB/VP interface represents a critical tribo-pair that integrates sealing, bearing, and lubricating functions [6]. As the contact interface with the largest surface area in EHM systems, it serves as the primary site for cavitation, leakage, and wear failures, thereby fundamentally constraining the performance ceiling of the entire system. The dynamically engaged region highlighted in red in Fig. 2.2(a) constitutes the core area of lubrication analysis.

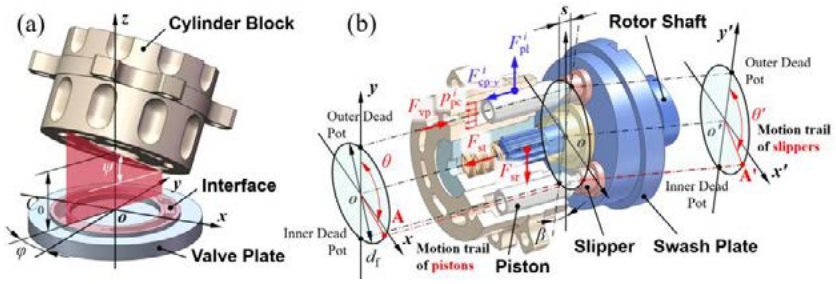


Figure 2.2. Geometric relationship of cylinder block/valve plate interface [22]

To characterize its hydrodynamic behavior, the dimensionless Reynolds equation is employed in this study. The pressure field is normalized using the contact zone's mid-radius (radial direction) and the nominal film thickness (normal direction) as characteristic length scales [13]. This approach ensures numerical stability and eliminates potential errors induced by dimensional inconsistencies during the computational process. Based on this framework, the following equation is established:

$$\frac{\partial}{\partial\varphi}\left(\rho H^3\frac{\partial P}{\partial\varphi}\right)+\frac{\partial}{\partial\lambda}\left(\rho H^3\frac{\partial P}{\partial\lambda}\right)=\Lambda\frac{\partial(\rho H)}{\partial\varphi}+2\Lambda\frac{\partial(\rho H)}{\partial T} \quad (2.1)$$

$$\Lambda=\frac{6\mu\omega}{P_a}\left(\frac{R}{C_0}\right)^2 \quad (2.2)$$

Where φ and λ denote the dimensionless circumferential and radial coordinates, respectively. T is the dimensionless operating time. P and H represent the dimensionless oil pressure and oil film thickness, respectively. ρ denotes the density of the hydraulic working medium. Λ denotes the scaling factor, and it is defined by the mid-radius of the interface R, rotational velocity ω, hydraulic medium viscosity η, nominal clearance C_0, and standard atmospheric pressure P_a.

Fig. 2.2(b) depicts the dynamic equilibrium of the cylinder block assembly. A lumped-parameter model representing the integrated cylinder block-piston-slipper-swash plate system was established in AMESim to characterize the multi-body dynamics. Detailed

modelling methodology is provided in Ref [25]. And tilting angle ψ of the cylinder block are incorporated into the characterization of H, as can be shown below.

$$H = \frac{C_0 + R\sin\psi\cos(\varphi + \varphi_0)}{R} \tag{2.3}$$

b) *Interaction Model*

The G-W model is a classical theory in contact mechanics, specifically employed to calculate elastic contact between rough surfaces. The core of this model lies in idealizing a rough surface as a series of elastic hemispheres with varying heights, and deriving the overall contact behavior by superposing the contact solutions of individual hemispheres. By quantifying the contact laws governing rough surfaces, this model provides a crucial theoretical framework for subsequent research in the field [24].

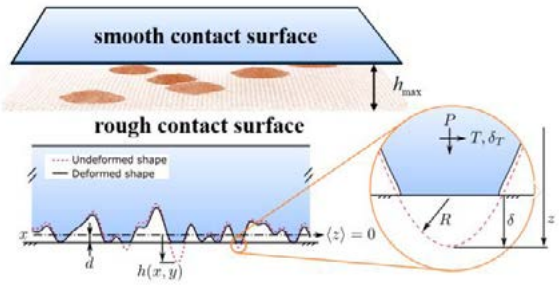


Figure 2.3. The Greenwood-Williamson contact mechanics model [24]

In this model, one of the rough surfaces is assumed to consist of an infinite number of independent hemispheres with identical radii, and the heights of these hemispheres follow a Gaussian distribution. The formula for the probability density function $f(h)$ is given as follows:

$$f(h) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(h-\mu)^2}{2\sigma^2}\right) \tag{2.4}$$

Where h is the height coordinate of an individual asperity, σ is the standard deviation of asperity height distribution, μ is the mean value of asperity height distribution.
Whereas the other surface is assumed to be perfectly smooth. Meanwhile, it is postulated that the contact between each individual hemisphere and the rigid plane is an independent event; the total contact force and contact area are thus the sum of contributions from all hemispheres that come into contact. The formula for the total contact force F is shown as follows:

$$F = \frac{4}{3} E * R^{\frac{1}{2}} n_0 \int_{h_c}^{\infty} (\partial + h - h_c)^{\frac{3}{2}} f(h)\,dh \tag{2.5}$$

Where E is the equivalent elastic modulus, R is the equivalent radius of asperities, n_0 is the asperity density, δ is the normal displacement, h_c is the contact threshold height.

2.2. Numerical Solution Methodology

The pressure distribution is solved using a circumferential difference scheme combined with an elliptic partial differential equation iterative configuration. The circumferential difference

equations ensure continuity of pressure and film thickness in the annular solution domain, while the elliptic formulation enables rapid convergence of the nonlinear fluid governing equations. This approach facilitates subsequent contact mechanics calculations. For specific implementation details, refer to [13].

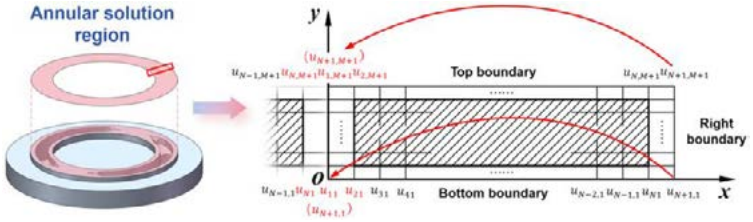


Figure 2.4. Circumferential finite difference discretization of the CB/VP Interface [13]

3. PINN Prediction Model via Mechanism-Data Integration

This study introduces a novel dual-domain nested PINN framework. Its loss function is innovatively constructed by integrating experimental data from EHM characteristic tests with multi-physics governing equations.

3.1. Deep Learning Prediction Framework

A dual-layer nested neural network architecture with mechanism-data fusion is developed in this research. The inner-layer network employs data-driven ANNs, which are trained on thermal characteristic test data from the CB/VP interface to predict its thermal characteristics. The trained ANN model is subsequently embedded to constitute a TEHD model. The outer-layer PINN framework incorporates TEHD governing equations and experimental data into its loss function, enabling accurate prediction of mixed friction characteristics at the CB/VP interface, as shown in Fig 3.1.

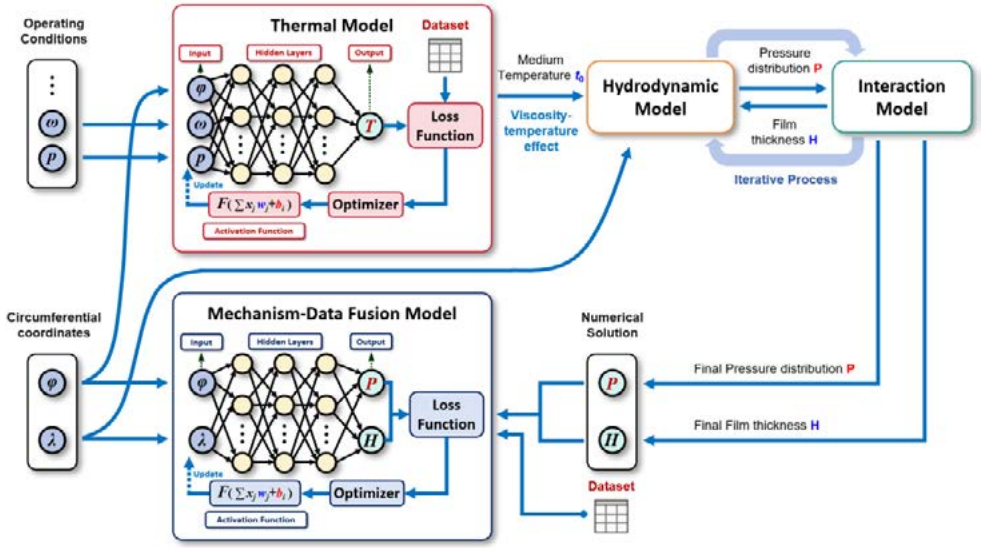


Figure 3.1. Mechanism-data fusion-based PINN prediction framework

Based on the EHD model for the CB/VP pair described in Section 2., where the fluid viscosity μ is initially a constant, an ANN-based temperature prediction model is now nested. Under specified outlet pressure and rotational speed boundary conditions, the ANN first computes the temperature distribution T across the CB/VP interface. This predicted temperature field is subsequently used to determine the spatially varying fluid viscosity μ_d via a temperature-viscosity effect. The updated, non-uniform viscosity field is then fed into the subsequent hydrodynamic and contact mechanics calculations, enabling a fully coupled TEHD simulation. Due to the fast computational speed of the ANN model, the overall computation time of this TEHD model remains comparable to the EHD model.

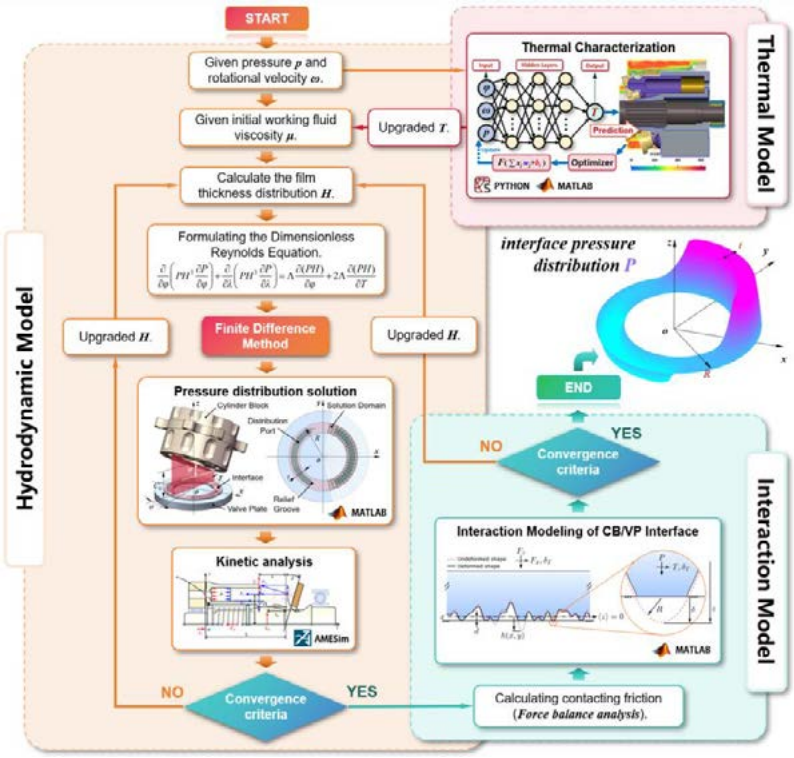


Figure 3.2. Thermal-elasto-hydrodynamic model based on PINN thermal prediction

3.2. Loss Function

The loss functions for the two nested neural networks are respectively defined, broadly categorized into theoretical loss and data loss.

a) *Implementation of Mechanism-Data Integration*

In the dual-layer nested architecture of this study, the neural networks in both the inner and outer layers share an identical internal structure. The distinction lies solely in the composition of their loss functions and the number of parameters in the input and output layers, as illustrated in Fig 3.3. and Fig 3.4. As depicted in Fig 3.3., the loss function directly computes the difference between the training set values and the predicted outputs, subsequently optimized by adjusting biases and weights through a solver.

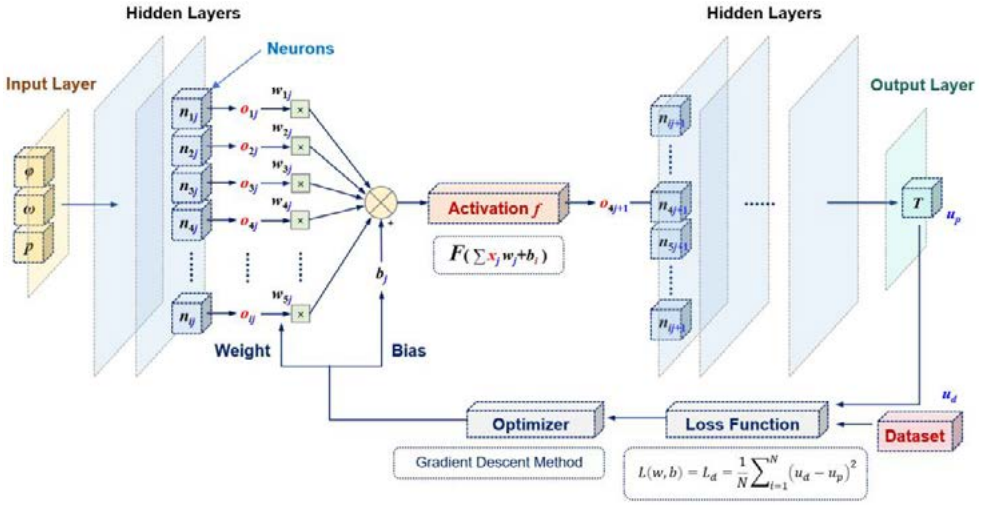


Figure 3.3. Training process of interface thermal prediction

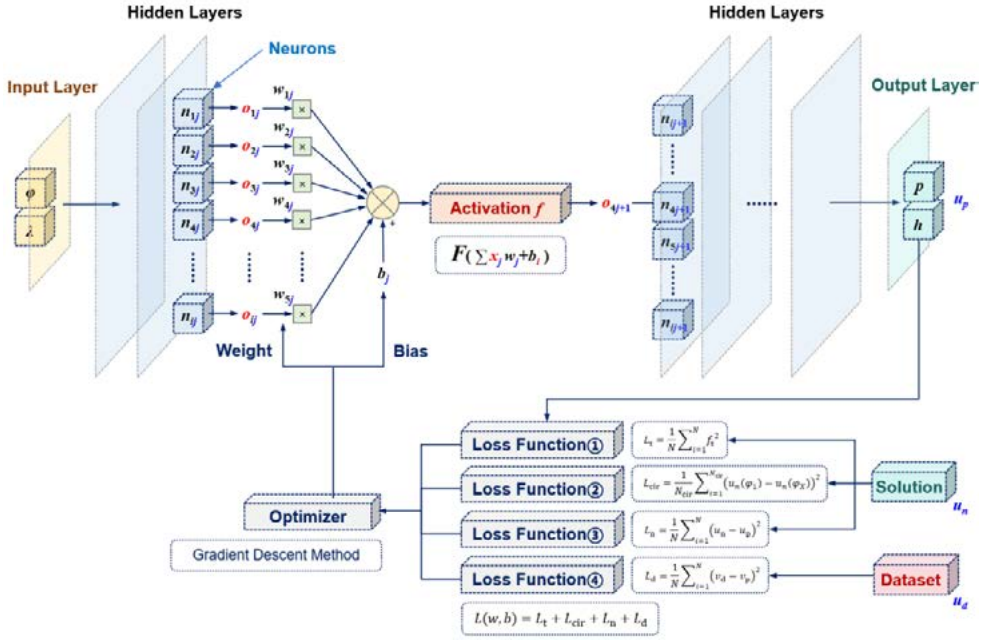


Figure 3.4. Training process of mixed lubrication prediction

In contrast, the loss function in the outer PINN framework shown in Fig 3.4. is considerably more complex. It incorporates the theoretical loss derived from the governing equations and data loss from numerical solutions. A key advantage of PINN architecture lies in its flexibility: unlike traditional equation systems that may suffer from over-constraint issues, the loss components in PINN can be seamlessly integrated without such limitations. However, it is important to note that simply stacking excessive loss terms does not necessarily enhance training accuracy and efficiency.

b) *Theoretical Loss*

In the nested architecture employed in this work, the theoretical error resides solely within the outer PINN framework and arises from the following three distinct sources, where u_{nn} denotes the prediction value. N denotes all types of collocation points (interior points set, boundary condition set, and circumferential boundary set):

1) the first component originates from physical equations, constructed by substituting the PINN outputs for dynamic pressure distribution P and film thickness distribution H into Eq. (2.1) and (2.5), then calculating the mean squared error between the left-hand and right-hand sides of these governing equations.

$$L_{t1} = \frac{1}{N_{in}} \sum_{i=1}^{N_{in}} \left(f_l(u_{nn}) - f_r(u_{nn}) \right)^2 \tag{3.1}$$

2) Second, the boundary condition error encompasses four segments: the inner and outer boundaries of the CB/VP interface, along with both kidney-shaped groove boundaries. The pressure values at the two kidney-shaped grooves are set to the high- and low-pressure zone values respectively, while the interface boundaries are assigned the chamber pressure.

$$L_{t2} = \frac{1}{N_b} \sum_{i=1}^{N_b} \left(u_{nn}(\varphi_{b1}, \lambda_{b1}) - P_b \right)^2 \tag{3.2}$$

3) Third, the circumferential continuity error arises from the annular solution domain of the CB/VP interface as mentioned in Section 2.2. This requires the nodal values, first-order partial derivatives, and second-order partial derivatives at the left and right boundaries of the nodal matrix to be consistent, ensuring periodicity across the computational domain.

$$L_{t2} = \frac{1}{N_c} \sum_{i=1}^{N_c} \left(u_{nn}(\varphi_{b1}, :) - u_{nn}(\varphi_{br}, :) \right)^2 \tag{3.3}$$

$$L_{t2}' = \frac{1}{N_c} \sum_{i=1}^{N_c} \left(\frac{\partial u_{nn}(\varphi_{b1}, :)}{\partial \varphi} - \frac{\partial u_{nn}(\varphi_{br}, :)}{\partial \varphi} \right)^2 \tag{3.4}$$

$$L_{t2}'' = \frac{1}{N_c} \sum_{i=1}^{N_c} \left(\frac{\partial^2 u_{nn}(\varphi_{b1}, :)}{\partial \varphi \partial \lambda} - \frac{\partial^2 u_{nn}(\varphi_{br}, :)}{\partial \varphi \partial \lambda} \right)^2 \tag{3.5}$$

c) *Data Loss*

Under this nested architecture, the two neural networks exhibit distinct data loss: 1) The data for the inner ANN model is entirely derived from EHM thermal tests, with its training set constructed from measured data. 2) The two output quantities of the outer PINN model (dynamic pressure distribution P and film thickness distribution H) have no direct correlation with temperature parameters. Consequently, measured temperature data cannot be directly used to build its training set. Instead, numerical solutions obtained from the TEHD model embedded with the ANN are employed as training data, which integrates thermal test data with mechanistic modelling. Nevertheless, both neural networks adopt identical loss function expressions, as shown below:

$$L_d = \frac{1}{N_t} \sum_{i=1}^{N_t} \left(u_{nn} - u_d \right)^2 \tag{3.6}$$

where u_d denotes the data value at a specific collocation point. Nt denotes all collocation points. The term L_d directly computes the mean squared error between the deep learning predicted values and the data values across all collocation points in the training set.

3.3. Network Training Configuration

Hyperparameter selection for neural networks remains predominantly experience-dependent, as these choices critically govern the model's convergence behavior, predictive accuracy, and computational overhead [14].

a) *Network Architecture*

Based on established research [26], the mesh configurations for the dual nested networks have been defined and are presented in Table 3.1. Notably, the ANN training set comprises 10 Operating Points (OPs) × 5 Measuring Points (MPs), whereas the PINN training set is constituted by 20% of the available numerical solution nodes.

Table 3.1. Parameter settings for network training

Networks	Hidden layers	Neurons	Training set size
1) Thermal prediction ANNs	4	20	10 OPs × 5 MPs = 50
2) Lubrication prediction PINNs	4	50	20% of numerical nodes

b) *Activation Function and Optimizer*

The Multilayer Perceptron (MLP) is constructed with 'Tanh' activation functions and trained using the 'Adam' optimizer, minimizing a mean squared error (MSE) loss that integrates numerical data with governing equations.

4. EXPERIMENTAL VALIDATION

This section details the experimental rig's modular architecture. A multi-level temperature monitoring system with distributed armoured thermocouples at the CB/VP interface enables automated data acquisition. The collected dataset is used to develop a PINN prediction model.

4.1. Overall Performance Test Platform of EHM

The experimental platform comprises three main modules. The first module is a high-frequency variable-condition motor-pump load simulation module, integrating a load simulation system, external control system for simulated loads, oil replenishment system for the tested servo pump, circulation filtration system, cooling system, residual oil recovery system, and hydraulic signal acquisition module. The second module is a human-machine interaction interface consisting of quick-connect components, safety monitoring systems, and display/control panels. The third module is an electronic control multi-dimensional real-time perception system incorporating signal acquisition and control units, with the physical setup shown in Fig 4.1.

Through coordinated operation of its modular systems, the platform performs essential functions including basic parameter testing, load simulation, system pressurization, oil temperature control, and data storage. A multi-level temperature monitoring system was established on the CB/VP interface using armored thermocouples. As shown in Fig. 4.2, a set of sensors were circumferentially arranged near the valve plate and embedded through radial drilling. The thermal testing module can automatically acquire temperature signals in real time and stored them in a database for the PINN training process.

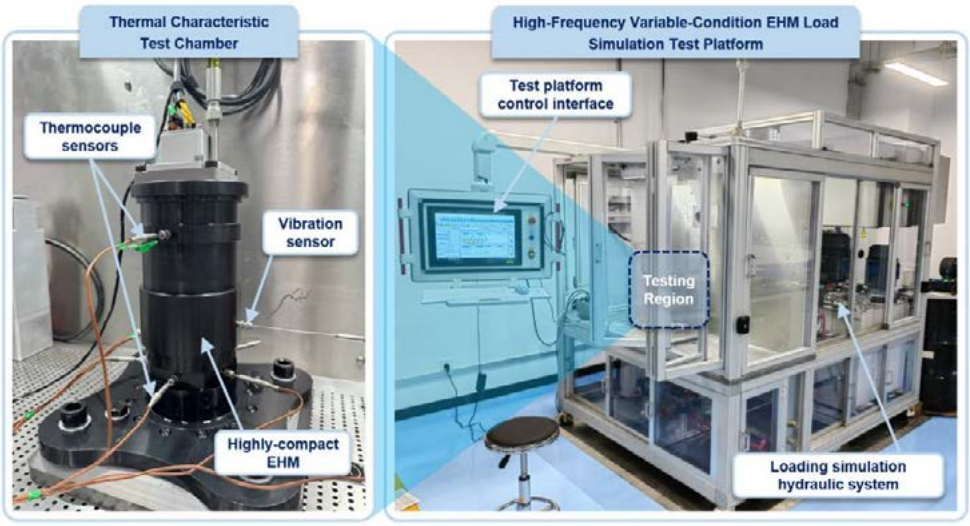


Figure 4.1. The overall performance test platform of EHM

4.2. Dataset Acquisition: Thermal Characteristic Test of CB/VP Interface

The dataset was constructed through the following three-step procedure: 1) First, temperature data were collected from all monitoring points during operation from initial startup to steady-state conditions at each designated working point. 2) Next, a mapping array was constructed using rotational speed ω, outlet pressure P, and initial temperature T_0 as inputs, with the corresponding steady-state temperatures as outputs. 3) And then, neural network architecture described in Section 3.2 was employed to train a temperature prediction model for the CB/VP interface using the developed dataset. 4) At last, the PINN prediction model of thermal characteristics was integrated into the TEHD framework. The curves shown in Fig 4.2. were measured under constant conditions of 14 MPa and 4000 r/min.

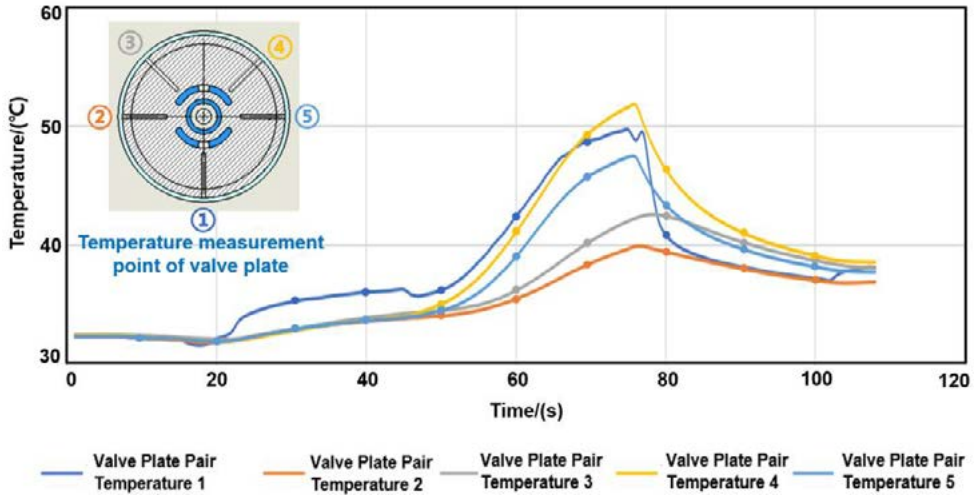


Figure 4.2. Thermal characteristic testing of EHM

5. RESULTS AND DISCUSSION

Model accuracy was assessed by comparing experimental data with EHD and PINN simulations. The influence of training epochs was then systematically analyzed to validate the model's precision. Ultimately, the PINN prediction on lubrication was finished.

5.1. Prediction Accuracy Comparison

As depicted in Figure 5.1, the leakage data and predictions are compared with experimental measurements across four operating points. The vertical axes denote the output pressure and rotational speed (represented by blue and red polylines), and the horizontal axis lists the operating points. The actual leakage, along with the predictions from both models, are visualized using bar clusters, with respective values detailed in Table 5.1.

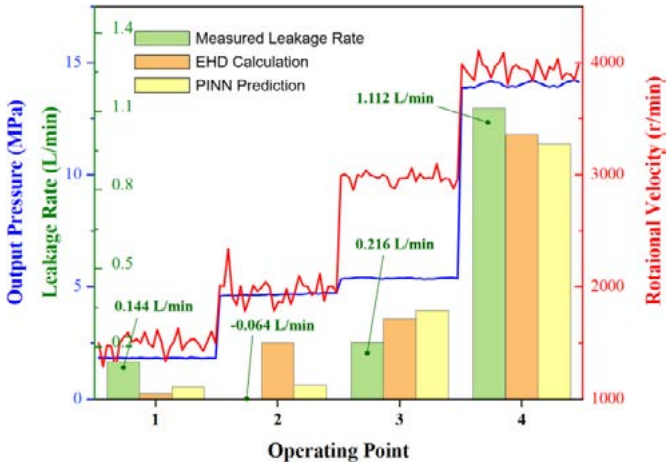


Figure 5.1. Comparisons of measured and predicted leakage rates

A comparative analysis of the relative errors shows a vastly wider distribution for the EHD model (13.9% to 332.8%) than for the PINN model (32.5% to 156.6%). The EHD model showed significant deviations, especially at Operating Points 2 and 3, whereas the PINN model yielded lower errors at all points except one. The significantly lower mean error of the PINN model confirms its superiority, indicating not only higher average precision but also better adaptability to different operating conditions and enhanced robustness.

Table 5.1. Operating point setting and leakage values

Operating Conditions	**Operating Points**			
	I	**II**	**III**	**IV**
Outlet Pressure (MPa)	1.842	4.679	5.390	14.03
Rotational Velocity (r/min)	1496	1968	2972	3938
Measured Leakage (L/min)	0.1438	-0.06440	0.2161	1.112
EHD Calculation (L/min)	0.02004	0.2143	0.3068	1.012
PINN Prediction (L/min)	0.04677	0.05293	0.3383	0.9748

The PINN model's training progress was analyzed by tracking its predictive accuracy across epochs. Detailed outputs, including numerical results and contour plots, were examined at eight key nodes. As shown in Fig. 5.2, the prediction error converged significantly from 1596.1% to 4.08%. This improvement was not entirely uniform, as a noticeable error rebound occurred at 40000 iterations.

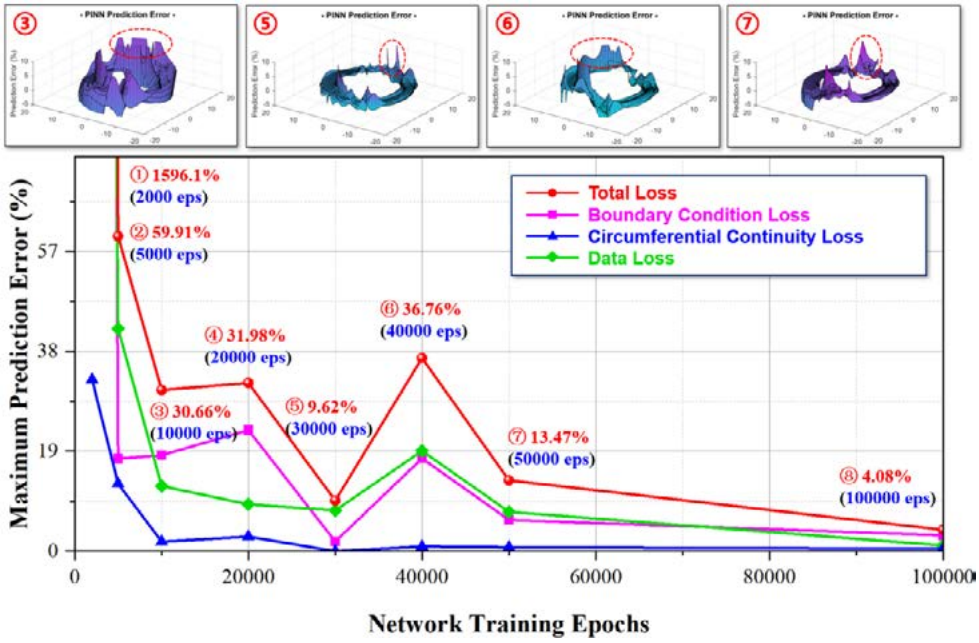


Figure 5.2. PINN prediction accuracy across various training epochs

By comparing the boundary loss, data loss, and circumferential continuity loss separately, it can be further observed that across different training epochs, the locations of maximum error consistently lie on the inner and outer boundaries of the solution domain, rather than at circumferential discontinuities. The influence of circumferential continuity loss is minimal, with errors being primarily driven by boundary loss and data loss. As the number of training epochs increases, the boundary loss shows a clear oscillatory decline, while the data loss fluctuates more irregularly. Additionally, errors near the two dead zones are larger than in other regions, with the locations of maximum error predominantly concentrated in these areas.

5.2. Mechanism-Data Fusion-Driven Prediction of Lubrication Characteristics

The PINN model demonstrates a decisive computational advantage by performing complete simulations and generating detailed contour plots within milliseconds. This stands in stark contrast to the conventional EHD model, which requires approximately 10 minutes to solve due to its inherent computational bottlenecks, a limitation that is further exacerbated by longer solve times and convergence challenges in its more complex TEHD formulation. The remarkable efficiency of the PINN approach enabled the extensive simulation of the interface's mixed lubrication state across a wide spectrum of operating conditions, as comprehensively detailed in Figure 5.3.

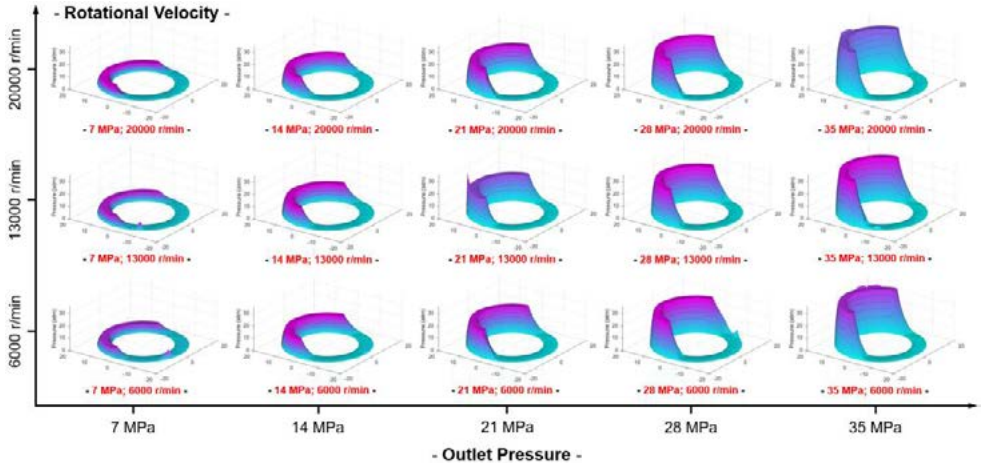


Figure 5.3. PINN prediction of mixed lubrication under various operating conditions

The pronounced growth of oil film thickness in the high-pressure zone under increasing load, at a constant speed, underscores the critical role of pressure in governing film formation. Meanwhile, variations in rotational velocity under a constant pressure condition exhibit minimal impact on the pressure profiles within critical areas such as the dead zone and kidney port, thereby maintaining the overall pressure landscape largely unchanged. Furthermore, the inherent operational tilt of the cylinder block is confirmed to generate an asymmetric oil film distribution, which sporadically triggers intense local pressure concentrations, manifesting as isolated, sharply defined peaks within the contour plots.

6. CONCLUSION

This study pioneers a mechanism-data fused PINN prediction method for analyzing the lubrication regime in aerospace EHM's CB/VP interface. The framework synergizes thermal experiments with EHD modelling, achieving unprecedented predictive accuracy over conventional methods. It significantly enhances the prediction accuracy and efficiency of mixed lubrication parameters while rigorously preserving physical consistency. The principal findings include:

- The PINN model demonstrates superior predictive accuracy over the conventional EHD approach, with a substantially narrower relative error range (32.5%-156.6% vs 13.9%-332.8%) and significantly lower mean error across four operating conditions.
- Training iterations critically influence model convergence, reducing prediction error from 1596.1% to 4.08%. The temporary error resurgence at 40,000 iterations represents a natural optimization phenomenon, while the overall downward trend confirms the value of extended training. Meanwhile, PINN enables millisecond-level simulations versus 10 minutes for EHD.
- The key lubrication mechanisms are revealed: pressure-dominated film formation, minimal speed impact on pressure profiles, and tilt-induced asymmetric film distribution with localized pressure peaks.

7. REFERENCES

[1] J.-C. Maré, J. Fu, 'Review on Signal-by-Wire and Power-by-Wire Actuation for More Electric Aircraft', Chin. J. Aeronaut., 30(3), pp. 857–870, 2017. doi: 10.1016/j.cja.2017.03.013.

[2] D. Lyu, et. al., 'Key Technologies and Challenges of High-Performance Servo Motor Pump', Acta Aeronaut. Astronaut. Sin., 45(15), p. 630225, 2024. doi: 10.7527/S1000-6893.2024.30225.

[3] J. Zhao, et. al., 'Review of Cylinder Block/Valve Plate Interface in Axial Piston Pumps: Theoretical Models, Experimental Investigations, and Optimal Design', Chin. J. Aeronaut., 34(1), pp. 111–134, 2021. doi: 10.1016/j.cja.2020.09.030.

[4] Y. Fu, et. al., 'Electrohydraulic Pump Development Status and Key Technology Review', Mach. Tool Hydraul., 40(1), pp. 143–149, 2012.

[5] S. Zhao, et. al., 'The Study on the Dynamic Capability of an Electro-Hydrostatic Actuator to Drive a Large Inertia Load', Proc. of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp. 836–841, 2016. doi: 10.1109/aus.2016.7748169.

[6] J. Zhao, et. al., 'Experimental Research on Tribological Characteristics of TiAlN Coated Valve Plate in Electro-Hydrostatic Actuator Pumps', Tribol. Int., 155, p. 106782, 2021. doi: 10.1016/j.triboint.2020.106782.

[7] J. Fu, et. al., 'Tribological Characteristics of Hard-to-Hard Matching Materials of Cylinder Block/Valve Plate Interface in Electro-Hydrostatic Actuator Pumps', Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci., 237(14), pp. 3320–3333, 2023. doi: 10.1177/09544062231189692

[8] D. Lyu, et. al., 'Experimental Research on Friction Characteristics of DLC Coated Cylinder Block/Valve Plate Interface in Aerospace Integrated Motor Pumps', 2023 Asia-Pacific International Symposium on Aerospace Technology (APISAT 2023) Proceedings, pp. 1141–1156, 2024. doi: 10.1007/978-981-97-4010-9-89.

[9] H. Deng, et. al., 'Lubrication Characteristics of Valve Plate Pair of Balanced Large Flow Pump Based on Elastic Deformation', Flow Meas. Instrum., 95, p. 102514, 2024. doi: 10.1016/j.flowmeasinst.2023.102514.

[10] Y. Lu, et. al., 'Comparative Study on Asymmetric and Symmetric Configurations of Valve Plates in Axial Piston Pumps', Proc. of the ASME/BATH Symposium on Fluid Power and Motion Control (FPMC), p. V001T01A056, 2024. doi: 10.1115/fpmc2024-140112.

[11] R. Chacon, M. Ivantysynova, 'Advanced Virtual Prototyping of Axial Piston Machines', Proc. of the 9th FPNI Ph.D. Symposium on Fluid Power, p. V001T01A026, 2016. doi: 10.1115/fpni2016-1561.

[12] L. Shang, M. Ivantysynova, 'An Investigation of Design Parameters Influencing the Fluid Film Behavior in Scaled Cylinder Block/Valve Plate Interface', Proc. of the 9th FPNI Ph.D. Symposium on Fluid Power, p. V001T01A027, 2016. doi: 10.1115/fpni2016-1540.

[13] D. Lyu, et. al., 'Development of Boundary/Mixed Friction Models for Cylinder Block/Valve Plate Interface in a Novel Aerospace Electro-Hydrostatic Module', Proc. of the ASME/BATH Symposium on Fluid Power and Motion Control (FPMC), p. V001T01A057, 2024. doi: 10.1115/fpmc2024-140246.

[14] H. Gao, et. al., ‘Physics-Informed Neural Network for Solving Hydrodynamic Lubrication Characteristics of Piston Pump Slipper Pair’, Tribol. Int., 181, p. 108290, 2025. doi: 10.13052/rp-9788743808251a11.

[15] X. Deng, et. al., ‘Fast Prediction of Flow Field in Scramjet Combustor Based on Physical Information Neural Network under Wide Mach Number’, Chin. J. Aeronaut., 38(7), p. 103482, 2025. doi: 10.1016/j.cja.2025.103482.

[16] F. Brumand-Poor, et. al., ‘Physics-Informed Neural Networks as Solvers for Elastohydrodynamic Lubrication Problems’, Proc. 13th Int. Fluid Power Conf. (IFK), 2022.

[17] F. Brumand-Poor, et. al., ‘Extrapolation of Hydrodynamic Pressure in Lubricated Contacts: A Novel Multi-Case Physics-Informed Neural Network Framework’, Lubricants, 12(4), p. 122, 2024. doi: 10.3390/lubricants12040122.

[18] F. Brumand-Poor, et. al., ‘Fast Computation of Lubricated Contacts: A Physics-Informed Deep Learning Approach’, Proc. 14th Int. Fluid Power Conf. (IFK), 2024.

[19] F. Brumand-Poor, et. al., ‘Physics-Informed Neural Networks for the Reynolds Equation with Transient Cavitation Modeling’, Lubricants, 2(11), p. 365, 2024. doi: 10.3390/lubricants12110365.

[20] A. Saleh, et. al., ‘Real-Time Prediction of Pressure and Film Height Distribution in Plain Bearings Using Physics-Informed Neural Networks (PINNs)’, Lubricants, 13(8), 360, 2025. doi: 10.3390/lubricants13080360.

[21] F. Yin, et. al., ‘Physics-Informed Machine Learning for Tribological Propertie s Prediction of S32750/CFRPEEK Tribopair under Seawater Lubrication via PISSA-CNN-LSTM’, Tribol. Int., 199, p. 109965, 2024. doi: 10.1016/j.triboin t.2024.109965.

[22] D. Lyu, et. al., ‘Investigations on Thermal-Fluid Coupling Characteristics of an Innovative Compact Electro-Hydrostatic Module (EHM) Considering Rotor Churning Effects’, Proc. of the 17th International Symposium on Fluid Power, pp. 1–10, 2024.

[23] B. Han, et. al., ‘Analysis of Friction Characteristics of Valve Plate Pair in an Axial Piston Pump Considering Cylinder Block Dynamics’, J. Tribol., 146(4), p. 041702, 2024. doi: 10.1115/1.4064116.

[24] Y. Xu, et. al., ‘An Asperity-Based Statistical Model for the Adhesive Friction of Elastic Nominally Flat Rough Contact Interfaces’, J. Mech. Phys. Solids, 164, p. 104878, 2022. doi: 10.1016/j.jmps.2022.104878.

[25] J. Zhao, et. al., ‘Flow Characteristics of Integrated Motor-Pump Assembly with Phosphate Ester Medium for Aerospace Electro-Hydrostatic Actuators’, Chin. J. Aeronaut., 35(12), pp. 256–271, 2022. doi: 10.1016/j.cja.2022.11.013.

[26] D. Lyu, et. al., ‘Extreme Variable-Speed Lubrication Dynamics of Cylinder Block/Valve Plate Interface Based on Neural Network Method’, Proc. of the 2025 Swedish International Conference on Fluid Power (SICFP2025), 2025. doi: 10.13052/rp-9788743808251a24.

Biographies



Lyu Dingchong received the bachelor's degree in mechanical engineering from Wuhan University of Technology in 2019, the master's degree in mechanical engineering from Xi'an Jiaotong University in 2022, respectively. He is currently pursuing the philosophy of doctorate degree in mechatronic engineering at the Laboratory of Aerospace Servo Actuation and Transmission, Beihang University. His research focuses on the dynamic and static characteristics of high-performance servo power components and their optimization design theories. He has contributed to multiple national-level research initiatives and has regularly presented his findings at international academic conferences.

Multi-Objective Optimization of Kfloid-Type Gerotor Pumps for High-Pressure Applications

Harsh Kulkarni*, Andrea Vacca*, Jochen Hebel, Benjamin Ginsberg**, Christian Kretzer**, Matthias Krug****

Maha Fluid Power Center, Purdue University*

Thomas Magnete GmbH**

E-Mail:hkulkar@purdue.edu

Abstract.

Gerotor pumps have been successfully used in low pressure systems (< 50 bar) such as in automotive and in charge systems, owing to their low cost, low acoustic noise, and high filling capability. These positive features make gerotor still attractive in modern applications that require higher energy efficiency or reduced spatial claim. Therefore there is still an interest in studying alternative geometries with respect to the conventional ones. This paper takes into consideration a gerotor morphology referred to as Kfloid, which is not well studied in literature, and compare its performance with respect to a reference industry design. The approach of study is based on a parametric numerical optimization of this morphology, where the objective functions relate to specific displacement of the unit, sealing capabilities, flow non-uniformity and mechanical stress and wear. To allow such analysis, a numerical model for the generation and the processing of Kfloid profile is presented, and then used in an optimization workflow. Results include Pareto fronts highlighting trade-offs between the considered objective functions. To highlight the potential of the optimized Kfloid profile, comparison with a design previously optimized by the research team, and based on a cycloidal morphology are presented.

Keywords. Pumps, Gerotors, Optimization

1. INTRODUCTION

A gerotor gearset consists of an internally toothed gear and a conjugate externally toothed pinion of one tooth difference and features continuous contact between the rotors. The name was coined by its inventor by the combination of the two words "generated" and "rotor" [1]. Gaps between the teeth are sealed by the contact points and create control volumes that increase and decrease in size throughout a revolution, which enables a pumping action in the axial direction. An example of a typical gerotor pump is shown in Figure 1.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

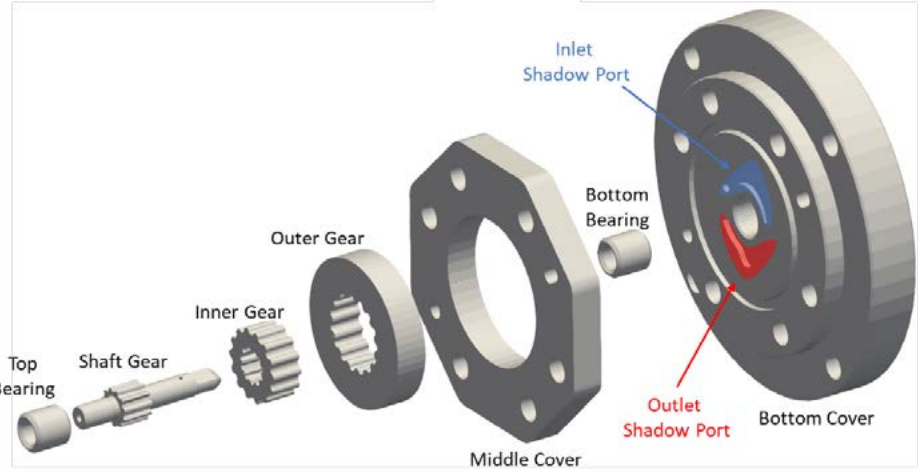


Figure 1: Typical Gerotor Assembly

Gerotor pumps are commonly employed in low-pressure applications such as oil, fuel, automatic transmission, and charge systems. They operate on a simple and elegant principle, requiring only a few components, which contributes to their robustness and cost-effectiveness. Their compact construction and suitability for powder metallurgy make them highly economical to manufacture. The most critical design consideration for a gerotor pump lies in the macro-geometry of its gears, as this directly influences the pump's displacement, flow ripple, efficiency, wear characteristics, and overall dimensions. Furthermore, the gears must maintain an appropriate drive relationship to ensure that the inner and outer rotors engage smoothly without the need for additional synchronization mechanisms.

In theory, for creating a gerotor profile any curve can be selected to form either the inner or outer gear, and a matching conjugate curve can be generated by applying the law of gearing [2]. An infinite number of possible profile families exist, each with an infinite number of profiles belonging to the family. This leads to the question, which is the best possible profile geometry? This is one of the main motivations behind undertaking this research for developing and understanding different profiles.

Much work has been done on multi-objective optimization of cycloidal gerotor profiles [5, 6, 12] and asymmetric profiles [11] has been done, but in these papers the main focus on the displacement, flow ripple and contact characteristics of the gerotor. The exploration of different composite profiles has been done by combining cycloidal and elliptical segments [13], cycloidal and circular arc curves [7], cycloidal and involute curves [8] and cycloidal curves with straight lines [3] has been conducted so far. But none of these works have tried to find an optimum pump geometry for high pressures at which the tip leakages become a very important consideration.

This work is motivated by the need to realize gerotor designs that operate reliably at higher pressures and speeds than conventional applications which in this study are considered to be 100 bars and 4000 RPM. A fast gear-generation algorithm is made and run within modeFRONTIER [16] to compute a Pareto set of candidate geometries. The paper first describes the gear-profile parameterization and generation procedure, then details the

optimization framework in modeFRONTIER and the objective functions (OFs). The resulting designs are compared against an a few industry standard pumps that were available, and a subsequent stage of optimization is performed to refine the porting span for each Pareto-optimal geometry. The optimal design is selected from the Pareto front through additional analytical screening of candidate geometries, with primary emphasis on leakage performance. The best-performing design is then subjected to a lumped-parameter analysis to optimize the port geometry while accounting for operating pressure and speed.

2. PROFILE DESCRIPTION

The mixed profile named taken as reference in this work is the kfloid profile morphology illustrated in the work by Liu et al [3] where the author shows that the kfloid morphology performs better at higher pressure in terms of volumetric efficiency. A kfloid half-tooth section can be very simply defined by three sections which are two shifted cycloidal sections connected by a straight line of length L which is by definition a common tangent to the two end points of the shifted cycloids as shown in figure 2.

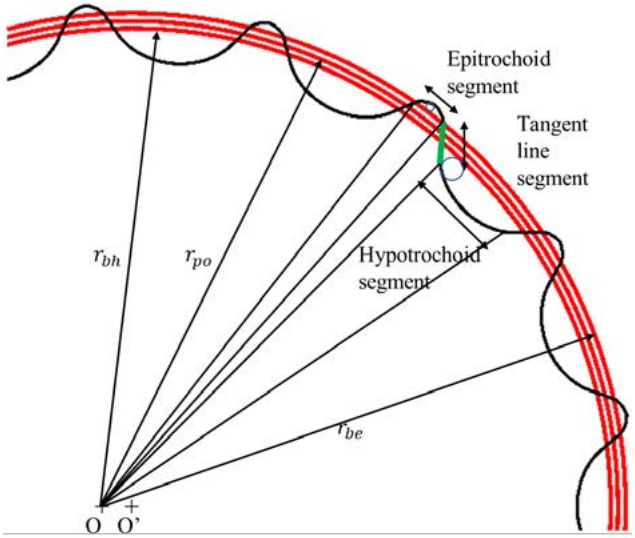


Figure 2: Geometric definition of quantities in a Kfloid profile, adapted from [3]

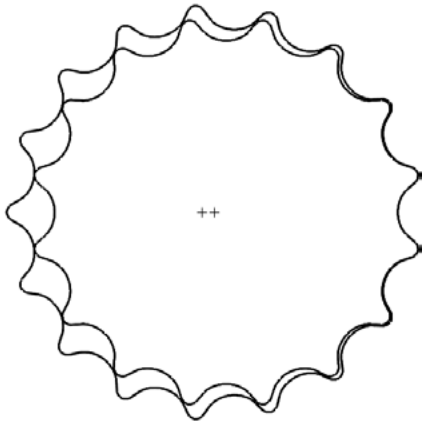


Figure 3: Example of a Kfloid gearset

The parametrization of the profile is done with the sets of uniquely defined parameters mentioned in table 1. This parametrization can be used to define either of the inner or outer gears.

In theory with the application of the law of gearing the generated Kfloid gear can be considered an inner or outer gear, but in this study the generated gear is defined as the outer gear and using the law of gearing the inner gear is generated from the defined outer gear. A resulting Kfloid inner and outer gearset is shown below in figure 3.

Parameter	**Value**	**Unit**
Number of Teeth	17	-
Eccentricity	3.190	mm
d	2.241	mm
b	1.344	mm
r_{po}	54.230	mm
s_{hypo}	0.848	mm
s_{epi}	5.231	mm
$\Delta\theta_1$.	315.573	deg
$\Delta\theta_2$.	287.965	deg

Table 1: Parametrization used to uniquely define the Kfloid gear in figure 3.

3. GEOMETRY CODE

The geometry code implements a numerical procedure for discretizing the pump domain into displacement chambers (DCs) by identifying minimum-distance segments between the inner and outer rotors that uniquely bound each chamber and preserve continuity of the

enclosed fluid volume over a full revolution. This method is based on the approach of Pelligri et al. [14], with targeted modifications to achieve smoother chamber evolution in the vicinity of the dead-volume region. In particular, the minimum-gap criterion is selectively deactivated over a small angular interval, and the DC boundaries are interpolated to provide a continuous and physically consistent transition through this region. The parameters extracted from the geometry code are summarized in Table 2, while Figures 5(a) and 5(b) illustrate the evolution of the DC near the dead volume at shaft angles of 320° and 340°, where the modified interpolation procedure replaces the standard minimum-distance algorithm to ensure a smoother transition.

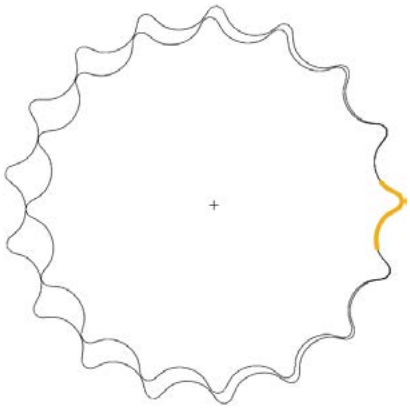


Figure 5 (a): DC definition at an angle of 320^{o} of the outer rotor rotation

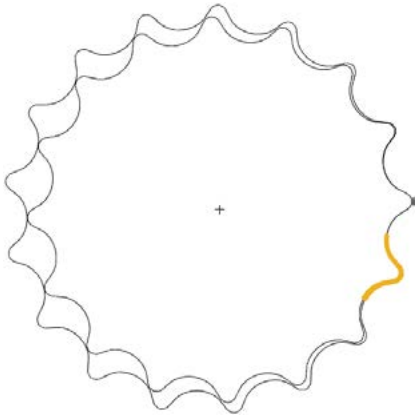


Figure 5 (b): DC definition at an angle of 340^{o} of the outer rotor rotation

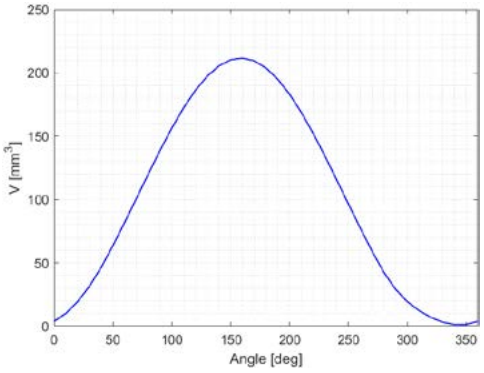


Figure 6: Resulting volume variation with the variation in outer rotor shaft angle

Parameter	Description
V	Displacement Chamber Volume (DCV) [mm^3]
h_t	Clearance gap height [mm]
Ω	DCV projected area [mm^2]
$\frac{dV}{d\theta}$	Rate of change of volume w.r.t rotation [mm^3/deg]
R_{eq}	Equivalent radius of curvature at contact point [mm]
$\hat{n}$	Contact point normal vector
v_x	Contact point velocity [m/s]

Table 2: Parameters obtained from the numerical geometry code

4. OPTIMIZATION METHODOLOGY

The gear generation and optimization framework uses the MOGA-II algorithm [17], which was selected for its suitability in handling multi-objective optimization problems through Pareto-based ranking rather than objective scalarization. This approach provides a broader understanding of the trade-offs between design parameters and helps identify an optimal family of solutions rather than a single design point. The first step of the design optimization was executed for a total of 350 generations, with the initial Design of Experiments (DOE) populated using a constraint-satisfaction sampling method to ensure that only geometrically feasible configurations were evaluated, the DOE table was populated by 100 designs in total. A numerical algorithm is used to numerically discretize the pump domain into isolated displacement chambers by drawing segments along the regions of minimum gap height, allowing for continuous variation in the computed chamber volume V(θ) along with other geometric properties of interest such as the tooth-tip gap height, the projected areas of the DCs and the curvatures of the gears in contact of the DC throughout the rotation for different rotational orientations.

Input Parameter	Min. Value	Max. Value	Unit
Number of Teeth	5	20	-
Eccentricity	0.5	5	mm
d	0.5	4.5	mm
b	0.5	4.5	mm
r_{po}	2.5	100.0	mm
s_{hypo}	0.0	10.0	mm
s_{epi}	0.0	10.0	mm
$\Delta\theta_1$.	300.0	420.0	deg
$\Delta\theta_2$.	240.0	330.0	deg

Table 3: Limits of variation of each geometric input parameter for the optimization

Parameter	Constraint	Value	Unit
Gear Root Radius	Lesser than	45	mm
Connecting Segment Length	Greater than	0	mm
Gear Intersection Area	Lesser than or equal to	0	mm^2
Minimum Chamber Volume	Equal to	10^{-9}	mm^3

Table 4: Constraints applied to each generated gear design for the analytical optimization

The algorithm selects geometric parameters which are described in table 3 within the prescribed design space and scales the pump dimensions to achieve the specified displacement of 3.365 cc/rev, which is maintained constant across all configurations for consistent comparison. The variables in table 3 describes the rolling circle parameters for the inner and outer cycloids namely as the rolling circle radius, base circle shift and angle of rotation for the inner circle as b, $s_{hypo,}$ and $\Delta\theta_1$ as well as the outer circle parameters d, $s_{epi,}$ and $\Delta\theta_2$ and the outer rotor root circle radius is r_{po}. Additional constraints are imposed on the maximum allowable radius and leakage rate to prevent unrealistic or non-physical designs in the analytical optimization along with a minimum chamber volume, these constrains are described in table 4. The optimization proceeds in two stages which are described in more detail further in section 4. The first stage involves analytical screening and ranking of designs based on leakage and stress performance described in the objective functions in section 4.1, while the second stage refines the selected candidate designs using a lumped-parameter simulation of the porting system to maximize both volumetric and hydromechanical efficiencies. Designs violating any geometric constraints described in table 4 are automatically discarded, and the MOGA-II evolutionary operators iteratively evolve the feasible population toward the Pareto-optimal front. This framework is flexible and can accommodate additional objective functions such as minimizing wear or stress

concentrations within the same multi-objective optimization procedure as described later in section 4.

5. OBJECTIVE FUNCTIONS

5.1. Analytical Design optimization

The objective of this optimization is to analytically identify the top-ranked gerotor designs with minimal leakage, so they can be selected from the Pareto front and advanced to the second optimization stage. Figure 7 gives a good overview of the complete optimization procedure. The feasibility of the geometry is obtained based on the constraints in table 4 and the design evaluation is performed by the numerical geometry code which extracts the parameters listed in table 2 in order to calculate the objective functions.

5.1.1. Gerotor sizing

$$OF1: minimize\ r_o \qquad (4.1)$$

For this project the expected delivery flow from the gerotor was fixed as per the requirements of the sponsor along with the width of the pump, a scaling algorithm for the gear radius was applied to ensure that the gerotor displacement is within the specification while the radius of the pump outer rotor r_o is minimized as per the objective function given in equation (4.1) and additionally a constraint on the outer radius was applied to ensure the sizing limits of the pump. The radius is defined as the distance of the farthest point on the outer rotor from the centre of the outer rotor.

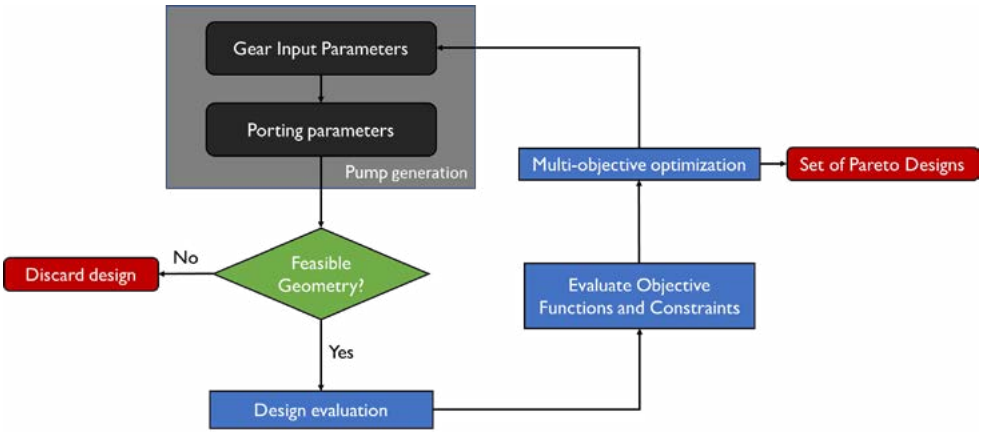


Figure 7: Algorithm for the first analytical step of the optimization

5.1.2. Flow ripple

The kinematic flow of the gerotor is defined as the sum of the outlet flows from all the DC, equations (4.2) and (4.3) describe the delivery flow rate from each DC which is defined as q_i for the ith DC. Whenever any DC is closing the negative of the volume change of the DC is defined as the kinematic delivery flow rate and when the DC is opening the delivery from the chamber is zero. The total kinematic flow (Q) is the sum of the delivery from each DC

as shown in equation (4.2) and the calculation is done assuming an incompressible fluid and no internal leakages within the pump geometry.

$$Q = \sum_i q_i \tag{4.2}$$

$$q_i = \begin{cases} -\frac{dV_i}{d\theta_i} & \frac{dV_i}{d\theta_i} < 0 \\ 0 & \frac{dV_i}{d\theta_i} > 0 \end{cases} \tag{4.3}$$

The flow pulsation and flow power are calculated and used as objectives to be minimized in the optimization, to improve flow pulsation performance and they are calculated as per equations (4.4), (4.5) and (4.6)

$$\bar{Q} = \frac{1}{N}\sum_{i=1}^{N} Q_i \tag{4.4}$$

$$Q_{peak} = \frac{Q_{max} - Q_{min}}{\bar{Q}} \tag{4.5}$$

$$OF2: minimize\ Q_{peak} \tag{4.6}$$

5.1.3. Tooth tip leakage

Tooth tip leakages are the most significant source of pump volumetric losses especially in this case where they dominate the leakage loss. To quantify the tooth-tip leakage for each design the tooth-tip leakage is calculated using the correlation provided by Rituraj et al [4] with the geometrical values obtained from the numerical discretization of the gearset. Here h is the tooth tip gap height and κ is the curvature of the tooth where the displacement chamber leakages are calculated.

$$Q_{tip} = \frac{m}{4\pi}\int_{\theta=\pi-\frac{2\pi}{m}}^{\theta=\pi+\frac{2\pi}{m}} h^{2.5}\kappa^{0.5} d\theta \tag{4.7}$$

$$OF3: minimize\ Q_{tip} \tag{4.8}$$

5.1.4. Adhesive wear

Due to the constant sliding contact that happens between gerotor teeth, quantification of wear experienced by the gerotor and optimizing the design for improved wear resisting performance is something that is important while designing a new gerotor profile. In this study the classical Archard wear model, expressed in Eq. (4.9), which states that the volume of worn material, W is proportional to the frictional work, where F is the contact force, S is the sliding distance, H is the material hardness, and K is an empirical wear coefficient. This model can be normalized with respect to operating pressure, face width, and material properties to yield a dimensionless wear rate as given in Eq. (4.10), where $\bar{F}$ and $\bar{S}$ represent normalized contact force and sliding speed, respectively, p is the operating pressure, and ω_i is the inner gear rotational speed. The normalized sliding speed can also be expressed in terms of the instantaneous velocities of the inner and outer gear contact points, v_i and v_o, and the unit tangent vector $\hat{t}$, as shown in Eq. (4.12) and the outer gear rotation speed ω_o is used to normalize the sliding velocity. The purpose of using the non-dimensional wear equation (4.10) is to express the tendency of wear irrespective of the material that is to be

used for the pump. Equation (4.11) describes the formulation of the force with is calculated using the pressure difference and the projected areas of the gear teeth and is normalized with the pressure difference and gear width. This objective function aims to minimize the maximum normalized wear rate over one full rotation, as defined in Eq. (4.13). A more detailed formulation and discussion is present in previous work of Robison et al [5-6].

$$W = \frac{KFS}{H} \tag{4.9}$$

$$\bar{W} = \bar{F}\bar{S} \tag{4.10}$$

$$\bar{F} = \frac{F}{\Delta p h} = \frac{\Delta p \sqrt{\Omega_x^2 + \Omega_y^2}}{\Delta p h} \tag{4.11}$$

$$\bar{S} = \frac{|v_i \cdot \hat{t} - v_o \cdot \hat{t}|}{\omega_o} \tag{4.12}$$

$$OF4: minimize \left\{ max\, \bar{W}(\theta_i),\ 0 \le \theta_i < \pi - \frac{2\pi}{m-1} \right\} \tag{4.13}$$

5.1.5. Contact stress

When force is transmitted between two curved surfaces, local elastic deformation occurs as the load is distributed over a finite contact region. This deformation generates both normal and shear stresses within the material, collectively referred to as contact stress. The maximum shear stress typically develops below the surface and alternates in sign across the contact plane, which may result in fatigue-induced surface damage such as pitting under repeated relative motion. In the case of a gerotor mechanism, the meshing teeth can be idealized using Hertzian contact theory as two cylinders in contact, with their radii corresponding to the local radii of curvature of the inner and outer gear flanks. The resulting peak contact pressure, p_0, is a key indicator of mechanical durability and can be normalized with respect to material properties and operating pressure as expressed in Eq. (4.14). In this formulation, E^* represents the combined modulus of elasticity defined in Eq. (4.15) as per [15], while κ_i and κ_o denote the curvatures of the inner and outer gear profiles at the point of contact. The Poisson's ratios of the respective gears, ν_i and ν_o, together with their elastic moduli E_i and E_o, determine E^*. The maximum value of p_0 observed over a full revolution is used as the objective function Eq. (4.16), further discussion can be found in earlier studies by Robison et al [5, 6].

$$\overline{p_0} = p_0 \sqrt{\frac{\pi}{E^* \Delta p}} = \sqrt{\bar{F} |\kappa_i + \kappa_o|} \tag{4.14}$$

$$\frac{1}{E^*} = \frac{1-\nu_i^2}{E_i} + \frac{1-\nu_o^2}{E_o} \tag{4.15}$$

$$OF5: minimize \left\{ max\, \overline{p_0}(\theta_i),\ 0 \le \theta_i < 2\pi + \frac{2\pi}{m-1} \right\} \tag{4.16}$$

5.1.6. Lateral gap leakage

Lateral gap leakages in gerotor pumps arise through the thin clearances between the rotor side faces and the pump casing or endplates. Each differential section of a tooth along its axial height can be modelled as a Poiseuille flow between parallel plates, where the pressure

difference acts laterally across the tooth width as is illustrated in figure 8. Assuming constant gap height, viscosity, and pressure difference, the leakage flow can be obtained by integrating the Poiseuille equation over the lateral direction of the tooth for each of the inner and outer gears, leading to a total leakage for each tooth as per equation (4.17), a more detailed description of the equation can be found in [6].

$$\overline{Q_L} = \frac{3Q_L\mu}{2h_z^3\Delta p} = \int_{x_0}^{x_1} \frac{1}{l} dx \qquad (4.17)$$

Where h_z is the lateral gap height, μ is the fluid viscosity, and l is the lateral flow length at a particular section. This expression can be normalized by the pressure difference to emphasize the geometric and fluid property dependencies, showing that longer, narrower teeth permit greater lateral leakage, whereas higher viscosity or improved axial sealing reduces it. Thus, while lateral leakages can theoretically contribute to volumetric losses, they remain negligible in well-sealed pumps with appropriately controlled axial clearances.

$$OF6: minimize\ \overline{Q_{L,i}} + \overline{Q_{L,o}} \qquad (4.18)$$

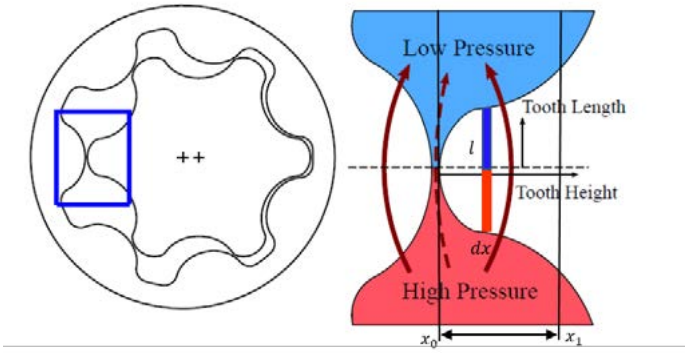
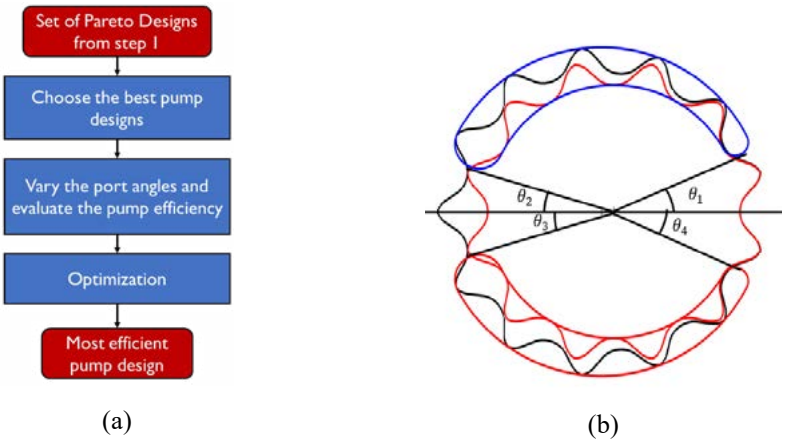


Figure 8: Representation of lateral gap leakage losses

5.2. Porting optimization

The porting optimization only has the objective to maximize the volumetric efficiency for every port angle of the design obtained from the multiple objective analytical optimization, multiple simulations were run where a lumped parameter based method was used to optimize the angular span of the porting considering the efficiency to be maximized in a manner similar to what is shown in figure 9 (a). The angular span of the porting geometry was parametrized as shown in figure 9 (b). To ensure the symmetric geometry for bidirectional operation of the pump the values of the four porting angles can be expressed with the parameters $\hat{\theta}_1$ and $\hat{\theta}_2$ as equations (4.19) and (4.20), while the volumetric efficiency was evaluated using equation (4.21) where Q_{sim} is the outflow predicted from the simulation in one revolution and V_{max}, V_{min} are the maximum and minimum values of the DCV that is predicted. For this study the parameters $\hat{\theta}_1$ and $\hat{\theta}_2$ were set to the range $\in [1, 2]$, while a 2D heat map could have been used to illustrate the performance variation of the pump with porting angle variation an optimizer which seeks the best combination was

considered to be a better alternative. Also, all the simulations for the optimization are performed at a high pressure difference Δp of 100 bar and a shaft speed of 4000 RPM.



(a) (b)

Figure 9: (a) Algorithm for the second step of the optimization and (b) Representation of the angular parametrization of the porting.

$$\theta_1 = \theta_4 = \hat{\theta}_1 \cdot \frac{180}{m} \quad (4.19)$$

$$\theta_2 = \theta_3 = \hat{\theta}_2 \cdot \frac{180}{m} \quad (4.20)$$

$$\eta_{vol} = \frac{Q_{sim}}{mN \cdot (V_{max} - V_{min})} \quad (4.21)$$

6. RESULTS

For the first step optimization, a total of greater than 35000 designs were evaluated out of which there were around 2700 pareto designs. The resulting pareto fronts are plotted as shown in figure 10 and an industry pump whose specification are provided by the sponsor is used as a reference for the comparison of the performance, here the reference pump is represented as a red dot with a black outline to highlight it. In the pareto fronts shown in figure 10 while the pumps with the best leakage performance are chosen from the leakage pareto front. The leakage performance of the new profile is seen to be within the range of the reference profile, but this profile can be seen to have a big advantage particularly when it comes to the wear performance in which the chosen pareto designs have almost an order of magnitude better performance in comparison to the reference cycloidal design. Additionally, design clusters can be observed with respect to the leakage performance which can be attributed to the clusters of similar geometrical designs with small variations in parameters, the designs shown in figures 11 (a) and (b) are two such best performing designs from design clusters having low leakages.

The figure 11 (a) and (b) show the geometry of the pumps chosen with the best leakage characteristics while (c) and (d) shows the maps of the efficiency plotted against the angular span of the ports. While table 5 describes the geometric parameters chosen for each gear design. The best performing design out of which has an efficiency (η_{vol}) of 97.5%.

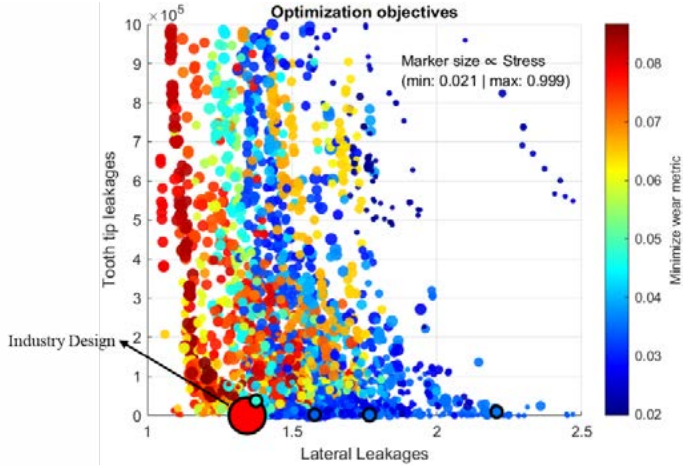


Figure 10: Pareto fronts from the analytical optimization

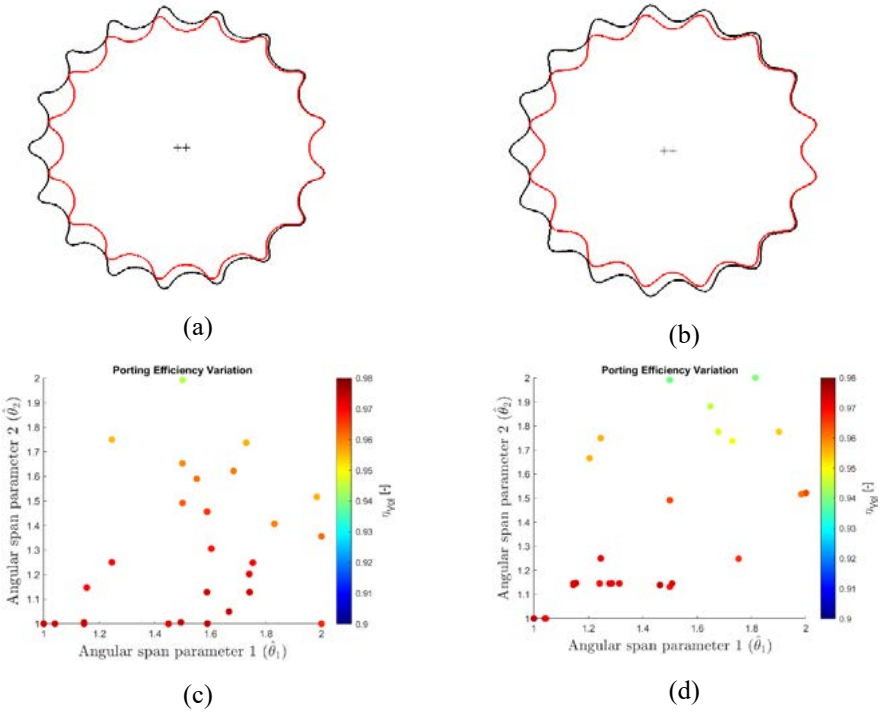


Figure 11: Results of the full simulation optimization for the ports.

Parameter	Value for (a)	Value for (b)	Unit
Number of Teeth	17	17	-
Eccentricity	2.623	3.085	mm
d	1.721	1.177	mm
b	0.977	1.599	mm
r_{po}	22.76	23.347	mm
s_{hypo}	7.955	9.493	mm
s_{epi}	1.977	0.064	mm
$\Delta\theta_1$.	346.33	343.18	deg
$\Delta\theta_2$.	288.45	264.60	deg

Table 5: Parametrization used to uniquely define the Kfloid gear in figures 11 (a) and (b).

7. CONCLUSIONS

A gear generation algorithm was developed to fully define the geometry, size, and other key parameters of a Kfloid-type gerotor pump. The Kfloid profile was subsequently optimized using modeFRONTIER in conjunction with Multics through a two-stage process aimed at maximizing volumetric efficiency for a specified delivery and pump size. A distinct Pareto front was established, and the optimized designs were compared with a reference industrial cycloidal type pump. The results indicate a modest improvement in volumetric efficiency, while the predicted trends suggest significantly better wear resistance and reduced flow ripple for the optimized Kfloid profiles. On comparing the resulting designs from the clusters that were formed as mentioned in section 6 we can see that design (a) has better lateral leakage performance compared to (b) due to the longer tooth geometry and on comparing the stress performance (b) performs better than (a) since geometry (b) has lower curvatures thus resulting in better stress performance.

Future work can focus on exploring mixed gerotor configurations with thinner teeth, similar to the Parafloid, Megafloid, and Geocloid profiles [6,7], with particular attention to hybrid geometries employing involute driving flanks due to their favourable drive characteristics [8].

8. ACKNOWLEDGEMENTS

The authors gratefully acknowledge ESTECO for providing access to the modeFRONTIER multi-objective optimization platform, which was used in this work to perform the numerical optimization studies.

9. NOMENCLATURE

Variable	Definition	Unit
m	Number of teeth of outer gear	-
e	Eccentricity	mm
d	Inner rolling circle radius	mm
b	Outer rolling circle radius	mm
r_{po}	Radius of outer gear base circle	mm
s_{hypo}	Shift in the hypocycloidal base radius	mm
s_{epi}	Shift in the epicycloidal base radius	mm
$\Delta\theta_1$.	Rotation of the rotating circle on the epitrochoid	deg
$\Delta\theta_2$.	Rotation of the rotating circle on the hypotrochoid	deg
L	Length of the connecting common tangent segment	mm
ν	Poissons ratio	-
E	Elastic modulus	MPa
$\hat{\theta}$	Porting span parameter for optimization	deg
κ	Curvature	mm^{-1}
H	Material Hardness	MPa
K	Empirical wear coefficient	$m^3/(Nm)$
S	Sliding distance	mm

10. REFERENCES

[1] Hill, M. F. (1927). Kinematics of Gerotors. The Peter Reilly Company, Philadelphia.

[2] Buckingham, E. (1949). Analytical Mechanics of Gears. Dover Publications, New York.

[3] Liu, H., & Lee, J.-C. (2017). Profile design and volumetric efficiency analysis of gerotor pumps for AWD vehicles. International Journal of Mechanical and Production Engineering, 5(4).

[4] Rituraj, F., & Vacca, A. (2020). An investigation of tooth tip leakages in gerotor pumps: modeling and experimental validation. Journal of Verification, Validation and Uncertainty Quantification, 5(1), 011001.

[5] Robison, A., & Vacca, A. (2018). Multi-objective optimization of circular-toothed gerotors for kinematics and wear by genetic algorithm. Mechanism and Machine Theory, 128, 150–168.

[6] Robison, A. J., & Vacca, A. (2021). Performance comparison of epitrochoidal, hypotrochoidal, and cycloidal gerotor gear profiles. Mechanism and Machine Theory, 158, 104228.
[7] Arinaga, S., et al. (2016). The latest trends in oil pump rotors for automobiles. SEI Technical Review, 82, 59–65.
[8] Choi, T. H., et al. (2012). Design of rotor for internal gear pump using cycloid and circular-arc curves. Journal of Mechanical Design, 011005.
[9] Jung, S. Y., Kim, M. S., Cho, H. Y., & Kim, C. (2009). Development of an automated design system for oil pumps with multiple profiles (circle, ellipse, and involute). Journal of the Korean Society for Precision Engineering, 26(3), 103–112.
[10] Nishida, Y., et al. (2018). Development of continuously variable discharge oil pump. SAE Technical Paper 2018-01-0932.
[11] Robison, A. J., & Vacca, A. (2020). Design of Asymmetric Gerotor Pumps. In International Scientific-Technical Conference on Hydraulic and Pneumatic Drives and Control (pp. 90-101). Cham: Springer International Publishing.
[12] Pareja-Corcho, J. C., Bartoň, M., Pedrera-Busselo, A., Mejia-Parra, D., Moreno, A., & Posada, J. (2024, August). On shape design and optimization of gerotor pumps. In Computer Graphics Forum (Vol. 43, No. 5, p. e15140).
[13] Jung, S. Y., Bae, J. H., Kim, M. S., & Kim, C. (2011). Development of a new gerotor for oil pumps with multiple profiles. International Journal of Precision Engineering and Manufacturing, 12(5), 835-841.
[14] Pellegri, M., Vacca, A., Devendran, R. S., Dautry, E., & Ginsberg, B. (2016). A Lumped parameter approach for gerotor pumps: Model formulation and experimental validation.
[15] Johnson, K. L. (1985). Contact Mechanics. Cambridge University Press.
[16] ESTECO. (2023). modeFRONTIER: Multidisciplinary Design Optimization and Process Integration Platform. ESTECO S.p.A., Trieste, Italy.
[17] Shah, K. Y., Cumberbatch, H., & Etchart, L. Optimization Techniques using modeFRONTIER-Process Review.

Beyond Pilot Control: A Prototype for Direct Actuation of Large Industrial Proportional Valves

Author: Tobias Vonderbank[1], Prof. Katharina Schmitz[1]

[1]*Institute for Fluid Power Drives and Systems(ifas), RWTH Aachen University, Aachen, Germany*

Corresponding Email: tobias.vonderbank@ifas.rwth-aachen.de

Abstract.

Recent advancements in electric drive technology are enabling new applications in hydraulics, particularly in valve actuation. New electric actuators based on stepper or BLDC motors offer higher forces and virtually unlimited stroke lengths, making them viable alternatives to conventional solenoids, especially for large hydraulic valves that have traditionally required hydraulic pilot systems. These actuators have already been successfully utilized in mobile machinery, where they enable the direct actuation of large valves, resulting in a simplified system architecture and improved efficiency. However, in stationary hydraulics, direct electromechanical actuation nowadays is limited to valves of NS10 due to the high forces required. Whereby the flow and spring forces constitute the most significant share of the total actuation force. This work investigates strategies for reducing flow forces to enable direct actuation of NS16 proportional valves with the actuators mentioned before, based on electric motors. Based on theoretical analysis, design modifications on the spool and housing/sleeve for flow force reduction are identified and evaluated using CFD simulations and measurements. The resulting optimized valve design significantly reduces actuation force requirements, providing a feasible pathway for direct electromechanical actuation of larger industrial valves and eliminating the need for pilot stages.

Keywords. Pilot Controlled Valves, Direct Valve Actuation, Flow Force Reduction, Experimental, CFD-Simulation

1. INTRODUCTION

The use of stepper motors for controlling hydraulic valves in mobile machinery has been explored since the 1990s. Started when the hydraulic pilot control of a tractor's lift valve was replaced by an electromechanical actuator based on a stepper motor back in 1995 [1]. Followed by the direct actuation of the front loader valve in 2000 [2]. One of the key motivations for this research was the potential to simplify the overall system by eliminating the separate pilot-control circuit that had been necessary up to that point [2]. Additionally, other significant advantages of stepper motor-based valve actuators were emphasized, including their high rigidity and force capabilities, as well as their ability to achieve stable and precise positioning [1].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

While the application of stepper motors was reported only sporadically in the 1990s and subsequent years, there has been a notable increase in activity over the last ten years regarding stepper motor-based electromechanical valve actuators. These developments have primarily been driven by the company SONCEBOZ, which has been publishing articles on the subject since 2014, presenting the latest research findings on their S42 and S40 actuators. That this research addresses market demands is evident, as Sonceboz is not the only company active in this field. For example, there was another actuator based on an electric motor presented by THOMAS MAGNETE at the 9th IFK. Both actuators the S40 by SONCEBOZ and the EMA by THOMAS MAGNETE are shown in **Figure 1**. [3]–[8]



Figure 1: S40 Actuator by Sonceboz [4], [9] and EMA-Actuator by Thomas Magnete [8]

Although the discussion about motor-based valve actuators is primarily led by individual companies that are active in the field of electromechanical actuation, while focusing their research on their specific actuators to enhance performance and extend their limits further, there has also been a growing momentum in university research in recent years. In contrast to the aforementioned company-led research, university studies focus on decreasing the necessary actuation forces [10]–[17] or fundamental analyses of pilot-controlled valves aiming to optimize their performance [18]–[20].

Overall, it is evident that direct control of large valves has been increasingly discussed in research and application for about a decade, and the first market-ready actuators are now available [4]–[6]. Compared to solenoid magnets, they are characterized by higher forces and a significantly longer stroke. Performance data of these actuators can be looked up at the individual homepages or in [8]. However, current in-field applications are mostly limited to the field of mobile machinery, where they have already been introduced in some product series on the market to directly control valves with flow rates of up to 100 l/min, see e.g. the LCV from Bucher Hydraulics [5]. In this case the actuators are used in combination with a pressure balance valve, which means that the differential pressure across the valve is always constant, limiting the maximum flow rate to the nominal flow rate chosen during the design process.

Even though the disadvantages of hydraulic pilot control, such as the sensitivity of the position regulation quality of pilot-operated valves to pressure fluctuations, contamination, and low temperatures, are particularly significant in the field of mobile machinery, it seems obvious to use these actuators also in industrial hydraulics to push the limits of direct pilot control above NS10. In contrast to applications with pressure balancing valves commonly

used in the field of mobile machinery, the flow rates in industrial valves can be much greater than the nominal flow rate, as a pressure balancing valve does not control the pressure difference over the valve. While valves of NS10 in the standard market range have a minimum nominal flow rate of around 30 l/min and a maximum flow rate of 75 l/min (see **Figure 2**), the flow rate increases to around 100 l/min and 460 l/min, respectively, for the next larger NS16 valves (see **Figure 3**). Consequently, the flow rates controlled by NS10 valves are smaller compared to those already achieved in known mobile applications. For NS16 valves, on the other hand, although the nominal flow rates remain comparable, the maximum flow rate can exceed 100 l/min by four times.

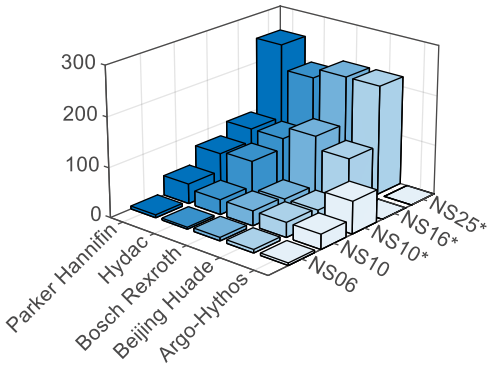


Figure 2: Nominal flow rate of proportional valves of different nominal sizes and manufacturers

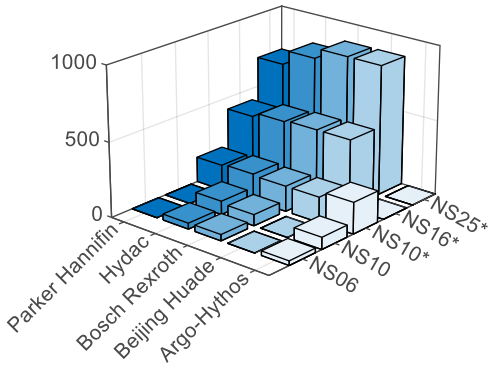


Figure 3: Maximum flow rate of proportional valves of different nominal sizes and manufacturers

In most cases, an increase in flow rate and valve size results in higher stroke and actuating force, unless the valve design has been specifically optimized to counteract these effects. For valves NS16, this typically results in forces and strokes that exceed the capabilities of conventional solenoids, which is why they are generally pilot-operated. Together with its industrial partner KMS Stoßdämpfer GmbH, ifas is currently working to demonstrate that the availability of previously presented actuators based on stepper or BLDC motors has opened up further potential for the direct control of larger valves in the field of industrial hydraulics. This is the first step towards shifting the application limit for direct control from NS10 to NS16 and making corresponding valves available on the market, thereby unlocking the associated potential for system optimization and performance enhancement.

This contribution not only aims to develop the initial step towards a directly controlled NS16 industrial valve but also employs systematic analysis to demonstrate how flow forces on spool valves can be reduced through appropriate geometric optimization. These findings can be applied to other valves from various manufacturers and in different sizes, facilitating a general evaluation of the feasibility of direct electromechanical control for larger valves using actuators based on stepper or BLDC motors.

2. NEW DIRECT DRIVEN PROPORTIONAL VALVE NS16

A central aspect of the project was the development of an optimized NS16 hydraulic main stage, characterized by low operating forces. To this end, the first step was to develop a functional hydraulic main stage with a nominal flow rate of 125 l/min and a maximum flow

rate of 400 l/min achieved at a pressure difference of max 70 bar over the metering notch. To keep the radial flow forces as low as possible and at the same time enable easy adjustment of the internal geometry, the valve was designed as a sleeve valve. This allows systematic investigation of different geometries for the spool and sleeve to demonstrate the potential for reducing flow forces. A cross-sectional view of the prototype is shown in **Figure 4**.

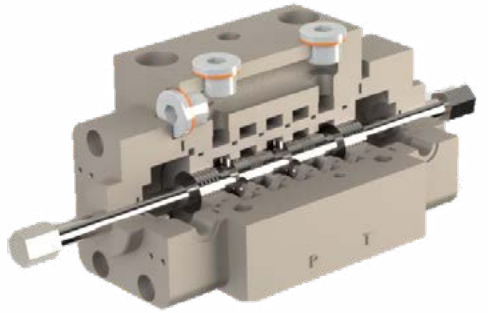


Figure 4: CAD of the developed direct driven main stage

2.1. Force Requirements

Since the valve cannot be actuated if the generable forces are insufficient, achieving an appropriate actuating force is a fundamental requirement for the implementation of direct electromechanical actuation. To determine the necessary actuating force, all forces acting on the spool must be analyzed. **Figure 5** shows a corresponding spool with all the forces acting on it. These are usually the spring forces F_C, friction forces F_R, flow forces F_{Flow}, and the actuating force F_{Act}.

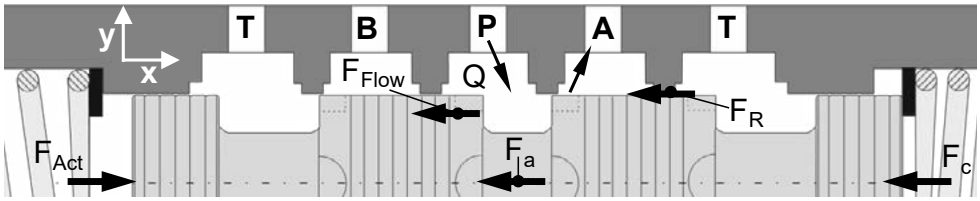


Figure 5: Forces acting on the slide of a globe valve according to [21]–[23]

Within the illustration, the spool is assumed to be deflected in the positive x-direction, with positive velocity and acceleration. Friction (F_R), inertial forces (F_a) and spring forces (F_C) therefore act in the negative x-direction, and the actuating force (F_{Act}) exceeds the sum of all these opposing forces. Since this is a standard design without any flow force compensation geometries, it can be assumed that the flow forces (F_{Flow}) act in the closing direction and thus in the negative x-direction. In this case, the force balance on the spool can be formulated according to **Equation 2.1**.

$$F_{Act} = F_a + F_C + F_R + F_{Flow} \tag{2.1}$$

To keep the required force as low as possible, all opposing forces must be minimized. However, each one can only be reduced to a limited extent due to their causes or functions.

2.2. Inertia Forces

Since inertial forces are defined by mass m and acceleration $\ddot{x}$ for a given acceleration profile, they can only be reduced by decreasing the mass. However, this is only possible to a limited extent, as both the shape and material of the spool cannot be chosen arbitrarily. For the valve considered in this contribution, the inertia force was estimated assuming that the spool must travel the full stroke of 7.5 mm within 50 ms, accelerating until the midpoint of the stroke followed by a subsequent constant deceleration. With a mass of 300 g this leads to an inertia force of 18 N.

2.3. Frictional Forces

In the context of pilot-operated valves, friction forces typically refer to those forces acting between the spool and the housing or sleeve. According to common models, these forces consist of a Newtonian and a Coulombian component. They always oppose the direction of movement and thus have a damping effect on the valve spool motion. The choice of material, the fit between the spool and the sleeve, and the radial flow forces influences friction. A careful trade-off must be made in terms of this fit. It must be tight enough to prevent excessive leakage but not so tight that frictional forces become too large.

2.4. Flow Forces

Flow forces are generally defined as the sum of the forces caused by the flow of fluid. These are usually forces resulting from uneven pressure distribution, viscous friction forces, and inertia and acceleration forces of the fluid [24], [25]. The law of conservation of momentum is usually used to derive a usable equation**.** Thereby, a control volume defined in the flow area between the spool, and the housing is used to calculate the flow force, see **Figure 6**. At the selected control volume, the flow Q enters the control volume at the P port at a velocity $\vec{v}_1$ at an angle of ε_1 and leaves the system at a velocity $\vec{v}_2$ at an angle of ε_2 at port A.

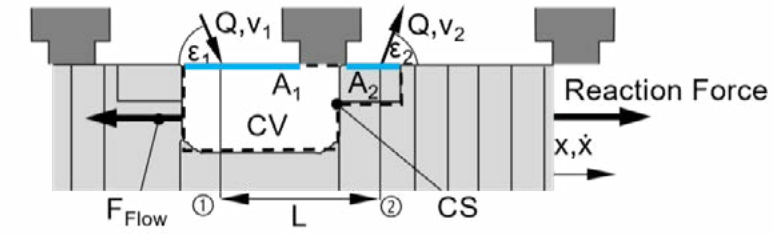


Figure 6: Derivation F_{Flow} using the law of conservation of momentum [22], [26]–[29]

Using Reynolds' transport theorem, the equation for the flow forces can be written according to **Equation 2.2**.

$$\sum \vec{F} = \frac{d}{dt}\left(\int_{CV} (\rho \cdot \vec{v}) dV\right) + \int_{CS} \rho \cdot \vec{v}(\vec{v} \cdot \vec{n}) dA \qquad (2.2)$$

The flow forces are derived analytically from the temporal derivative of the integral of density ρ and the velocity vector $\vec{v}$ over the volume V of the control volume CV, added to the integral of the product of density ρ, velocity vector $\vec{v}$ and the scalar product of velocity vector $\vec{v}$ with direction vector $\vec{n}$ integrated over the area A of the whole control surface (CS). The first part of the equation describes the unsteady flow force and thus forces related to the

inertia and acceleration of the fluid, while the second term describes the steady state flow forces. While steady state axial flow forces have been well researched today, the number of studies dealing with the unsteady flow force is considerably lower. However, these studies show that the unsteady flow force for slow and moderately switching valves is considerably lower than the steady state flow forces. The assumption of quasi-steady state that is usually made, therefore, can generally be considered valid [30].

2.4.1. Axial Flow Forces

Using the orifice equation and assuming constant fluid density, ideal rotational symmetry, sharp metering edges, and a negligible clearance between the spool and the sleeve, the equation for the steady-state axial flow force for one metering edge can be expressed as follows [27], [31].

$$F_{Flow,\mathrm{axial,stat}} = 2\alpha_D^2 \cdot \pi \cdot d \cdot x \cdot \Delta p \cdot \frac{\cos(\varepsilon_1)}{\sin(\varepsilon_1)} \tag{2.3}$$

While measuring the axial flow forces, frictional forces typically superimpose on them. As a result, plotting the measured force against position usually produces a hysteresis loop (see **Figure 7**).

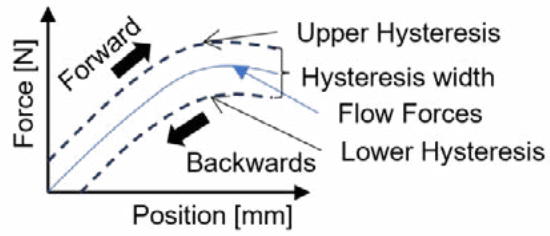


Figure 7: Illustration of measured forces: Hysteresis and flow forces

Assuming the frictional forces are equal in both the opening and closing directions, the axial flow forces are centered within the hysteresis, with half the width of the hysteresis corresponding to the frictional force. [21], [22], [27], [32]

2.4.2. Radial Flow Forces

For the radial flow forces, Equation 2.2 must be integrated over the circumference of the spool. With an ideally rotationally symmetrical geometry and an ideally rotationally symmetrical flow, these forces would completely compensate each other, resulting in a total net force of 0 N. In reality, however, perfect rotational symmetry of geometry and flow profile is rarely achieved, so that radial flow forces usually still occur. An increase in the nominal forces acting on the spool results in higher Coulomb friction forces; consequently, variations in the hysteresis width may reflect changes in the radial flow force [32].

2.4.3. Spring Forces

For 4/3-way valves, the rest position defined by the springs can be either one of the two switched positions (A or B) or the center position (0). In Figure 5, a valve with a blocked center position is shown, in which both springs push the spool into the center position. Since

a secure rest position must always be ensured and reached within a specified time, the springs must be designed accordingly. In the case of pilot-controlled valves, the required actuating force usually does not need to be considered, as forces in the kN range can typically be achieved. Therefore, in proportional valves, springs are chosen to have high stiffness in order to achieve high accuracy within the control loop [33].

However, stepper motors are inherently capable of moving to a defined position. Even when operated in a closed-loop position control system via the motor controller to improve accuracy, spring stiffness has only a minor influence on position control. Given the limited forces available from electromechanical actuators, it is therefore advisable to design the spring with the lowest possible stiffness while selecting the highest possible preload force. Selecting a suitable spring within the available installation space presents a separate optimization problem. Nevertheless, to achieve the highest possible spring force and thus minimize return time while considering the limited actuation forces, all other opposing forces must be kept as low as possible, which is why flow forces need to be reduced.

3. PRINCIPLES FOR FLOW FORCE REDUCTION

The reduction of flow forces by optimizing spool geometries is a widely researched topic and is regularly discussed in the technical literature [10], [13], [15]–[17], [34]–[38]. However, most studies focus on servovalves, with optimizations that are difficult to adapt to proportional valves with metering notches. Therefore, this contribution investigates geometric optimization methods for reducing flow forces that can be applied to proportional valves with metering notches. Based on theoretical, analytical, and literature-based considerations [11], [12], [14], [16], [35], [37], [39]–[41], possible geometric optimizations for the spool and sleeve were identified. Their potential to enable direct actuation by electromechanical actuators through the reduction of flow forces is investigated in a systematic study on the NS16 main stage developed as part of this work. For the evaluation of this potential, a non-optimized proportional valve, shown in **Figure 8**, serves as the reference. The geometric optimizations investigated are illustrated in **Figure 9**.

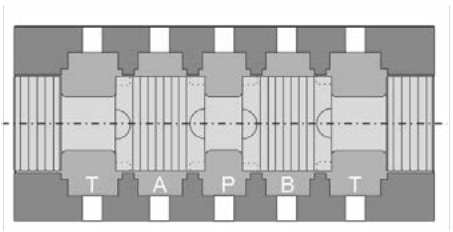


Figure 8: Spool without features

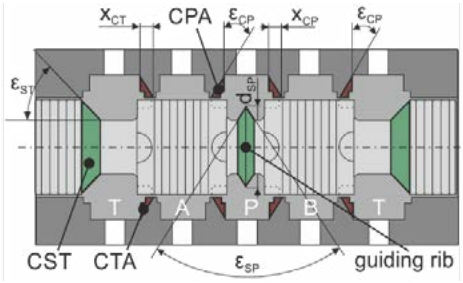


Figure 9: Spool with features

Compared to the non-optimized geometry, Figure 9 shows areas in green where additional material needs to be added, and areas in orange where material must be removed. For the sleeve, adding a chamfer to the partition between the valve chambers was identified as a

promising way to achieve sufficient flow guidance and thereby reduce flow forces. Two geometric parameters are required to fully define this chamfer the angle ε and the chamfer length x. At the Port P, it is referred to as chamfer CPA and defined by the parameters x_{CP} and ε_{CP}. At the Port T, it is referred to as chamfer CTA and is defined by x_{CT} and ε_{CT}. For the spool, a centrally located guiding rib at the geometric center and an additional chamfer at the outer shoulder of the T-Port were identified as possible optimizations for reducing flow forces. The guiding rib can be fully described by d_{SP} and ε_{SP}, while the chamfer CST is characterized by the angle ε_{ST}. Whereas the guiding rib and chamfer CST are established features used in prior studies and industrial applications to reduce flow forces (see [11], [12], [14], [16], [35], [37], [40], [41]), the sleeve chamfers represent an optimization derived from theoretical considerations.

In order to achieve the most efficient reduction in flow forces based on the reference geometry, the influence of the geometric features was analyzed in a three-step process consisting of two simulation steps carried out with ANSYS Fluent CFD-Software and one experimental step. The first step involved a statistical analysis whereby a parameterized CAD model was imported into the simulation software, and simulation points were generated and solved using an optimally space-filling design. Response surfaces were then derived from these simulations to assess the influence of each feature. Subsequently, each feature was simulated individually to determine suitable parameters. As a result, four valve geometries incorporating the identified measures were manufactured and tested experimentally.

To keep the simulation effort manageable, it was important to minimize the computational domain in the CFD simulations. Therefore, the valve was divided into two separate models, each examined individually. The first model, system P, includes the P-port and one working port (A), allowing investigation of flow from P to A. The second model, system T, includes the other working port (B) and the T-port, enabling analysis of flow from B to T. This division made it possible to specifically study both flow scenarios and to investigate the features separately, while also reducing computational complexity.

3.1. Validation of CFD Simulation

To validate the CFD simulations, the reference geometry was first manufactured and measured. The ifas has a measuring device and a test bench available for this purpose. The measuring device consists out of a stepper motor, whose rotational motion is converted into linear motion via a threaded rod and nut assembly with a pitch of 3 mm. A force sensor is integrated into the connection between the actuation unit and the spool to measure the forces which are acting on the spool. To minimize the influence of transverse forces on the measurement, a combination of clevis and head joints was implemented. Additionally, prior to the measurement, the device was carefully aligned with the valve spool. Therefore, the actuation mechanism was moved over the full stroke to ensure that resulting friction forces were as low as possible. Additionally, a position sensor was incorporated to monitor the position of the spool. A similar actuation device also designed by ifas has been described in [42].

The circuit diagram of the test rig can be found in [43]–[45]. While this configuration enables measurement of the forces acting on the spool during flow from the pressure port

(P) to either of the working ports (A or B), as well as during overall flow through the valve (PABT/PBAT), it does not permit individual measurement of flow from either working port (A or B) to the tank port (T). Consequently, separate analysis of the return flow paths (AT or BT) is not possible.

For the reference geometry, the target flow rate of 400 l/min was achieved at a pressure difference of 60 bar, while at a pressure difference of 5 bar, a flow rate of 125 l/min was achieved. In line with theoretical expectations, an increase in pressure difference resulted in a corresponding increase in flow force. Consequently, operation at a pressure difference of 60 bar led to the highest observed flow forces within the pressure range (5 – 60 bar). The results of the simulations and measurements for a pressure difference of 60 bar for the system P (Flow from P to A) of the reference geometry are shown in **Figure 10**. Within the simulation the mesh consists of around 13 Mio Cells with local refinement within relevant areas like the metering notches or the inlet and outlet. The average orthogonality was 0.9662 and the skewness 0.03381.

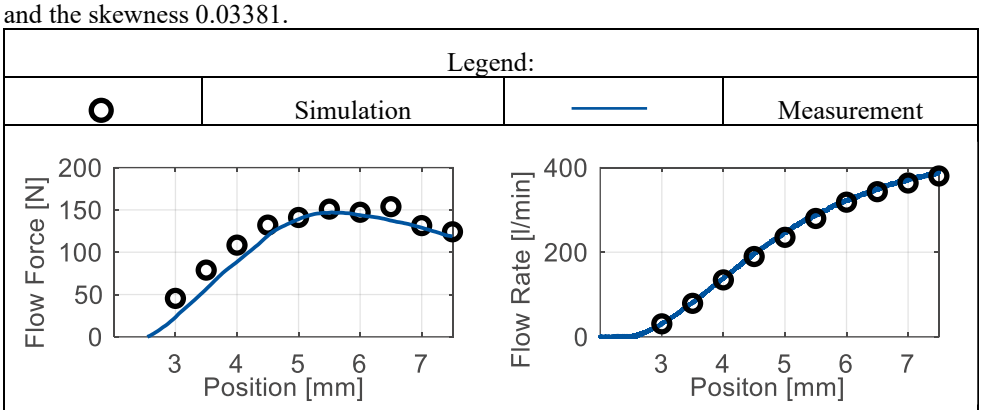


Figure 10: Comparison Simulation Measurement for Flow Rate and Flow Force

As can be seen in the graphs, the simulation results correspond very well with the measurement data. The CFD simulation can therefore be used to generate reliable results in terms of both flow rate and flow forces for the considered system P of the reference geometry.

3.2. Statistical Evaluation of Principles (CFD-Simulation)

The following **Figure 11** to Figure 14 show examples of the response surfaces for the guiding rib feature, illustrating the influence of the parameters diameter d_{SP} and angle ε_{SP}. Since both the flow force and the flow rate must be considered over the entire range of spool positions x, the visualizations are presented as three-dimensional characteristic maps. The resulting flow force F_{Flow} and flow rate Q are plotted on the Z-axis, spool position x on the Y-axis, and the value of each geometric feature under investigation (e.g., d_{SP} and ε_{SP}) on the X-axis. This form of representation enables a detailed analysis of how spool position and geometric features interact to affect flow force and flow rate.

As shown in **Figure 11**, the guiding rib is particularly effective for large valve openings (X) and large diameters (d_{SP}). For diameters greater than 20 mm, even negative forces are achievable, which means that the system is overcompensated and the flow force is acting in opening direction. However, selecting a large diameter d_{SP}, may negatively affect the flow

rate characteristic curve. This can be seen in **Figure 12**, where the flow rate drops for diameters larger than 20 mm. By appropriately selecting the diameter d_{SP}, it is possible to reduce the flow force without adversely impacting the flow rate characteristic for example with diameters of maximum 20 mm. **Figure 13** and Figure 14 show the response surfaces for the angle ε_{SP} of the guiding rib. As can be seen the angle ε_{SP} of the guiding rib has a relatively minor influence on the flow force. Angles around 100° result in a maximum flow force, whereas both smaller and larger angles have a slightly positive effect by reducing the flow force. At the same time, the impact of angle ε_{SP} on the flow rate characteristics is very small.

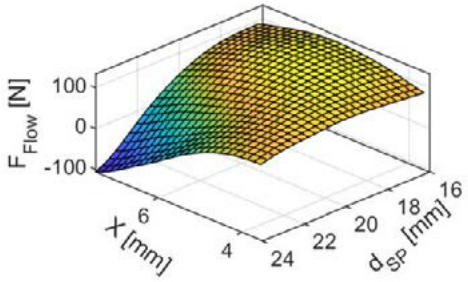


Figure 11:Response Surface F_{Flow}-X-d_{SP}

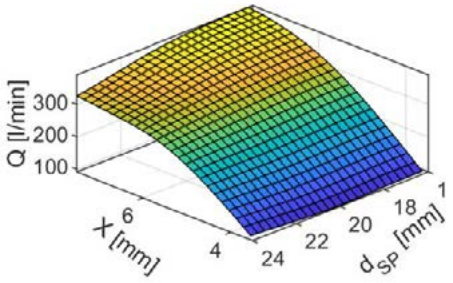


Figure 12: Response Surface Q-X-d_{SP}

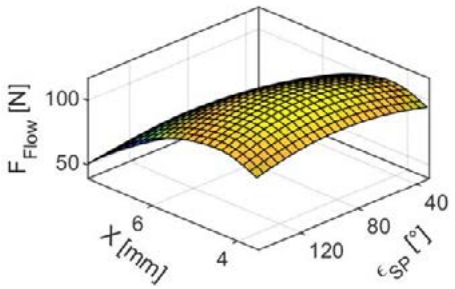


Figure 13: Response Surface F_{Flow}-X-ε_{SP}

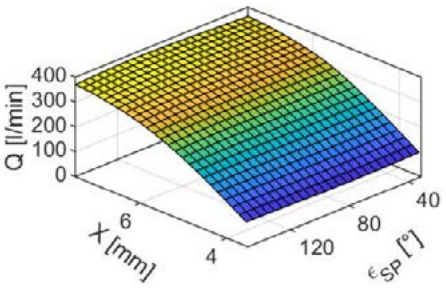


Figure 14: Response Surface Q-X-ε_{SP}

3.3. Simulation of the various concepts

In statistical analysis, the influence of individual features can be evaluated quickly and efficiently. However, assessing combined effects is significantly more complex. After it was proven that all features can have a positive effect, their influence was systematically analyzed in a second step to identify optimal design parameters. As the key objective of this investigation was to identify the potential of each individual design feature for reducing flow forces, a targeted exploration of the design features was preferred over deriving a statistical model in the following. This approach enables a direct and transparent assessment of each feature's contribution to flow force reduction. Initial parameters for each feature were selected based on the prior results, and then individual parameters were systematically varied to identify suitable values for each feature.

From the results, two promising configurations for both the spool and the sleeve were selected, each with and without optimization features. This allowed four different valve

configurations to be formed. $Spool_{Ref}$-$Sleeve_{Ref}$ without any feature, $Spool_{Opt}$-$Sleeve_{Ref}$ with features in the spool, $Spool_{Ref}$-$Sleeve_{Opt}$ with features in the sleeve and $Spool_{Opt}$-$Sleeve_{Opt}$ with features applied to both the spool and the sleeve. By the different combinations, the effectiveness of each feature could be examined separately and in combination. The selected geometric values were as follows. The angles ε_{SP}, ε_{ST}, ε_{CP} and ε_{CT} were set to 45°, the removal depths $X_{CP\ and}$ X_{CT} to 5 mm and the diameter d_{SP} to 22 mm.

Simulation results for these geometries at a pressure level of 60 bar are presented in **Figure 15** and **Figure 16** (system P) and **Figure 17** and **Figure 18** (system T). The simulation results of system P show that all features are effective in reducing flow forces. However, the guiding rib feature can negatively affect the flow rate characteristic, as the flow rate decreases for strokes greater than 5 mm. This negative effect can be avoided when the guiding rib is combined with the chamfer in the sleeve.

For the system T, it can also be concluded that all features have a positive effect, with no negative impact on the flow rate characteristic curve. Notably, in the $Spool_{Opt}$-$Sleeve_{Ref}$ and $Spool_{Opt}$-$Sleeve_{Opt}$ configurations, the chamfer at the spool leads to an opposing flow force for strokes greater than 4 mm respectively 6.5 mm, resulting in overcompensation.

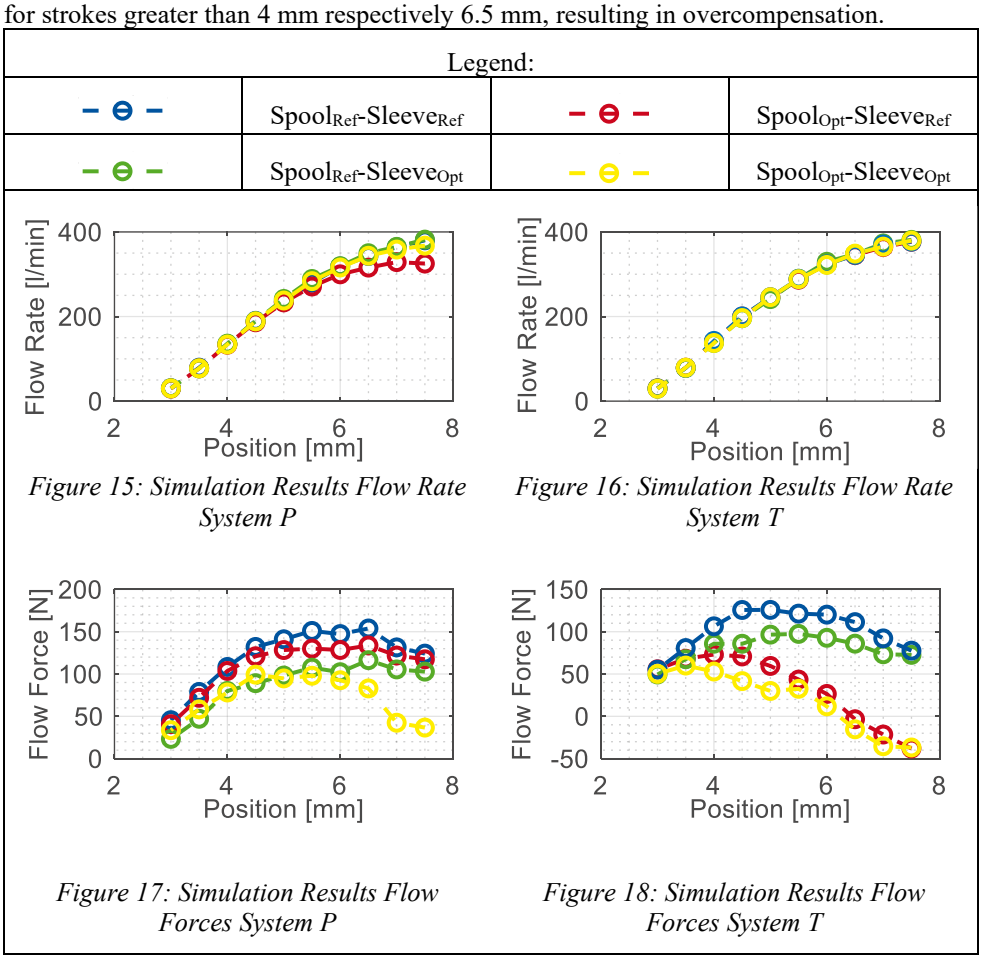


Figure 15: Simulation Results Flow Rate System P

Figure 16: Simulation Results Flow Rate System T

Figure 17: Simulation Results Flow Forces System P

Figure 18: Simulation Results Flow Forces System T

4. EXPERIMENTAL EVALUATION OF COMBINED PRINCIPLES

Once the effectiveness of the features had been identified through simulations, all versions were evaluated experimentally. To this end, the two spool and sleeve variants were manufactured and measured in their respective combinations. Measurements were conducted for system P, where flow occurs from P to A, as well as for the combined system PT, where flow occurs from P to B and from A to T, using the available test bench. Unfortunately, it was not possible to measure system T separately due to limitations of the existing test rig. Therefore, only the simulations result of system A can directly be compared with measurement data. The simulations result of system T can only be analyzed using the measurements of system P and PT, as by comparison between both systems the effectiveness of the features of system T can be identified. The measurement results are shown in **Figure 19** to **Figure 22.**

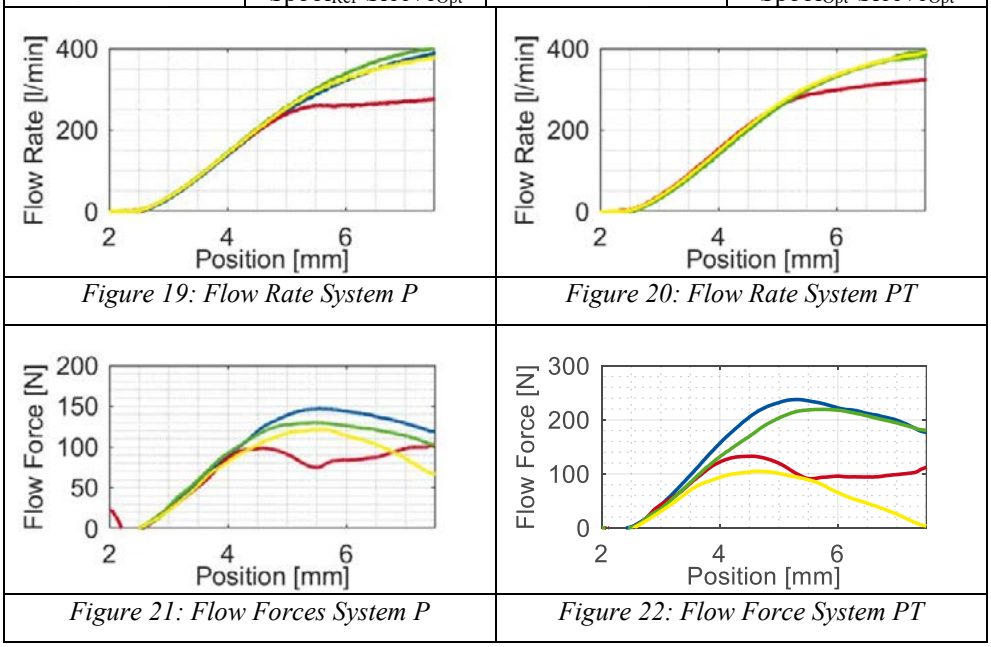


Figure 19: Flow Rate System P

Figure 20: Flow Rate System PT

Figure 21: Flow Forces System P

Figure 22: Flow Force System PT

The measurement data show that the configuration without optimization features exhibits the highest flow forces, and that applying the features produces the expected effects, as anticipated from the CFD simulations. Implementing the features on the spool alone leads to a reduction in flow force; however, this is accompanied by a decrease in flow rate, as assumed from the simulation whereby the impact is underestimated within the simulation. Within simulations conducted at a pressure difference of 70 bar, the effect was considerably more pronounced. It remains questionable whether the observed reduction in flow force is genuinely due to the optimization or simply a result of the lower flow rate. Optimizations to the sleeve show a positive effect, with a reduction of 12 %, although this is significantly less

than predicted by simulation. The same applies to the combination of both measures. In this case, flow forces were reduced by 18 %, indicating a clear positive effect but still falling short of the optimistic simulation results. When only the system P is flowed through, the proposed features lead to a reduction in flow force of 18 %. When both PB and AT are evenly flowed through, an even greater reduction of 65 % is achieved.

Since simulations appear to overestimate the potential of sleeve optimization, further analysis is required to determine the reason for this discrepancy and improve the simulation models. In this study, CFD was employed as a design tool to efficiently investigate the effects of various geometric features on flow forces. By systematically screening a large number of design variants through simulation, it was possible to identify promising designs for experimental validation without having to manufacture each individual parameter set. This approach allowed for a significant reduction in the number of required measurements and enabled efficient identification of improved geometries. Consequently, further refinement of the simulation models be desirable for future valve improvements, however such optimization of the simulation model lies beyond the scope of this contribution

5. SUMMARY AND OUTLOOK

Over the past ten years, new actuators based on stepper and BLDC motors have been tested for use in mobile hydraulics and have also been introduced to the market in some products. Compared to conventional solenoids, these actuators are characterized by higher forces and strokes, making them a viable alternative to hydraulic pilot control. Direct control offers several advantages over hydraulic pilot control, especially in mobile machines where highly fluctuating loads, operation with cold fluid during startup, and heavy fluid contamination are common conditions. Due to their increased forces and strokes, these actuators also have potential applications in stationary hydraulics. In this context, they could help shift the system limit for direct control from NS10 to NS16.

However, the maximum possible flow rates in stationary hydraulics are considerably greater than those in mobile applications, which increases demands on the actuator. While spring forces should be minimized overall, maximizing them within the limits of the actuator's maximum capacity helps to minimize reset times and enhance system safety. Therefore, reducing flow forces is essential.

Within the scope of this contribution, various flow force compensation design features that can be implemented on conventional proportional spools and sleeves/housings were identified, analyzed and investigated by simulations and measurements for their potential to reduce flow forces. Overall, flow forces could be reduced by 18 % percent compared to a reference spool without optimization measures. In particular, the chamfer in the sleeve (ε_{CP}, x_{CP}) and the additional angle ε_{ST} on the spool demonstrated high potential for reducing flow forces while having only a minor impact on the flow rate characteristic curve. The addition of a guiding rib also resulted in a reduction in flow forces, however, if not correctly designed, it can negatively affect the flow rate characteristic curve.

The reduction in flow forces achieved helps to keep the total required actuating force as low as possible. Depending on the selected spring, the necessary actuating forces may be matched to the maximum output of modern actuators. Therefore, reducing flow force is a decisive step toward enabling direct electromechanical control of large valves.

6. REFERENCES

[1] Max-Eyth-Gesellschaft für Agrartechnik: 'Landtechnik 1995 - Tagung Braunschweig, 12. und 13. Oktober 1995', VDI-Berichte, VDI-Verl., Düsseldorf, 1995.

[2] Manfred Becker: 'Schrittmotoren als Aktuator für Hydraulik-Wegeventile'.

[3] Schwander, P., Flückiger, M. 'Aktuator für Schieberventile in der mobilen Hydraulik', ATZoffhighway, Vol. 7, Nr. 3, pp. 70–77, 2014. DOI: 10.1365/s35746-014-0194-3.

[4] Lux, J., Habegger, C. 'New generation of intelligent electromechanical valve actuation', 11TH INTERNATIONAL FLUID POWER CONFERENCE, pp. 394–402, 2018.

[5] Bucher Hydraulics: 'Proportionalwegeventile - Baureihe LCV', https://www.bucherhydraulics.com/datacat/files/Katalog/Ventile/Zwischenplattenventile_Sandwichventile/LCV/LCV_100-P-000161-de.pdf, 2022.

[6] Sonceboz: 'Elektromechanische Ventilantriebe ergänzen die Hydraulik'.

[7] Norbert Menden: 'Punktgenau in allen Klimazonen - Ventil mit Can-Aktuator für mobile Anwendungen', https://www.fluid.de/abo/heftarchiv/ausgabendetail/fluid-01-02-2018.html, 2018.

[8] Ermert, M. 'Electromechanical actuator concept for the controlled and direct actuation of a hydraulic main stage', 10TH INTERNATIONAL FLUID POWER CONFERENCE, pp. 503–513, 2016.

[9] Sonceboz: 'SMART LINEAR ACTUATOR', https://sonceboz.com/product/smart-linear-actuator/.

[10] Schuster, G. 'CFD-gestützte Maßnahmen zur Reduktion von Strömungskraft und Kavitation am Beispiel eines hydraulischen Schaltventils', Reihe Fluidtechnik D, Shaker, Aachen, 2005.

[11] Tamburrano, P., Plummer, A. R. et al. 'A Review of Direct Drive Proportional Electrohydraulic Spool Valves: Industrial State-of-the-Art and Research Advancements', Journal of Dynamic Systems, Measurement, and Control, Vol. 141, Nr. 2, 2019. DOI: 10.1115/1.4041063.

[12] Amirante, R., Moscatelli, P. G., Catalano, L. A. 'Evaluation of the flow forces on a direct (single stage) proportional valve by means of a computational fluid dynamic analysis', Energy Conversion and Management, Vol. 48, Nr. 3, pp. 942–953, 2007. DOI: 10.1016/j.enconman.2006.08.024.

[13] Jackson, R. S., Belger, M. M. et al. 'CFD SIMULATION AND EXPERIMENTAL INVESTIGATION OF STEADY STATE FLOW FORCE REDUCTION IN A HYDRAULIC SPOOL VALVE WITH MACHINED BACK ANGLES', BATH/AMSE 2017 Symposium on Fluid Power and Motion Control, 2017.

[14] Chen, L., Zhang, C. et al. 'Simulation and experimental research on steady flow force compensation for a servo proportional valve', Flow Measurement and Instrumentation, Vol. 94, pp. 102457, 2023. DOI: 10.1016/j.flowmeasinst.2023.102457.

[15] Ding, X., Li, R. et al. 'Study on flow force compensation characteristics and optimization design of jet guiding groove', Flow Measurement and Instrumentation, Vol. 86, pp. 102194, 2022. DOI: 10.1016/j.flowmeasinst.2022.102194.

[16] Gui, S., Zhang, S. et al. 'Fluid-dynamic analysis and multi-objective design optimization of piezoelectric servo valves', Flow Measurement and Instrumentation, Vol. 85, pp. 102157, 2022. DOI: 10.1016/j.flowmeasinst.2022.102157.

[17] Li, R., Wang, Z. et al. 'Design and optimization of hydraulic slide valve spool structure based on steady state flow force', Flow Measurement and Instrumentation, Vol. 96, pp. 102568, 2024. DOI: 10.1016/j.flowmeasinst.2024.102568.

[18] Zhou, D., Jin, B. et al. 'Modeling, simulation and optimization of a pilot-operated threaded cartridge proportional relief valve', Flow Measurement and Instrumentation, Vol. 107, pp. 103071, 2026. DOI: 10.1016/j.flowmeasinst.2025.103071.

[19] Yang, M., Lian, K. et al. 'Simulation analysis and optimization of pilot type electro-hydraulic proportional directional valve based on multi field coupling', Flow Measurement and Instrumentation, Vol. 100, pp. 102726, 2024. DOI: 10.1016/j.flowmeasinst.2024.102726.

[20] Elsaed, E., Linjama, M. 'A Review of Pilot-operated Hydraulic Valves – Development, Challenges, and a Comparative Study', International Journal of Fluid Power, 2023. DOI: 10.13052/ijfp1439-9776.2443.

[21] Backé, W., Tatar, H. 'Untersuchung des Einflusses von Störkräften auf den Schaltvorgang bei Wegeventilen der Hydraulik', Forschungsberichte des Landes Nordrhein-Westfalen, Westdt. Verl., Opladen, 1975.

[22] Bordovsky, P. 'Evaluation of steady-state flow forces in spool valves - = Evaluierung stationärer Strömungskräfte in Schieberventilen', Reihe Fluidtechnik. D, [1. Auflage], Shaker Verlag, Düren, 2019.

[23] Schrank, K., Murrenhoff, H. 'Beschreibung der Strömungskraft in Längsschieberventilen mittel Impulserhaltung', O+P Journal, Nr. 4, pp. 4–15, 2013.

[24] Bordovsky, P., Murrenhoff, H. 'Investigation of steady-state flow forces in spool valves of different geometries and at different oil temperatures with the help of measurements and cfd simulations', BATH/AMSE 2016 Symposium on Fluid Power and Motion Control, 2016.

[25] Bordovsky, P., Schmitz, K., Murrenhoff, H. 'CFD Simulation and Measurement of Flow Forces Acting on a Spool Valve', 10TH INTERNATIONAL FLUID POWER CONFERENCE, pp. 473–484, 2016.

[26] Herakovič, N., Duhovnik, J., Šimic, M. 'CFD simulation of flow force reduction in hydraulic valves', Tehnički vjesnik - Technical Gazette, Vol. 22, Nr. 2, pp. 453–463, 2015. DOI: 10.17559/tv-20141128090939.

[27] Bordosky, p., Murrenhoff, H. 'VALIDATION OF AN ENHANCED MODEL OF STEADY-STATE FLOW FORCES FOR SPOOL VALVES', The 10th JFPS International Symposium on Fluid Power, 2017.

[28] Bordovsky, P., Murrenhoff, H. 'Analysis of Flow Angles and Flow Velocities in Spool Valves for the Calculation of Steady-State Flow Forces', The 15th Scandinavian International Conference on Fluid Power, pp. 371–379, 2017. DOI: 10.3384/ecp17144371.

[29] Ye, Y., Yin, C.-B. et al. 'Effects of groove shape of notch on the flow characteristics of spool valve', Energy Conversion and Management, Vol. 86, pp. 1091–1101, 2014. DOI: 10.1016/j.enconman.2014.06.081.

[30] Del Vescovo, G., Lippolis, A. 'A review analysis of unsteady forces in hydraulic valves'.

[31] Schmitz, K., Murrenhoff, H. 'Grundlagen der Fluidtechnik', Reihe Fluidtechnik. U, Vollständig neu bearbeitete Auflage, 2018.

[32] Hucko, S., Vonderbank, T., Schmitz, K. 'Investigation of Frictional Forces in Hydraulic Spool Valves and Their Effect on the Estimation of Axial Flow Forces', ASME/BATH 2023 Symposium on Fluid Power and Motion Control, 2023. DOI: 10.1115/FPMC2023-111904.

[33] Lisowski, E., Filo, G., Rajda, J. 'Pressure compensation using flow forces in a multi-section proportional directional control valve', Energy Conversion and Management, Vol. 103, pp. 1052–1064, 2015. DOI: 10.1016/j.enconman.2015.07.038.

[34] Ristic, M. 'Dreidimensionale Strömungsberechnungen zur Optimierung von Hydraulikventilen bezüglich der stationären Strömungskräfte', Reihe Fluidtechnik D, Shaker, Aachen, 2000.

[35] Amirante, R., Distaso, E., Tamburrano, P. 'Sliding spool design for reducing the actuation forces in direct operated proportional directional valves: Experimental validation', Energy Conversion and Management, Vol. 119, pp. 399–410, 2016. DOI: 10.1016/j.enconman.2016.04.068.

[36] Borghi, M., Milani, M., Paoluzzi, R. 'Stationary Axial Flow Force Analysis on Compensated Spool Valves', International Journal of Fluid Power, Vol. 1, Nr. 1, pp. 17–25, 2000. DOI: 10.1080/14399776.2000.10781079.

[37] Herakovič, N. 'Flow-force analysis in a hydraulic sliding-spool valve', Strojarstvo Casopis Za Teoriju I Praksu U Strojarstvu, Vol. 51, pp. 555–564, 2009.

[38] Bordovsky, P., Schmitz, K., Murrenhoff, H. 'Anwendung der statistischen Versuchsplanung zur Parametidentifiaktion für Strömungskraftberechungen', 2018.

[39] Borghi, M., Milani, M. 'Influence of Notch Shape and Number of Notches on the Metering Characteristics of Hydraulic Spool Valves', International Journal of Fluid Power, Vol. 6, pp. 5–18, 2005. DOI: 10.1080/14399776.2005.10781216.

[40] C. Qianpeng, J. Hong et al. 'Proposal for Optimization of Spool Valve Flow Force Based on the MATLAB-AMESim-FLUENT Joint Simulation Method', IEEE Access, Vol. 6, pp. 33148–33158, 2018. DOI: 10.1109/ACCESS.2018.2846589.

[41] Lugowski, J. 'FLOW FORCE IN A HYDRAULIC SPOOL VALVE'.

[42] Aengenheister, S., Liu, C. et al. 'Functional proof of a flat slide valve as a 4/3-way proportional valve', 12TH INTERNATIONAL FLUID POWER CONFERENCE, pp. 163–170, 2020. DOI: 10.25368/2020.24.

[43] Vonderbank, T., Laßl Chavez, P. M., Schmitz, K. 'Direct Actuation of Large Sized Valves by a Hydraulically Relieved Electromechanical Actuation System', International Journal of Fluid Power, 2021. DOI: 10.13052/ijfp1439-9776.2312.

[44] Vonderbank, T., Schmitz, K. 'ELECTROMECHANICAL ACTUATOR FOR LARGE SIZED DIRECTIONAL CONTROL VALVES', 10th International Conference on Fluid Power Transmission and Control, ICFP, pp. 312–317.

[45] Vonderbank, T., Laßl Chavez, P. M., Schmitz, K. 'Novel hydraulically relieved electromechanical direct actuation system for large scale switching valves', The 17th Scandinavian International Conference on Fluid Power, pp. 173–185.

Biographies



Tobias Vonderbank received the Bachelor of Engineering from Hochschule Niederrhein University for Applied Sciences in 2016 and the Master of Science from RWTH Aachen University in 2018. He is currently working as a Research Associate at the Institute for Fluid Power Drives and Systems, RWTH Aachen University. His research areas are Fluid Power, valves and electromechanical valve actuation systems.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

1.

Innovative Approach to Cleaning and Cleanliness Control of Newly Manufactured Hydraulic Assemblies

Nejc Novak, Ana Trajkovski, Jan Pustavrh, Mitjan Kalin, Franc Majdič

Faculty of Mechanical engineering, University of Ljubljana,

nejc.novak@fs.uni-lj.si, ana.trajkovski@fs.uni-lj.si, jan.pustavrh@fs.uni-lj.si, mitjan.kalin@fs.uni-lj.si, franc.majdic@fs.uni-lj.si

Abstract.

Initial cleanliness of hydraulic components is vital for system reliability and longevity. Existing on-line oil condition monitoring technologies, though advanced, cannot verify baseline cleanliness before assembly. To address this limitation, a novel device was developed that combines component flushing and cleanliness evaluation in a single process. The system features a dual-pump configuration: one pump performs high-flow flushing and sampling, while the other circulates and filters a glycol-water fluid. Flow is controlled through valves, and contaminants are collected under vacuum onto membrane filters for microscopic evaluation per ISO 4407 and rated using ISO 4406. A key innovation is vibration-assisted flushing, which effectively dislodges embedded particles. Tests showed cleanliness improvements of 3–5 ISO codes within five minutes. The device provides a practical, repeatable, and ISO-compliant solution for verifying initial hydraulic cleanliness, with planned developments including automated particle analysis and integration into digital quality management systems.

Keywords. Cleanliness, membrane filters, IS0 4406, ISO 4407, filtration, particles.

2. INTRODUCTION

Ensuring the cleanliness of hydraulic components prior to assembly is crucial for system reliability, wear prevention, and extended operational life. Contaminants such as metal particles, dust, or residues from manufacturing can cause premature wear, clog valves, and reduce the efficiency of hydraulic systems, making initial cleanliness a key determinant of long-term performance.

Particles that are in the fluid can cause abrasive wear [1,2] and cause damage to surfaces. For instance degradation of hydraulic gear pump is shown in increase of microparticles in the fluid, which causes a decrease in the quality of the fluid, which is a class that does not meet the prescribed standards. Solid particles cause specific wear mechanisms of the contact surfaces (abrasion, erosion, material fatigue, etc.) [3,4]. Karanović et al. [3] describe the influence of solid particle impurities in the radial gap between the control piston and the

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

directional valve housing: due to three-body abrasion, they carry away material and increase the gap. As the gap increases, the flow through the gap also increases (internal leakage), which causes even more particles to travel through the gap. This further increases friction and stimulates the wear process. Pressure losses in the gaps and friction of the particles cause local heating. Optimal fluid cleanliness is required to reduce the intensity of wear. Jung-Hun et al. [5] studied the influence of particle composition and particle quantity. It is proven that size distribution also have an impact on damaging the surfaces because that are so-called "critical size" particles [6,7], that come between two surfaces, causing three body abrasion. This phenomenon is also shown in tribological studies in micro-scale abrasive tests [8].

Fiber filtration technology is used due to its versatility, adaptability and superior filtration efficiency [9]. Multilayer nonwoven fiber filter media has superior flirtation efficiency, lower pressure drop and higher dust holding capacity [10,11]. The filtration velocity and fiber diameter along with surface energy determines the pressure drop through filter material [9].

Although real-time, on-line oil condition monitoring has seen rapid advances in recent years—supported by sensor fusion, microfluidic technologies, and inductive-resistive detection systems [12–14] —these systems are inherently limited in their ability to assess the initial cleanliness of hydraulic assemblies. On-line particle counters are most effective during early system operation but quickly lose relevance as contaminants either settle or disperse. This limitation makes it particularly difficult to verify the baseline cleanliness of components immediately after manufacturing, creating a gap in quality assurance processes.

Current standard practices for cleanliness assessment rely heavily on manual sampling and laboratory analysis, which can be time-consuming, subjective, and insufficiently integrated with the cleaning process. Furthermore, complex geometries within hydraulic assemblies often trap particles in hard-to-reach areas, reducing the effectiveness of traditional flushing or filtration methods. To address these challenges, there is a need for a methodology that combines efficient cleaning with objective, reproducible measurement of particulate contamination. Integrating vibration-assisted flushing with ISO-compliant membrane filtration and microscopic particle counting for measuring cleanliness offers a promising approach to overcome these limitations. The purpose of this paper is to critically evaluate such a methodology for cleaning hydraulic pipe assemblies and to demonstrate its effectiveness in achieving the required cleanliness standards, providing a reliable and repeatable workflow for pre-assembly quality assurance.

3. MATERIALS AND METHODS

The objective of the developed device is to provide a repeatable procedure for achieving and verifying the required cleanliness level of pipe assemblies (PA) prior to their integration into hydraulic systems. To reach this goal, the device combines controlled flushing, vibration-assisted particle excitation, and standardized cleanliness measurement within a single workflow. Cleanliness verification is performed by sampling contaminants onto membrane filters using a vacuum filtration unit, followed by microscopic analysis in accordance with ISO 4407 [15] and cleanliness classification according to ISO 4406 [16]. This approach enables both effective removal of contaminants and reliable assessment of the resulting cleanliness level.

The working principle of the device is based on two hydraulically separated but functionally complementary circuits, driven by two electric motors and centrifugal pumps. The first circuit is dedicated to flushing and cleanliness measurement of the PA, while the second circuit continuously filters the working fluid to maintain a stable and low background contamination level. A glycol–water mixture (DOWFROST™ LC 25) is used as the flushing medium due to its suitable viscosity, corrosion protection, and thermal stability.

3.1. Flushing function and performing cleanliness measurements

The device consists of a centrifugal pump (MovitecV 006/11) which pushes the liquid through the throttle, further along the flexible pipe to the upper ball valve and further to the pipe assembly. The liquid flows through the lower ball, through the coarse filter (25 µm) and the fine filter (3 µm) into the tank. The flow rate is determined by the centrifugal pump characteristics and the overall pressure drop of the system. A throttle valve was used to modify the system resistance (control the pressure) and thus flow rate. At full throttle opening, the pump outlet pressure was 6.7 bar and the flow rate was 150 L/min, whereas at an outlet pressure of 9.2 bar (with semi-closed throttle) the flow rate decreased to 50 L/min. Pressure was measured at the pump outlet and therefore does not directly represent the pressure drop across the pipe assemblies, which remains constant for a fixed configuration and is governed by pipe dimensions and quick couplings. Pump maximum pressure is 11 bar. In case of higher pressure-resistance pump will not provide flow because of the limitations of centrifugal pump. The viscosity decrease due to the increase of temperature is negligible.

In the case of performing cleanliness measurements, the upper ball valve is bypassed, so that the vent filter is in function. The liquid passes through the quick coupling to the PA via the bypassed lower ball valve and the measuring tube to the measuring ball valve and the vacuum device for separating particles on the membrane. The measuring ball valve supplies the appropriate volume of liquid (100 ml) to the upper container of the vacuum device for separating particles on the membrane, which must be provided for the cleanliness measurement.

3.2. Liquid filtration function and heating

The second circuit also contains a centrifugal pump, which sucks the liquid from the bottom of the tank and pushes the liquid through flexible pipes via a fine filter (3 µm) into the tank. This circuit is intended exclusively for filtration and maintaining a high level of liquid cleanliness.

A heater is located in the tank, for heating the liquid in the tank. The temperature switch and the level switch constitute a safety module that switches off the entire device if the liquid has a higher temperature than that set on the temperature switch or if the liquid level is lower than the minimum permissible level (determined by the float mounted on the level switch in the tank). The vibration motor (Netter Vibration NEG 5050 Electric-vibrator, external 400V, 3000 rpm) is important in vibrating the pipe assembly during the flushing function and is controlled by a frequency regulator mounted in the electrical cabinet. Vibration is defined as the applied mechanical oscillation of the system, quantified by its frequency (Hz). The frequency converter may regulate the frequency between 20 and 60 Hz at a constant torque load (linear volt-hertz-curve). It was set to 40 Hz (stage 7 out of 14 on frequency potentiometer control rolling button).

4

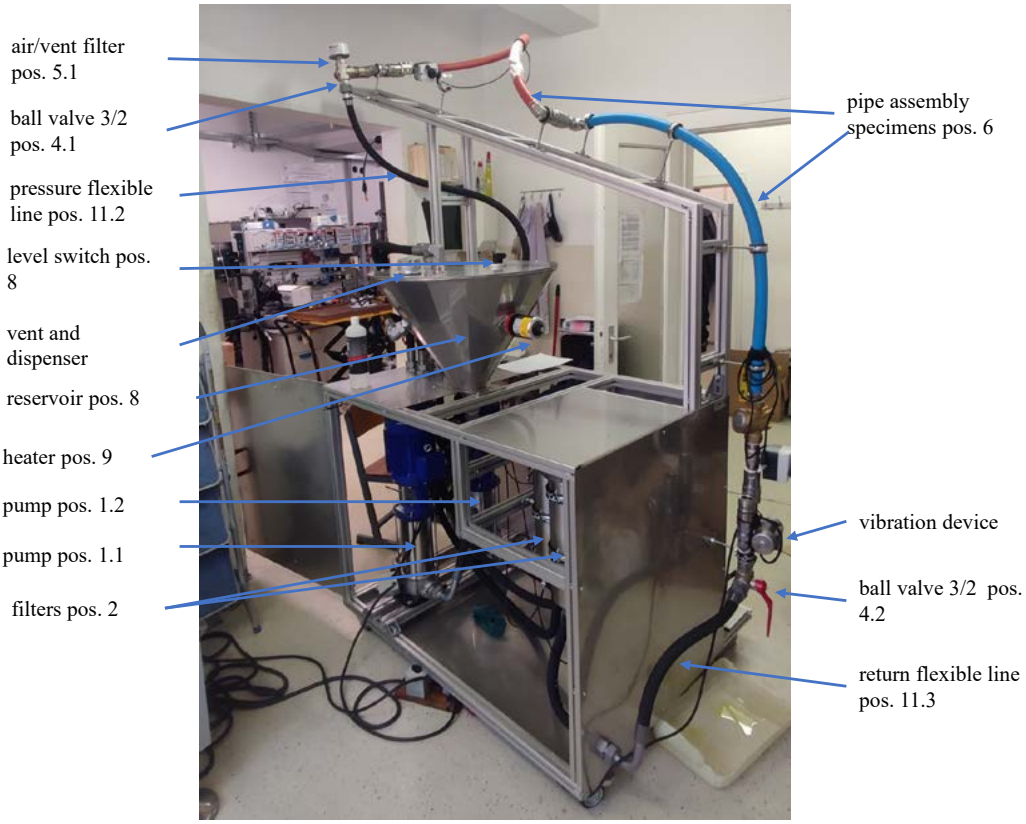


Figure 1. Test rig for cleaning and performing cleanliness measurements on pipe assemblies (PA)

Table 1. Hydraulic scheme (Figure 2) of components BOM

Position	Description		Position	Description
1	centrifugal pumps		8	tank
2	fine filters 3 µm		9	heater
3	coarse filters 25 µm		10	throttle valve
4	3/2 directional ball valve		11	flexible lines
5	breather filter 3µm		12	thermostat
6	pipe assembly (PA)		13	level switch
7	laboratory accessories		14	quick couplings

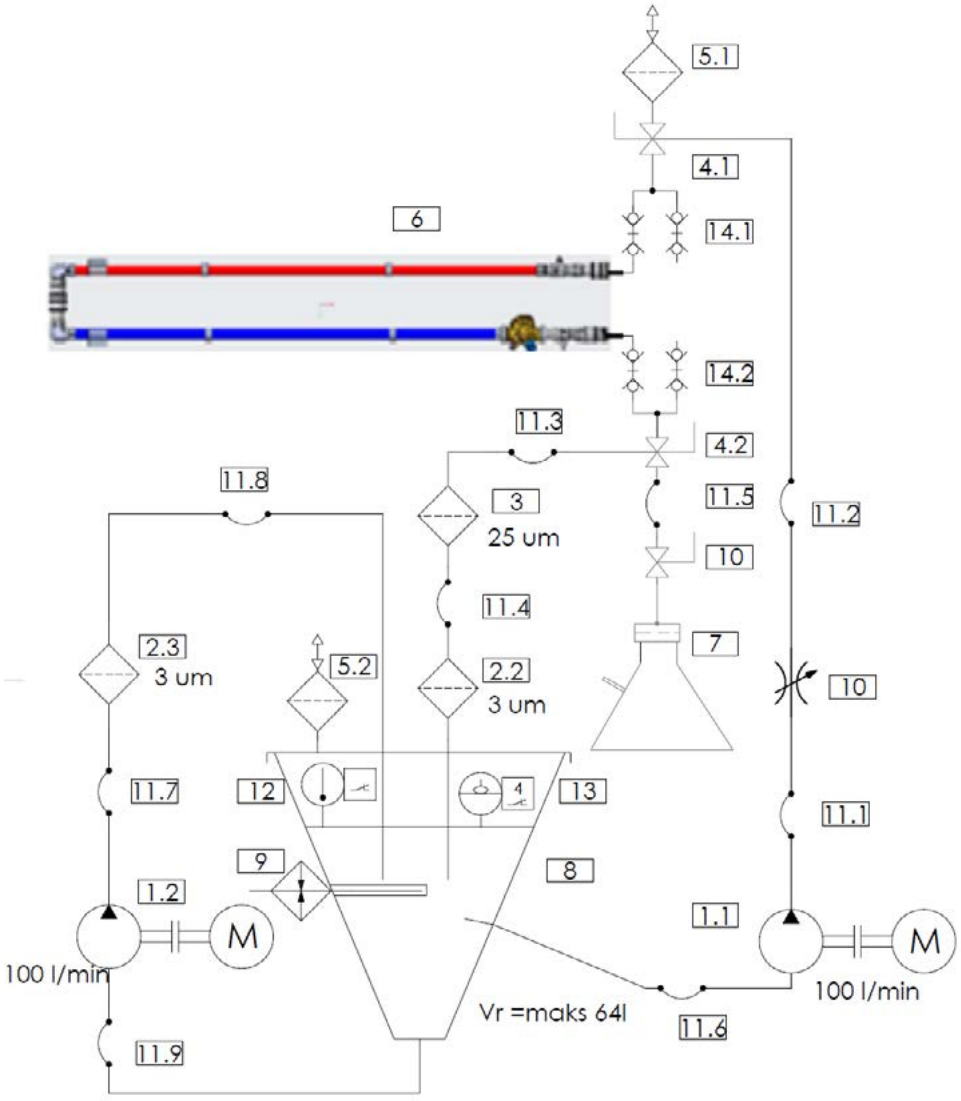


Figure 2. Hydraulic scheme of cleaning device for PAs

3.3. Procedure for flushing and measuring cleanliness

The cleaning and cleanliness assessment of pipe assemblies (PAs) were performed using a dedicated device integrating controlled fluid flow, vibration excitation, and membrane filtration. Prior to measurement, the hydraulic connections were configured so that the lower 3/2 ball valve established a link between the PA and the measuring tube, while the upper 3/2 valve connected the supply line to the PA. The assembly was mounted using quick couplings to ensure a sealed, contamination-free connection.

To define the initial liquid volume, the pump was actuated intermittently, introducing approximately 1600 ml of glycol–water mixture into the PA. A one-minute vibration cycle was applied to release loosely bound contaminants. The flow path was then redirected toward the vent filter to enable sampling. Cleanliness evaluation was conducted by vacuum-filtering 100 ml of the working fluid through a membrane, capturing suspended particles for microscopic examination according to ISO 4407. Multiple samples were taken to ensure statistical reliability.

Following the initial measurements, the rinsing/flushing phase start by redirecting the flow between the PA, inlet, and outlet lines. The rinsing pump was operated for approximately five minutes, accompanied by vibration excitation to enhance particle removal from internal surfaces. After rinsing, the flow configuration was restored for sampling, and the cleanliness measurement was repeated under identical conditions.

Upon completion, the PA could be disconnected either filled with fluid or vented, depending on subsequent handling requirements. This procedure provides a repeatable method for both cleaning and ISO-compliant cleanliness verification of hydraulic pipe assemblies, enabling a quantitative assessment of contamination removal efficiency.

Test coolant DOWFROST LC 25 with fluorescent yellow-green colour was used. It can be diluted with water. The liquid also supports bio-oxidative degradation at concentration between 15 % and 25 % glycol volume. Viscosity of fluid is 2.5 mm^2/s at 20 °C and 1.5 mm^2/s at 50 °C. In Table 2 there are given characteristics of the liquid [17].

Table 2. Characteristics of DOWFROST LC 25 [17]

Component	Concentration
Water	65.0 – < 75.0 %*
Propylene glycol	> 25.0 – < 26.0 %
Dipotassium hydrogen phosphate	1.0 – < 5.0 %*
Sodium tolyltriazole	≥ 0.1 – < 0.25 %*

Initial cleanliness, measured from the plastic container was 14/14/12. Temperature was varying from room temperature due to the insufficient cooling system and at times reached temperature up to 100 °C.

3.4. Measurements according to ISO 4407

Solid particle contamination in hydraulic fluids was quantified following ISO 4407, which specifies a standardized microscopic counting procedure for determining fluid cleanliness. Maintaining low particulate levels is critical for the operational reliability of hydraulic systems. Although the method was once constrained by manual counting and slow analysis, recent progress in automated microscopy and digital imaging has re-established it as an efficient and accurate technique for contamination assessment.

For all cleanliness evaluations, 100 ml of DOWFROST LC 25 working fluid was vacuum-filtered through a membrane to collect particulate contaminants released from the pipe assembly (PA). The initial cleanliness, referred to as pouring, was determined by introducing

clean fluid into the PA through short pump cycles, followed by one minute of vibration excitation. The resulting 100 ml sample was withdrawn and filtered through a membrane. Subsequent measurements were performed after different flushing periods, where highest cleanliness improvements were observed. During filtration, the membrane filter was carefully positioned on the vacuum apparatus to prevent airborne contamination (figure 3). The upper container was secured with metal clips, clamping the membrane between the container and the porous support.

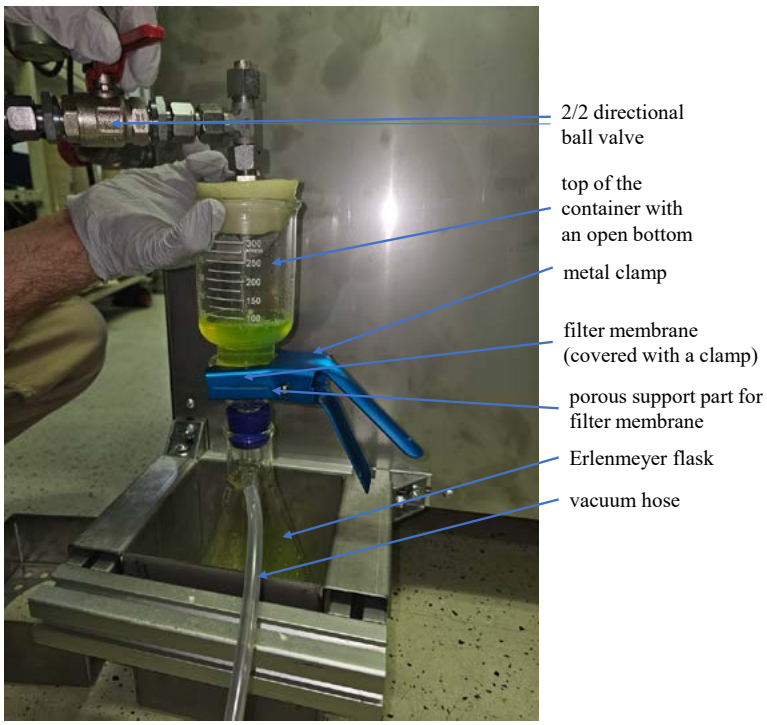
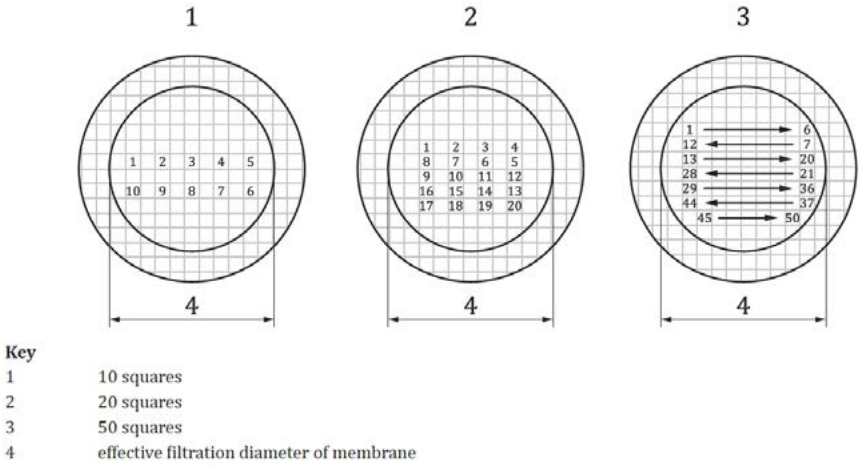


Figure 3. Taking a sample of the cleanliness of the liquid on a vacuum device for membrane filtration

A 100 ml fluid sample was introduced through the measuring tube and filtered under vacuum. After filtration, the membrane was removed with its holder and prepared for analysis. Samples were then placed in an oven at 50–60 °C for two hours to ensure uniform drying without thermal damage. The dried membranes were subsequently examined under a microscope to determine particle count and size distribution in accordance with ISO 4407.

In Figure 4, the number 4 indicates the effective diameter through which the liquid is filtered. The filter membranes are Φ47 mm in size. The first set of filters was made of mixed cellulose (MCE) material with 0.45 µm pores and 2.69 mm x 2.69 mm squares etched into the membrane. The second set of membranes was made of the same material with 1.2 µm pores without squares.



Key

1	10 squares
2	20 squares
3	50 squares
4	effective filtration diameter of membrane

Figure 4. Particle counting in 10 area squares using method 1 according to ISO 4407 [15]

The number of particles was determined in 100 ml of liquid according to the ISO 4407 standard using the equation (2.1) [15]:

$$N = \frac{A \cdot n \cdot 10^5}{f \cdot L \cdot W \cdot V} \quad \text{in 100 ml} \qquad (2.1)$$

where:

A [mm^2] effective filtration area

n [/] number of counted particles larger than the selected size

f [/] number of counted squares

L [mm] length of the observed rectangle, or mesh size

W [µm] width of the observed rectangle,

V [mL] volume of filtered liquid

After drying the membranes, the particles were examined under a Nikon Eclipse Ti-U microscope (Japan) equipped with an IDS u-Eye camera for classical transmitted light technique. It allows for combination with an EMCCD Andor Ixon 888 Ultra camera, a Carin Optosplit III image splitter system and a CoolLed pE-4000 illumination source for epifluorescence microscopy. Only classical transmitted light technique with an E Plan 10X/0.25 objective and WD 7.0 was used.

Since an area of 1 mm x 1 mm was imaged due to magnification and the standard states that the entire square area is analysed, it was necessary to take several images and then combine them in the Image J program and appropriately determine the particle analysis field [18]. This was followed by determining the particle detection intensity threshold, since the analysis works on the bitmap principle and the program counts automatically. The program allows for measurement of the Feret length, which is described in more detail in the article by Walton [19]. Count and classification were done for the Feret particle lengths in Excel into size categories specified by ISO 4406. The oil cleanliness rating consists of three

numbers, the first indicating the class of particles greater than 4 µm, the second the class of particles greater than 6 µm, and the third the class of particles greater than 14 µm per milliliter of liquid.

4. RESULTS AND DISCUSSION

The fluid cleanliness was measured based on the filtration of 100 ml of DOWFROST LC 25. The results were divided into 3 sets. The first set presents the cleanliness measurements of the flow pipe assembly (blue) labelled PA1F and the return pipe assembly (red) labelled PA1R at the beginning before rinsing (pouring) and after different flushing time intervals.

The second set presents the measurements of the combined flow-return pipe assembly labelled PA2FR and determines the cleanliness for both pipe assemblies simultaneously. The cleanliness results are presented at the beginning before rinsing (pouring) and after different flushing time intervals. The last set presents the results of the combined flow-return pipe assembly sets labelled PA5FR and PA6FR using the same procedure with slight changes in regard to vibrations and flow.

4.1. Determining cleanliness measurement from particle counting method

The procedure for determining cleanliness measurement by microscopic particle counting and translation to ISO 4406 is shown using the example of a PA1F pipe assembly. Figure 5 shows the beginning (pouring) before rinsing. First, an image is obtained from the microscope. Figure 6 shows the threshold setting for automatic particle detection software [20]. All red zones are recognized as particles that are determined according to the threshold, as can be seen in the box in Figure 6.

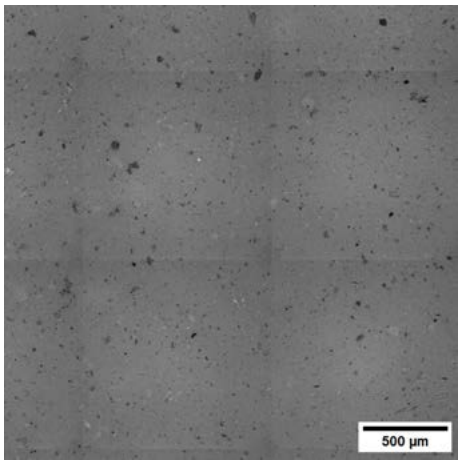


Figure 5: Particles in one particle analysis square for sample PA1F at the beginning (pouring) before rinsing

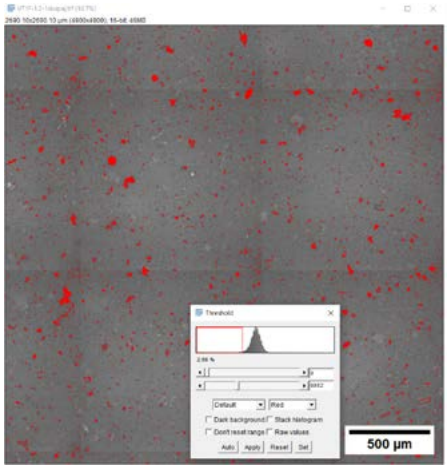


Figure 6: Determination of the particle detection threshold for the PA1F sample at the beginning (pouring) before rinsing

When a threshold is selected, the program automatically converts the image into a 2-bit image (Black and White) (Figure 7). Then it is analysed with the Analyse Particles function in Image J. In this function, it is possible to specify the values that are measured in the image [20]. The size of the particle is measured according to Feret diameter [21].

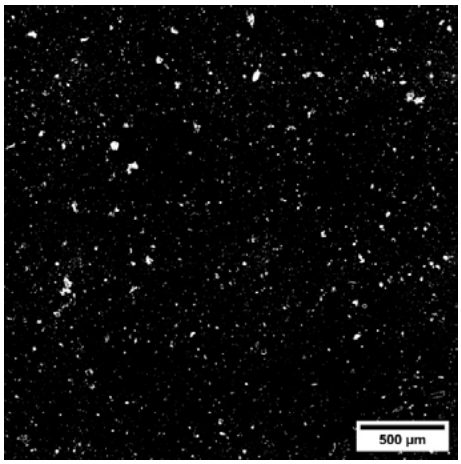


Figure 7. Bitmap for PA1F particle count analysis at the beginning (pouring) before rinsing

4.2. Cleanliness measurements

The first set of measurements served as a preliminary evaluation of the cleanliness of the hydraulic fluid itself, without the influence of any specialized cleaning procedures or treatments. For this purpose, samples from the flow pipe assembly PA1F (blue) and the return pipe assembly PA1R (red) were analysed separately. Microscopic images were

obtained using MCE membranes with 0.45 µm pores, and a grid of drawn squares on the membrane facilitated systematic particle counting. This approach allowed for a detailed "blank" test, establishing a baseline of particulate contamination present in the fluid prior to any flushing. The results, summarized in Table 3, provide the largest particle detected in each of the five analysed squares, along with the corresponding number of particles per millilitre, calculated according to ISO 4406 standards. This initial assessment was essential for establishing reference values, against which the effectiveness of subsequent flushing and vibration-assisted cleaning could be quantitatively evaluated. It also allowed identification of inherent variability in particle distribution across the membrane, ensuring that later measurements would be both reliable and statistically meaningful.

Table 3: Cleanliness measurements on 5 different squares (areas) of a membrane filter of 100 ml of clean liquid (from the packaging container)

Internal	>4µm	>6µm	>14µm
0-1	28	20	2
0-2	29	22	4
0-3	98	69	15
0-4	121	91	28
0-5	71	54	33
	>4µm	>6µm	>14µm
Measured particle count 1 ml	127	93	30
ISO 4406	**14**	**14**	**12**

On PA1F measurements were performed at the beginning (pouring) and after 10 min, 20 min and 30 min of rinsing. Results after 20 min of rinsing showed high cleanliness (14/14/13) (Table 4).

Table 4: Collected together for sample PA1F according to ISO 4406

	>4µm	>6µm	>14µm	Largest particle, µm
PA1F-pouring (start)	**20**	**19**	**17**	128
PA1F-10 min	**16**	**15**	**13**	130
PA1F-20 min	**14**	**14**	**13**	87
PA1F-30 min	**14**	**13**	**11**	60

In the pipe assembly sample PA2FR, the total cleanliness of the joined "flow" and "return" pipe assemblies with quick couplings were measured at the same time. 4 repetitions of cleanliness measurements (cleanliness removal on the membrane) before rinsing (pouring

(start)) and after 1 min, 2 min, 3 min, 4 min, 5 min and 27 min of rinsing (Table 5) were made (together 28 measurements). The speed of improvement of cleanliness in terms of determining the optimal time of cleaning the pipe assembly . For greater clarity, only the cleanliness according to the ISO 4406 standard is given.

Table 5: Total collected for the PA2FR sample according to ISO 4406

	>4µm	>6µm	>14µm	Largest particle, µm
PA2FR-pouring (start)	**19**	**18**	**16**	299
PA2FR-1min	**17**	**17**	**15**	295
PA2FR-2min	**16**	**15**	**11**	64
PA2FR-3min	**16**	**15**	**13**	508
PA2FR-4min	**15**	**14**	**12**	252
PA2FR-5min	**14**	**14**	**12**	113
PA2FR-27min	**14**	**13**	**11**	117

For the other samples (PA5FR pipe assembly), vibration was activated during rinsing and at the beginning, but during the actual collection of the liquid for the cleanliness measurement, the tube assembly was not vibrated. From initial cleanliness 18/17/16 the particles were reduced due to 5 min filtration to cleanliness 15/14/12. The last PA6FR pipe assembly was measured with a reduced flow rate of 50 L/min and a pressure of 9.2 bar. This measurement was made to observe the influence of the different flow rates to cleanliness and it is concluded that higher flow rates have slightly higher cleaning capabilities. The other pipe assemblies were measured with a flow rate of 150 L/min and a pressure of 6.7 bar upstream of the pump. Cleanliness went from 17/16/15 to 14/13/11 after 5 min filtration.

4.3. Examination of particles

Particles retained on the membrane from sample PA1F-pouring (start) were examined using a Hirox HRX1 optical microscope with the objective of demonstrating the capability of the proposed method to classify particle types and potential contamination sources following standardized membrane filtration. In Figure 7, an elongated red fibre was identified and measured, with a length of approximately 950 µm and a width of 20 µm. Similar fibrous particles (black, dark blue, and light blue) were repeatedly detected, indicating a systematic contamination mechanism rather than an isolated occurrence, most likely associated with tube handling or mechanical damage. Such fibres were classified as low-hardness contaminants and are therefore expected to have limited abrasive impact on hydraulic components.

A golden-yellow particle with a reflective surface was observed in Figure 8. Based on its optical appearance, this particle was categorized as a metallic fragment, potentially originating from copper-based alloys such as brass. In contrast, angular particles with dimensions of approximately 52 µm, morphologically similar to standardized test dust, were also detected. These particles were classified as potentially critical due to their sharp edges and higher likelihood of causing abrasive wear. Their presence is indicative of either

airborne dust ingress, insufficient removal of sediments from the working fluid, or residual particles introduced during manufacturing processes, such as sawing or cutting operations. The identification and differentiation of these particle classes demonstrate that the examination serves not to document random particles, but to validate the method's ability to systematically detect, size, and interpret contamination relevant to hydraulic reliability.

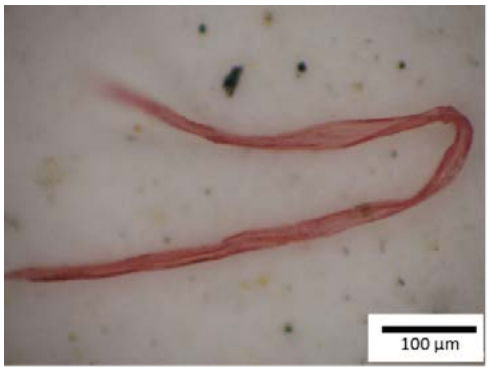


Figure 8. Red fiber on membrane pouring at 600x magnification



Figure 9. Golden-yellow particle on membrane pouring at 800x magnification

5. CONCLUSIONS

This study proposes a method for assessing and improving the initial cleanliness of hydraulic pipe assemblies that complements, rather than replaces, existing standard procedures. Conventional approaches—such as flushing and filtering with parallel use of on-line particle counters or post-flush oil sampling followed by ISO 4407 analysis—are well established and effective during system commissioning or operation. However, these methods provide limited insight into the true baseline cleanliness of newly manufactured components, as

particle counters are sensitive to transient contamination and oil samples are often taken after particles have already settled or been partially removed.

The primary benefit of the proposed method lies in its systematic integration of cleaning and standardized cleanliness verification into a single, controlled workflow. By coupling defined flushing conditions with immediate, ISO-compliant membrane filtration, the method ensures that contaminants released during cleaning are directly captured and quantified. This eliminates uncertainties associated with delayed sampling, fluid dilution, or background contamination that can affect conventional flushing-and-measurement strategies.

All cleanliness evaluations are performed according to ISO 4407 and ISO 4406, ensuring full comparability with established standards. The novelty of the approach does not lie in the measurement principle itself, but in the methodological arrangement that enables reproducible, pre-assembly cleanliness verification under laboratory-like conditions. Optional mechanical excitation was introduced as an auxiliary means to enhance particle release in complex geometries; however, the core advantage of the method remains the controlled flushing and immediate, standardized assessment rather than the excitation mechanism itself.

Furthermore, the use of image-based microscopic analysis improves traceability and repeatability compared to purely manual particle evaluation, while remaining fully compliant with ISO requirements. Overall, the proposed method provides a practical and scientifically justified extension of standard cleanliness procedures, specifically targeted at the manufacturing stage. It enables manufacturers to verify and document initial component cleanliness prior to assembly, thereby closing a critical gap between production and operational monitoring. The methodology is readily transferable to other hydraulic components and can be integrated into existing quality assurance frameworks.

Funding

The presented research results were funded by Slovenian Research and Innovation agency (ARIS), grant No. P2-0231 and No. L2-50083

6. REFERENCES

[1] M.G. Hamblin, G.W. Stachowiak, Description of Abrasive Particle Shape and Its Relation to Two-Body Abrasive Wear, Tribology Transactions 39 (1996) 803–810. https://doi.org/10.1080/10402009608983598.

[2] G.W. Stachowiak, A.W. Batchelor, 11 - Abrasive, Erosive and Cavitation Wear, in: G.W. Stachowiak, A.W. Batchelor (Eds.), Engineering Tribology (Third Edition), Butterworth-Heinemann, Burlington, 2006: pp. 501–551. https://doi.org/https://doi.org/10.1016/B978-075067836-0/50012-2.

[3] V. Karanović, Development of a solid particle influence model on performance of piston cylinder contacting pairs for hydraulic components, University of Novi Sad, 2015.

[4] D. Wooton, The Lubricant's Nemesis-Oxidation, Practicing Oil Analysis 5 (2007) 5–6.

[5] S. Jung-Hun, J. Kyung-Ryeol, K. Hyoung-Eui, A Study on the Effects of Contaminant Types on the Wear Degradation Characteristics in Internal Gear Pumps, Tribology and Lubricants 27 (2011) 134–139.

[6] R. Frith, A model of gear pump wear due to solids contamination, PhD, Queensland University of Technology, 1994.

[7] R.H. Frith, W. Scott, Control of solids contamination in hydraulic systems — an overview, Wear 165 (1993) 69–74. https://doi.org/https://doi.org/10.1016/0043-1648(93)90374-U.

[8] P.J. Esteves, V. Seriacopi, M.C.S. de Macêdo, R.M. Souza, C. Scandian, Combined effect of abrasive particle size distribution and ball material on the wear coefficient in micro-scale abrasive wear tests, Wear 476 (2021) 203639. https://doi.org/https://doi.org/10.1016/j.wear.2021.203639.

[9] H. Jin, G. Peng, J. Chai, Z. Wu, H. Cheng, S. Hu, C. Gui, S. Li, Research on the filtration and deposition characteristics of mixed particles in nonwoven fiber filter media, Sep Purif Technol 379 (2025) 135084. https://doi.org/https://doi.org/10.1016/j.seppur.2025.135084.

[10] W. Jabri, P. Vroman, A. Perwuelz, Study of the influence of synthetic filter media compressive behavior on its dust holding capacity, Sep Purif Technol 156 (2015) 92–102. https://doi.org/https://doi.org/10.1016/j.seppur.2015.09.068.

[11] J. Wu, Y. Rao, B. Lin, H. Zhou, Q. Zhou, C. Xu, S. Feng, Z. Zhong, W. Xing, Design of PTFE/PP multilayer composite membrane as mask filter layer with steady filtration efficiency and high dust holding capacity, Sep Purif Technol 326 (2023) 124782. https://doi.org/https://doi.org/10.1016/j.seppur.2023.124782.

[12] Z. Yuan, X. Wu, Z. Li, J. Yuan, Y. Zhao, Z. Li, S. Qin, Q. Ma, X. Shi, Z. Zhao, J. Li, S. Zhang, W. Jing, X. Wang, L. Zhao, An integrated micromachined flexible ultrasonic-inductive sensor for pipe contaminant multiparameter detection, Microsyst Nanoeng 10 (2024) 111. https://doi.org/10.1038/s41378-024-00734-0.

[13] H. Shi, H. Zhang, L. Ma, L. Zeng, A multi-function sensor for online detection of contaminants in hydraulic oil, Tribol Int 138 (2019) 196–203. https://doi.org/https://doi.org/10.1016/j.triboint.2019.05.025.

[14] T. Wu, H. Wu, Y. Du, Z. Peng, Progress and trend of sensor technology for on-line oil monitoring, Sci China Technol Sci 56 (2013) 2914–2926. https://doi.org/10.1007/s11431-013-5400-5.

[15] ISO 4407:2025 Hydraulic fluid power — Fluid contamination — Determination of particulate contamination by the counting method using an optical microscope, 2025.

[16] ISO 4406:2017 Hydrulic fluid power-Fluids-Method for coding the level of contamination by solid particles, 2017.

[17] DOWFROSTTM LC 25 Heat Transfer Fluid, (2025). https://www.dow.com/en-us/pdp.dowfrost-lc-25-heat-transfer-fluid.497419z.html#overview (accessed October 7, 2025).

[18] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat Methods 9 (2012) 671–675. https://doi.org/10.1038/nmeth.2089.

[19] W.H. WALTON, Feret‘s Statistical Diameter as a Measure of Particle Size, Nature 162 (1948) 329–330. https://doi.org/10.1038/162329b0.

[20] S. Preibisch, S. Saalfeld, P. Tomancak, Globally optimal stitching of tiled 3D microscopic image acquisitions, Bioinformatics 25 (2009) 1463–1465. https://doi.org/10.1093/bioinformatics/btp184.

[21] H.G. Merkus, Particle size measurements : fundamentals, practice, quality, 1st ed., Springer, Dordrecht, 2009.

Biographies



Nejc Novak received the PhD in Mechanical Engineering from University of Ljubljana in 2025. He is currently working as researcher at the Department of Tribology and maintenance systems, Laboratory for Fluid Power and Controls, Faculty of Mechanical Engineering, University of Ljubljana. His research areas include analyses of hydraulic components wear mechanisms and lifetime, hydraulic fluid properties and filtration.



Ana Trajkovski received the PhD in Mechanical Engineering from University of Ljubljana in 2015. She is currently working as teaching assistant at the Department of Tribology and maintenance systems, Laboratory for Fluid Power and Controls, Faculty of Mechanical Engineering, University of Ljubljana. Her research areas include inovative materials and lubricants for sustainable hydraulics, smart control design and optimization of hydraulic components, water hydraulics. She has written many scientific papers and collaborated in many national and international conferences and research projects.



Jan Pustavrh received the M.Sc. in Mechanical Engineering from University of Ljubljana in 2022. He is currently working as researcher at the Department of Tribology and maintenance systems, Laboratory for Fluid Power and Controls, Faculty of Mechanical Engineering, University of Ljubljana. His research areas include design and control of multi-axis servo hydraulic systems, smart control and design of hydraulic components.



Mitjan Kalin is full professor and the head of the Department of Tribology and maintenance systems, at the Faculty of Mechanical Engineering, University of Ljubljana. He has written more than 200 international peer-review papers, gave more than 80 invited lectures at international conferences and is project leader of more than 50 major projects. His main research areas are in wear and friction mechanisms of advanced materials, green lubrication technologies and nanoscale interface phenomena. In particular, he is recognised in academia and industry as a pioneer in boundary lubrication and tribochemistry of diamond-like carbon (DLC) coatings. Prof. Dr. Kalin was elected as a member of Slovenian Academy of Engineering, a Member of the International Academy of Engineering, and in 2024 to European academy of science and arts. He received a 2025 ASME Mayo D. Hersey Award, one of the highest academic recognitions in tribology.



Franc Majdič received the PhD in Mechanical Engineering from University of Ljubljana in 2010. Hi is currently working as assoc. professor at the Department of Tribology and maintenance systems, at the Faculty of Mechanical Engineering, University of Ljubljana and is the head of the Laboratory of Fluid power and controls. His research areas include water hydraulics, hydraulic components development and testing. He has written many scientific papers and collaborated in many international conferences. He is a project leader of more than 50 industrial and research projects and collaborated in several industrial patents.

Automated Processing of FMEA-Induced Fault-Injection Simulation for Safety Functions in Dangerous Machinery

Dipl.-Ing. Christa Düsing, Dipl.-Ing. Benjamin Erzberger

Authors affiliation, duesing@xcmg-erc.com, benjamin.erzberger@fluidon.com

Abstract.

Safety-critical and hazardous mobile machines must not only be technically safe, but also require a validation process with increasing demands on process quality, traceability, and reproducibility as the required safety level rises. Since safety functions (SF) often operate close to safety limits, validation by physical testing is associated with high effort and late availability; therefore, systematic simulation-based testing is considered as an early, safe and repeatable alternative. This paper presents a systematic simulation-based validation approach using a digital model that represents the aspects of machine structure and kinematics, drive and control subsystems, and safety-related software algorithms within a physics-based simulation environment. The method is derived from a Safety Function-FMEA (Failure Modes & Effects Analysis) [4, 10] in combination with Failure Modes, Effects & Diagnostic Analysis (FMEDA) [3, 10] to define relevant failure modes, test conditions, and diagnostic expectations. These failures are implemented in the digital model as targeted fault injections via virtual sensors and manipulators, enabling controlled and adjustable experiments for machine-level assessment of system response, failure detection, and diagnostic performance. The implementation is carried out using a digital model [1, 14] in DSHplus® [1, 2] (1D simulation of hydraulics, mechanics and controls) By integrating the digital model into the Cube® platform [1,2], the simulation environment is extended to a digital twin, enabling automated parameter management, test execution, and result evaluation in a reproducible workflow according to established definitions. In addition to verifying logic and algorithms, the approach is aligned with relevant functional safety requirements and processes such as ISO 13849 [8], ISO 19014 [7] and ISO 26262-8, Clause 11 [9], supported by tool assessment and complemented by model validation and plausibility checks. The approach is demonstrated using the safety function for working range limitation of an excavator and shows how digital-twin-based fault injection testing can provide traceable evidence for early, standard-compliant validation before physical prototypes are available or dangerous tests have to be carried out.

Keywords. Simulation-based testing, Fault-Injection-Test, FMEA-induced Failure Modes, Digital model, Digital twins, Cube Platform, Verification & Validation, Safety Assessment

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non-Commercial 4.0 International

1. INTRODUCTION

The development of safety-related parts of control systems (SRP/CS) [8] in mobile machinery with increasing levels of automation requires a systematic evaluation of safety functions to ensure reliable operation in the event of failures. Functional safety is not only a regulatory requirement defined by standards such as ISO 13849 [8], ISO 19014 [7], but also a key contributor to the operational performance and acceptance of automated and (semi-) autonomous mobile machines. With increasing automation, safety functions are required to operate close to physical and functional limits, which significantly increases the complexity of their development and validation. In conventional development processes, the validation of safety functions is often time-consuming, labour-intensive, and usually only feasible in late development stages. This late availability delays the detection of safety-relevant design weaknesses and increases development effort and costs, particularly when design iterations become necessary. At the same time, safety standards demand traceable and reproducible evidence that safety functions meet the required Performance Level (PL) [8], which further increases the pressure on high-quality development and validation processes, especially for highly dangerous machines according to new machinery directive in which notified bodies have to be involved along with the development. In current industrial practice, these challenges are primarily addressed through extensive hardware-based testing and analytical safety assessments such as FMEA [4,10] and FMEDA [3, 10], which are however largely limited to late development stages and component-level evaluation. To address these challenges, automated and simulation-based validation methods are increasingly considered as a means to enable early, repeatable, and cost-efficient testing of safety functions. In addition, such methods must not only reproduce nominal machine behaviour, but also allow systematic investigation of failure conditions in order to support functional safety validation. In this context, safety functions are distributed across multiple technical domains, including machine structure and kinematics, hydraulic or electric actuators, electronic hardware such as sensor, controllers and software algorithms. Ensuring functional consistency across these domains requires validation methods that operate at machine level rather than at isolated subsystem or component level. The approach presented in this paper follows a structured validation process in which Safety Function-FMEA [4, 10] and FMEDA [3, 10] results are transferred into a physical-based digital model of the machine. This digital model is used to execute systematic fault-injection simulations, enabling automated evaluation of safety function behaviour at machine level. By embedding the simulation model into an automated workflow environment, parameter variations, test execution, and result evaluation can be managed in a traceable and reproducible manner. Figure 1-1. illustrates the structured validation process followed in this work.

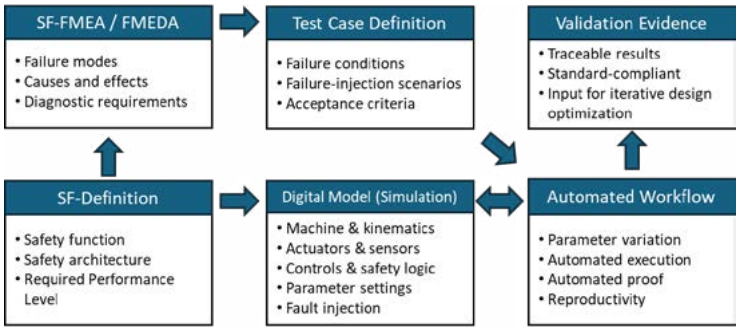


Figure 1-1. Process Flow of the proposed Simulation-based Validation approach for SF

2. EXCAVATOR USE CASE AND SAFETY FUNCTION SCENARIO

The proposed approach is applied to a representative use case of a (semi-)autonomous excavator equipped with the safety function *"Avoid bucket collision with detected obstacle"* for a hydraulic excavator. This function ensures that the bucket movement is automatically stopped before a potential impact occurs and requires a performance level PL = d. The presented simulation model represents the safety-relevant machine behaviour and includes the subsystems kinematics, the hydraulic system, sensors and the control algorithms required to analyse the selected safety function. Figure 2-1 shows in the top the excavator test without failure injections. The sequence illustrates the excavator bucket approaching the obstacle (a), detecting it (b), and stopping automatically before contact (c). This test verifies the correct operation of the implemented safety function under nominal conditions.

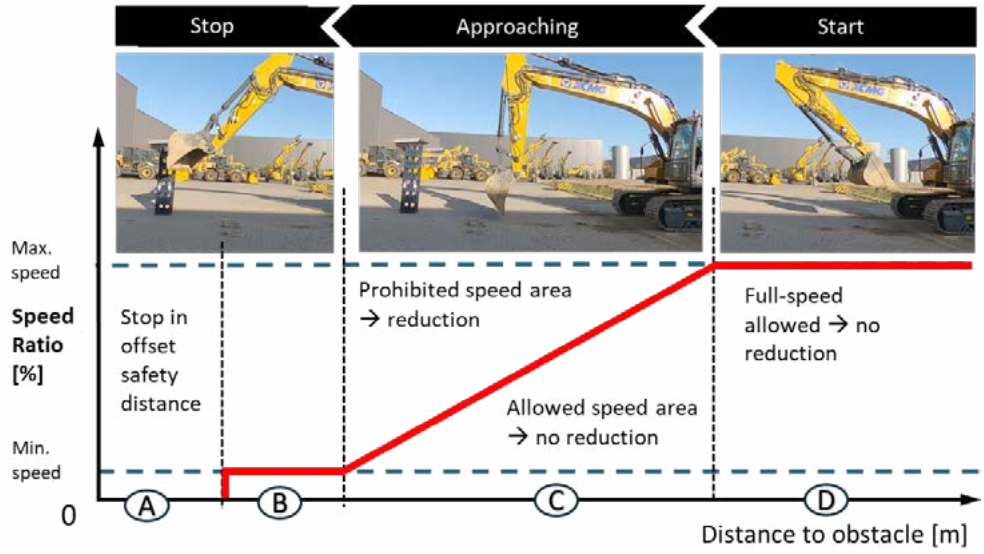


Figure 2-1. Excavator movement limit function: (c) Bucket approaching, (b) entering critical Obstacle Distance, (a) Bucket stopped [11, 12]

The bottom area of Figure 2-1 illustrates the determination of the speed scaling factor for the movement limiter. The speed scale on Y-axis is tracked against the distance to the obstacle on the X-axis. The red line shows the maximum speed curve for each distance. In section D, full speed is allowed and now speed reduction is applied. In the section C, set speed is limited along the red line. Values above the red line are reduced accordingly, values below the red line pass without reduction. Minimum speed is applied in the section B to approach obstacle slowly until safe stop is achieved in a safe offset distance to the obstacle in section A. Based on the operator-defined boundary limits, potential collision points of the work equipment are identified and extended by circular safety offsets this way. The maximum and minimum x- and z-position of these points are assigned to the corresponding boundaries. If multiple boundaries are active, the smallest scaling factor is applied to all drives of the excavator. The resulting scaling factor is multiplied with the set speed input and transmitted to the valve solenoid via the main PLC.

2.1. Possible Failures and Effects

During approach of an obstacle, various failure conditions can affect the safe execution of the movement limiter function. The following failure modes are typical examples derived from the Safety Function-FMEA [4, 10] and do not constitute a complete list:

(A.) **Sensor drift or sensor offset,** i.e. a gradual or constant deviation of the sensor signal from the actual value, can delay obstacle detection, increase the stopping distance or cause a collision;

(B.) **Hydraulic valve malfunction**, e.g. a stuck spool, can prevent the actuator from stopping even when the stop command is issued;

(C.) **Signal loss or signal degradation,** i.e. a temporary or intermittent reduction in signal quality (e.g. noise, dropouts or reduced resolution), can lead to incorrect interpretation of the machine status and uncontrolled movements of the actuator;

These effects directly influence the reliability of the stop function and must be validated to ensure compliance with ISO 13849 [8] and the associated functional safety standards.

2.2. Objective of the Investigation

The objective of this investigation is to evaluate the diagnostic behaviour of the safety function *"Avoid bucket collision with detected obstacle"* implemented by the movement limiter under selected failure conditions derived from the Safety Function-FMEA [4,10]. The safety function limits hazardous machine movements in the relevant directions, with the bucket–obstacle collision used as a representative use case. The focus is on assessing failure detection, system response, and diagnostic coverage using simulation-based fault-injection tests. Test success is evaluated by comparing the achieved diagnostic coverage (DC) with the required DC larger than 90% for PL d in a Category 2 safety function structure.

3. SIMULATION MODEL

3.1. Machine Model

The simulation model represents the excavator including its hydraulic actuators, control logic and virtual sensing system. All safety-relevant components are connected according to a Category 2 structure, consisting of a functional channel and an independent test channel for periodic diagnostic checking, as defined in ISO 13849 [8]. Figure 3-1 shows the simulation model consisting of an excavator model (2) that is moved by the hydraulic control system (1). The excavator model provides information about coordinates of the bucket tip (2B), but also outputs the angular values for each steel part of the attachment, so that sensor raw data can be generated and signals used in the test equipment (4). Sensor data feeds the movement limiter function (3) that can scale the control command (7). The test equipment (B) compares deviations between the redundant sensors signals and can deactivate a hydraulic pilot pressure switch (6) if critical deviations outside the tolerance band of +/- 2°are detected. The model also features an integrated manipulator (A) used for failure injection and signal variation of the arm angle sensor. The functional channel processes the sensor signals of the safety function, including redundant sensor inputs, whereas periodic diagnostic checking is realised via an independent test path in accordance with the applied Category 2 architecture. The manipulator (A) is used exclusively in the simulation to inject

defined failures derived from the Safety Function-FMEA [1, 10] into the signal paths, shown in Figure 3-1 to extend the simulation model for the representation of sensor-related failure scenarios. The hydraulic enable block (C) represents the actuation path for implementing and observing failure effects on the physical system behaviour. This setup supports the systematic assessment of the diagnostic performance by analysing failure detection behaviour and achievable diagnostic coverage under controlled failure conditions, following established principles of model-based diagnosis and failure-tolerant control [15, 16]. The simulation model as such represents a digital model of the machine, as it reproduces structure, behaviour, and failure responses without direct coupling to a physical asset. Figure 3-1 shows the simulation model structure with manipulator (A), test equipment (B), and hydraulic enable path (C) according to a Category 2 architecture.

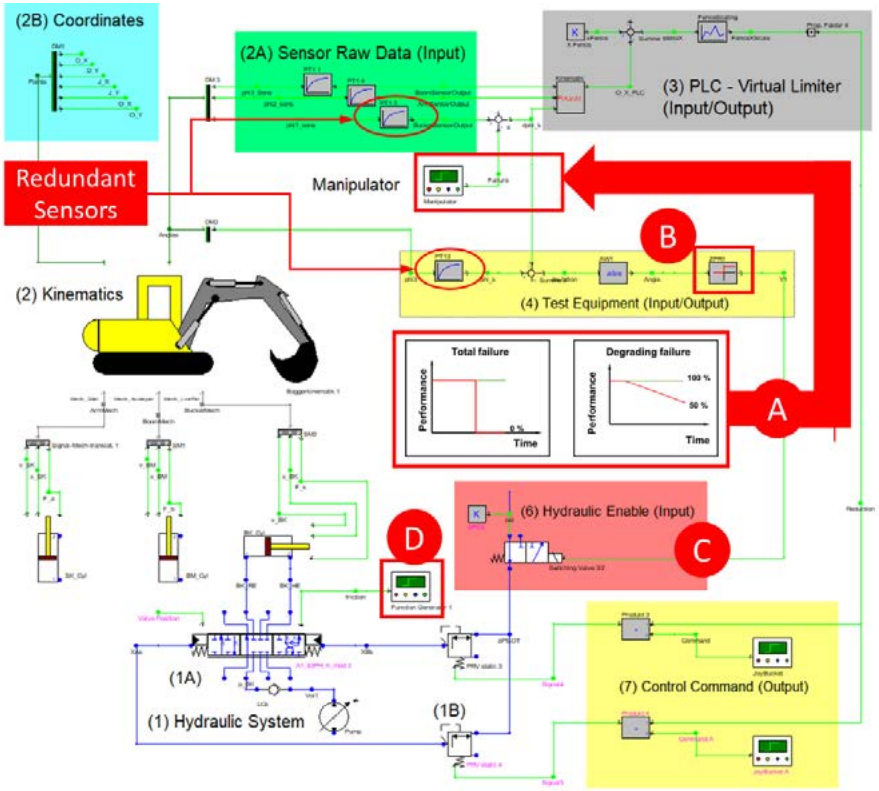


Figure 3-1. Simulation Example of Digital Model with Manipulator for an Excavator Arm representing Category 2 Structure according to ISO 13849 [8, 11, 12]

3.2. Valve Model

The directional control valve (1A in Figure 3-1) is modelled in detail to analyse the influence of mechanical failures on system behaviour. The focus is on the main valve stage, where effects such as increased friction, temperature-dependent spool motion, or a stuck valve spool can occur. Such failures can lead to unsafe machine states if the hydraulic actuation continues even though a stop command is requested. For this purpose, the valve position is derived from the equilibrium of forces on the spool that includes also terms for viscous

friction in the housing bore and a dry friction force that can be externally applied by the force manipulator D. The valve is actuated via electro-hydraulic pilot valves, which generate a differential pilot pressure acting on the ends of the main valve spool. This pilot pressure shifts the spool against spring forces, which in turn controls the hydraulic flow to the actuator. In the simulation model, external manipulation of the spool friction is used to reproduce an undesired stuck condition of the valve spool. The simulation example Figure 3-1 illustrates a failure scenario in which the valve spool remains stuck after a stop command. Although a corresponding closing signal is activated, the hydraulic flow is not fully interrupted and the actuator does not stop as intended. Reproducing this malfunction in the simulation allows the evaluation of control logic responses and virtual sensing mechanisms, as well as the assessment of potential countermeasures such as additional stroke sensing or parameter monitoring.

3.3. Test Procedure derived from FMEA

This section describes the execution of the simulation-based fault-injection tests and the associated evaluation workflow. The excavator simulation starts with arm and bucket cylinder at total extension, so that bucket is close to the cabin and bucket teeth. Both, arm and bucket cylinder are simultaneously retracted, so that arm moves away from the cabin and bucket turns leading to a bucket moving towards the obstacle. Under normal conditions, a movement limiter reduces speeds of arm and bucket with regard to the distance and stops without colliding. Defined failure scenarios are introduced at specific simulation times, and the system response is analysed using consistent evaluation criteria. An overview of the considered failure scenarios and their corresponding test objectives is provided in Table 3.1.

Table 3.1. Simulation-based Fault-Injection Scenarios and Evaluation Objectives for the Safety Function “Avoid bucket collision with detected obstacle” [10, 11, 12, 15, 16]

Injected failure scenario	**Description**	**Diagram reference**
Calibration / degrading failure	Evaluation of the influence of sensor drift on failure detection time and final stop position during bucket motion.	Failure Detection – Drift vs. Movement / Time
Performance test of bucket movement	Assessment of start and stop dynamics and their effect on the timing accuracy of failure detection and controlled stopping behaviour.	Failure Detection – Drift vs. Stop Position over Time
Total failure of inclination sensor	Analysis of the impact of complete inclination sensor signal loss on stop distance, safety margin to the obstacle, and diagnostic response.	Failure Detection – Inclination Sensor Offset
Stop behaviour	Evaluation of the influence of increasing angular offsets on stop distance and diagnostic tolerance within the specified sensor accuracy range.	Failure Detection – Inclination Sensor Angular Offset

Each virtual experiment follows a defined sequence answering the key questions of the test design:

(A.) **Baseline simulation:** The machine is simulated under nominal conditions to establish reference curves for pressure, flow and velocity. This serves as a benchmark for later comparison with failure scenarios;

(B.) **Fault-injection:** A selected failure mode, such as sensor drift, offset, or valve friction is activated at a predefined simulation time using the manipulator module. All test parameters are managed and executed automatically within Cube® [1, 2], ensuring repeatability and efficient data handling.;

(C.) **Detection analysis:** The system response is analysed to determine the safe failure detection time t_{detect}, defined as the interval between failure activation and detection by the diagnostic logic. Diagnostic performance is evaluated using two complementary coverage measures: $DC_{t_failure}$, addressing time-based detection behaviour, and $DC_{x_position}$ addressing the remaining permissible positional margin at the time of detection. This combined assessment supports a differentiated evaluation of diagnostic effectiveness with respect to temporal detection and positional safety, particularly for slowly developing failures;

(D.) **Diagnostic evaluation:** Based on ISO 13849 [8], the Diagnostic Coverage (DC) [8] is used to assess the effectiveness of the implemented diagnostic logic, while the time-resolved evaluation of failure detection follows system-level validation principles as defined in ISO 26262 [9]. The critical detection time $t_{detect.critical}$ is defined as the maximum allowable detection time interval derived from the safety function requirements. DC is calculated as the ratio of this critical detection time to the reference cycle time $t_{cycle,ref}$ and expresses the proportion of failures that can be detected within the allowable detection time interval;

The described evaluation approach enables a quantitative basis for comparing diagnostic quality across different failure types. The systematic approach enables direct comparison of failure scenarios and yields quantitative results on detection speed, reaction behaviour, and overall robustness. By combining DSHplus® [1, 2] with Cube® [1, 2] automated evaluation functions, a large number of test cases can be executed efficiently, forming the basis for the result analysis presented in Chapter 4.

3.4. Data management system structure for automated simulation

The virtual test setup for evaluating the excavator's stop function was realised in a combined simulation and automation environment. The excavator system model, including the machine structure, kinematics hydraulic components, and safety-related logic, was implemented in DSHplus® [1, 2], while parameter management, simulation control, and result evaluation were executed in the Cube® platform [1, 2]. The interaction between the simulation model and the workflow environment is based on a structured exchange of parameters, simulation control signals and result data, enabling automated, repeatable and traceable fault-injection testing of safety-relevant functions. Based on the system structure and data flow shown in Figure 3-2, the automated testing approach relies on a systematic parameterisation of test conditions applied within the simulation model. Test-relevant parameters are provided to the DSHplus® [1, 2] simulation model, where the fault-injection tests are executed and the dynamic system behaviour is calculated. The resulting simulation data are then transferred to the Cube® [1, 2] workflow environment for automated processing, evaluation, and visualisation, ensuring that every test case is executed under

clearly defined and reproducible conditions. This separation between simulation execution and result evaluation allows fault-injection tests to be automated and repeated without manual interaction with the simulation model.

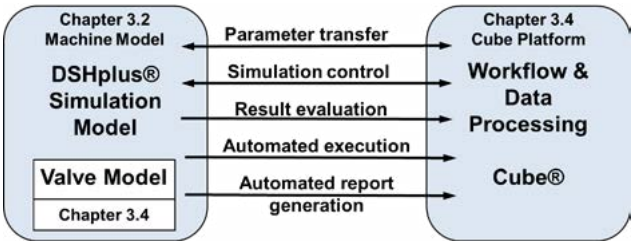


Figure 3-2. System Structure and Data Flow between DSHplus® and Cube® [1, 2]

The parameter set used for automated testing comprises several categories. Failure-related parameters describe the type and characteristics of the injected failure, such as the failure mode, activation time, magnitude, and duration. These parameters are directly derived from the Safety Function-FMEA [4, 10] and FMEDA [3, 10] results and represent the analytical basis for fault-injection testing. Machine operating conditions define the system state at the start of each test, including machine position, velocity, load conditions, and actuator states, ensuring that safety functions are evaluated under representative operating scenarios. In addition, environment-related parameters may be considered, for example the presence of obstacles or moving objects within the working area, if relevant for the investigated safety function. Sensor-related parameters, such as sensor placement or signal availability, allow the influence of sensing configurations on failure detection and diagnostic behaviour to be analysed. Finally, software-related parameters define internal settings of the safety function implementation, including limit values, timing parameters, and logic configurations used for monitoring and failure detection. All parameter variations are defined at workflow level and transmitted automatically from Cube® [1 ,2] to DSHplus® [1, 2] as input data for simulation execution. Simulation control, parameter transfer, and result feedback are handled within the automated workflow, allowing a large number of test cases to be executed efficiently. The resulting data, including system response, detection times, and diagnostic states, are returned to Cube® [1, 2] for structured processing and evaluation.

4. SIMULATION RESULTS

The conducted simulations show how the digital machine model and the automated workflow can be used to analyse functional safety aspects under defined failure conditions. The results illustrate how individual component and system failures influence the diagnostic behaviour of the safety function *"Avoid bucket collision with detected obstacle"*. The evaluation is performed both at virtual machine level and at system level with a focus on hardware-related failures. Based on simulation-based fault-injection, the diagnostic coverage is assessed in a structured and reproducible manner. The automated Cube workflow enables a process-safe and transparent evaluation, comparable to established SIL (Software-in-the-Loop) or HIL (Hardware-in-the-Loop) test procedures and allows efficient iterations in the process, e.g. when modifications in the algorithms become necessary. It can be used to automatically generate diagrams and key-figures that support the approval process. In addition, the simulation results are compared with the original data-

sheet-based evaluation approach, allowing additional insights and recommendations for the assessment of diagnostic coverage to be derived.

4.1. Investigation of Sensors Failures

Sensor-related failure effects on the safety function are analysed based on time-dependent signal behaviour. Characteristic signal trajectories of sensor output, hydraulic enable signal, cylinder position and actuator pressure are evaluated to illustrate the interaction between sensor failures, diagnostic logic, and system response. This allows a clear distinction between normal operation, detected failures and undetected failures, as well as the assessment of delayed failure recognition and its impact on the stop response. To represent safety-relevant sensor failures derived from the Safety Function FMEA, one of the redundant inclination sensor signals is manipulated. The applied failure patterns include time-dependent drift and step-like deviations, while the second redundant sensor provides the nominal reference signal. The diagnostic logic monitors the deviation between both signals and classifies a failure as detected once the predefined diagnostic threshold is exceeded. The resulting system behaviour is illustrated in Figure 4-1.

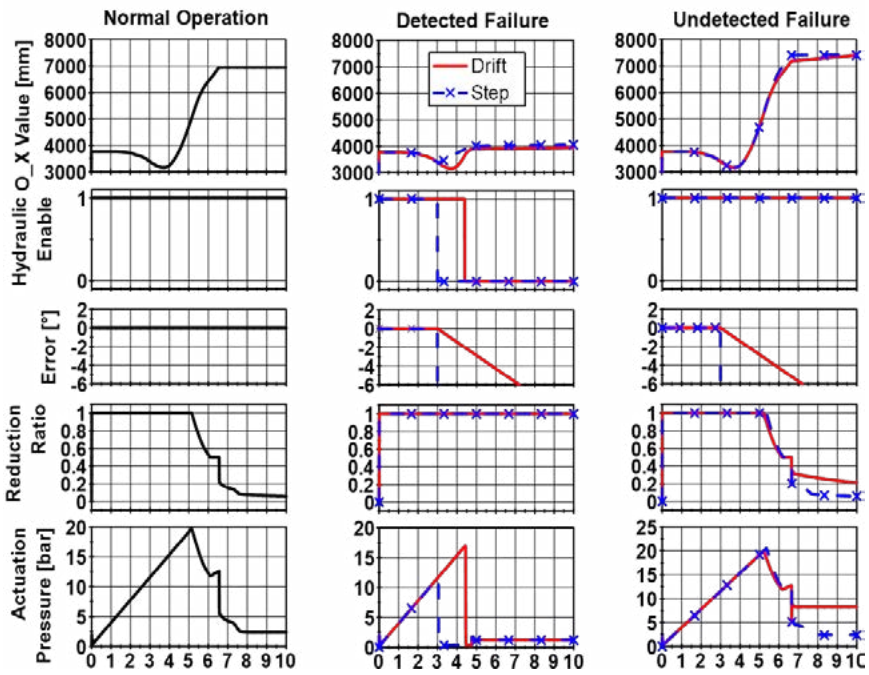


Figure 4-1. Sensor Drift, normal Operation, detected & undetected Failures [11, 12, 15, 16]

Normal operation has no error and applies the reduction ratio from the movement limiter to stop the bucket at 7000 mm in a reference cycle time $t_{cycle,ref} = 6.6\,sec.$ In the detected failure case, the injected deviation exceeds the diagnostic threshold, leading to deactivation of the hydraulic enable signal and thus, immediately stops movement at 4000 mm. Undetected failures remain below the threshold and therefore do not trigger the stop command, despite deviation of the manipulated sensor signal from the nominal signal. The failure causes a stop exceeding 7m and thus causes a collision with the obstacle. The results demonstrate the influence of failure severity on detection capability and confirm the

suitability of the simulation-based approach for assessing diagnostic coverage for sensor-related failures.

4.1.1 Sensor Drift

Figure 4-2(a) shows the bucket position at failure detection and the corresponding stop position for different sensor drift rates. Higher drift rates lead to earlier detection and shorter stopping distances, ensuring collision avoidance. Medium drift rates are detected sufficiently early but operate closer to the defined safety limit. Very small drift rates result in late detection or remain within tolerance, reducing the available safety margin. The results highlight the relevance of defined tolerance bands for achieving sufficient diagnostic coverage. The left-shift of the −0.5 °/s curve results from late detection at small bucket displacement.

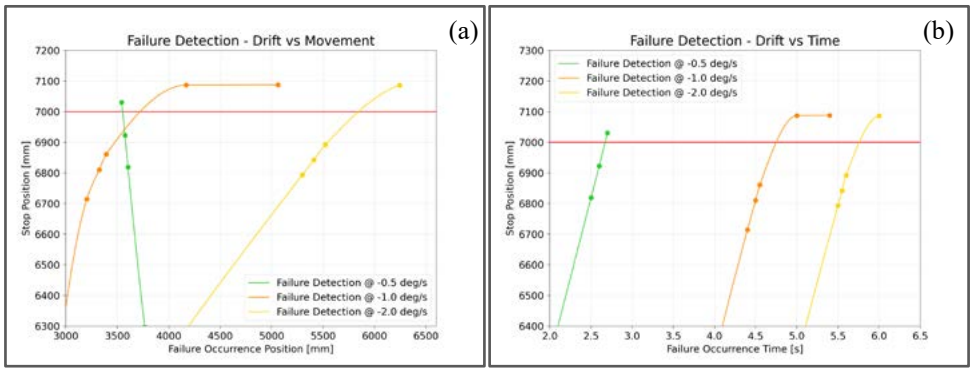


Figure 4-2. Failure Detection Characteristics [11, 12, 15, 16]

Figure 4-2(b) relates the time of fault-injection to the bucket position at stop. Faster drift rates cause early detection shortly after failure onset, whereas slower drifts lead to delayed detection closer to the obstacle. This behaviour demonstrates the time-dependent nature of drift faults and underlines the criticality of slow sensor drifts for diagnostic algorithms. For the automated data processing, the time-based diagnostic coverage $DC_{t_failure}$ is calculated according to equation (4.1) using the simulation-derived detection timing for the failure setting crashing into the obstacle and the reference cycle time $t_{cycle,ref}$:

$$\mathrm{DC_{t_failure}} = \frac{\mathrm{t_{failure\ occurance,crash}} + \Delta \mathrm{t_{detect}}}{\mathrm{t_{cycle,ref}}} \times 100 \qquad (4.1)$$

In addition, a position-based diagnostic coverage $DC_{x_position}$ is introduced. The $DC_{position}$ describes the ratio of the position at which failure occurrence leads to a crash and the position of the obstacle and is defined as equation (4.2):

$$\mathrm{DC_{x_position}} = \frac{\mathrm{x_{failure\ occurance,crash}}}{\mathrm{x_{obstacle}}} \times 100 \qquad (4.2)$$

The quantitative results for the investigated drift cases are summarised in Table 4.2. The results show that the two DC values differ from each other, but also that the demanded rate of 90% and higher could not be performed by the current design and algorithm of the movement limiter. A higher absolute drift rate corresponds to a more severe sensor failure, leading to a faster deviation of the perceived inclination from the actual machine state,

whereas smaller drift rates represent slowly developing failures that are typically more challenging to detect in time. The selected drift rates of −2°/s, −1°/s, and −0.5°/s therefore allow the diagnostic behaviour to be evaluated under clearly differentiated failure dynamics, while all other boundary conditions remain unchanged. This ensures that differences in detection time and diagnostic coverage can be directly attributed to the severity of the injected sensor drift.

Table 4.1. Quantitative Results for different Inclination Sensor Drift Rates obtained from Simulation-based Fault-Injection Test for a reference Cycle Time of 6.6 sec

Case	Drift rate [°/s]	Δt_{detect} [s]	Δt_{detect} [s]	$DC_{tfailure}$ [%]	x [mm]	$DC_{xposition}$ x [%]
D1	-2°/s	5.6	0.89	98.33	5750	82.14
D2	-1°/s	4.6	1.42	91.21	3600	51.43
D3	-0,5°/s	2.6	2.93	83.78	3600	51.43

The results in Table 4.1 show a clear relationship between drift severity, detection time, and diagnostic coverage. Larger drift rates lead to shorter detection times, as the sensor signal reaches the diagnostic decision limits more rapidly. Consequently, the proportion of the admissible reference cycle time during which the failure is detected increases, resulting in higher DC_{time} values. In contrast, smaller drift rates result in a more gradual signal deviation and longer detection times. This reduces the available safety margin before the permissible displacement is exceeded and leads to lower diagnostic coverage. The complementary metric $DC_{position}$ further highlights how early failure detection preserves spatial safety margins, underlining the criticality of slow sensor drifts. The result illustrates that the diagnostic effectiveness is not only determined by the implemented logic itself, but also strongly depends on the temporal characteristics of the failure itself. This underlines the importance of time-resolved, system-level fault-injection testing for evaluating diagnostic coverage beyond static or datasheet-based assumptions.

Figure 4-3 illustrates the relationship between the sensor drift rate and the time to failure detection Δt. Failure detection occurs once the predefined ±2° tolerance band is exceeded (to be verified in the simulation). The results show that higher drift rates are detected earlier, while slow drift rates lead to significantly delayed failure detection. The depicted trend follows a nonlinear inverse relationship.

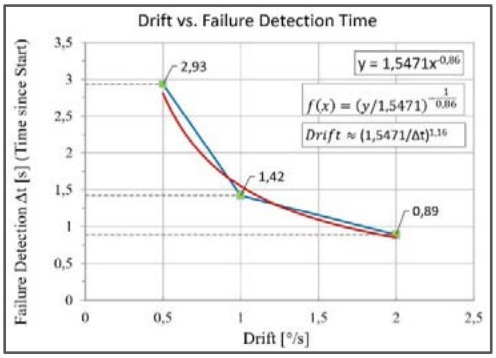


Figure 4-3. Failure Detection Characteristics [11, 12, 15, 16]

In addition, Figure 4-3 shows clearly that high drift rates are uncritical, as detection occurs early and sufficient safety margins remain. In contrast, slow sensor drifts are critical, since failure detection happens late and close to the safety limit, reducing the available stopping distance. This behaviour of slowly developing sensor drifts motivates the introduction of an intermediate Zone B in the safety algorithm, where operation continues at reduced speed before a safe stop in Zone A with a defined distance to the obstacle is initiated. Figure 4-3 therefore emphasizes the importance of defined tolerance bands and staged safety reactions to ensure adequate diagnostic coverage for slow sensor drifts.

4.1.2 Total and offset failures

Figure 4-4 illustrates the influence of inclination sensor offset failures on the diagnostic behaviour of the safety function in accordance with the principles of ISO 13849 [8] and ISO 19014 [17]. In contrast to drift failures, an offset represents a static deviation and leads to a constant discrepancy between the redundant inclination sensor signals. Figure 4-4(a) shows the resulting bucket stop position for angular offsets of 3° started at different time instants during the bucket approaching the obstacle. As the offset does not increase over time, failure detection is determined by the absolute deviation relative to the defined tolerance band of +/- 2°. Variations in the resulting stop position are caused by the system state at the time of fault-injection. The distance shown represents the absolute bucket position relative to the obstacle highlighted by the red line. Figure 4-4(b) presents the evaluation for increasing offset magnitudes. Offset values within the permissible angular tolerance band of +/- 2° are handled by plausibility monitoring in the test equipment of the safety function and do not lead to a safety reaction. The system fails to react for negative offset values as stop positions are beyond the limit of 7000 mm. Positive offset values bring the system to the safer side and stop in front of the obstacle instead. Once the offset exceeds the predefined diagnostic threshold, the inconsistency between both sensor signals is detected and the stop command is initiated. This behaviour is consistent with the diagnostic principles defined for Category 2 architectures according to ISO 13849 [8], where systematic sensor failures are detected by periodic or continuous plausibility checks. The results indicate that the applied tolerance band method in the movement limiter may lead to stop position deviations of up to 100 mm, therefore an additional safety offset distance in stop area Zone A, Figure 2-1, is recommended.

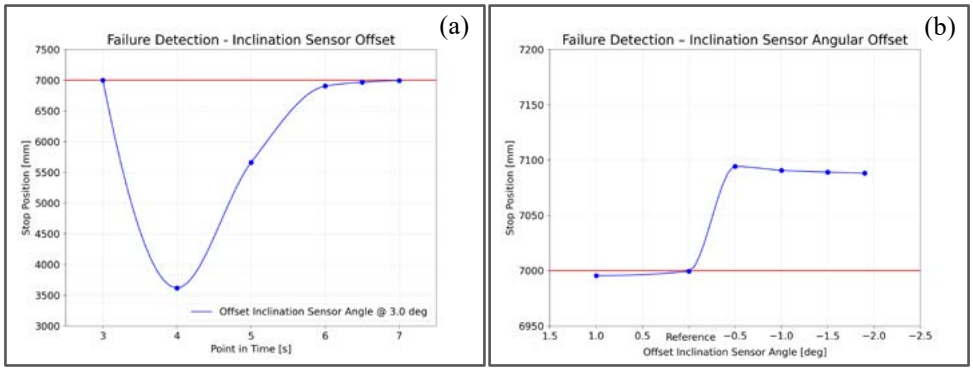


Figure 4-4. Total Failure of Inclination Sensor & Inclination Sensor Angular Offset [11, 12, 15, 16]

4.2. Valve spool behaviour

The results for wrong valve actuation behaviour and an active movement limiter are illustrated in Figure 4-5. In the friction scenario shown in Figure 4-5a, the valve spool opens to 40% that only allows a slow movement of the bucket while approaching the obstacle. The friction in the valve spool gap delays the valve closing after the stop demand at 4 sec. As a consequence, the valve spool stroke decreases progressively until the actuator motion comes to a standstill at an effective stroke of approximated 20% that is below the positive overlap of the metering edge geometry. The valve closing time of few seconds indicate, that the such frictions can become more serious when the valve already has higher openings before stop command is initiated. In contrast, the stuck piston failure illustrated in Figure 4-5b represents a mechanical blockage of the valve spool that avoids the valve closing. The stop command at 5 sec cannot be executed by the valve spool. The valve opening remains at 60%, so that the hydraulic flow is still further transmitted to the actuator. As a result, the bucket does not stop in front of the obstacle. Both behaviours indicate a total loss of controllability and a decoupling between the control signal and the physical actuator response.

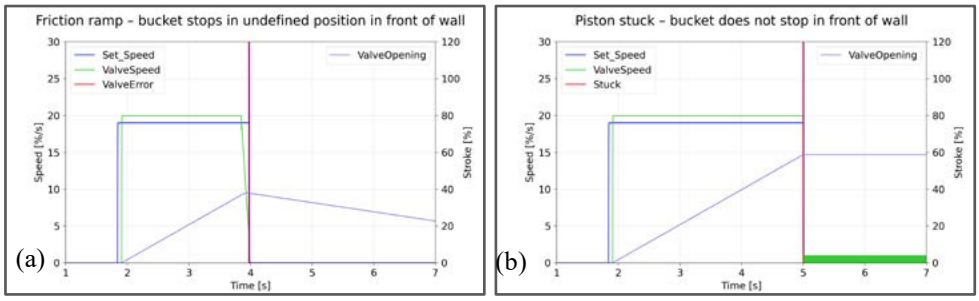


Figure 4-5. Comparison of friction-induced and mechanically blocked piston behaviour in the hydraulic valve section [11, 12, 15, 16]

The resulting bucket stop behaviour is shown in Figure 4-6(c) and Figure 4-6(d). In the friction ramp scenario shown in Figure 4-6(c), the reduced actuator speed leads to a controlled deceleration of the bucket, which comes to rest in front of the obstacle.

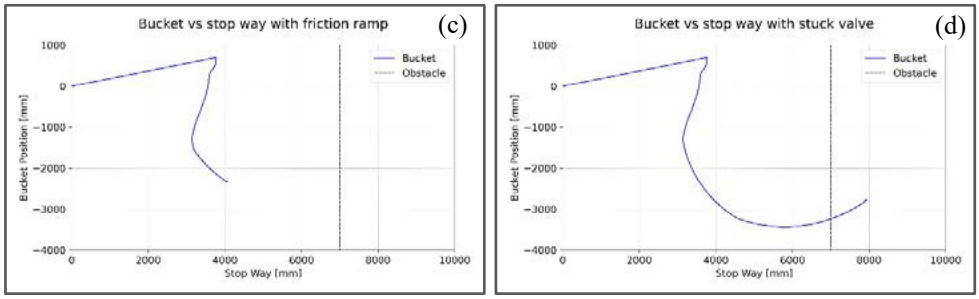


Figure 4-6. Bucket stopping behaviour under valve failure behaviour

In contrast, the stuck valve failure shown in Figure 4-6(d) causes continued bucket motion beyond the defined stopping position. The movement stops in the mechanical end position of the excavator attachment. This behaviour indicates that the hydraulic flow through the

valve is not interrupted, despite the presence of a stop command, and illustrates a safety-relevant failure mode in which unintended motion persists due to clamping of the valve.

4.3. Evaluation

The simulation results show that the implemented diagnostic concept reliably detects sensor drifts, signal offsets and hydraulic malfunctions before the defined safety limit is exceeded. The Category 2 architecture [8] with redundant sensors which is consistent with the functional safety objectives for earth-moving machinery as defined in ISO 19014 [7], increases failure tolerance and enables early detection of degrading behaviour. The combination of DSHplus® [1, 2] and Cube® [1, 2] allows systematic execution and evaluation of extensive test series while maintaining traceability. To place the obtained diagnostic coverage results in a broader context, Table 4.2 compares the proposed simulation-based DC evaluation with conventional datasheet-based approaches. The comparison highlights differences in scope, validation depth, and the integration of FMEA-derived failure modes [4, 10], illustrating the added value of time-resolved, system-level fault-injection-tests for assessing diagnostic effectiveness.

Table 4.2. Comparison of datasheet-based & simulation-based DC evaluation

Aspect	**Datasheet-based Evaluation**	**Simulation-based Evaluation (Fault-Injection-Test)**
Scope	Component Level	System level (including software logic)
Basis	Static manufacturer data (MTTFD, DC values)	Dynamic failure behaviour under realistic conditions
Focus	Hardware reliability	Algorithmic failure detection and reaction
Detection proof	Assumed via design parameters	Verified through time-resolved detection response
Performance Level Validation	Theoretical	Practical validation within virtual environment
Limitation	Cannot assess control algorithm or reaction time	Requires modelling and simulation effort
Added value	Quick estimation	Realistic validation of functional safety behaviour of the machine
Integration of FMEA results	Not considered - datasheet assumes ideal operation	FMEA-identified failure used for targeted fault-injection tests
Phase in V-Model	Hardware validation	System and machine validation

4.4. Cube-Based Evaluation

The simulated test cases are evaluated using the automated Cube® [1, 2] workflow environment, which coordinates simulation execution, data processing, result visualisation, and documentation in a predefined and reproducible sequence. In combination with the automated Cube® workflow for execution, parameter management, and result evaluation, the virtual twin is extended to a digital twin [1, 14] representation in accordance with the concepts described in the literature [14] and their implementation in an automated simulation and workflow [1]. Coupling the DSHplus® [1, 2] simulation model with the Cube® workflow enables automatic execution of parameter variations and failure scenarios

under consistent test conditions. The evaluation shows a substantial reduction in manual engineering effort, while the numerical simulation runtime per test case remains comparable for manual and automated execution. The observed efficiency gain is achieved by eliminating repetitive manual tasks such as parameter configuration, data storage, post-processing, and result documentation rather than by accelerating the simulation itself. Individual simulation runs require only a few seconds, allowing parameter studies with numerous failure scenarios to be completed within minutes. In contrast, equivalent real-machine tests typically require several minutes per test case due to machine setup, repeated execution, and data acquisition, which limits the number of scenarios that can be investigated. In a manual simulation workflow, the surrounding engineering activities typically dominate the total effort per test case, whereas the Cube®-based execution performs these steps deterministically without user interaction. In addition, the automated workflow improves reproducibility and result quality by applying all test parameters consistently and eliminating handling errors associated with repeated manual testing. The simulation environment further allows the investigation of critical or hazardous scenarios, such as collision conditions or complete failure of safety functions, which cannot be systematically explored in real-machine experiments. Overall, the Cube-based evaluation provides a reproducible framework for system-level assessment of diagnostic performance, enabling higher test coverage with significantly reduced engineering effort compared to conventional manual or real-machine testing. The benefit of the Cube®-based approach therefore lies not in faster simulation, but in the systematic reduction of manual effort and variability in the evaluation process.

5. SUMMARY AND CONCLUSION

This work shows that simulation-based fault-injection testing can be used to evaluate diagnostic behaviour of safety-related control functions in a time-resolved and system-level manner. The purpose of the presented approach is not the simulation itself, but the generation of early and reproducible evidence for diagnostic performance under defined failure conditions, which is difficult to obtain by conventional testing methods. The results demonstrate that diagnostic coverage is strongly influenced by the temporal characteristics of failures and by the domain in which failures occur. Slowly developing sensor degradations and hydraulically induced effects may lead to delayed detection, even if the implemented diagnostic logic is functionally correct. This highlights that diagnostic effectiveness cannot be assessed solely on a structural or datasheet level, but requires dynamic evaluation under realistic operating conditions. From a development perspective, the approach supports decision-making by shifting extensive diagnostic validation from late-stage physical testing to an earlier simulation-based phase. This enables systematic comparison of design variants, monitoring strategies, and parameter settings before hardware availability, while maintaining traceability to FMEA-derived failure assumptions. In summary, the presented methodology provides a means to validate diagnostic concepts with respect to timing, coverage, and system interaction, thereby supporting robust safety design and reducing the risk of late-stage design iterations. Although the movement limiter already provides a high level of operational safety, the results indicate that the dynamic diagnostic coverage identified by the FMEA-based digital model cannot yet fully satisfy the performance requirements. Consequently, the simulation results reveal the need for further refinement of both the movement limiter algorithms and the associated diagnostic strategy.

6. REFERENCES

[1] R. Alt, et. al., 'Seamless Integration of Device and Field Data into the System Simulation of a Hydraulic Servo-Press using AAS', 14th Int. Fluid Power Conf. (14th IFK), pp. 553-569, Dresden, March 2024.

[2] H. Baum, 'Cube – A SaaS development environment for technical system', Drivetrain and System Engineering Conf. (DSEC), Aachen, March, 2025.

[3] International Standard, 'DIN EN ISO 61508-5:2011-02. Functional safety of electrical/ electronic/ programmable electronic safety-related systems – Part 5: Examples of methods for the determination of safety integrity levels;

[4] Ch. Düsing, D. Prust, 'Supplementary failure mode and effect analysis (FMEA) for safety application standards DIN EN ISO 135849 Safety Function-FMEA', 12th Int. Fluid Power Conf. (12th IFK), pp. 349-358, Dresden, Oct. 2020.

[5] Ch. Düsing, 'Integration of a Safety Function-FMEA into a Professional FMEA-Software Tool to meet the New Draft of ISO 13849', 13th Int. Fluid Power Conf. (13th IFK), pp. 297-308, Aachen, June, 2022.

[6] International Standard, 'DIN EN ISO 12100:2011-03, Safety of machinery - General principles for design – Risk assessment and risk reduction', DIN Media Verlag, Berlin, March 2011.

[7] International Standard, 'DIN EN ISO 19014-1:2025-5 Earth-moving machinery – Functional safety – Part 5: Tables of performance levels', DIN Media Verlag, Berlin, Sept.2025

[8] International Standard, 'ISO 13849-1:2023-04, Safety of machinery – Safety related parts of control systems – Part 1: General principles of design', DIN Media Verlag, Berlin, April 2023.

[9] International Standard, 'ISO 26262-4:2018-12 Road Vehicles – Product development at the system level', DIN Media Verlag, Berlin, December 2018

[10] International Standard, 'IEC 60812:2018-08, Failure modes and effects analysis (FMEA and FMECA)', DIN Media Verlag, Berlin, August 2018

[11] Ch. Düsing, M. Inderelst, 'Evaluation of the Diagnostic Coverage for safety-relevant components in automated drive systems for mobile construction machinery', 9th Symposium Construction Machinery, pp. 199-217, Dresden, Sept. 2022.

[12] Ch. Düsing, F. Will, 'Safety Function Failure Mode and Effect Analysis (S-FMEA) – A Novel Approach of FMEA for Safety Application in Mobil Working Machinery, Global Fluid Power Society (GFPS), pp. 157-182, Springer Nature, 2025. ISBN 978-3-031-84505-5

[13] IFA Report 1/2025, 'Functional safety of machine controls – Application of DIN EN ISO 13849' Online Version, Germany, May 2025.
ISBN-978-3-948657-66-6

[14] Werner Kritzinger, Digital Twin in manufacturing. A categorical literature review and classification, IFAC Papers OnLine, June, 2018, Bergamo, Italy, pp. 51-11, DOI: 10.1016/j.ifacol.2018.08.474

[15] R. Isermann, Fault-Diagnosis Systems – An Introduction from Fault Detection to Fault Tolerance, Springer Verlag, Berlin, 2006. ISBN-978-3-540-30368-8, https://doi.org/10.1007/3-540-30368-5

[16] Blanke et al., Diagnosis and Fault-Tolerant Control, Springer Verlag, Berlin, October, 2016. ISBN-978-3-662-47943-8

Biographies



Christa Düsing received a diploma in mechanical engineering from Rheinische Hochschule Köln (RH Köln). She is currently FMEA coordinator and certified FMEA moderator (DGQ) at XCMG European Research Center GmbH in Krefeld. She is responsible for conducting FMEA analyses of mobile working machines, including mobile cranes, graders, and loading cranes. In addition, she provides FMEA training for the XCMG team across the group worldwide. She has extensive experience as a development engineer focusing on the optimization and further development of hydraulic systems and components for mobile machinery. Since September 2020, she has been pursuing an external doctorate at the Technical University of Dresden.



Benjamin Erzberger received a diploma in mechanical engineering from RWTH Aachen University in 2012, and has since worked for FLUIDON GmbH. He is mostly involved in hydraulic system simulation and pressure pulsation problems. This involves stationary systems like forging presses or articulated bridges as well as mobile applications, reaching from cranes, excavators and agricultural machinery to airplanes. Apart from enjoying fixing stability issues of load control valves he has developed a special fondness of solving acoustic problems in all kinds of hydraulic systems.

Hydraulic Spool Valve Contamination Susceptibility Test and Visualization Study of Clearance Particle Ingress-Aggregation Phenomenon

Chen Qianpeng, Ji Hong, Xie Huan, Ding Ling

School of Energy and Power Engineering, Lanzhou University of Technology, Lanzhou, 730050, China

Key Laboratory of Special Pumps, Valves and Flow Control Systems, Ministry of Education, Lanzhou 730050, China

chenqianpenglut@163.com

Abstract.

During hydraulic equipment service, solid particles infiltrate the oil due to component wear or external contamination, leading to oil contamination. The ingress of contaminated particles into spool valve clearances can cause abrupt increases in spool friction force, resulting in control valve malfunctions and even catastrophic hydraulic system failures. This study focuses on a threaded cartridge directional valve, constructing a contamination susceptibility test rig to investigate the effects of particles with varying contamination levels (10-300 mg/L) and size ranges (0-3 μm, 0-5 μm, 0-10 μm, 0-20 μm, 0-30 μm) on spool friction. It was found that for a unilateral clearance of 5-10 μm, "sensitive particles" (0-10 μm, comparable to the clearance size) induced the maximum electromagnetic driving force on the spool, reaching 150.5 N at 200 mg/L—5 times greater than in clean oil. To unveil the microscopic mechanism of friction escalation caused by sensitive particles, a visualized model experiment was designed, capturing the complete process of particle ingress and aggregation within the clearance using high-speed cameras. Finally, the particle ingress rate in the clearance was calculated via the Discrete Phase Model (DPM), and three optimized structures were proposed based on visualization and simulation results, reducing particle ingress by 80% and mitigating spool sticking risks. This study elucidates the macro-phenomena and micro-mechanisms of "sensitive particle"-induced spool sticking while providing improvement strategies, offering experimental methodologies and theoretical foundations for high-reliability hydraulic spool valve design and contamination control.

Keywords. hydraulic spool valve, friction force, sensitive particles, visualization, structural improvements.

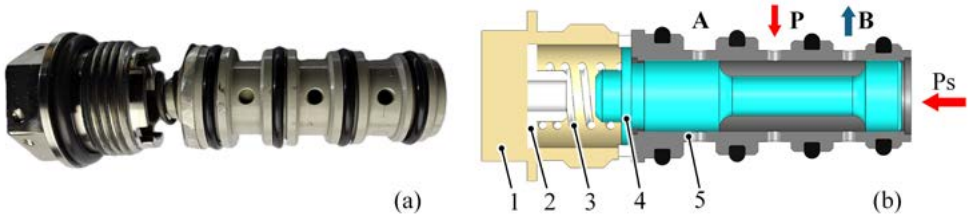
All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Cartridge valves are widely used in hydraulic control systems of engineering machinery and marine equipment due to their compact structure and ease of installation [1]. Among them, spool-type cartridge valves employ high-precision spools with a typical unilateral clearance of 5–10 μm, which provides sealing and guidance functions. As a typical configuration of control valves, the spool features a long cylindrical shaft structure that regulates the flow direction and throttling area of the fluid, thereby determining key parameters such as flow rate and pressure[2-5]. Under high temperature, high pressure, and heavily contaminated conditions, the spool is susceptible to sticking caused by multiphysics coupling effects, which severely compromises control accuracy and response performance[6-9]. Therefore, conducting experimental studies on contaminant-induced jamming and visualizing particle migration and accumulation in spool valves will contribute to improving the theoretical framework of spool design.

Research indicates that spool valve sticking is primarily caused by radial unbalanced forces, non-uniform thermal deformation, and particle intrusion into the clearance. Lin et al. systematically analyzed the mechanisms of these three factors and proposed corresponding suppression strategies; their numerical study based on a solid-liquid two-phase flow model further revealed the migration and sticking patterns of particles within the clearance[10, 11]. Qian et al. found that particles tend to accumulate in the pressure balancing groove regions, and the groove geometry affects the magnitude of the sticking force[12]. Among different groove shapes, the triangular groove demonstrated the best anti-sticking performance due to its minimal fluid deflection and the tendency for particles to settle at the bottom of the groove. Brazhenk experimentally determined the influence of particle size and concentration on the sticking and wear of hydraulic valves, noting that aerospace hydraulic valves are particularly sensitive to certain particle sizes[13]. Furthermore, numerous scholars have experimentally and numerically investigated particle migration, retention, dynamics, and force characteristics in fluid media[14-17].

In terms of visualization studies, Jianjun Zhang et al. employed fluid-structure interaction simulations and high-speed photography of enlarged models to reveal the dynamic process of micron-sized particles entering the clearance, validating the effectiveness of the numerical approach[18]. Xinqiang Liu et al. conducted microscopic observation and simulation analysis to systematically investigate particle trajectories and erosion characteristics at the throttle orifice of servo valves, clarifying the influence of pressure difference, particle shape, and size on erosion behavior. These studies provide important insights for understanding the microscopic movement mechanisms of particles and for structural optimization [19][20].



1 screw cap; 2 spring seat; 3 spring; 4 valve spool;5 valve sleeve
(a) Physical diagram of the spool-type directional control valve (b) 3D Structural diagram of the spool-type directional control valve

Figure 1.1. Physical and Three-Dimensional Structural Diagrams of the Spool-Type Directional Control Valve

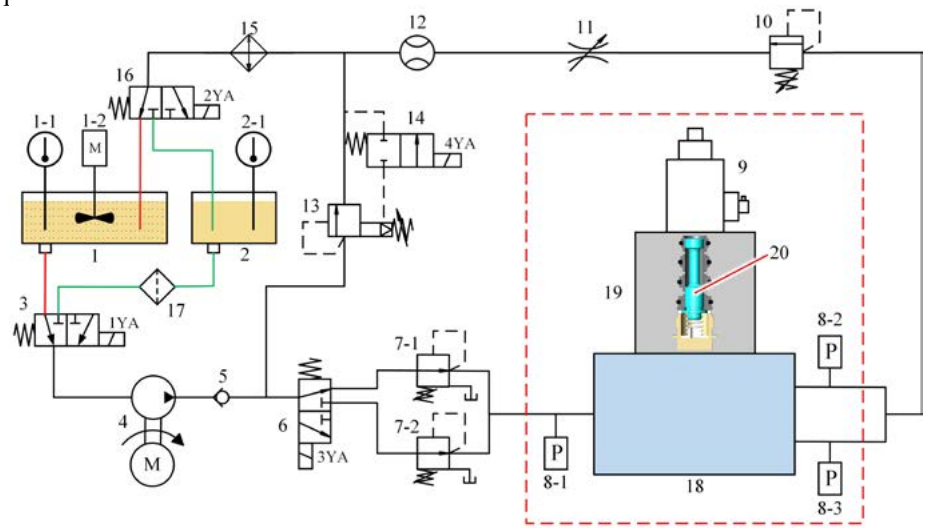
The studied object is a spool-type directional valve used to regulate hydraulic circuits in a hydraulic system. Its structure and operating principle are shown in Figure. 1.1. The valve consists of a screw cover, spring seat, spring, sleeve, and spool. It is a two-position, three-way valve with a pressure inlet (P) and two outlet ports (A and B). When the actuation force (Ps) is not applied, the spool remains in the right position due to the spring force, connecting P to B. When Ps is applied, the spool moves to the left against the spring force, thereby connecting P to A.

2. MEASUREMENT OF CONTAMINATION-INDUCED FRICTION FORCE IN SPOOL-TYPE DIRECTIONAL CONTROL VALVES

2.1. Design of the Experimental Plan

2.1.1. Test Principle, Working Medium, and Selection of Critical Components

Figure 2.1 shows the schematic diagram of the component contamination susceptibility test. Pressure and flow rate can be adjusted by controlling the throttle valve and the electric motor. A switching mechanism between the contaminated oil tank and the clean oil tank is adopted: the contaminated tank is used to evaluate the external characteristics of the valve under working media of different contamination levels, while the clean tank serves to flush out solid particles remaining in the system, thereby facilitating subsequent test cycles. A water-cooled oil cooler is employed to provide forced heat exchange, dissipating heat generated during system operation and minimizing the impact of temperature on the experimental results.



1 Contaminate oil tank; 2 Clean oil tank; 1-1, 2-1Temperature sensor; 1-2 Particle agitator; 3,6,16 Two-position three-port directional control valve; 4 Motor-pump unit; 5 Check valve; 7-1,7-2 Fixed pressure reducing valve; 8-1, 8-2, 8-3, Pressure sensor; 9 Proportional solenoid + displacement sensor (PM + LVDT); 10 Relief valve; 11 Adjustable throttle valve; 12 Flow meter; 13 Safety valve; 14 Unloading valve; 15 Cooler; 17 Filter; 18 Base manifold; 19 Test manifold; 20 Test valve

Figure 2.1. Schematic diagram of the external characteristic test for contamination sensitivity

For other key components in the system, such as the hydraulic pump, an Italian HAWK high-pressure piston pump, model NHD1525L, was employed. This pump features ceramic pistons, which offer high hardness and strong contamination resistance. Its inlet filtration accuracy is 150 μm, allowing it to accommodate the particle sizes introduced in the experiments. The high contamination resistance makes it suitable for contamination susceptibility tests on this platform, as the addition of particles to the working fluid has minimal impact on the pump's performance and does not affect the final test results of the valve. To further minimize the influence of particles introduced into the system during experiments, all tubing and valve components were made of stainless steel.

Given the physical properties of aviation kerosene at 20°C – density range 775–830 kg/m³ and kinematic viscosity 1.25 mm²/s – an ethylene glycol aqueous solution was formulated to match these characteristics. Experimental curve fitting of kinematic viscosity versus glycol concentration at 20°C (Figure 2.2) revealed that an 11.37% ethylene glycol solution by mass exhibits the closest match to the kerosene's properties. Consequently, this 11.37% concentration was ultimately selected as the working fluid substitute.

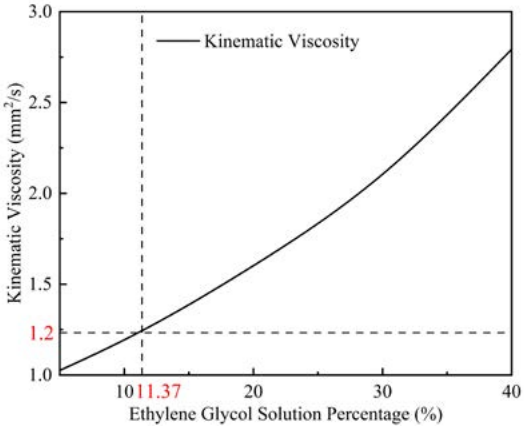


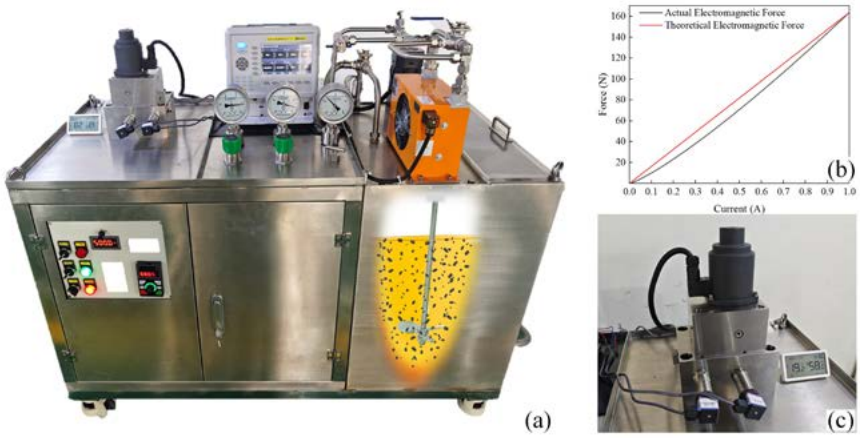
Figure 2.2. Kinematic Viscosity of Ethylene Glycol Aqueous Solutions at Varying Concentrations

2.2. Test Bench Construction and Experimental Results Analysis

2.2.1. Test Bench Construction

The test bench model is shown in Figure. 2.3(a). It features an integral frame structure with a triple-compartment tank arranged on the right side. From back to front, these are the contaminated oil tank, the clean oil inlet tank, and the clean oil return tank. The contaminated oil tank is equipped with an agitator operating at 60–200 r/min, suitable for mixing low-viscosity fluids. The agitator shaft and impeller are made of 304 stainless steel, and a thermocouple is installed in the tank to monitor the oil temperature in real time. Components such as the cooler, agitator, and manual directional valve are mounted on top of the tank. The operating panel is equipped with an adjustable shut-off valve, two pressure-reducing valves, pressure gauges, and a paperless recorder. The test valve block is placed separately on the left side for easy disassembly and installation. The lower cabinet contains the electric motor, frequency converter, flow meter, and solenoid directional valve. To measure the friction force during the movement of the spool in the directional valve, a proportional

solenoid with a rated stroke of 8 mm was selected. Within its rated stroke range, the thrust force is proportional to the current supplied to its coil. The stroke-force characteristic curve of the solenoid is shown in Figure. 2.3(b). Within the rated stroke range, the variation in stroke has almost no effect on the force.



(a) Test Bench Model; (b) Solenoid Current-Force Characteristic Curve; (c) Physical Views of the Base and Test Manifolds

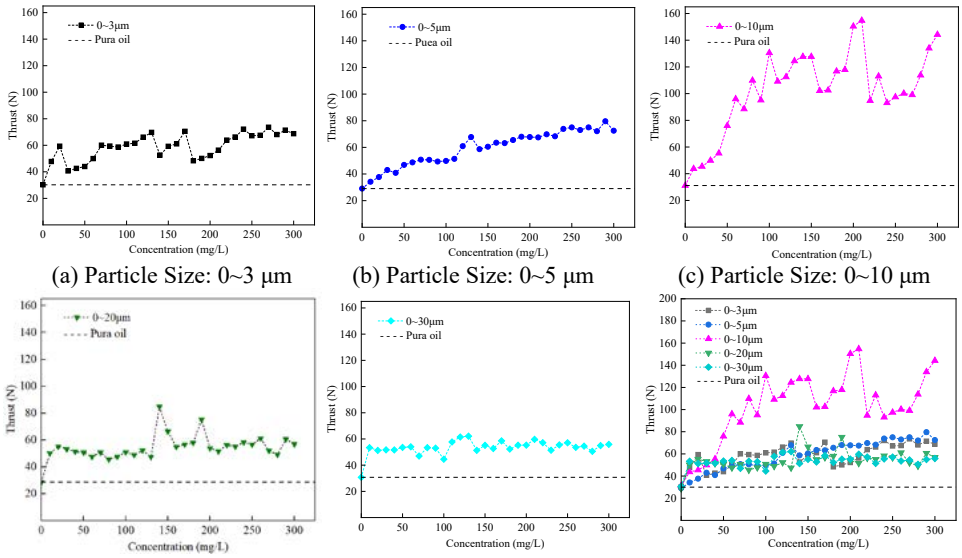
Figure 2.3. Test Bench Model and Physical Views of the Base and Test Manifolds

2.3. Analysis of Experimental Results

The experiment maintained an inlet pressure of 2 MPa and an outlet pressure slightly below 2 MPa, with a flow rate of 3–4 L/min. The directional valve spring preload was set to 22.4 N. After 3 minutes of continuous operation, force was gradually applied to the right end of the valve spool via the solenoid. Pressure transients in both outlet branches were monitored while solenoid current values were recorded to indirectly determine the stiction force magnitude. To investigate variations in the valve's external characteristics under contamination, particle size ranges were selected: 0~3 μm, 0~5 μm, 0~10 μm, 0~20 μm, and 0~30 μm. To prevent frequent spool seizure from high initial concentrations, particle concentration was incrementally increased from 10 mg/L to 300 mg/L in 10 mg/L steps. For each particle size, 31 distinct concentration levels were tested. Stiction force was measured every 3 minutes per condition, with four consecutive measurements averaged per data point.

As shown in Figure 2.4, the line plots illustrate variations in solenoid thrust force under different particle sizes and concentrations. Initially, the solenoid thrust force in particle-free working medium was 26.6 N. For particle sizes of 0~3 μm, the maximum solenoid thrust force reached 73.54 N; for 0~5 μm particles, it increased to 79.63 N; and for 0~10 μm particles, the thrust force surged to 127.7 N. A pronounced stiction phenomenon was observed with rising particle concentrations. Notably, once the concentration exceeded 300 mg/L, the solenoid-driven spool movement could not be reset by the spring force after actuation. For 0~20 μm particles, the maximum solenoid thrust force decreased to 84.73 N, representing a 33.6% reduction compared to the 0~10 μm particle case. With 0~30 μm particles, the maximum thrust force further dropped to 62.14 N, and no stiction occurred during testing. These results indicate that 0~10 μm particles act as critical particles (sensitive particles) in this system. Particles smaller than 0~10 μm are flushed out of the spool valve clearances by the fluid medium, while larger particles (exceeding 0~10 μm) cannot enter the

narrow clearances, thereby failing to generate significant stiction forces. 0~10μm particles caused larger thrust fluctuations, while other sizes also caused fluctuations but with smaller amplitude than 0-10μm particles.



(a) Particle Size: 0~3 μm (b) Particle Size: 0~5 μm (c) Particle Size: 0~10 μm

(d) Particle Size: 0~20 μm (e) Particle Size: 0~30 μm (f) Solenoid Thrust Force Variation with Particle Size and Concentration

Figure 2.4. Line Plots of Solenoid Thrust Force Variation with Particle Size and Concentration

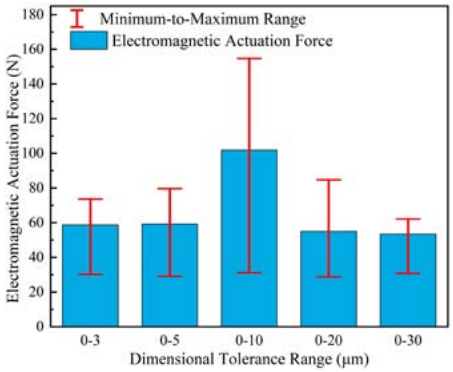


Figure 2.5. the maximum, minimum, and average solenoid forces under different particle sizes

As shown in Figure 2.5, the maximum, minimum, and average solenoid forces under different particle sizes are presented. The stiction force increases with particle size until the particle size exceeds the spool valve clearances, beyond which the stiction force diminishes accordingly.

As shown in Figure 2.6, during spool valve operation, solid particles smaller than the valve mating clearance enter the clearance with the fluid medium. A portion of these particles exit with the fluid flow, while others remain trapped within the clearance. Over time, particles gradually accumulate in the clearance. For an individual particle, when its upper and lower

surfaces simultaneously contact the spool and valve body, the particle rotates under torque due to spool movement. To further investigate particle retention within valve clearances, subsequent sections present simulation and experimental visualization studies of particle ingress dynamics.

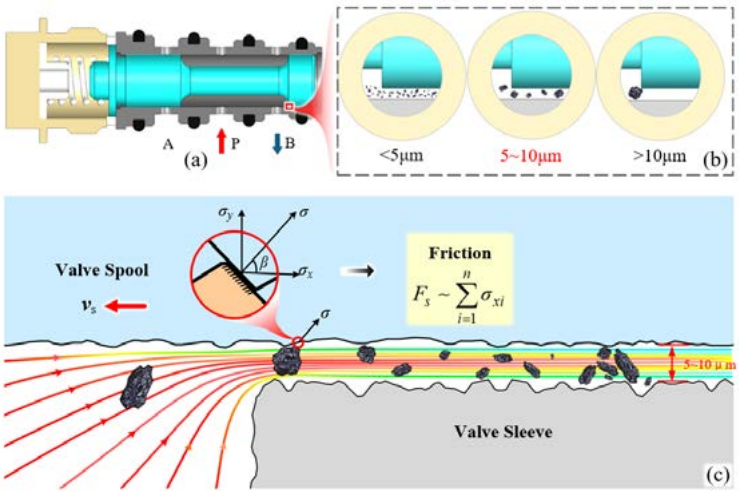


Figure 2.6. Schematic of Particle Invasion into Clearances

3. VISUALIZATION STUDY OF PARTICLE INVASION INTO CLEARANCES

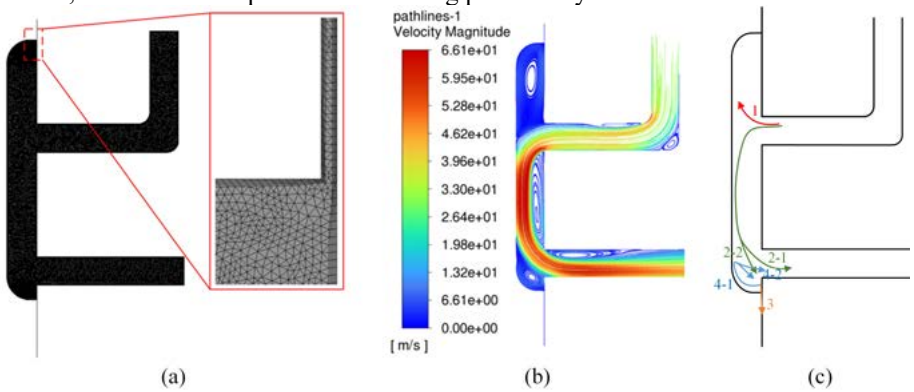
3.1. Flow Field Simulation and Results Analysis

3.1.1. Simulation Model and Parameter Settings

Given the substantial number of particles invading the clearance and the requirement for computational efficiency, the three-dimensional spool valve model with clearances was simplified to a two-dimensional model, with simulations conducted exclusively for the P-B operating condition (P port connected to B port). The spool-sleeve clearance range of 5~10 μm was represented by a 10 μm clearance in the simulation model. Local mesh refinement ensured grid skewness below 0.6, yielding a final mesh of 500,000 cells; the partitioned fluid domain mesh is illustrated in Figure 3.1 (a).

The fluid phase is fuel, with boundary conditions set as a pressure inlet (1 MPa) and pressure outlet (0 MPa). Three outlets were defined in the simulation: two clearance outlets and one main flow path outlet. For the particle phase, iron particles were selected for trajectory tracking, capturing their motion from injection to escape from the valve chamber fluid domain. Three particle sizes were tested: 5 μm, 8 μm, and 10 μm. Particles were injected from the inlet surface with incident directions spanning an angular range of 30° to 150°. Each simulation cycle injected 10 particles of each size (30 particles per cycle), repeated for 20 cycles (totaling 600 particles). The inlet and outlet boundaries were defined as escape boundaries, while the walls were set as rebound boundaries. Figure 3.1(b) displays flow streamlines within the valve chamber, revealing distinct vortices in the upper and lower regions where particles become entrained in rotational motion.

Figure 3.1(c) illustrates the particle trajectory diagram. A minority of particles follow Path 1 into the upper chamber, while most migrate downward via Path 2. Particles along Path 2-1 flow directly to the outlet, whereas those on Path 2-2 collide with the wall. Post-collision, some particles invade the lower clearance via Path 3, while others enter Path 4, entrained in the vortex flow. Vortex-driven particles either escape through the outlet or invade the lower clearance. Over time, more particles accumulate in the lower clearance than in the upper clearance, with numerous particles circulating persistently in the lower vortex.



(a) Simulation Model Mesh Generation; (b) Flow Trace Diagram; (c) Particle Trajectory Diagram

Figure 3.1. 2D Simulation Model and Mesh Generation

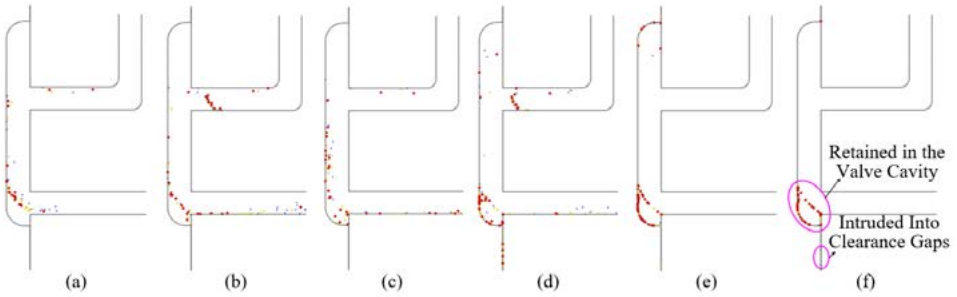


Figure 3.2. Particle Position Diagrams at Different Time Instants

Figure 3.2 depicts particle positions in the valve chamber at different simulation stages. For clarity, particles are magnified 30 times. Blue, yellow, and red represent 5 μm, 8 μm, and 10 μm particles, respectively. Initially, particles are dispersed randomly. As the simulation progresses, their migration patterns align with flow dynamics. In later stages, 34 particles remain trapped in the chamber, rotating along the lower vortex traces.

3.1.2. Particle Migration and Invasion Analysis

The particle retention rate (mechanical trapping ratio) is defined as the ratio of particles remaining in the valve chamber at the end of the simulation to the total injected particles, while the invasion rate represents the ratio of particles invading the clearances to the total injected particles. For the P-B operating condition, the particle retention rate in the valve chamber was 34/600 = 5.6%, with a total invasion rate of 35/600 = 5.9%. Specifically, the

invasion rate into the lower clearance was 22/600 = 3.7%, and into the upper clearance was 13/600 = 2.2%. The retention and invasion rates are summarized in Figure 3.3.

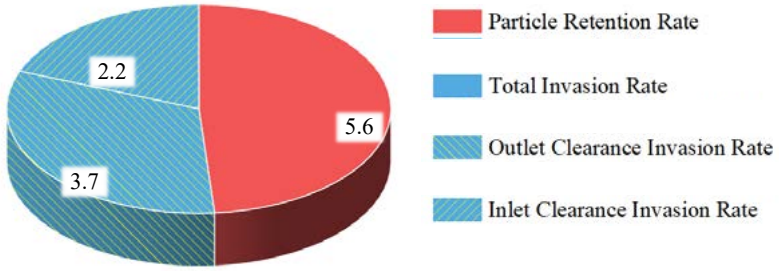
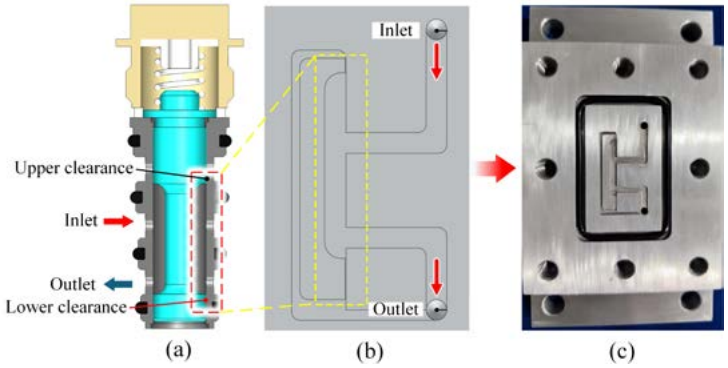


Figure 3.3. Retention and Invasion Rates of Particles in the Valve Chamber

3.2. Experimental Model Design

To investigate particle invasion dynamics within valve clearances, the 3D spool valve chamber was simplified to a 2D planar model. Flow channels were machined into an aluminum alloy plate with a uniform groove depth of 0.6 mm. Given the practical impossibility of directly observing sub-micron particle motion within the actual 5~10 μm spool-sleeve clearance and manufacturing constraints for experimental models at this scale, the clearance was geometrically scaled up by 10× to 0.1 mm (100 μm), with contaminant particles proportionally enlarged to 70~120 μm. Figures 3.4(a) and (b) show the 3D spool valve configuration and its corresponding 2D planar flow channel equivalent.



(a) 3D Directional Control Valve Model; (b) Simplified Experimental Model Schematic; (c) Physical Assembly of the Experimental Model

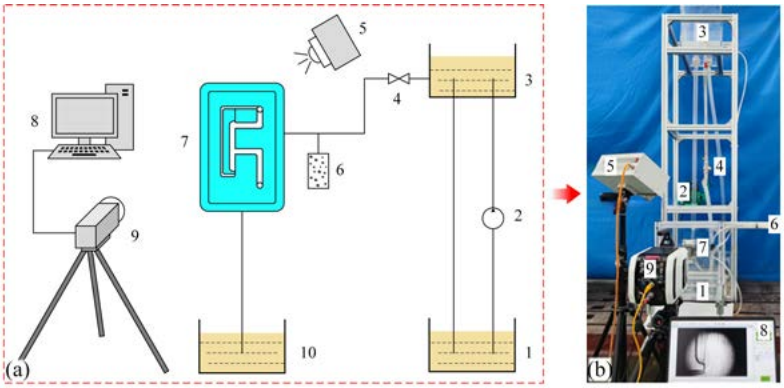
Figure 3.4. 3D Valve Model, Simplified Experimental Model, and Experimental Assembly

The 2D visualization setup comprises two layers: Lower layer: Aluminum alloy base plate featuring machined flow channels and perimeter sealing grooves, with side-mounted threaded ports for fluid connections. The clearance outlet connects to the flow channel at the end of the plate. Upper layer: A transparent acrylic plate for real-time observation of particle motion. The two layers are sealed using a rubber sealing gasket and fastened with bolted connections. The assembled experimental model is illustrated in Figure 3.4(c).

3.3. Test Bench Design

Considering the hazardous nature of fuel and the need for high imaging clarity, a water-based medium with comparable density and viscosity was selected as the working fluid. The

schematic of the visualization system is shown in Figure 3.5(a). The experimental setup consists of a hydraulic water system and a high-speed camera system. A constant pressure is supplied to the system by an elevated water tank 3. The water pump 2 draws fluid from the supply tank 1 to feed the elevated tank 3. The outlet of the elevated tank 3 is connected to the test model 7 via a pipeline with a shut-off valve 4. Contaminant iron particles are injected into the flow upstream of the test model 7 using a particle injector 6. The fluid then returns to the return water tank 10. During high-speed camera recording, lighting 5 illuminates the test section. The motion of particles within the flow field of the test model 7 is captured by the high-speed camera 9 and transmitted to the computer 8 for analysis.



1 Supply water tank; 2 Water pump; 3Elevated water tank; 4 Shut-off valve; 5 Light source; 6 Particle injector; 7 Test model; 8 Computer; 9 High-speed camera; 10 Return water tank
(a) Schematic of the Visualization Test System; (b) Physical View of the Visualization Test System
Figure 3.5. Schematic and Physical View of the Visualization Test System

As shown in Figure 3.5(b), the physical experimental system is displayed. The test frame is constructed entirely from aluminum alloy profiles. Both the low-level and high-level water tanks are fabricated as bonded assemblies of transparent acrylic panels, enabling real-time monitoring of water levels. The 1.6 m height difference between the water level and the manifold inlet corresponds to an inlet pressure of 15.68 kPa (calculated via hydrostatic pressure: P=ρgh$P=\rho gh$).

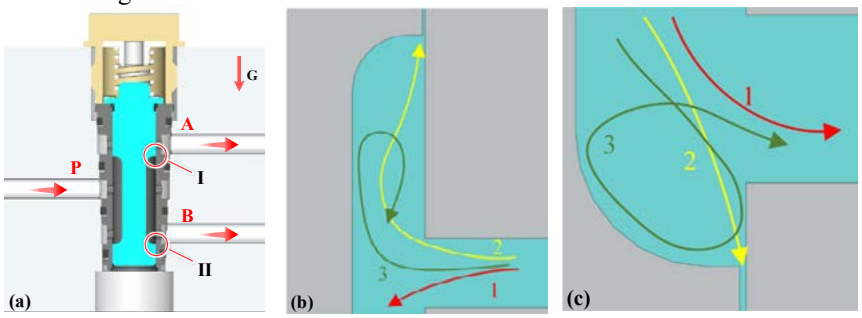
3.4. Analysis of Experimental Results

During testing, the two clearance regions of the spool valve were primarily imaged to observe particle invasion into the clearances. Since post-shutdown particle accumulation is a critical factor in spool stiction, particle deposition within the valve chamber and clearances was recorded 5 minutes after system shutdown. The high-speed camera captured 10,000 frames per second (fps), with a maximum recording duration of 3 seconds per session. To focus on critical dynamics, the motion trajectories of select particles within the valve chamber and invading the clearances were selectively recorded. Due to the camera's limited field of view, imaging was divided into the upper and lower valve chamber regions.

Figure 3.6(a)(b) shows the 3D schematic of the spool valve and the imaged upper chamber and upper clearance. Three primary particle trajectories are marked: Path 1: Particles migrating toward the outlet. Path 2: Particles invading the clearance. Path 3: Particles colliding with the wall and subsequently flowing toward the outlet. The majority of particles follow Path 1 toward the outlet, while a portion of particles collide with the wall along Path

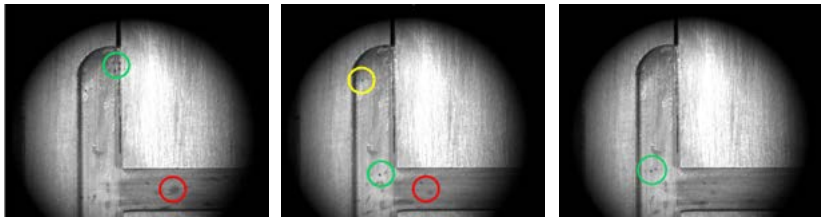
2, either invading the upper clearance or moving along the wall surface. Another subset of particles collides with the upper spool wall along Path 3 before redirecting toward the outlet.

Figure 3.6(a)(c) shows the 3D schematic of the spool valve and the imaged lower chamber and lower clearance. Three primary particle trajectories are marked: Path 1: Particles migrating toward the outlet. Path 2: Particles invading the clearance. Path 3: Particles colliding with the wall, rotating within the chamber, and subsequently flowing toward the outlet. The majority of particles follow Path 1 toward the outlet. A subset collides with the wall along Path 2 and invades the clearance, while others collide with the wall along Path 3 before redirecting to the outlet.



(a) Upper and Lower Valve Chamber (b) I: Particle Trajectory Diagram
(c) II: Particle Trajectory Diagram
(G: Direction of gravity, I: Capture images of the upper valve chamber and upper clearance region, II: Capture images of the lower valve chamber and lower clearance region)

Figure 3.6. Upper and Lower Valve Chamber Section and Marked Particle Trajectories



(a) 0ms (b) 1.3ms (c) 4.6ms

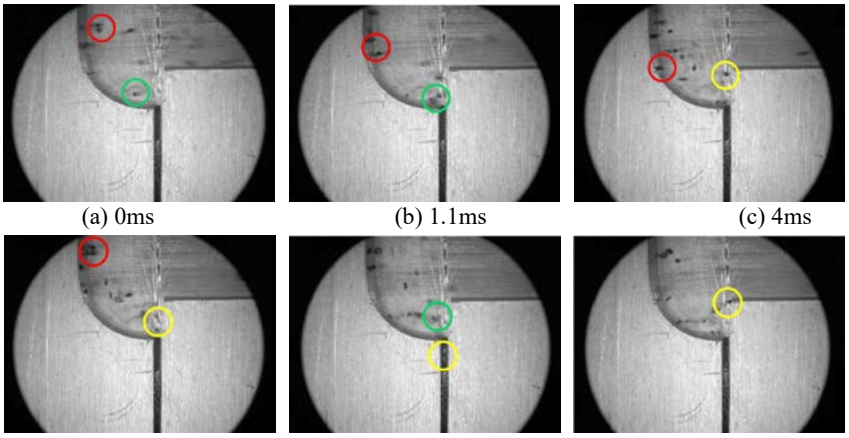


(d) 22.2ms (e) 26.2ms (f) 33.8ms

Figure 3.7. Particle Positions in the Upper Valve Chamber at Different Time Instants

Figure 3.7 illustrates the particle positions in the upper valve chamber at different time instants. Key observations include: Particles along Path 1 exhibit significantly higher velocities as they flow toward the outlet. Particles along Paths 2 and 3 demonstrate lower velocities. Specifically, particles along Path 3 ascend briefly under fluid forces before

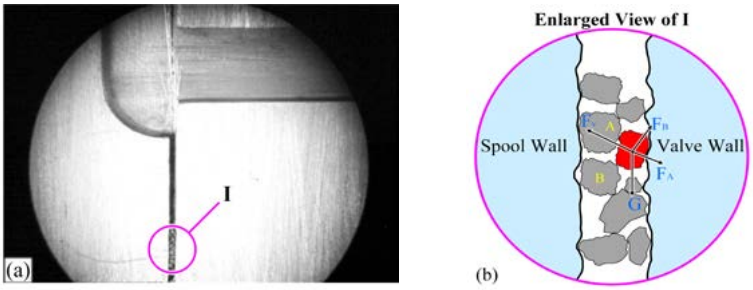
descending under the influence of gravity. Particles invading the clearances move slowly within the valve chamber and are fewer in number.



(a) 0ms (b) 1.1ms (c) 4ms

(d) 5.1ms (e) 6.2ms (f) 9.7ms

Figure 3.8. Particle Positions in the Lower Valve Chamber at Different Time Instants



(a) Lower Valve Chamber and Clearance (b) Force Analysis on a Single Magnified Particle

Figure 3.9. Particle Retention After System Shutdown

Figure 3.8 illustrates particle positions in the lower valve chamber at sequential time instants. Observations reveal: Particles following Path 1 exhibit significantly higher velocities toward the outlet. Particles along Path 2 collide with the sleeve wall surface, subsequently descending to invade the lower clearance. Particles traversing Path 3 rebound upward upon collision with the sleeve wall, ultimately exiting along the wall surface.

Figure 3.9. illustrates the particle retention after shutting off the shut-off valve and system shutdown. Significant particle accumulation is observed in the lower clearance. This phenomenon arises because the interaction forces between aggregated particles, frictional forces exerted by the clearance walls, adhesive forces between particles and walls, and the gravitational force collectively cause the particles within the clearance to reach force equilibrium, thereby remaining trapped.

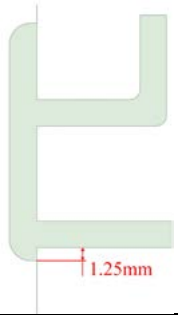
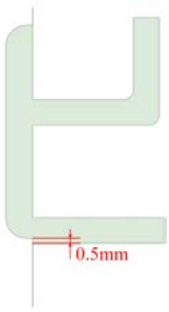
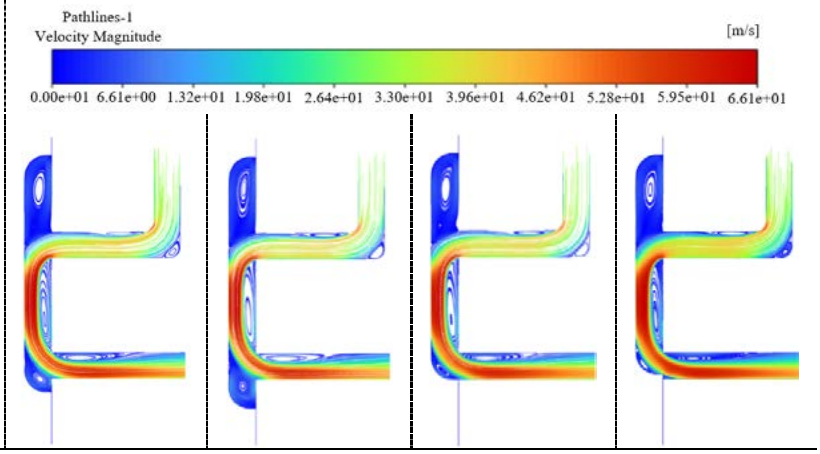
4. VALVE STRUCTURAL OPTIMIZATION

4.1. Structural Design Variations

For the P-B operating condition, three distinct structural designs were developed: (1) Lowering the spool shoulder by 1.5 mm, (2) Aligning the spool shoulder flush with the outlet

port, (3) Raising the spool shoulder by 1.75 mm. The geometric configurations of these designs are detailed in Table 1.

Table 1 Modified Spool Shoulder Models and Post-Modification Flow Trace Diagrams

Designation	Original Model	Lowering the Spool Shoulder	Spool Shoulder-Aligned	Raising the Spool Shoulder
Model Number	**O**	**I**	**II**	**III**
Geometric Model				
Flow Trace Diagrams				

4.2. *Simulation of Different Structures*

A comparative simulation analysis was conducted for the three designs. Performance was evaluated based on particle retention rate and clearance invasion rate, leading to the identification of the optimal design. The flow trace diagrams for each structure are summarized in Table 1.

Structure I: 40 particles remained in the valve chamber at the end of the simulation. Total invasion rate: 44/600 = 7.3% (29 particles in the lower clearance (29/600 = 4.8%); 15 particles in the upper clearance (15/600 = 2.5%)).

Structure II: 92 particles were retained within the valve chamber at the end of the simulation. Total invasion rate: 7/600 = 1.2% (3 particles in the upper clearance (3/600 = 0.5%); 4 particles in the lower clearance (4/600 = 0.6%)).

Structure III: 24 particles remained in the valve chamber at simulation completion. Total invasion rate: 10/600 = 1.7%, with all invading particles exiting through the upper clearance outlet. Lower clearance invasion rate: 0/600 = 0% (no particles invaded). Upper clearance invasion rate: 16/600 = 2.6%.

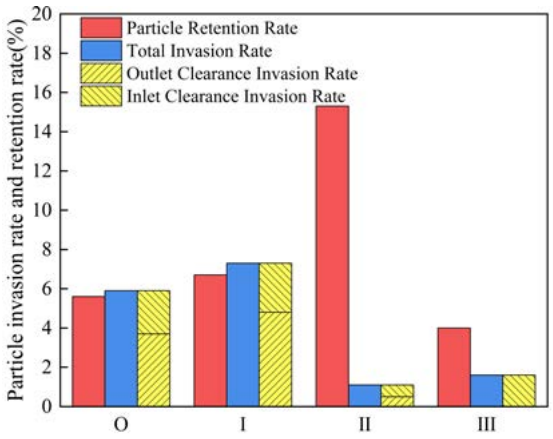


Figure 4.1. Particle Retention and Invasion Rates for Different Structural Models

Figure 4.1 illustrates the particle retention and invasion rates for different structural models. A comparative analysis of four models — the original model, spool shoulder-lowered model, spool shoulder-aligned model, and spool shoulder-raised mode — reveals the following:

- (1) Retention Rate Analysis: The spool shoulder-raised model reduced particle retention by 1.7% compared to the original model. The spool shoulder-lowered and spool shoulder-aligned models exhibited increased retention rates of 1.1% and 9.6%, respectively.
- (2) Invasion Rate Analysis: The spool shoulder-aligned and spool shoulder-raised models reduced invasion rates by 4.8% and 4.3%, respectively. Notably, the lower clearance invasion rate for the raised model was 0%. The spool shoulder-lowered model increased the invasion rate by 1.4%.

Conclusion: Considering both retention and invasion rates, the spool shoulder-raised model demonstrates optimal performance.

5. CONCLUSIONS

(1) Experimental results demonstrate that particles within the 0~5 μm, 0~10 μm, 0~20 μm, and 0~30 μm size ranges induce maximum stiction forces of 79.63 N, 127.7 N, 84.73 N, and 62.14 N, respectively, in the 5~10 μm clearance between the spool and sleeve. The 0~10 μm particles generated the highest stiction force, confirming that critical particles (sensitive particles) with sizes comparable to the spool clearance are the primary contributors to spool friction.

(2) Simulation analysis revealed particle migration and clearance invasion dynamics. At simulation completion, the retention rate in the valve chamber was 5.6%, with a total invasion rate of 5.8%. Smaller particles dominated invasion into the inlet-side clearance.

The vortex phenomenon significantly influenced particle trajectories, causing some particles to circulate within the chamber and eventually become trapped.

(3) Visualization experiments showed that most particles entering the valve chamber with the fluid were flushed out. A small fraction remained trapped in vortex-driven rotations near the outlet or invaded the upper/lower clearances. Post-shutdown, particles invading the clearances were not expelled by gravity but aggregated due to particle-wall interaction forces. Particles rotating in vortices settled at the chamber base after shutdown.

(4) Three spool shoulder optimizations were tested: lowering, aligning, and raising the shoulder. Simulations yielded retention rates of 6.7%, 15.3%, and 4%, and total invasion rates of 7.3%, 1.1%, and 1.6%, respectively. The spool shoulder-raised design demonstrated superior performance when balancing retention and invasion metrics.

6. REFERENCES

[1] M. A P R, A. C G R, R. L, et al. Water Microturbines for Sustainable Applications: Optimization Analysis and Experimental Validation[J]. Water Resources Management,2023, 38(3):1011-1025.

[2] Jia-yi W, Yang Y, Zhen-bo L, et al. Modal and structural analysis on a main feed water regulating valve under different loading conditions[J]. Annals of Nuclear Energy, 2021, 159.

[3] Jie D, Yipan D, Wenbin C, et al. Wear failure analysis of suction valve for high pressure and large flow water hydraulic plunger pump[J]. Engineering Failure Analysis, 2022, 134.

[4] Liang L, Peng Y X, Ru M L, et al. Analysis of fretting wear behavior of unloading valve of gasoline direct injection high pressure pump[J]. Journal of Zhejiang University-SCIENCE A, 2022, 23(4): 314-328.

[5] Hao T, Jianbo L, Yongjun G. Rapid diagnosis of proportional valve pilot stage stiction fault via dynamic model-based temporal features[J]. Proceedings of the Institution of Mechanical Engineers, 2025, 239(2): 373-388.

[6] Kazama T, Aoki S, Kobessho M. Cavitation Erosion Characteristics of High Bulk Modulus Oils: Preliminary Evaluation Experiment for Hydraulic Equipment[J]. Materials Performance and Characterization, 2020, 9(1): 20200003.

[7] Borivoj N, Ljiljana R, Mića Đ, et al. Analysis of the Behaviour of the Hydraulic Fluid HM VG 32 and HV 32 at Different Operating Conditions of the Axial-Piston Pump[J]. Tehnički vjesnik, 2024, 31(5): 1523-1529.

[8] Yang W, Li R, Xu J, et al. CFD-based axial steady-state hydrodynamic study and structural optimization of miniature switching valve[J]. Flow Measurement and Instrumentation, 2024, 100102744-102744.

[9] Li J, Peng C, Guo Q, et al. Study on Flow Field Characteristics of Micro-Flow Control Valve Based on Fluid-Thermal-Solid Coupling[J]. Arabian Journal for Science and Engineering,2024,50(12):1-18.

[10] Lin Z.H. Mechanisms of Spool Jamming in Hydraulic Valves and Mitigation Strategies [D]. Hangzhou: Zhejiang University; 2023. Dissertation Number: 2023.000863.

[11] Lin Z.H., Nie J.X., Hua T.F., et al. Solid-Liquid Flow Dynamics and Jamming in Hydraulic Spool Valves: A Clearance Analysis[J]. Chinese Hydraulics & Pneumatics, 2024, 48(3):115-120.

[12] Qian J, Xu J, Zhong F, et al. Solid-liquid flow characteristics and sticking-force

analysis of valve-core fitting clearance[J]. Journal of Zhejiang University-SCIENCE A, 2023, 24(12): 1096-1105.
[13] N. V B, V. E M, Cai C J. Mechanical Admixture Influence in the Working Fluid on Wear and Jamming of Spool Pairs from Aircraft Hydraulic Drives[J]. Journal of Friction and Wear, 2021, 41(6): 526-530.
[14] Bian L, Guo J, Pan J. Micromechanics analysis for thermal expansion coefficients of three-phase particle composites[J]. Archive of Applied Mechanics, 2019, 89(8): 1641-1654.
[15] Sajjadi H, Salmanzadeh M, Ahmadi G, et al.Lattice Boltzmann method and RANS approach for simulation of turbulent flows and particle transport and deposition[J]. Particuology, 2016, 3062-72.
[16] Liu X, Ji H, Min W, et al. Erosion behavior and influence of solid particles in hydraulic spool valve without notches[J]. Engineering Failure Analysis, 2020, 108104262-104262.
[17] Kharoua N, Alshehhi M, Khezzar L. Effects of fluid flow split on black powder distribution in pipe junctions[J]. Advanced Powder Technology, 2016, 27(1):42-52.
[18] Zhang J, Ji H, Zhao W, et al. Numerical Investigation of Micrometer-Sensitive Particle Intrusion in Hydraulic Valve Clearances and Its Impact on Valve Performance[J]. Processes, 2024, 12(5): 864.
[19] Qiang X L, Fei L, Hong J, et al. Particle erosion transient process visualization and influencing factors of the hydraulic servo spool valve orifice[J]. Flow Measurement and Instrumentation, 2023, 89.
[20] Xinqiang L, Hong J, Fei L, et al. Particle motion and erosion morphology of the spool orifice in an electro-hydraulic servo valve under a small opening[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2022, 236(6): 3160-3173.

Biographies



Chen Qianpeng received his Ph.D. degree from Lanzhou University of Technology, China in 2022. He is currently serving as an Associate Professor and Master's Supervisor in the Department of Fluid Transmission and Control, College of Energy and Power Engineering, Lanzhou University of Technology. He is a recipient of the "Hiwin" Outstanding Ph.D. Thesis Award from the Chinese Mechanical Engineering Society and has been awarded research funding from the China National Natural Science Fund. His long-term research focuses on precision measurement and high-reliability design of hydraulic components, hydraulic system applications, and the development of specialized machinery.

Performance analysis of rotary mechanical pulsation compensators for different types of pumps

Gudrun Mikota[1], David Holzer[2]

Institute of Machine Design and Fluid Power, Johannes Kepler University Linz, Altenbergerstrasse 69, A-4040 Linz, Austria, [1]gudrun.mikota@jku.at, [2]david.holzer@jku.at

Abstract.

Rotary mechanical pulsation compensators require a torsionally compliant coupling attached to the pump rotor. To quantify their performance for different types of pumps, the pump outlet pressure magnitude minimum at the compensation frequency and the bandwidth of the compensation effect are considered. For these performance criteria, closed-form solutions are found with an undamped hydraulic oscillator or a simple capacitance at the pump outlet. Experiments are carried out with a compensator coupling and an internal gear pump attached to different hydraulic systems. Measurement results are compared with frequency domain response from appropriate theoretical models. Compensator performance is derived from experiments to validate closed-form solutions. Using these, pressure magnitude minimum and compensator bandwidth are predicted for an axial piston pump.

Keywords. Piston pump, gear pump, pressure pulsation, compensator coupling, pressure magnitude minimum, bandwidth.

1. INTRODUCTION

A novel pulsation compensation principle for hydraulic pumps and motors was introduced in [1]. This principle uses the mass moment of inertia of the machine rotor and requires a torsionally compliant coupling in the drivetrain. Assuming a suitable stiffness of such a coupling, numerical studies were provided in [2]. The design of a compensator coupling was described in [3], where the novel assembly was classified as rotary mechanical pulsation compensator. Measurements with a radial piston pump were presented in [4]. Although the desired effect could be demonstrated, better performance was expected for other types of pumps. In [2-4], theoretical predictions on compensator performance were made by plotting frequency response functions of the coupled hydraulic-mechanical system for particular sets of parameters.

In this paper, performance criteria for rotary mechanical pulsation compensators shall be specified. They address narrow-band pulsation compensation for constant speed applications and broadband pulsation reduction for relevant speed ranges. With respect to the attached hydraulic system, resonant and non-resonant conditions are distinguished. The

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

coupled hydraulic-mechanical assembly is modelled by an analogue hydraulic system, which is transformed into a dual mechanical system [5] to apply the existing theory of mechanical vibration absorbers [6]. Closed-form solutions are derived for the relationships between system parameters and performance criteria. These solutions are verified by measurements on an internal gear pump attached to different hydraulic systems, which are compared to a coupled hydraulic-mechanical model in the frequency domain. Besides flow rate excitation as in previous models, torque excitation from geometric pulsation is also included. Finally, closed-form solutions are applied to predict compensator performance for an axial piston pump.

2. PERFORMANCE MEASURES

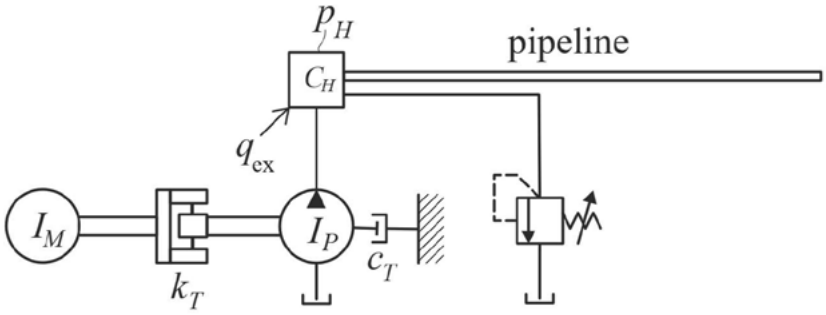


Figure 2.1. Schematic of setup with radial piston pump and pipeline

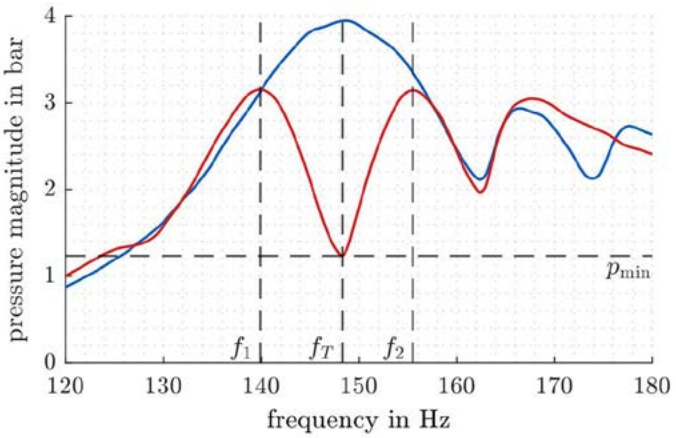


Figure 2.2. Pressure pulsations of radial piston pump. Blue: without compensator, red: with compensator

In [4], an assembly according to Figure 2.1 was used for first measurements on a rotary mechanical pulsation compensator. A compensator coupling with torsional stiffness k_T was arranged between the driving motor and a radial piston pump with mass moments of inertia I_M and I_P, respectively. The inevitable torsional damping on the pump rotor was denoted by c_T. Pump pulsations were considered as an external flow rate excitation q_{ex} into the

hydraulic capacitance C_H at the pump outlet, which was connected to a pressure relief valve. A closed pipeline was attached at the pump outlet to obtain a hydraulic resonator. The pressure relief valve was known to be heavily damped with cut-off frequency far below the first pipeline resonance. In an alternative configuration, the compensator coupling was replaced by a comparatively stiff connection between the driving motor and the pump. The magnitudes of measured pressures p_H are shown in Figure 2.2. With compensator coupling, a minimum p_{min} occurred at the compensation frequency f_T between the side resonance frequencies f_1 and f_2.

Pulsation compensators can be used in different application cases. For constant speed applications, narrow-band compensation may be sufficient unless overshooting side resonances are passed during run-up or run-down. A range of pump speeds requires an appropriate bandwidth of the compensation effect. The pressure magnitude minimum p_{min} at the pump outlet constitutes an obvious performance criterion in all application cases. A second criterion is given by the compensator bandwidth $b = (f_2 - f_1)/f_T$, which roughly describes the pulsation reduction range. With respect to both criteria, compensator performance is rather limited for the radial piston pump. Therefore, pressure magnitude minimum and compensator bandwidth shall be analyzed for other types of pumps.

3. SYSTEM TRANSFORMATION

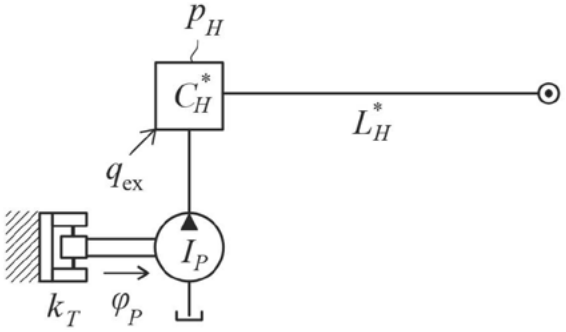


Figure 3.1. Coupled hydraulic-mechanical assembly with hydraulic oscillator

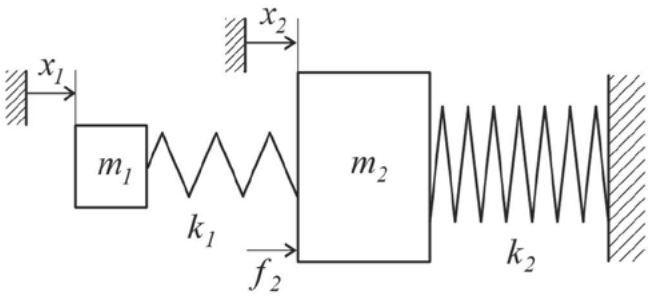


Figure 3.2. Dual mechanical system with fixed boundary

A simplified hydraulic-mechanical setup according to Figure 3.1 can be transformed into a mechanical system as follows. With the displacement V_P of the pump, the transmission ratio

between the angle φ_P of the pump shaft and the displaced volume at the pump outlet is $2\pi/V_P$. In an analogue hydraulic system, the compliance of the compensator coupling becomes a hydraulic capacitance

$$C_T = \left(\frac{V_P}{2\pi}\right)^2 \frac{1}{k_T}, \tag{3.1}$$

and the pump inertia turns into a hydraulic inductance

$$L_T = \left(\frac{2\pi}{V_P}\right)^2 I_P. \tag{3.2}$$

Hydraulic capacitance C_H^* and hydraulic inductance L_H^* represent an undamped hydraulic oscillator at the pump outlet. The inductance L_H^* is connected to a constant pressure boundary. Considering the pressures

$$p_T = -\frac{2\pi}{V_P} k_T \varphi_P \tag{3.3}$$

and p_H as state variables, this hydraulic system corresponds to the dual mechanical system in Figure 3.2 with

$$m_1 = C_T \tag{3.4}$$

$$k_1 = \frac{1}{L_T} \tag{3.5}$$

$$m_2 = C_H^* \tag{3.6}$$

$$k_2 = \frac{1}{L_H^*} \tag{3.7}$$

and a fixed end boundary condition [5]. The state variables of the dual mechanical system are the velocities $\dot{x}_1 = p_T$ and $\dot{x}_2 = p_H$. It can be seen that the classical mechanical vibration absorber setting corresponds to a rotary mechanical pulsation compensator with resonant conditions of the attached hydraulic system.

Non-resonant conditions are obtained for a simple hydraulic capacitance at the pump outlet as shown in Figure 3.3. This setting corresponds to the dual mechanical system in Figure 3.4. With C_H instead of C_H^* Equations (3.1) to (3.6) still apply, free boundaries result, and state variables are defined as above.

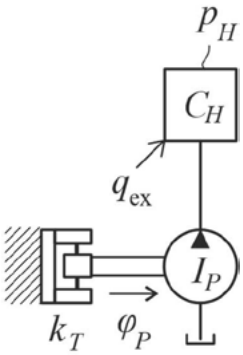


Figure 3.3. Coupled hydraulic-mechanical assembly with hydraulic capacitance

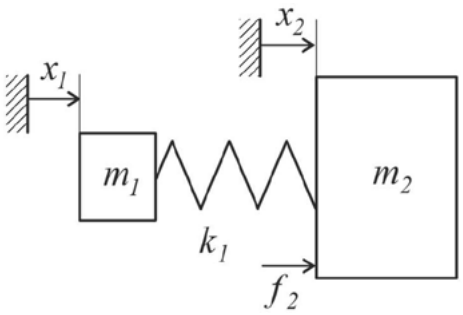


Figure 3.4. Dual mechanical system with free boundaries

4. CLOSED-FORM SOLUTIONS

For mechanical vibration absorbers, bandwidth considerations are available in [6]. Referring to Figure 3.2,

$$\frac{k_1}{m_1} = \frac{k_2}{m_2} \tag{4.1}$$

is assumed, which becomes

$$\frac{k_T}{I_P} = \frac{1}{C_H^* L_H^*} \tag{4.2}$$

with Equations (3.1 – 3.2) and (3.4 – 3.7). The bandwidth b of the absorber is obtained from the difference of the side resonance frequencies divided by the compensation frequency

$$f_T = \frac{1}{2\pi}\sqrt{\frac{k_1}{m_1}}. \tag{4.3}$$

An eigenvalue analysis of the system in Figure 3.2 yields the side resonance frequencies f_1 and f_2. With the mass ratio

$$\mu = \frac{m_1}{m_2}, \tag{4.4}$$

the bandwidth results as

$$b = \frac{f_2 - f_1}{f_T} = \sqrt{\frac{2+\mu+\sqrt{4\mu+\mu^2}}{2}} - \sqrt{\frac{2+\mu-\sqrt{4\mu+\mu^2}}{2}}. \tag{4.5}$$

For the hydraulic-mechanical assembly in Figure 3.1, Equations (3.1), (3.4), and (3.6) are used to calculate

$$\mu = \left(\frac{V_P}{2\pi}\right)^2 \frac{1}{k_T C_H^*}. \tag{4.6}$$

The compensator bandwidth b increases with μ, which means that the compensator coupling should be as compliant as possible.

For non-resonant conditions as in Figures 3.3 and 3.4, the lower side resonance does not exist. The compensator bandwidth can be calculated as

$$b_n = \frac{f_2\text{-}f_T}{f_T} = \sqrt{1+\mu}\text{-}1 \tag{4.7}$$

with μ according to Equations (4.4) or (4.6).

To determine the pump outlet pressure magnitude minimum p_{min} under resonant conditions as in Figure 3.1, torsional damping c_T on the pump rotor must be taken into account. According to the assumption (4.2), the attached hydraulic system is capable of autonomous pulsations at the compensation frequency

$$f_T = \frac{1}{2\pi}\sqrt{\frac{k_T}{I_P}}. \tag{4.8}$$

The remaining torsional single-degree-of-freedom oscillator yields the relationship

$$(-\omega^2 I_P + i\omega c_T + k_T)\frac{2\pi}{V_P}Q_{ex}(i\omega) = -i\omega\frac{V_P}{2\pi}P_H(i\omega) \tag{4.9}$$

with the angular frequency ω and the Laplace Transforms Q_{ex} and P_H of q_{ex} and p_H, respectively. At $\omega = \omega_T = 2\pi f_T$, this becomes

$$P_H(i\omega_T) = -c_T\left(\frac{2\pi}{V_P}\right)^2 Q_{ex}(i\omega_T). \tag{4.10}$$

For the fundamental harmonic of the flow rate excitation,

$$Q_{ex}(i\omega_T) = \frac{V_{ex}}{2\pi}\cdot\frac{\omega_T}{z} \tag{4.11}$$

results with the displacement amplitude V_{ex} and the number z of pulsations per revolution of the pump shaft. As the pressure magnitude minimum will occur near $\omega = \omega_T$,

$$p_{min} \approx \frac{2\pi\omega_T c_T V_{ex}}{zV_P^2} \tag{4.12}$$

is obtained as an approximation.

Under non-resonant conditions, it is possible to calculate a pressure magnitude reduction compared to the pressure magnitude p_C resulting form the flow rate excitation q_{ex} into the closed hydraulic capacitance C_H. For the assembly in Figure 3.3, it can be shown that

$$\frac{p_{min}}{p_C} \approx \frac{1}{\sqrt{1+\frac{1}{\left(c_H\left(\frac{2\pi}{V_P}\right)^2\omega_T c_T\right)^2}}}. \tag{4.13}$$

Apparently, the torsional damping on the pump rotor should be as small as possible.

5. EXPERIMENTS WITH INTERNAL GEAR PUMP

The existing compensator coupling is combined with an internal gear pump and different hydraulic systems at the pump outlet. Figure 5.1 shows the experimental setup. The drivetrain consists of an electric motor, a torque measurement shaft, the compensator coupling, and the internal gear pump. The pipeline at the pump outlet can be closed at

different lengths or disconnected completely. The mean pressure is adjusted by a pressure relief valve. The compensator coupling is realized by a spoke wheel according to Figure 5.2, which also shows one of the rotation angle sensors mounted on each side of the compensator coupling.

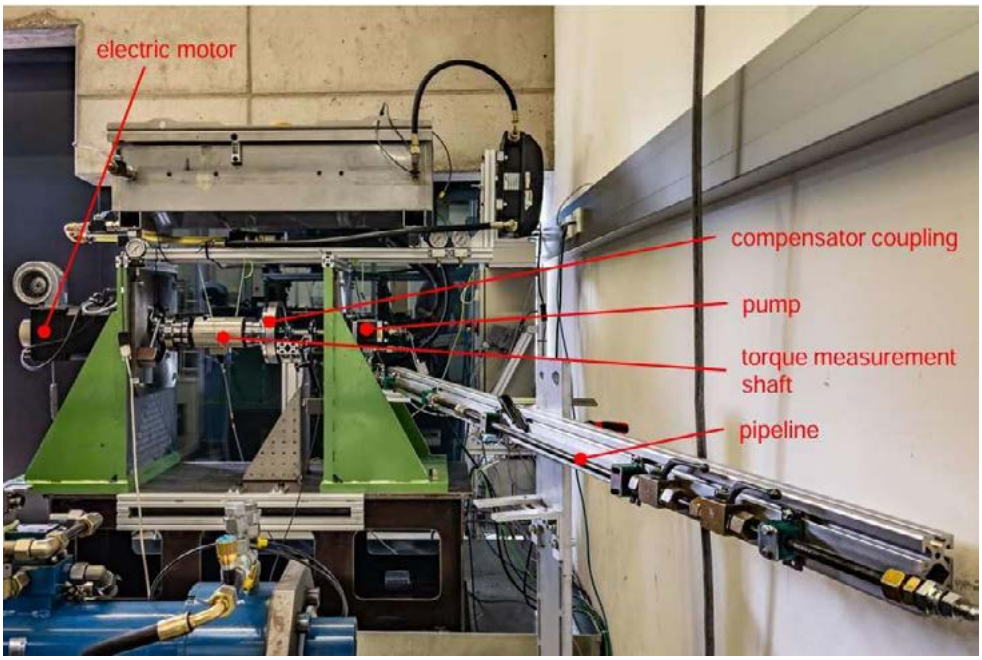


Figure 5.1. Experimental setup

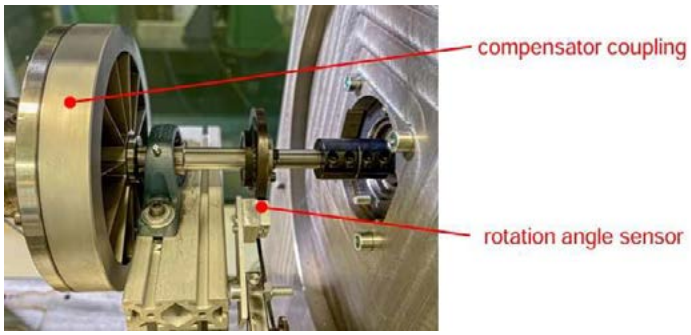


Figure 5.2. Detail with compensator coupling

The hydraulic systems at the pump outlet consist in pipelines of different lengths and the inevitable capacitance of the fittings between pump, pressure relief valve, pressure sensor, and pipeline attachment. If the pipeline is disconnected, this capacitance remains. The rotational speed of the driving motor is increased stepwise from 500 to 2000 rpm. Ten full rotations are performed after each step of 4 rpm. For each rotational speed, rectangular

windows of measured signals are considered with a trigger on the pump shaft angle, and Fourier coefficients of the fundamental pulsation frequency are calculated. In this way, the phase with respect to flow rate excitation can also be obtained. All experiments are repeated without compensator coupling.

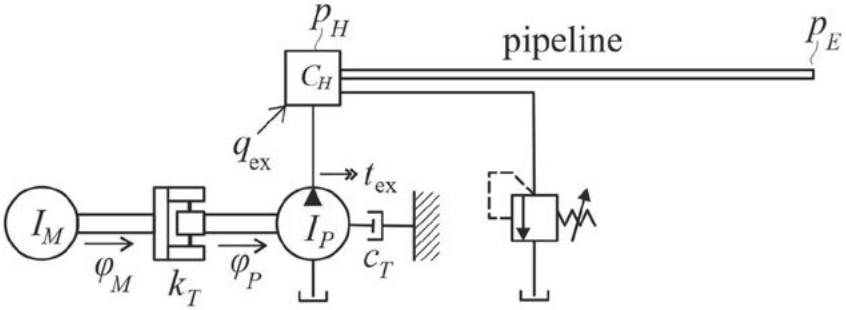


Figure 5.3. Schematic of setup with internal gear pump and pipeline

The coupled hydraulic-mechanical model in Figure 5.3 is implemented for comparison with measurements. In addition to previous models, an excitation torque t_{ex} is considered on the pump rotor. This excitation applies for geometric pulsations and originates from the fluctuating transmission ratio between the mean pressure and the torque on the pump rotor, which is determined by the respective displacement amplitude $V_{ex,g}$. In contrast, dynamic pulsations do not imply an excitation torque and only contribute a flow rate excitation.

Flow rate pulsations of external gear pumps have been studied extensively [7]. Comparable information on internal gear pumps can be found for a particular tooth profile [8]. It is concluded from literature that for both types of gear pumps, geometric and dynamic pulsations can be relevant. Therefore, q_{ex} and t_{ex} are considered as independent excitations.

The model further includes the dynamic behavior of the pipeline and the resistance R of the pressure relief valve. It is assumed that this resistance depends on the mean flow rate, which is proportional to the fundamental pulsation frequency. The state variables of the model are the rotation angles φ_M and φ_P on motor and pump side of the compensator coupling, the pressure p_H at the pump outlet, and the pressure p_E at the closed end of the pipeline. The respective Laplace transforms are denoted by Φ_M, Φ_P, P_H, and P_E, while Q_{ex} and T_{ex} denote the Laplace transforms of flow rate excitation and excitation torque, respectively.

In the frequency domain, the model can be formulated as

$$\left(-\omega^2\begin{bmatrix} I_M & 0 & 0 & 0 \\ 0 & I_P & 0 & 0 \\ 0 & -\frac{V_P}{2\pi} & C_H & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + i\omega\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & c_T & 0 & 0 \\ 0 & 0 & \frac{1}{R(i\omega)} + S(i\omega) & T(i\omega) \\ 0 & 0 & T(i\omega) & S(i\omega) \end{bmatrix} + \right.$$

$$\left. + \begin{bmatrix} k_T & -k_T & 0 & 0 \\ -k_T & k_T & \frac{V_P}{2\pi} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right)\begin{bmatrix} \Phi_M(i\omega) \\ \Phi_P(i\omega) \\ P_H(i\omega) \\ P_E(i\omega) \end{bmatrix} = \begin{bmatrix} 0 \\ T_{ex}(i\omega) \\ i\omega Q_{ex}(i\omega) \\ 0 \end{bmatrix} \qquad (5.1)$$

with the functions

$$S(i\omega) = -\frac{iA}{\sqrt{E\rho}\, f(i\omega)} \cdot \cot\left(\frac{f(i\omega)\omega l}{\sqrt{E/\rho}}\right) \tag{5.2}$$

and

$$T(i\omega) = \frac{iA}{\sqrt{E\rho}\, f(i\omega)} \cdot 1/\sin\left(\frac{f(i\omega)\omega l}{\sqrt{E/\rho}}\right) \tag{5.3}$$

describing the dynamic behaviour of the pipeline with

$$f(i\omega) = \sqrt{-\frac{J_0(ir\sqrt{i\omega/\nu})}{J_2(ir\sqrt{i\omega/\nu})}}, \tag{5.4}$$

where J_0 and J_2 are Bessel functions of first kind [9]. For an internal gear pump, the fundamental frequencies of geometric and dynamic pulsations coincide and the corresponding angular frequency is denoted by ω. The resistance model is formulated as

$$R(i\omega) = p_{ref} / \left(\frac{V_P}{2\pi} \cdot \frac{\omega}{z}\right), \tag{5.5}$$

where pump leakage is considered by choosing the reference pressure p_{ref} higher than the pump outlet pressure. In the frequency domain, the flow rate excitation from geometric pulsations reads

$$Q_{ex,g}(i\omega) = \frac{V_{ex,g}}{2\pi} \cdot \frac{\omega}{z}, \tag{5.6}$$

in which z is the number of teeth of the gear on the pump shaft. The corresponding excitation torque becomes

$$T_{ex}(i\omega) = -\frac{V_{ex,g}}{2\pi} \cdot p_m, \tag{5.7}$$

where p_m denotes the mean pressure at the pump outlet. In an operating point, the flow rate excitation from dynamic pulsations can be written as

$$Q_{ex,d}(i\omega) = \frac{V_{ex,d}}{2\pi} \cdot \frac{\omega}{z} \cdot e^{i\alpha} \tag{5.8}$$

in analogy to Equation (5.6) with a phase shift α which is assumed to be frequency independent for simplicity. Then, the total flow rate excitation becomes

$$Q_{ex}(i\omega) = Q_{ex,g}(i\omega) + Q_{ex,d}(i\omega) = \frac{V_{ex}}{2\pi} \cdot \frac{\omega}{z} \cdot e^{i\beta} \tag{5.9}$$

with a resulting phase shift β.

Model parameters are summarized in Table 1. For the reference case without compensator coupling, a rigid connection of electric motor and pump is assumed so that the angles φ_M and φ_P coincide. In this case, the excitation torque t_{ex} acts directly on a high mass moment of inertia with negligible effect on the resulting motion. Therefore, the parameter V_{ex} could be identified by comparing measured and calculated pressure magnitude maxima under resonant conditions of the hydraulic system. With compensator coupling, the parameters I_P, c_T, $V_{ex,g}$, and β have been adjusted by respective comparisons of compensation frequencies, pump outlet pressure magnitudes at compensation and side resonance frequencies, and phase curves of the pump outlet pressure.

Table 5.1. Model parameters

motor sided mass moment of inertia	I_M	0.16 kg m^2
compensator coupling stiffness	k_T	2350 Nm/rad
pump sided mass moment of inertia	I_P	0.0008 kg m^2
pump displacement	V_P	25 cm^3
number of teeth	z	13
total displacement amplitude	V_{ex}	0.025 V_P
geometric pulsation displacement amplitude	$V_{ex,g}$	0.0125 V_P
phase shift	β	80°
torsional damping constant	c_T	0.01 Nms/rad
dead volume at pump outlet	V_d	1.8 V_P
bulk modulus	E	16000 bar
hydraulic capacitance at pump outlet	C_H	V_d/E
reference pressure	p_{ref}	150 bar
mean pressure	p_m	100 bar
pipeline length	l	221 / 189 cm
pipeline radius	r	6 mm
fluid density	ρ	870 kg/m^3
kinematic viscosity	υ	50 mm^2/s

In addition to the pressure p_H at the pump outlet, a torque equivalent pressure

$$p_T = \frac{2\pi}{V_P} k_T(\varphi_M - \varphi_P) \tag{5.10}$$

in the compensator coupling is considered in both measurements and calculations. Frequency domain measurements for 221 cm pipeline length are shown in Figure 5.4. Figure 5.5 shows the respective calculations. The blue curves apply for the case without compensator coupling, where the torque equivalent pressure cannot be calculated. With compensator coupling, the torque equivalent pressure p_T reaches much higher values than the pressure p_H at the pump outlet and is therefore downscaled by a factor of 5. Measured and calculated pressures match fairly well, which is also the case for 189 cm pipeline length (see Figures 5.6 and 5.7).

For 221 cm pipeline length, the compensation frequency and the resonance frequency of the hydraulic system coincide at 258 Hz, which corresponds to a pump speed of 1191 rpm. With compensator coupling, both measurement and model result in a pump outlet pressure magnitude minimum of 0.15 bar, which corresponds to a reduction to 5 % of the magnitude without compensator coupling. The closed-form solution (4.12) yields a value of 0.083 bar. This solution does not account for the excitation torque, which explains the deviation from the measured value. It should be noted that the closed-form solution relies on the identified displacement amplitude V_{ex} and the adjusted torsional damping constant c_T.

According to measurements, the compensator bandwidth amounts to 24 %, while 27 % are obtained from the model. Increasing the dead volume V_d by a third of the pipeline volume to estimate the hydraulic capacitance C_H^* in Equation (4.6), the closed-form solution (4.5) for the bandwidth becomes 29 %, which is in fair agreement with measurements.

The compensator exhibits excellent performance at the compensation frequency and is robust with respect to small variations in pump speed. However, an overshooting upper side resonance occurs, and the respective pulsation frequencies should be avoided.

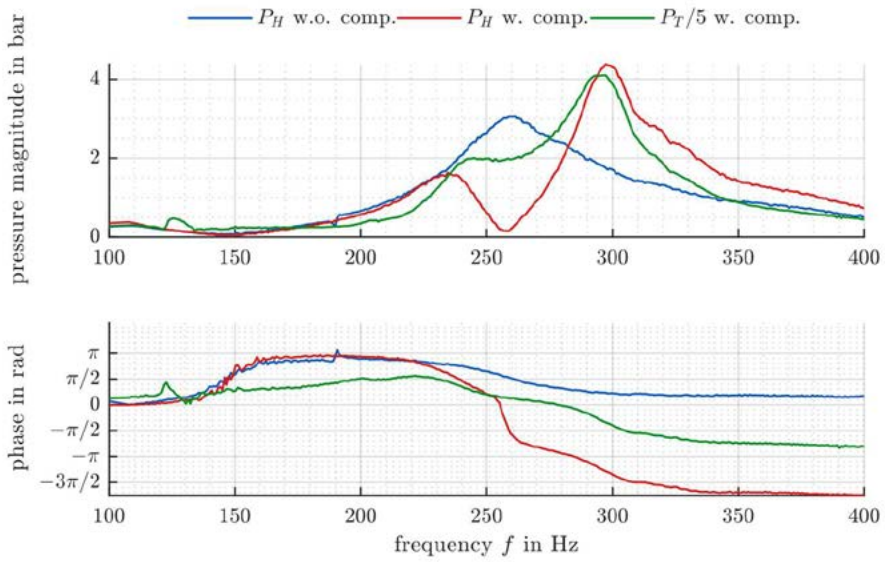


Figure 5.4. Measured pressures for 221 cm pipeline length

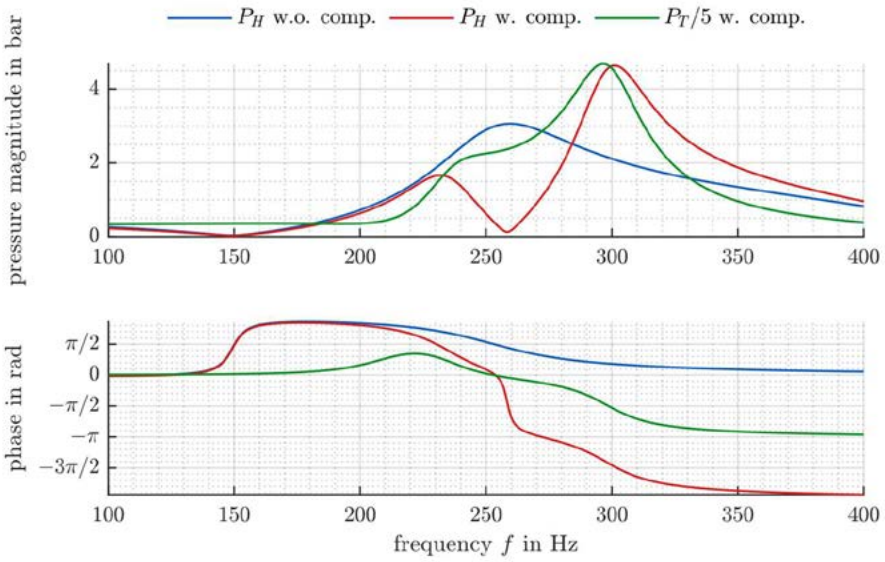


Figure 5.5. Calculated pressures for 221 cm pipeline length

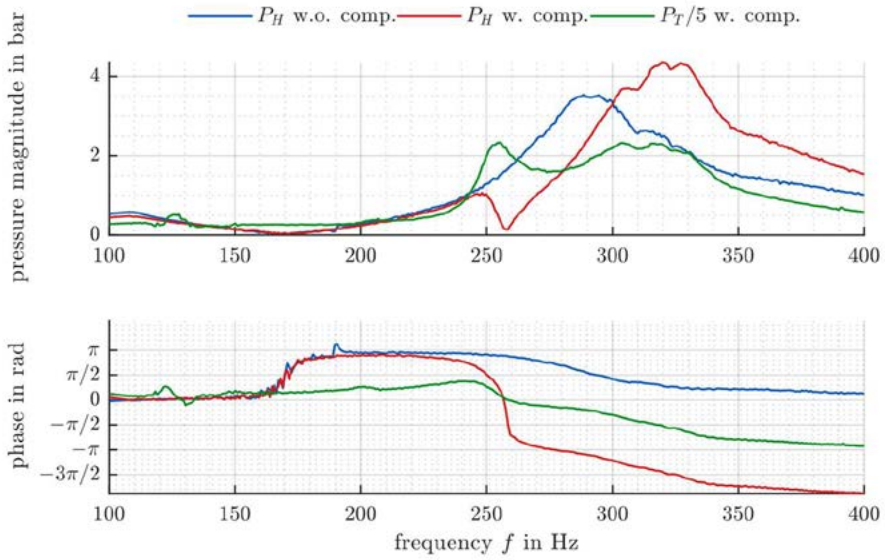


Figure 5.6. Measured pressures for 189 cm pipeline length

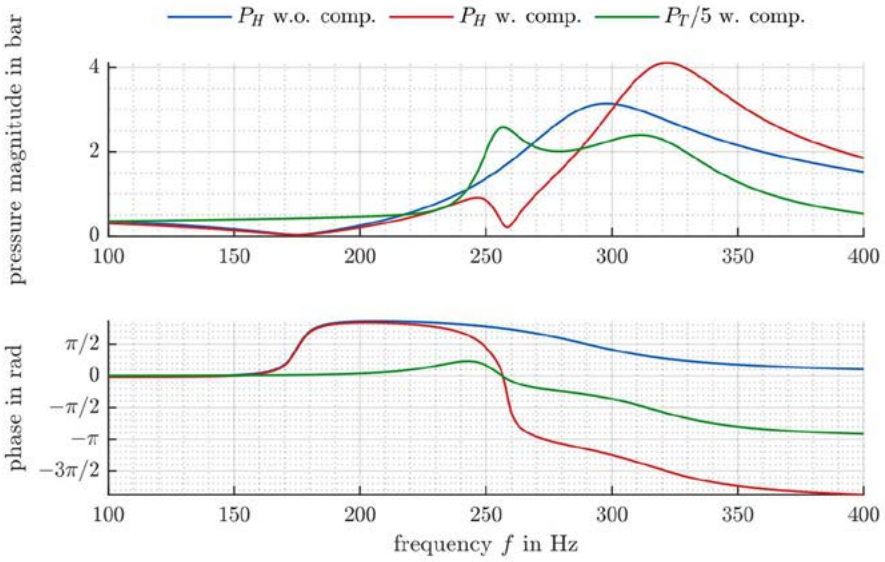


Figure 5.7. Calculated pressures for 189 cm pipeline length

By reducing the pipeline length to 189 cm, the resonance frequency of the hydraulic system is increased and no longer coincides with the compensation frequency. Around the compensation frequency, the compensator still reduces the pump outlet pressure magnitude.

Condition (4.2) is no longer fulfilled so that the closed-form solutions are not applicable in this case.

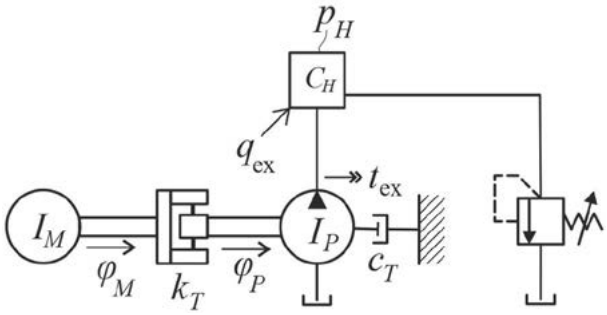


Figure 5.8. Schematic of setup with internal gear pump and outlet capacitance

If the pipeline is disconnected from the pump outlet, the model in Figure 5.8 applies. The respective frequency domain formulation reads

$$\left(-\omega^2\begin{bmatrix} I_M & 0 & 0 \\ 0 & I_P & 0 \\ 0 & -\frac{V_P}{2\pi} & C_H \end{bmatrix} + i\omega\begin{bmatrix} 0 & 0 & 0 \\ 0 & c_T & 0 \\ 0 & 0 & \frac{1}{R(i\omega)} \end{bmatrix} + \right.$$

$$\left. + \begin{bmatrix} k_T & -k_T & 0 \\ -k_T & k_T & \frac{V_P}{2\pi} \\ 0 & 0 & 0 \end{bmatrix}\right)\begin{bmatrix} \Phi_M(i\omega) \\ \Phi_P(i\omega) \\ P_H(i\omega) \end{bmatrix} = \begin{bmatrix} 0 \\ T_{ex}(i\omega) \\ i\omega Q_{ex}(i\omega) \end{bmatrix}. \quad (5.11)$$

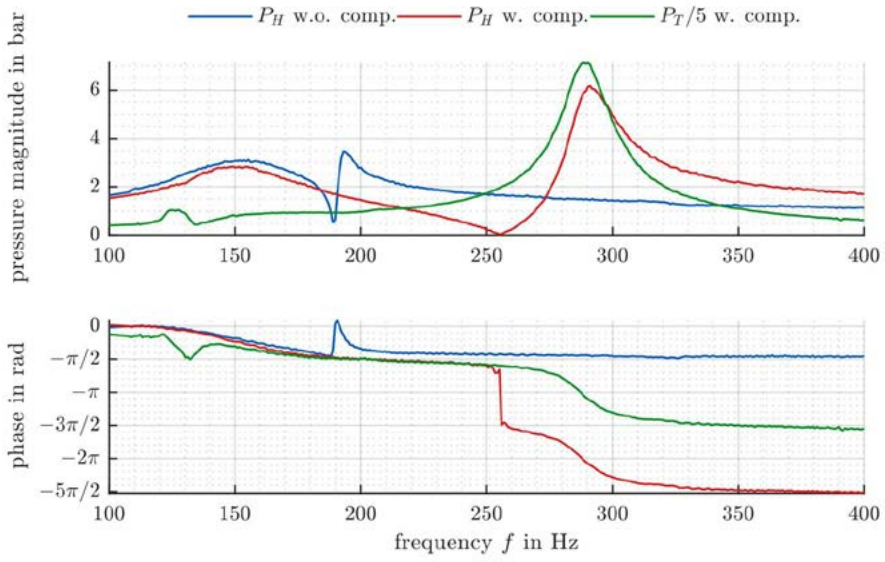


Figure 5.9. Measured pressures with outlet capacitance

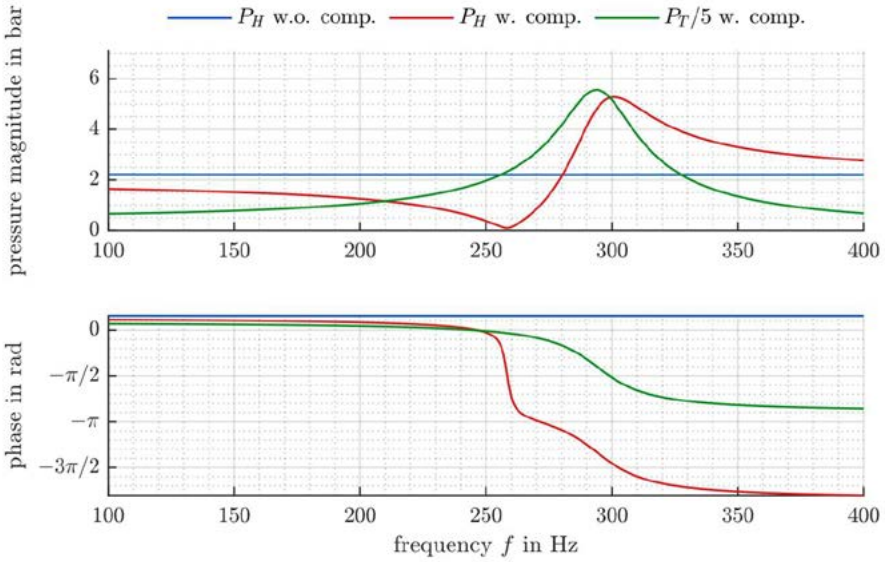


Figure 5.10. Calculated pressures with outlet capacitance

Figure 5.9 shows the corresponding measurements. Around 190 Hz, they contain an additional effect from the torque measurement shaft, which is not included in the model. Calculated results are shown in Figure 5.10. At the compensation frequency, the pump outlet pressure magnitude is almost cancelled. The closed-form solution (4.13) results in a pressure magnitude reduction to 3 %. There is no lower side resonance, and broadband pulsation reduction is achieved below the compensation frequency. Above this frequency, a heavily overshooting side resonance is observed. The measured bandwidth amounts to 13.7 %, and the frequency domain model yields a bandwidth of 16.7 %. For comparison, the closed-form solution (4.7) results in a bandwidth of 11.3 %.

6. PREDICTIONS FOR AXIAL PISTON PUMP

Under non-resonant conditions, closed-form solutions for compensator performance shall be evaluated for an axial piston pump with seven pistons, 25 cm^3 displacement, and 0.0017 kg m^2 mass moment of inertia at a rotational speed of 1500 rpm. In a first approach, the hydraulic capacitance at the pump outlet and the torsional damping constant are taken from the internal gear pump with the same displacement. The compensator is tuned to the fundamental frequency of dynamic pulsations, which occurs at 175 Hz. Equation (4.8) can be used to calculate the required torsional stiffness of the compensator coupling. With Equation (4.6), Equation (4.7) yields a bandwidth of 12.9 %. Equation (4.13) results in a pressure magnitude reduction to 2.0 %. At 3000 rpm, a bandwidth of 3.4 % and a pressure magnitude reduction to 3.9 % are obtained. Smaller bandwidth is due to higher stiffness of the compensator coupling required for 350 Hz compensation frequency. If damping is increased by a factor of five, pressure magnitude reductions to 9.7 % and 19.2 % result for 1500 rpm and 3000 rpm, respectively.

7. CONCLUSIONS

For the investigated internal gear pump, the rotary mechanical pulsation compensator shows excellent performance at a mean outlet pressure of 100 bar and a constant rotational speed near 1200 rpm. Due to sufficient bandwidth, it is robust with respect to small speed variations. Similar performance is expected from an equally sized axial piston pump at a rotational speed of 1500 rpm. At higher pump speeds, the remaining pulsation will be higher and the bandwidth will decrease.

For external gear pumps, geometric pulsations may dominate and cause a significant excitation torque. In this case, the closed-form solutions for compensator performance are no longer applicable. Appropriate investigations of external gear pumps will be presented in the future.

8. REFERENCES

[1] G. Mikota, 'Vorrichtung zur Reduktion von Druckpulsationen in einem Hydraulik-system', Austrian Patent AT 526449 B1 2024-05-15.

[2] G. Mikota, B. Manhartsgruber, 'A novel pulsation compensator for displacement machines', Proc. 14th International Fluid Power Conference (14th IFK), pp. 395-406, Dresden, 19th – 21th March, 2024.

[3] G. Mikota, 'Design and simulation of a drivetrain integrated pulsation compensator for pumps', Proc. BATH/ASME 2024 Symposium on Fluid Power and Motion Control (FPMC 2024), FPMC2024-140002, Bath, 11th – 13th Sept., 2024.

[4] D. Holzer, G. Mikota, 'Measurement of a rotary mechanical pulsation compensator under hydraulic resonance', Proc. 19th Scandinavian International Conference on Fluid Power (SICFP'25), Paper No. 43, Linköping, 2nd – 4th June, 2025.

[5] G. Hametner, R. Scheidl, 'Duality and its application for the dynamic analysis of hydraulic drive systems', Proc. Bath Workshop on Power Transmission and Motion Control (PTMC 2003), pp. 65-76, Bath, 10th – 12th Sept., 2003.

[6] H. Dresig, F. Holzweißig, 'Dynamics of Machinery', Springer, 2010.

[7] X. Zhao, A. Vacca, 'Theoretical investigation into the ripple source of external gear pumps', Energies 12 (2019) 535.

[8] H. Chai, G. Yang, G. Wu, G. Bai, W. Li., 'Research on flow characteristics of straight line conjugate internal meshing gear pump', Processes 8 (2020) 269.

[9] H. Theißen, 'Die Berücksichtigung instationärer Rohrströmung bei der Simulation hydraulischer Anlagen', Dissertation, RWTH Aachen, 1984.

Biographies



Gudrun Mikota received the diploma in mechanical engineering from Vienna University of Technology in 1994, the doctorate in mechanical engineering from Vienna University of Technology in 2001, and the habilitation in dynamics of machines from Johannes Kepler University Linz in 2021. She has worked with companies Jenbacher, LCM, and Eisenbeiss. In 2023, she was appointed Professor of Fluid Power at Johannes Kepler University Linz. Her research areas include fluid power dynamics, modal analysis of hydraulic and rotating systems, energy efficient hydraulic drives, and the dynamic behaviour of the human arterial system.



David Holzer received the bachelor's degree in mechatronics from Johannes Kepler University Linz in 2021 and the master's degree in mechatronics from Johannes Kepler University Linz in 2023. He is currently working as a Scientific Assistant at the Institute of Machine Design and Fluid Power, Johannes Kepler University Linz. His research area is the dynamic behaviour of efficient hydraulic machinery.

Enhancing Volumetric Efficiency and Productivity of External Gear Pumps through Optimized Run-In Strategies.

Alfonso Rosario Apuzzo[1*,2], Sara Bonati[2], Emma Frosina[1] and Luca Riccò[2]

[1]*Department of Engineering, University of Sannio, Piazza Guerrazzi, 82100 Benevento, Italy.*

[2]*Duplomatic MS S.p.a., Via Mario Re Deapolini, 24, 20015, Parabiago Milan, Italy.*

** E-mail: a.apuzzo@studenti.unisannio.it*

Abstract.

In the final stage of production, External Gear Pumps (EGPs) must be run-in following a specific working cycle in which the pump body is subject to material removal resulting in a consequent increase in energy demand. The run-in cycle impacts the pump performance and the production costs. In this study, five new run-in cycles are defined with the objective to analyse the impact on the power absorbed during and after the cycle, the material removed from the body and the impact on pump performance. The work shows that the use of different run-in cycles can result in a higher pump volumetric efficiency, up to + 2%, combined with a cost reduction of 1.74% and an increase of productivity of 78%. Subsequently, experimental data were used to define a preliminary AI based algorithm to predict the run-in cycles parameter. The AI algorithm, which was trained and validated on the experimental data, demonstrated a high predictive capability with a maximum error of 6%.

Keywords. External Gear Pump, Run-in cycle, Efficiency, AI Algorithms.

1. INTRODUCTION

External Gear Pumps (EGPs) are positive displacement machines widely used in fluid power applications across both mobile and industrial sectors. EGPs are widely chosen, having a good combination of performance, reliability and cost-effectiveness. The EGP under study, manufactured by Hydreco (HY2.16 pump), is a fixed positive displacement machine of 16 cm^3/rev capable of operating at a maximum pressure of 260 bar and a rotational speed range of 500 – 3200 rpm. It finds application in a wide range of fields, from agriculture to material handling and aeronautics devices. In Figure 1. 1, an exploded view of HY2.16, a pressure compensated (PC) EGP, is shown.

Before being delivered to customers, EGP is end-of-line (EoL) tested with a specific run-in cycle in which the pressure forces cause the gears translation and inflection towards the suction side. In this condition, the tips of gears' teeth interfere with the pump body. Due to these micro-movements of the gears, typically made of hardened steel, the body on the low-pressure side undergoes a milling phase. The gears dig their optimal seats into the body and,

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

as a result, leakage flows can be minimized (maximization of the volumetric efficiency) at that operating condition.

The run-in cycle has a relevant importance for EGPs, as it defines the internal profile of the pump housing (or body) and, consequently, its efficiency. This cycle is usually performed at a constant speed within the operative pump range, but different levels of speed can be adopted [1, 2].



Figure 1. 1– Exploded view of Hydreco HY2 pump (aluminium housing)

The amount of removed material is related to the operative working conditions and the clearance variability [1]. Once the material removal phase is complete, a performance check (testing) phase is conducted to verify the pump's performance. The duration time of the cycle depends on the pump tested, it can vary from a couple of minutes [1] until to 15 minutes [3]. This time impacts on the total cost of an EGP which is given by the summation of the components, assembly and EoL testing (run-in and performance check) costs. Specifically, the assembly and EoL testing costs are strictly dependent on the hourly operating cost, which typically ranges between the 30 – 60€/h.

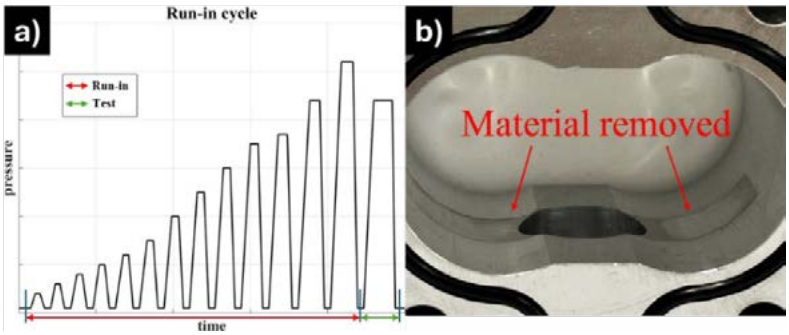


Figure 1. 2 – a) Standard run-in cycle; b) Gear seats with removed material after the run-in cycle.

Figure 1. 2.a shows an example of a typical run-in cycle in which the load pressure has a step variation; the run-in and testing phase are highlighted. Figure 1. 2.b shows how the gear seats appear after the cycle in Figure 1. 2.a.

In the open literature, there are only a limited number of studies focused on the run-in cycles and correlated effects on the pump performance. The first work was published by M. Borghi et al. [2]. The authors developed a simple analytic model to predict the internal profile of the pump body for different run-in pressures and rotational speeds. The gear position was defined by two contributions, the recovery of radial clearances, as in [4] and the eccentricity contribute evaluated with the Sommerfield theory. In [2], the impact for different run-in speeds in range 500 – 3000 rpm was presented. In [2] and [4], the calculation of the internal body profile is integrated into lumped-parameter models, thus enabling the correlation of post-run-in geometrical variations with the unit performance. In [1], a lumped parameter model for EGPs was adapted to predict the run-in cycle and the amount of material removed was calculated in each pressure step. The authors verified experimentally and numerically that the major amount of material was removed in the first low pressure ramps, and they proposed a shorter cycle removing four out of thirteen ramps at the middle-hight pressures and extending two ramps at low pressures. In [3], the authors proposed a mathematical model for EGPs that can predict the volume of material removed under different operating conditions and define a run-in cycle for applications up to 400 bar. The proposed cycle has been defined to have the same depth of asportation for each pressure ramp. In these studies, the authors refer to cycles with a standard shape as reported in Figure 1. 2.a. and no references to different shapes are present. More recently, there has been an ever-increasing diffusion and interest in the AI algorithms, including their application in the study of EGPs. In [5] neural networks were used to predict the output flow-rate of an EGP, while in [6], machine learning algorithms were used for predictive maintenance. However, based on the best knowledge, studies using AI algorithms to investigate the run-in cycle on EGPs are not present in the open literature.

In the present work, starting from a baseline run-in cycle, used as a reference, and the authors' expertise, five new cycles have been defined. A sample of thirty EGPs was tested and for each pump, its performance characterization was performed to evaluate in detail the impact of the new run-in cycles. In addition, torque and pressure have been acquired during the run-in cycle to estimate the absorbed energy and to identify the phases in which the material removal occurs. Moreover, the wear profile on the gear seats was measured using a 3D Coordinate Measuring Machine (CMM). Finally, a preliminary model based on AI algorithm has been developed. The acquired experimental data were used in part to train AI algorithm and in part to test it, showing a very good alignment.

The aim of the present work is to define new run-in cycles that allow to have a shorter test time and to preserve and improve the pump performance. Cycles with an innovative pressure shape which differ from [1, 3] will be presented and analysed.

2. EXPERIMENTAL ACTIVITY – RUN-IN CYCLES AND PUMPS DEFINITION

Five new run-in cycles, with a shorter time duration, were defined to reduce both the pump cost and energy consumption to have as resulting also economic and environmental benefits. For each cycle, five pumps have been properly assembled and tested.

2.1. Run-in cycles

The run-in cycle is defined as a load cycle performed at a specified rotational speed (N), which can be either constant or variable over time. The term load cycle refers to the evolution of pressure over time. In this work six run-in cycles have been defined through five different load cycles that can be grouped in three different families identified as A, B and C. Table 2. 1 summarises the parameters that define each cycle.

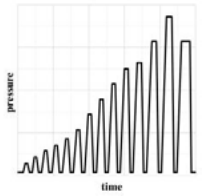
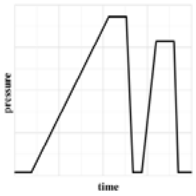
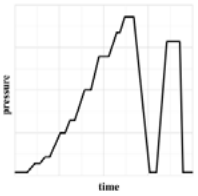
A	B	C
		
$\#\boldsymbol{S}\,;\boldsymbol{p_i^{max}}\,;\boldsymbol{t_i^j}\,;\boldsymbol{p_{test}}\,;\,\boldsymbol{t_{test}}$	$\boldsymbol{p^{max}}\,;\boldsymbol{t^j};\boldsymbol{p_{test}}\,;\,\boldsymbol{t_{test}}$	$\#\boldsymbol{S}\,;\boldsymbol{p_i^{max}}\,;\boldsymbol{t_i^j}\,;\boldsymbol{p_{test}}\,;\,\boldsymbol{t_{test}}$
$i = 1, 2, \dots, \#S$ $j = 1, 2, \dots, 4$	$j = 1, 2, \dots, 4$	$i = 1,2, \dots, \#S$ $j = 1, 2, 3\ if\ i = 1, 2, \dots, \#S - 1$ $j = 2, 3, 4\ if\ i = \#S$

Table 2. 1 – Load cycles definition

The variable $\#\boldsymbol{S}$ is the number of pressure steps, $\boldsymbol{p_i^{max}}$ the pressure of the i-th ramp and $\boldsymbol{t_i^j}$ one of the four times that define the ramp i-th. The times $\boldsymbol{t^j}$ used to define the pressure ramps are: 1) t^1 : time at low pressure; 2) t^2: time to pass from low to high pressure; 3) t^3 : time at high pressure; 4) t^4 : time to pass from high to low pressure. The parameters $\boldsymbol{t_{test}}$ and $\boldsymbol{p_{test}}$ define the test ramp that for this study has been maintained constant for all the cycles. Another common parameter for all the cycles is the $\boldsymbol{p_S^{max}}$, which value is the one used to run-in the pump for the maximum operative pressure. The defined run-in cycles are reported in the Table 2. 2. Although the actual values of these parameters are not disclosed for confidentiality reasons, it can be stated that $N_2 = 2 * N_1$, $S_{New} < S_{STA}$, $t^2_{1STEP\ 3} = 4 * t^2_{1STEP\ 1}$ and the t^3_S (time at maximum pressure) is the same for all cycles.

The last two columns of Table 2.2 show the percentage variations in the new cycles compared to the current reference, called STA N1.

Table 2. 2 - Run-in cycles defined

ID cycle	Load cycle	$\#\boldsymbol{S}$	$\boldsymbol{N_{Run-in}}$	$\boldsymbol{N_{test}}$	$\boldsymbol{t_{cycle}}$	$\boldsymbol{E_{cycle}}$
STA N1	*A*	S_{STA}	N_1	N_1	t_{STA}	E_{STA}
STA N2	*A*	S_{STA}	N_2	N_1	*+0.0%*	*+100%*
STAIR	*C*	S_{STA}	N_1	N_1	*-34.6%*	*-17.4%*
New STA	*A*	S_{New}	N_1	N_1	*-29.5%*	*-32.9%*
1STEP 1	*B*	*1*	N_1	N_1	*-57.4%*	*-49.3%*
1STEP 3	*B*	*1*	N_1	N_1	*-39.7%*	*-23.4%*

As shown in Table 2. 2, all new cycles, except for STA N2, are better than STA N1 in terms of cost reduction, thanks to their shorter duration and lower theoretical energy demand (equation 2.1).

$$E\ [J] = \int_{t=0}^{t_{cycle}} T_{th} * \omega * dt \qquad (2.\ 1)$$

The new cycles, with a reduced run-in time compared to STA N1 provide an increment of the test rig productivity $\#Pumps/h$, defined as the number of pumps tested hour (Table 2. 3). It was considered constant the time to allocate the pump on the test rig as the cycle changes.

Table 2. 3 – Test rig productivity compared to the STA N1 cycle

	STA N2	**STAIR**	**New STA**	**1STEP 1**	**1STEP 3**
$\#\boldsymbol{Pumps/h}$	*+0.00%*	*+36.12%*	*+29.29%*	*+78.61%*	*+43.99%*

A shorter run-in time allows for reducing the production cost and finally, combined with the lower energy demand, impacts on the greenhouse gas emissions (GHG) due to the production process, an aspect not negligible considering the objective of a zero net GHG by 2050 established by the Daikin Group, of which Hydreco is a member.

2.2. Samples assembly

Five pumps have been tested for each run-in cycle. To avoid that experimental results could be influenced by the manufacturing capability, pumps have been assembled combining the components to keep basically constant the average radial clearance between gears and body $\boldsymbol{\delta_r^{av.}}$, and the maximum radial shift of the shafts $\boldsymbol{\epsilon_{max}^{av.}}$. Five kits composed of two thrust plates, driven and driving gears were defined and kept constant. Consequently, thirty pump bodies have been selected, ensuring the smaller possible variability of radial interference $\boldsymbol{int}$. Refer to appendix for the definition of $\boldsymbol{\delta_r^{av.}}$, $\boldsymbol{\epsilon_{max}^{av.}}$ and $\boldsymbol{int}$.

The relative standard deviation (RSD) of $\boldsymbol{int}$, defined as the ratio between the standard deviation and the mean value, was 4.6%. This corresponds to a value within the measurement accuracy of the CMM used for measuring the pump bodies (section 3.3), and it was considered acceptable. As concerns axial clearance, there was more variability, with an RSD of 24.6%; nevertheless, for the purposes of this study, this variability was considered not to impact and thus negligible.

2.3. Test bench layout

Figure 2. 1 shows the hydraulic scheme of the test bench used for the dynamic characterization of the run-in cycles and the acquisition of pump performances in steady-state conditions. During the run-in cycle were acquired the suction (1) and delivery (2) pressures, the torque (3) and the shaft speed (4). While the performance test also acquired the delivery flow rate (5).

Table 2. 4 - Sensors used for the experimental characterization

#ID	**Symbol**	**Component**	**Range**	**Accuracy**
1	p_1	Trafag 8845.73	0 - 1.6 barA	±0.5% of span
2	p_2	Trafag 8254.84	0 - 400 bar	±0.4% of span
3	T_{eff}	NCTE Series 3000-Model 250	0 - ±250 Nm	±0.2% of span
4	N	Rotary Encoder	0 - ±3000 r/min	±0.05% of span
5	Q_{eff}	Flow meter KRACHT - size 3	0 - 120 l/min	±0.3% of span

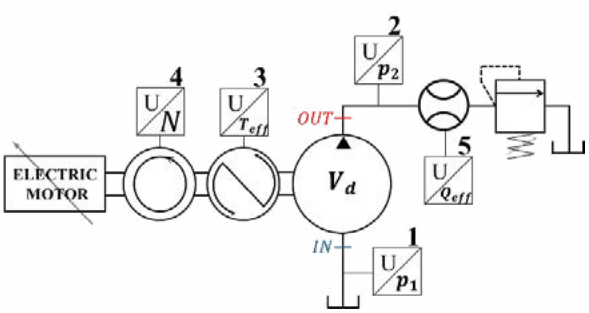


Figure 2. 1 - Test rig layout

The specifications of the used transducers are reported in Table 2. 4. For the dynamic acquisition of the run-in cycles, a sample frequency of 1kHz and a saving frequency of 20Hz were set. The steady state acquisition of the efficiencies, in the pressure range [60 - 240 bar] and in the speed range [1000 - 3000 rpm], has been performed with a sample and saving frequencies of 1kHz. An ISO VG-46 oil with a temperature setting of $39 \pm 2°C$ was used for the tests. The working operative point is varied by means the pressure control valve and the electric motor.

3. EXPERIMENTAL RESULTS AND DISCUSSION

A complete characterization of the new run-in cycles was executed and their impacts on the pump performance were analysed. Three separate acquisitions have been performed: 1) Run-in cycle; 2) Performance characterization; 3) Measurement of the material removal.

3.1. Run-in cycles – Instantaneous torque

The continuous acquisition of all the variables of interest marked from 1 to 4 in Figure 2. 1 was performed for the entire run-in cycle. The total efficiency, defined as the product of the volumetric and hydromechanical efficiencies, was calculated during the constant pressure portion of the test phase; refer to the last ramps of Figure 3. 1.

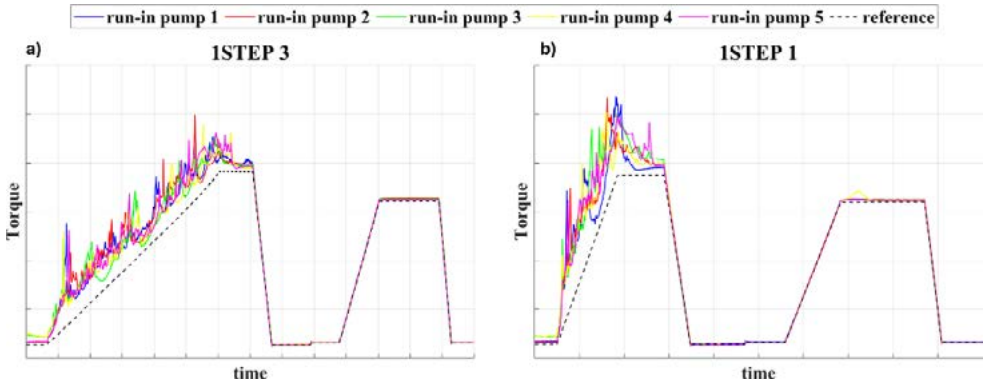


Figure 3. 1 - Torque absorbed during the run-in cycles #1: a) 1STEP 3 cycle; b) 1STEP 1 cycle. The scale factor of the time scale of graph "b" with respect to the graph "a" is 7:5.

Figure 3. 1 shows the torque during the run-in cycles of five pumps tested with 1STEP 1 and the 1STEP 3 run-in cycles (Table 2. 2). The black dashed line represents the torque absorbed when does no material removal occurs (after run-in is completed). It depends only on the pressure load and the pump mechanical losses, and it is identified as a reference curve. All the torque curves during the run-in phase show a higher mean value than the reference curves, as well as spikes and fluctuations due to interactions and contact between internal components during the material removal.

Table 3. 1 – Run-in cycles acquisition

ID cycle	δE^{th}[%]	δE^{eff}[%]	$E^{eff-mill}$ [kJ]	T_{av}^{add}[Nm]	δT_{max}[%]
STA N1	*0%*	*0%*	*51.8*	*2.8*	*0%*
STA N2	*+100%*	*+81.4%*	*32.1*	*0.8*	*-13.8%*
STAIR	*-17.4%*	*-14.3%*	*49.4*	*4.0*	*-14.8%*
New STA	*-32.9%*	*-26.3%*	*48.3*	*3.6*	*-6.8%*
1STEP 1	*-49.3%*	*-45.7%*	*32.4*	*4.0*	*-7.8%*
1STEP 3	*-23.4%*	*-19.6%*	*54.2*	*4.7*	*-15.9%*

Table 3. 1 shows the average values of the most relevant physical quantities, calculated on the five pumps tested for each cycle. The first two columns show the percentage deviation from the reference cycle of the theoretical $\boldsymbol{E^{th}}$ and the effective $\boldsymbol{E^{eff}}$ energy required for the new cycles. The effective energy reduction (except for the cycle *STA N2)* of the new run-in cycles results smaller than the theoretical one. The relationship between $\boldsymbol{\delta E^{th}}$ and $\boldsymbol{\delta E^{eff}}$ is described by equations (3. 1) and (3. 2); refer to the appendix for the simplifying assumptions adopted. Equation (3. 2) was obtained assuming that $\boldsymbol{E^{eff-mill}}$ (energy difference between the run-in cycles with and without material removal) is constant, however, experimental results have demonstrated that this is only true for four out of six cycles. The $\boldsymbol{E^{eff-mill}}$ required by the New STA and 1STEP 1 cycle is generally lower than that required by the other cycles. When $\boldsymbol{E^{eff-mill}}$ is not constant, equation (3. 2) becomes equation (3. 3).

$$\delta E_{i,j}^{eff}[\%] = \frac{\left(E_{i,j}^{eff-runin} + E_{i,j}^{eff-mill}\right) - \left(E_{STA\ N1,j}^{eff-runin} + E_{STA\ N1,j}^{eff-mill}\right)}{\left(E_{STA\ N1,j}^{eff-runin} + E_{STA\ N1,j}^{eff-mill}\right)} * 100 \qquad (3.\ 1)$$

With $i = \{STA\ N1, STA\ N2, STAIR, New\ STA, 1STEP\ 1, 1STEP\ 3\}$; $j = 1,2, \dots, 5$

$$\delta E_{i,j}^{eff}[\%] = \delta E_i^{th} * A \quad ; \ 0 < A < 1 \qquad (3.\ 2)$$

$$\delta E_{i,j}^{eff}[\%] = \delta E_i^{th} * A + B \ ; \ 0 < A < 1 \qquad (3.\ 3)$$

The parameter $\boldsymbol{T_{av}^{add}}$ represents the average additional torque required by the pump during the cycle. It is defined by equation (3. 4).

$$T_{av}^{add} = \frac{1}{t_{cycle}} \int_{t=0}^{t_{cycle}} \left[T_{Run-in\ 1}^{eff}(t) - T_{Run-in\ 6}^{eff}(t)\right] * dt \qquad (3.\ 4)$$

In the last column of Table 3. 1, the maximum torque values obtained in the selected cycles are reported. The data show that the maximum torque in the new cycles is lower than that of the reference cycle.

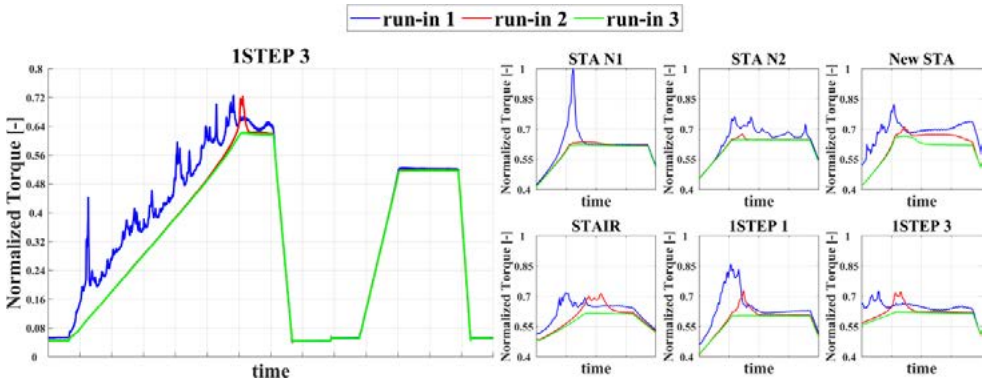


Figure 3. 2 - Torque absorbed by the 1STEP 3 run-in cycles 1, 2 and 3 of pump 1 with a zoomed view of the area at constant pressure of all the cycles.

The trend of $\boldsymbol{T_{av}^{add}}$ in Table 3. 1 is also confirmed in Figure 3. 2 in which the torque scales are normalized to the maximum torque of the STA N1 run-in 1. The left-hand side chart shows three consecutive run-in cycles of one pump during 1STEP 3 cycle. On the right-hand side, zoomed views of the area at the maximum constant pressure for all the six cycles are shown. It should be noted that the temporal extension shown is the same for all the zoomed views. Torque spikes were observed during the second and third consecutive run-in cycles for all tests except for STA N1. This seems to indicate that the material removal process was incomplete after the first run-in, suggesting that the selected $\boldsymbol{t_S^3}$ value was not optimal. Since, typically during the testing, the run-in cycle is performed only one time, it must be ensured that a complete material removal is obtained. Consequently, future run-in cycles will be designed with an increased value of $\boldsymbol{t_S^3}$ to be aligned to the actual one (STA N1).

For the EGP under investigation, the run-in phase (baseline cycle) represents the 3.95% of the production cost composed of: components, assembly phase and run-in cycle costs. In addition, the reduction of the run-in time allows not only to increase productivity (Table 2. 3) but, combined with a minor energy demand (Table 3. 1), allows for further reduction of the costs up to 1.74% (Table 3. 2).

Table 3. 2 – Pump cost compared to the STA N1 cycle

	STA N2	STAIR	New STA	1STEP 1	1STEP 3
Pump cost	*+0.03%*	*-1.04%*	*-0.90%*	*-1.74%*	*-1.20%*

Table 3. 2 shows as the major cost reduction is related to the reduction of the run-in time, in fact for STA N2 cycle a +81.4% of energy demand (Table 3. 1)corresponds to only +0.03% of the pump cost. In the present work, an energy cost of 0.12€/kWh was considered. Figure 3. 3 shows the efficiency values of each pump at the end of the run-in cycles. These efficiencies, referred to as the check operative point defined by the test phase, exhibit variation across different run-in cycles. The new run-in cycles seem to improve the pump performances; the benefit of the new run-in cycles is mainly impacting the volumetric efficiency that reaches values up to +1.5% respect to the one obtained with the standard run-in cycle.

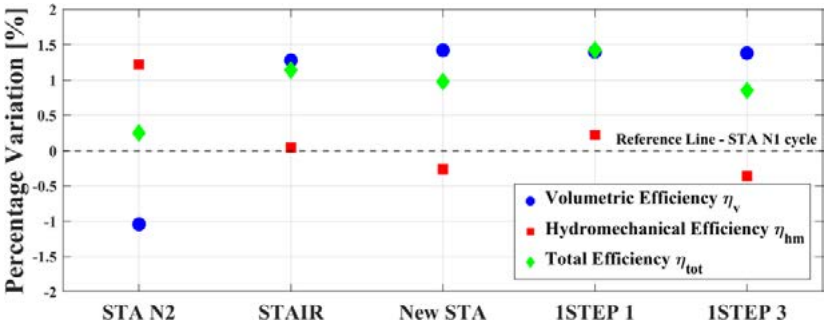


Figure 3. 3 - Run-in 1 efficiencies

3.2. Pump efficiencies

To confirm the results shown in the Figure 3. 3 a complete characterization of the pump efficiencies, acquiring #21 operative points, was performed. The characterization was carried out after repeated run-in, therefore the material removal phase can be considered complete for all bodies. The results pointed out nonsignificant variation for the hydromechanical efficiency (η_{hm}) while a more significant variation for the volumetric efficiency (η_v). Thus, Figure 3. 4 reports only the values of η_v in the different operating points.

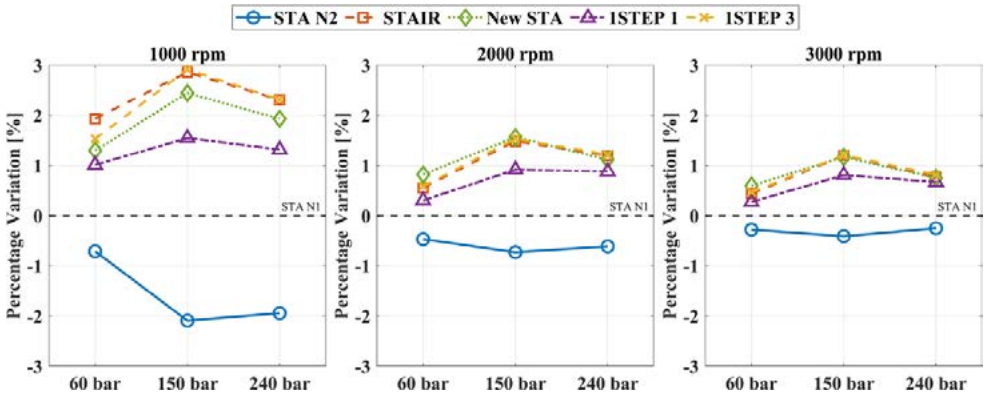


Figure 3. 4 - Volumetric Efficiency

The results show an effective improvement in volumetric efficiency for all cycles except STA N2, confirming the trend reported in Figure 3.3. The cycles defined in Table 2. 2 differ in shape, cycle time, and pump speed. The p_S^{max} value was the same for all the run-in cycles, but a reduction in cycle time corresponds to an increased pressure gradient. From the analysis of the volumetric efficiencies, it is possible state the following: 1) with fixed pressure gradient, an increase in run-in speed negatively impacts volumetric efficiency of the pump (STA N2 compared to STA N1); 2) with fixed run-in speed, an increased pressure gradient positively affects volumetric efficiency (STAIR, New STA, 1STEP 1 and 3 compared to STA N1). However, this effect reduces with excessive cycle time reduction (1STEP 1 compared to 1STEP 3); 3) with a fixed pressure gradient and rotational speed, a change of the cycle profile does not seem to have an impact on the volumetric efficiency

(STAIR compared to 1STEP 3). Cycles STAIR and 1STEP 3 have comparable cycle times and therefore comparable pressure gradients.

Regarding the cost reduction, the run-in cycle 1STEP 1, that is the shortest one, represents the optimal solution. However, considering the results reported in Figure 3. 4 the cycles 1STEP 3 and STAIR represent the best choices.

3.3. Measurement of the pump bodies

A CMM with an accuracy of 2 μm has been used to measure material removal. Results are presented in Figure 3. 5 and Table 3. 3, where the internal profiles measured in the central section of the pump body are reported.

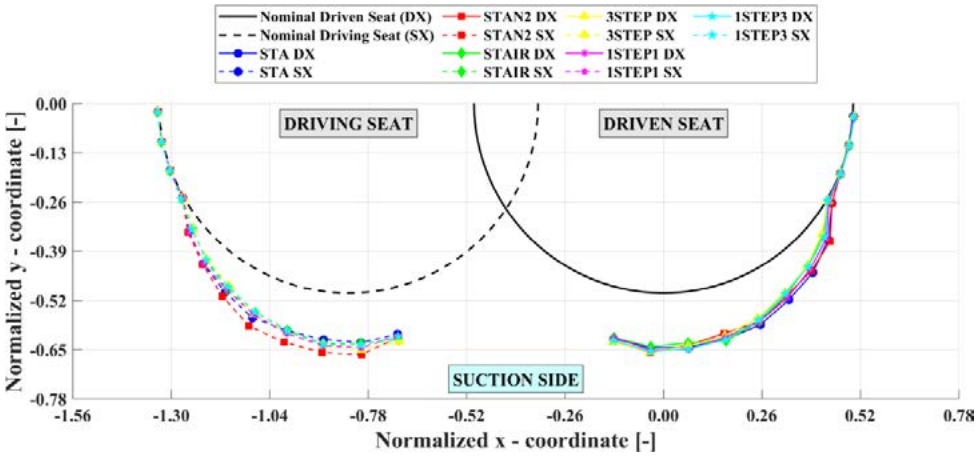


Figure 3. 5 - Measured material milled on the Internal pump body. DX identifies the driven gear while SX the driving one.

For confidential reasons, in Figure 3. 5 the axes are normalized to the tip diameter. The measured profiles differ slightly, with a maximum volume variation of 6%. The measured profile of the bodies run-in with cycle STA N2 shows a volume of material removed on the driving gear side comparable with the driven gear side (δV_{DX-SX}^{mill} below 4%). This can be directly correlated to the observed reduction of the volumetric efficiency (Figure 3. 4).

Table 3. 3 - Volume of material milled

	STA N1	**STA N2**	**STAIR**	**New STA**	**1STEP 1**	**1STEP 3**
$\boldsymbol{V_{DX}^{mill} + V_{SX}^{mill}}$	$V_{STA\,N1}^{mill}$	+5.94%	- 6.12%	-5.93%	-1.99%	-5.44%
$\boldsymbol{\delta V_{DX-SX}^{mill}}$	-15.82%	-3.82%	-16.56%	-13.93%	-16.53%	-17.43%

The use of a CMM permitted to obtain a 3D map of the milled regions, as showed in Figure 3. 6, and to estimate the volume of removed material (as reported in Table 3. 3). Figure 3. 6 and Figure 3. 7 show the 2D shape of the milled areas along the axial direction parallel to the rotation axis of gears and the 3D maps for the run-in cycles STA N2 and STAIR. For confidential reasons, the z-axes in Figure 3. 6 and Figure 3. 7 are normalized to the pump width. As previously observed in Figure 3. 5, for the STA N2 run-in cycle the depth of milling for the driving and driven gears seats are aligned. While for the STAIR run-in cycle the milling depth of the driven gear side is deeper than the one on the driving side.

Comparing the two cycles, a greater height of milling is observable for the STA N2 cycle with respect to the STAIR one, which returns a better volumetric efficiency.

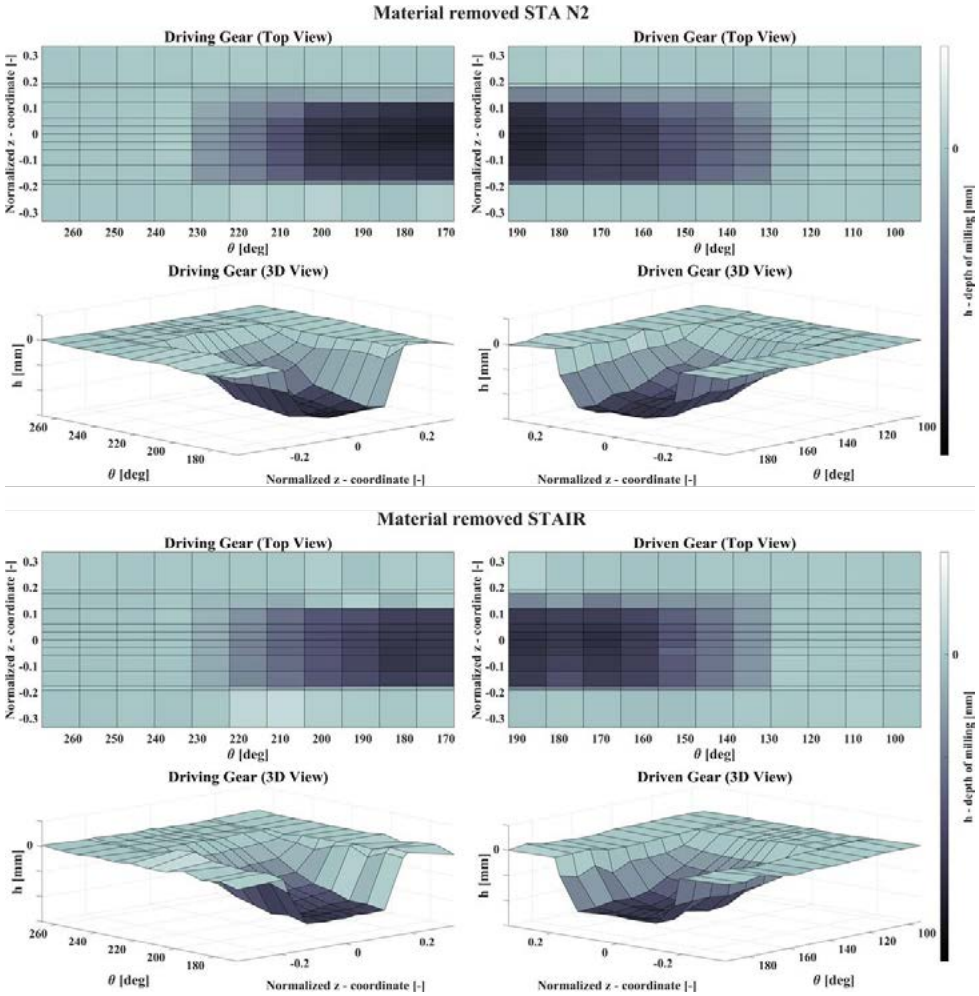


Figure 3. 6 - Top View and 3D view of the material removed on the suction side for the STA N2 and STAIR cycles

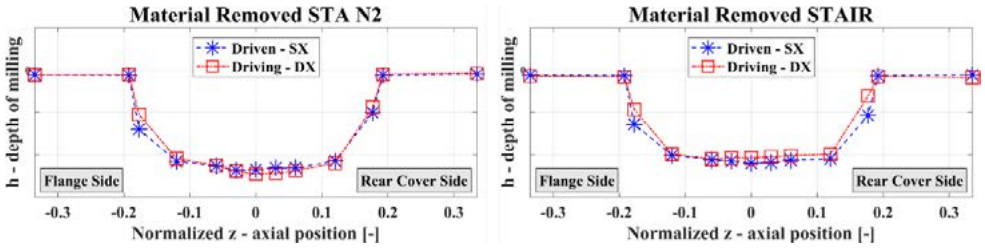


Figure 3. 7 - Milling line in the axial direction

Significant improvements in volumetric efficiency were observed with almost all the new cycles (Figure 3. 4), but the STA N2 with higher shaft speed and volume of material removed showed worsening performance. The pressure gradient or the run-in cycle seems to have an impact on the material removal. Varying the run-in cycles but maintaining constant p_S^{max} and t_S^3, the internal profile shape changes slightly. From the analysis of the removed volume, it is possible state the following: 1) under a fixed pressure gradient, as the run-in speed increases, the material removal on the driving and driven gears becomes comparable between the driving and the driven gears and a greater total volume of material is removed with a negative impact on the volumetric efficiency (cycle STA N2 compared to STA N1); 2) contrarily, with a fixed rotational speed, as the pressure gradient increases (reduction of the time cycle) the total removed material decreases impacting positively on the volumetric efficiency (cycles STAIR, New STA, 1STEP 1 and 3 compared to STA N1). The increase in volumetric efficiency (Figure 3. 4) is directly correlated with the reduction of the removed material volume in gears seats (as shown Table 3. 3), where a lower volume corresponds to a reduced milling height (as shown in Figure 3. 6 and Figure 3. 7). In conclusion, the combination of p_S^{max} and N defines the final internal profile shape [1] and the maximum depth of milling [2].

4. AI ALGORITHM: TORQUE PREDICTION

With the future goal of optimize the run-in cycle, a preliminary model based on AI algorithm has been developed. The use of an AI-based tool is motivated by the need to model a complex, non-linear and transient behaviour of an external gear pump, which is difficult to describe accurately using purely physics-based approaches. An EGP, equipped with sensors, generates data in the form of time series that represent the inputs and outputs for the model. For this activity, at this initial stage, the commercial software NeurEco® [7], with a specific dynamic template for processing time series data, has been used for the creation of a predictive model. The tool defines a neural network through a controlled topological enrichment process that uses parsimony to reach optimal accuracy [8]. The procedure starts from a minimal architecture that is progressively enriched by adding the most valuable neurons and connections as needed. This iterative expansion of the architecture allows the model to reach the global optimum for the problem. When no further enrichment improves the prediction, the process stops. This approach provides the model with the requested complexity. NeurEco® enforces a minimal model size, which eliminates overfitting by construction.

This commercial tool was used to develop a preliminary AI-based predictive model for the EGPs run-in cycle. The first application of the model was focused on predicting of the torque needed during the 1st and 6th run-in, having the delivery pressure profile as input. The analysis presented in this section refers to the cycles with the N1 speed. Considering that the 6th run-in was acquired only for one pump for each cycle type, the total data set available was composed of five continuous acquisitions (STA N1, STAIR, New STA, 1STEP 1 and 1STEP 3), while for the first run-in cycle all the five samples were available for each cycle.

Firstly, a dynamic physical model, called Model0, with the pressure profile as input, was built for predicting the 6th run-in (with no material removal). The model outputs are torque, which allows for estimating the hydromechanical efficiency, and the absorbed energy of the EGP. Three samples, STA N1, 1STEP 1 and New STA, were used for the training phase and

the remaining two, STAIR and 1STEP 3, were used for the blind test phase. Figure 4. 1 shows the acquired (*exp.*) and predicted (*pre.*) torque for the 6th run-in 1STEP 3. The average error on the torque prediction was 0.93% and, on the energy (calculated using equation 2.1) was 0.68%. Despite the small training sample size, the model showed a very good capability in the prediction of the torque in the absence of material removal (in fact, during run-in 6th there is no interference between gear teeth and pump body). In the second phase, the pressure profile and the predicted torque by the Model0 were used as inputs to a second model for predicting the torque needed during the run-in with material removal. Thus, the Model0 output was used as input for the improved model, called Model1. The inputs for Model1 were the pressure profile, as per Model0, and the torque predicted by Model0. The results of Model1 are reported in Figure 4. 1 and compared to Model0 and the experimental curve. Even if the *pre.* curve (red dashed line) associated to run-in 1 does not capture the spikes and fluctuations of the experimental torque curve, the results of Model1 can predict the average value of the torque during the material removal of the run-in cycle with an average error of 5.87% and an error less than 0.11% on the required energy (lower than Model0 as expected). Model1, with respect to the Model0, was trained on a sample of fifteen acquisitions. Starting from the duty cycle (pressure profile), the combinate use of Model0 and Model1 (Figure 4. 2.a) allows for predicting with a small error the run-in torque, energy and hydromechanical efficiency. In the present work, the AI model represented by Figure 4. 2.a was tested for a unique pump displacement and for load cycles belonging to the families defined in Table 2. 1.

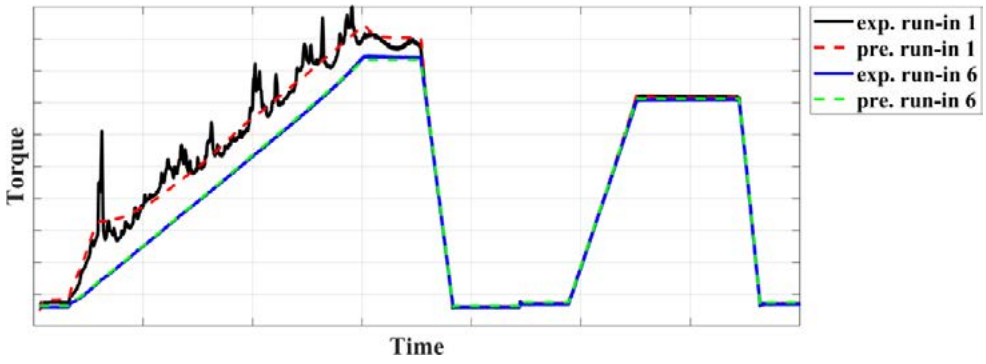


Figure 4. 1 – Experimental (exp.) and Predict (pre.) torque for the 1STEP 3 cycle

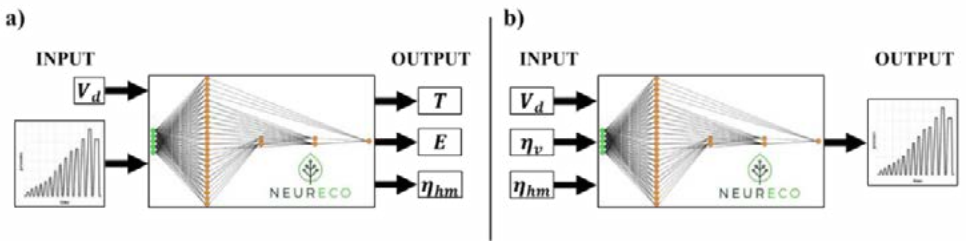


Figure 4. 2 – a) Actual predictive AI model; b) Desired predictive AI model

The present AI model represents a first step toward optimizing run-in cycles (as shown in the flowchart in Figure 4. 2.b); consequently, in future works, it will be tested on various pump displacements and duty cycles to better understand its behaviour far away the training

conditions. As said, at the present, the model is not able to predict the gears seats profile, torque spikes and fluctuation (but can be considered negligible, having the purpose of optimizing the run-in cycle) and the volumetric efficiency. The last can be included by recording the flow rate during run-in cycle.

5. CONCLUSION

The paper presents a deep analysis focused on the run-in cycle of EGPs made thought an extended experimental investigation and the definition of a preliminary AI-based model. Thirty pumps were tested with six different types of run-in cycles, showing interesting results in terms of pump performance improvement and cost reduction. For all the pumps, the same maximum load (p_S^m) was set and the same shape of the internal body profile, with slight differences in the height of milling and in the total volume removed, was obtained. One interesting result is that the final profile shape of the driver gears seats results deeper and aligned with the driven gear when the run-in speed increases, keeping constant the maximum pressure, impacting negatively on the volumetric efficiency. Moreover, adopting new cycles with the same maximum pressure but reduced cycle time the final profile shape of gears seats changes slightly (Table 3. 3) and a minor height of milling is achieved with significantly impacts on the pump efficiency (Figure 3. 4). For the shorter cycles, it was experimentally verified that, the complete material removal requires a time at the maximum load greater than the reference cycle STA N1; spikes of torque appeared on the second and third run-in cycles. Thus, for future activities, a variable t_S^3 will be considered to find the minimum time necessary to have the finished internal profile. The adoption of a shorter cycles ensured a pump cost reduction of 1.74% and an increase in the testing capability until to 78% for the 1STEP 1 cycle. At the same time, the shorter cycles, in particular the 1STEP 3 and the STAIR, showed an increment of pumps performances thanks to a higher volumetric efficiency. The latter benefit was predominant at the low speeds and at the medium-low pressures; working conditions in which the percentage weight of the tip leakages is more relevant. In the STAIR, 1STEP 1 and 1STEP 3 cycles, the run-in phase did not include time intervals for depressurisation, in which the body surface milled by the gears could cool down and the material could be easily removed, and this could have affected the surface quality. In future works, a detailed analysis of the surface asperities in the wear zone will be provided. Therefore, optimising the run-in cycle can ensure economic benefits and improved pump performance. A preliminary AI model for the prediction of the run-in process variables (torque) was developed using the commercial tool NeurEco®. Despite the small size of the training sample, the model was able to predict the absorbed torque, with an error below 6% for the run-in 1 and below 1% for the run-in 6. In future activities, the prediction of other physical quantities, such as the flowrate (therefore volumetric efficiency) and the material removed.

6. APPENDIX

6.1. Definition of $\delta_r^{av.}$, $\epsilon_{max}^{av.}$ and int

The three quantities $\delta_r^{av.}$, $\epsilon_{max}^{av.}$ and int are defined as:

$\delta_r = r^{body} - r^{gear}$; $\delta_r^{av.} = \left(\delta_r^{driving} + \delta_r^{driven}\right)/2$ where r identifies a radius and δ a radial clearance.

$\epsilon_{max} = \left(r^{body} - r_{ex}^{thrust\ plate}\right) + (r_{in}^{thrust\ plate} - r^{shaft})$ where ex and in identify the external and internal radius. Considering the two pump sides, flange and rear cover, and the two gears, driven and driving is obtained $\epsilon_{max}^{av.} = \left(\epsilon_{max}^{driving-flange} + \epsilon_{max}^{driving-cover} + \epsilon_{max}^{driven-flange} + \epsilon_{max}^{driven-cover}\right)/4$. Combining $\delta_r^{av.}$ and $\epsilon_{max}^{av.}$ the quantities int is defined as $int = \epsilon_{max}^{av.} - \delta_r^{av.}$. This last identifies the radius interference between the teeth of the gears and the pump body under the condition of max radial shift of the shaft ϵ_{max}.

6.2. Effective Energy Variation

The passage from equation (3. 1) to equation (3. 2) is possible under various simplifying assumptions.

Assuming that all the pumps with the same RB have the same hydromechanical efficiency, $\eta_{hm,j}^{STA\ N1} = \eta_{hm,j}^{STA\ N2} = \eta_{hm,j}^{STAIR} = \cdots = \eta_{hm,j}^{1STEP\ 3} = \eta_{hm,j}$ for $j = 1,2, \dots 5$, the equation. (3. 1) can be rewritten as:

$$\delta E_{i,j}^{eff} = \frac{\left(\frac{1}{\eta_{hm,j}}\right)\left[\left(E_{i,j}^{th} + \eta_{hm,j} * E_{i,j}^{eff-mill}\right) - \left(E_{STA\ N1,j}^{th} + \eta_{hm,j} * E_{STA\ N1,j}^{eff-mill}\right)\right]}{\left(\frac{1}{\eta_{hm,j}}\right)\left(E_{STA\ N1,j}^{th} + \eta_{hm,j} * E_{STA\ N1,j}^{eff-mill}\right)} * 100$$

Under the assumption that the energy required for the material remotion is the same for all the pumps tested (thirty), $E_{i,j}^{eff-mill} = E^{eff-mill} > 0$ for $j = 1,2, \dots 5$ and $i = \{STA\ N1, STA\ N2, STAIR, New\ STA, 1STEP\ 1, 1STEP\ 3\}$, the above equation becomes:

$$\delta E_{i,j}^{eff} = \frac{E_{i,j}^{th} - E_{STA\ N1,j}^{th}}{\left(E_{STA\ N1,j}^{th} + \eta_{hm,j} * E_{STA\ N1,j}^{eff-mill}\right)} * 100 = \delta E_i^{th} * \frac{E_{STA\ N1,j}^{th}}{\left(E_{STA\ N1,j}^{th} + \eta_{hm,j} * E^{eff-mill}\right)}$$

Defined A as:

$$A = \frac{E_{STA\ N1,j}^{th}}{\left(E_{STA\ N1,j}^{th} + \eta_{hm,j} * E^{eff-mill}\right)} < 1$$

The equation. (3. 2) is obtained. It is important to highlight that the assumption of a unique value of $E^{eff-mill}$ considers that for all the pumps there is the same amount of material removed. This hypothesis can be valid for two reasons: 1) All the run-in cycles have the same maximum pressure (defines the gears' inflection and shift towards the inlet); 2) The thirty pumps have been assembled, maintaining control of the variability of the radial clearances.

If $E_{i,j}^{eff-mill}$ is not constant the parameter B can be defined as:

$$B = \frac{E_{i,j}^{eff-mill} - E_{STA\ N1,j}^{eff-mill}}{\left(\frac{1}{\eta_{hm,j}}\right)\left(E_{STA\ N1,j}^{th} + \eta_{hm,j} * E_{STA\ N1,j}^{eff-mill}\right)} * 100$$

Using the parameters A and B the equation (3. 1) becomes the equation (3. 3). The sign of B depends by the relationship between $E_{i,j}^{eff-mill}$ and $E_{STA\ N1,j}^{eff-mill}$.

Acknowledgment and Data availability

The authors would like to thank ADAGOS, BSim Engineering, and Hydreco Hydraulics for their technical support. For confidential reasons, the data are not available.

REFERENCES

[1] Dalpiaz, G.; Fernández, A.; Mucchi, E. Modeling run in process in external gear pumps. In Proceedings of the ASME 8th Biennial Conference on Engineering Systems Design and Analysis, Torino, Italy, 4–7 July 2006; pp. 927–936. DOI:10.1125/ESDA2006-95466

[2] M.Borghi, M. Milani, F. Paltrinieri, M. Guidetti. Influenza Del Rodaggio Sulle Condizioni Di Funzionamento Di Macchine Volumetriche Ad Ingranaggi Esterni. 56° Congresso Nazionale ATI.

[3] Osiński, P.; Bury, P.; Cieślicki, R. Optimizing the Break-in Process of High-Pressure Gear Pumps. In Advances in Hydraulic and Pneumatic Drives and Control 2020 (NSHP 2020), Poland; Stryczek, J., Warzynska, U., Eds.; Lecture Notes in Mechanical Engineering; Springer: Cham, Switzerland, 2020. DOI: 10.1007/978-3-030-59509-8_6.

[4] S.Manco, N. Nervegna, Simulation of an external gear pump and experimental verification, Proceedings of the JFPS International Symposium on Fluid Power, 1989(1), 147–160. doi:10.5739/isfp.1989.147.

[5] B. Peric; M. Engler; M. Schuler; P. Woias; Using Neural Networks as a Data-Driven Model to Predict the Behavior of External Gear Pumps; Processes 2024, 12(3), 526; https://doi.org/10.3390/pr12030526

[6] K. Lakshmanan, A. J Gil, F. Auricchio, F. Tessicini; A fault diagnosis methodology for an external gear pump with the use of Machine Learning classification algorithms: Support Vector Machine and Multilayer Perceptron. DOI: 10.17028/rd.lboro.12097668.v1.

[7] Neureco – AI and Digital Twin data-based for engineers | Bsim Engineering

[8] Bashtova, K., Causse, M., James, C., Masmoudi, F., Masmoudi, M., Turki, H., & Wolff, J. "Application of the topological gradient to parsimonious neural networks." Computational Sciences and Artificial Intelligence in Industry: New Digital Technologies for Solving Future Societal and Economical Challenges. Cham: Springer International Publishing, 2021. 47-61.

Biographies



Alfonso Rosario Apuzzo received the bachelor's degree in aerospace engineering from University of Naples Federico II in 2021, the master's degree in mechanical engineering from the University of Naples Federico II in 2024. In November 2024 started the PhD program born from the collaboration of the University of Sannio and the Duplomatic MS company. His research includes the study of the external gear pumps via numerical and experimental approaches.



Sara Bonati received her MSc degree in mechanical engineering from the University of Parma in 2016. At the same university, she carried out a research grant in 2016 concerning the development of lumped-parameter fluid dynamics models of positive displacement compressors. She has been working since 2017 in the field of Fluid Power, focusing on research and product development activities through CAE tools. She joined the Duplomatic Group in 2023 as R&D drive system engineer.



Dr. Emma Frosina is an Associate Professor in the Department of Engineering at the University of Sannio, Benevento. She received her M.S. degree in Mechanical Engineering from the University of Naples Federico II in 2012 and her Ph.D. degree in Mechanical Systems Engineering in 2016. She is the author of over 80 scientific papers published in peer-reviewed international journals and conference proceedings. Her research interests are in the field of fluid power, where she specializes in the modeling and optimization of components (machines, valves, etc.) using numerical approaches (both lumped parameter and three-dimensional CFD) and experimental testing.



Dr. Luca Riccò received his MSc degree in mechanical engineering in 2012 and his PhD in industrial engineering in 2015 from the University of Parma. After, he started his career in the field of fluid power with different R&D roles. Starting from July 2023 he leads the R&D drive system and global simulation team of the Duplomatic Group, a member of Daikin Group.

Influence of surface structuring on the hydrostatic and hydrodynamic pressure distribution of slipper sealing lands

Svenja Horn, Jürgen Weber

Institute of Mechatronic Engineering, TU Dresden, svenja.horn@tu-dresden.de

Abstract.

Modern operating conditions and upcoming lead-free material regulations challenge the tribological performance of axial piston pumps, particularly at low speeds where hydrodynamic support is weak. While meso-structures on slipper surfaces have shown practical success, a fundamental understanding of their working principles is lacking. This paper presents a systematic analysis of how specific geometric parameters of step and parabolic structures influence the hydrostatic and hydrodynamic performance of the slipper-swashplate interface. Simplified analytical models based on the Reynolds equation are used to evaluate the impact on hydrostatic balance and hydrodynamic tilting moments. The findings are qualitatively compared with results from a comprehensive Caspar FSTI simulation. Results show that a smaller gap height at the inner radius reduces the hydrostatic balance but generates a stabilizing tilting moment in the direction of movement. A reduced gap at the outer side of the sealing land by contrast, increases the hydrostatic balance but at the cost of a destabilizing negative tilting moment. The analytical trends are confirmed by the dynamic simulations, providing a foundational understanding for designing structured surfaces to improve pump reliability and efficiency.

Keywords. Surface structures, slipper geometry, axial piston pump, tribology

1. INTRODUCTION

New operating ranges motivated by electrification, start-stop applications or speed-variable drives challenge traditional hydraulic machines with respect to mixed friction conditions, low swash plate angles and partial load operation. At these critical conditions, the pumps are exposed to reduced hydrodynamic pressure in the lubricating gaps, leading to increased friction losses, wear, and a higher likelihood of failures. Traditionally, to ensure good run-in behavior and satisfactory dry-running performance under mixed friction, a combination of leaded bronze or brass and steel materials has been used. However, future EU regulations will prohibit the use of lead-containing materials. In light of these upcoming restrictions, it

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

becomes imperative to find new solutions to withstand the harsh operating conditions without the use of lead-containing metals.

The main difficulty centers the lubrication gaps of the pump's rotating group. Illustration 3 and 4 in Figure 1.1 explain the principal correlation between surface geometry and power losses for the slipper / swash plate interface. There are two forces, that mainly determine the gap height between slipper and swashplate. The piston force F_{SK} from the pressure in the displacement chamber and the opposed lift force F_{fz} due to the pressure distribution of the fluid film lubrication in the gap. A trade-off arises due to the conflicting requirements on the gap height to allow good bearing properties while reducing leakage at the same time. Figure 1.1-4 shows the optimization problem between leakage and friction depending on the gap height. The larger the gap, the higher the leakage, with a simultaneous decrease in friction losses. As the curves have different optima depending on the operating point and design, there is an optimum slipper design for each operating point.

This optimization challenge is addressed by the design of meso-structures. Meso-structures are surface designs that exhibit small changes in height (in the µm range) over larger distances (in the mm range). The principal mechanism of surface structuring on power losses is illustrated in Figure 1.1, using results of chapter 4. Meso-structures allow to change the pressure field in the gap in a targeted way and thus to build up additional hydrodynamics, depicted exemplary for a step structure in Figure 1.1-1 and Figure 1.1-2. This again changes the balance of forces and moments (Figure 1.1-3), which ultimately determines the gap height and slipper's micro motion and associated power losses (Figure 1.1-4).

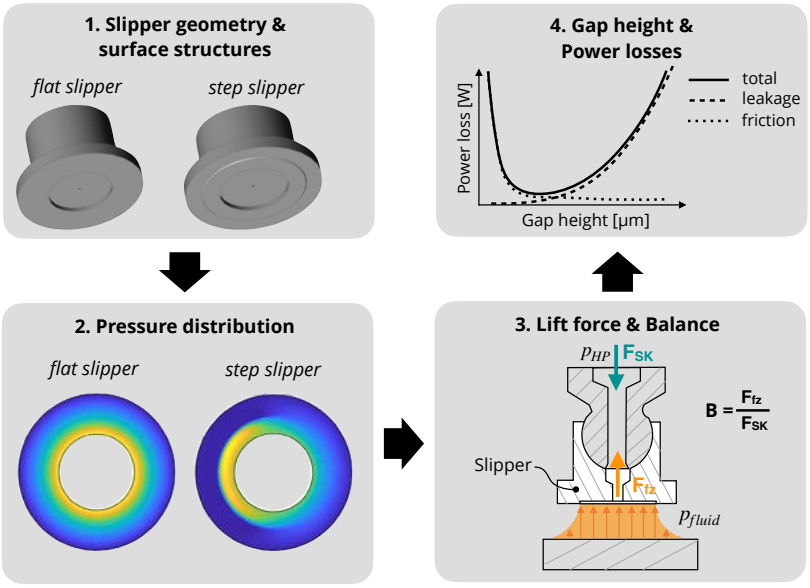


Figure 1.1. Principal mechanism of surface structuring on power losses

Our earlier publications have demonstrated the practical success of these structured surfaces. Several structures on the slipper sealing land have been investigated and optimized using the pump simulation software Caspar FSTI in [1]. The two most promising designs from

simulation were manufactured and validated on a dedicated pump test-rig, showing almost no wear [2] and an increase in efficiency for low speeds, low pressures as well as partial load conditions [3]. They were further investigated on the tribometer in [4] and [5] being exposed to extreme thermal conditions and different pressure pulsations.

While these results confirm the significant potential of meso-structures, a fundamental understanding of how specific geometric parameters influence the underlying hydrostatic and hydrodynamic mechanisms is still lacking. The objective of this work is to draw a systematic analysis to bridge this knowledge gap, which will not only enable more efficient design of new slipper geometries but also describes the effect of wear-induced surface changes. The findings are expected to be transferrable to other bearings within axial piston pumps and also to a broader class of plain bearings.

Chapter 3 establishes the theoretical foundation for pressure build-up in lubricating gaps, based on the Reynolds equation, and introduces the two selected surface structures for analysis. In chapter 4, the influence of key geometric parameters on both the hydrostatic balance and the hydrodynamic tilting moments are investigated in detail for each design. The findings for the first structure are also compared with results from a comprehensive simulation to contextualize the analytical approach in chapter 5. The paper concludes with a summary of the key findings and their implications for the design and optimization of tribological interfaces in axial piston pumps.

2. LITERATURE

The performance of the slipper-swashplate interface has been a subject of extensive research, with a particular focus on the influence of the slipper's surface geometry on the lubricating film. Early analytical investigations by Hooke established that a convex slipper profile is crucial for ensuring stable fluid film lubrication [6]. This finding was later reinforced by Borghi et al., who used simulations to show that an initially flat slipper is inherently unstable and will eventually wear into a convex shape to achieve stable operation [7]. However, the ideal profile is not straightforward. Manring, for instance, found that while a linear concave profile could increase load-carrying capacity, it came at the cost of higher leakage and power loss. His work further revealed that profile deformations have a greater impact on flow rate than on load-carrying capacity, highlighting the complex trade-offs inherent in slipper design [8].

More recent studies employing Elastohydrodynamic (EHD) models have provided deeper insights into the role of deformation. Hashemi et al. developed a "tribo-slipper-pump" test rig and showed through both measurement and simulation that a slightly concave slipper can deform elastically under high pressure to form a convex shape, thereby improving its hydrodynamic characteristics [9]. This underscores the importance of considering material deformation in any realistic analysis. Beyond general surface profiles, specific geometric structures have been proposed to actively manipulate the pressure field. A notable example of using a micro-step is found in the patent US 5,983,781 by Ivantysyn, from 1999 [10]. It is relevant to note that this patent has since expired, placing the described invention in the public domain. The patent describes a shallow recess at the inner diameter of the sealing land, which is designed to create a self-adjusting load-bearing capacity that changes with the gap height [10]. While the patent focuses on this specific self-adjusting mechanism, the present work undertakes a broader, systematic analysis of such step structures. It investigates

the fundamental effects of varying key parameters, including step height and position, on the hydrostatic and hydrodynamic performance. This allows for a more comprehensive understanding of the design trade-offs, building on the concepts introduced in earlier work. Another promising approach involves sinusoidal or wave-like structures. Baker and Ivantysyn investigated sinusoidal meso-structures on the sealing land of a valve plate, using simulation to optimize the wave's frequency, amplitude, and phase [11], [12]. The most promising design was manufactured and tested, demonstrating an increase in the total pump efficiency of up to 10 % at low displacement and pressure [11]. While this technology was patented in the United States in 2009, it was not extended elsewhere, making the application and further development freely available in other markets, including Europe [13]. The success of this wave profile on the valve plate provides a strong motivation for investigating a similar radial wave structure on the slipper surface, as is done in the previous studies.

Other approaches include surface texturing. Ye et al. used an EHD model to identify an optimal dimple area density and depth-to-diameter ratio to maximize load-carrying capacity [14]. Similarly, Wang et al. simulated micro-textures with varying dimple shapes and depths, also finding improvements in load-carrying capabilities and friction losses, though these results were not experimentally validated [15].

A significant contribution to surface optimization on the slipper sealing land was made by Darbani et al., who proposed an optimized convex shape derived from typical wear profiles [16]. Their work aimed to minimize total energy loss across various operating conditions, reporting a theoretical efficiency improvement of around 51%, mainly decreasing friction losses at high speeds [16]. However, the study has limitations: the results were not experimentally validated, and crucial geometric data such as the slipper balance factor were whether provided nor changed, making the findings difficult to transfer or compare directly. Despite these gaps, the promising results from this optimization approach motivate the inclusion of a similar convex profile in the present analysis. More recently, Mo et al. also analyzed the influence of different convex wear profiles on lubrication performance, though the analysis was limited to a single operating point and did not include an optimization perspective [17].

A key challenge in analyzing these complex interactions is to distinguish the contributions of hydrostatic and hydrodynamic effects. A recent methodological advancement by Ernst and Vacca demonstrated a technique for separating the hydrostatic and hydrodynamic force components from the simulated pressure distribution [18]. Their findings revealed that while the load is primarily carried hydrostatically, the hydrodynamic component can account for over 20 % of the total lift under certain conditions, playing a critical role in balancing the system [18]. This approach of separately analyzing hydrostatic and hydrodynamic effects is adopted in the present work to provide a fundamental understanding of how specific geometric parameters influence each mechanism.

3. THEORETICAL BACKGROUND AND METHODOLOGY

3.1. Pressure build-up in small lubricating gaps

The pressure distribution within the lubrication gaps of axial piston pumps determines the forces and moments acting on the components and can be described by the Reynolds equation, first derived from the Navier Stokes momentum and continuity equations by

Osborne Reynolds in 1886 [19]. Considering a very small fluid gap height in the range of micrometer together with a large gap length in the range of millimeter, as it is the case for the lubricating gaps in axial piston pumps, several assumptions can be made, that simplify the complex Navier Stokes equation. These include assuming constant pressure over the fluid film thickness $\partial p/dz = 0$, neglecting inertial and body forces as they are very small compared to the viscous and pressure forces, and considering the lubricant to be a Newtonian fluid, where shear stress is proportional to the shear rate. For the analytical approach in this work, the fluid is further assumed to be incompressible and of constant viscosity. Under these assumptions, the generalized Reynolds equation is given by [18]

$$\underbrace{\frac{\partial}{\partial x}\left(\frac{\rho h^3}{12\mu}\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{\rho h^3}{12\mu}\frac{\partial p}{\partial y}\right)}_{\text{Poiseuille}}=\underbrace{\frac{\partial}{\partial x}\left(\frac{\rho h(u_\mathrm{a}+u_\mathrm{b})}{2}\right)+\frac{\partial}{\partial y}\left(\frac{\rho h(v_\mathrm{a}+v_\mathrm{b})}{2}\right)}_{\text{Couette}}$$
$$+\underbrace{\rho\left(w_\mathrm{a}-w_\mathrm{b}-u_\mathrm{a}\frac{\partial h}{\partial x}-v_\mathrm{a}\frac{\partial h}{\partial y}\right)}_{\text{Squeeze}}+\underbrace{h\frac{\partial \rho}{\partial t}}_{\text{Expansion}}. \tag{3.1}$$

The equation consists of several terms that represent different physical effects. The *Poiseuille* term describes the flow rates due to pressure gradients and is used to calculate the hydrostatic pressure field in the gap. The *Couette* term describes the hydrodynamic pressure build up due to the relative sliding motion of the surfaces. This hydrodynamic effect is primarily generated by a "physical wedge," where a converging geometry in the direction of motion creates positive pressure [20]. The squeeze term accounts for the cushioning effect when surfaces move towards or away from each other, generating pressure in response to a changing film thickness. The local expansion term describes the impact of mechanical or thermal deformations on the fluid film thickness. This paper focuses on the hydrostatic (Poiseuille) and hydrodynamic (Couette) effects to understand the fundamental influence of surface geometry.

A key metric for evaluating the hydrostatic design is the balance factor, calculated by dividing the hydrostatic lift force by the opposing piston force. A balance of 100 % indicates that the forces are in equilibrium. In industrial practice, slippers are often designed with a balance around 90 % to allow for additional hydrodynamic lift at higher speeds. However, this under-balance can lead to high friction and wear during start-up and at low speeds when hydrodynamic effects are weak. Conversely, a balance factor near or above 100 % can ensure separation but may lead to increased leakage losses.

The primary mechanism for generating hydrodynamic pressure is the wedge effect, which can be created by a tilted surface or a geometric feature like a step, as illustrated in Figure 3.1. The magnitude of hydrodynamic pressure generated is dependent on the ratio between the minimum gap height and the height of the geometric feature (i.e., step or inclination height). For the specific case of a parallel-step bearing, an optimal configuration that maximizes hydrodynamic lift was first derived by Lord Rayleigh [21]. As presented in detail in [20], this optimum is defined by the following dimensionless parameters for the step height ratio, H_0, and the step position ratio, n_s:

$$H_0 = \frac{h_0}{s_h} = 1.155,\ n_s = 0.7182. \tag{3.2}$$

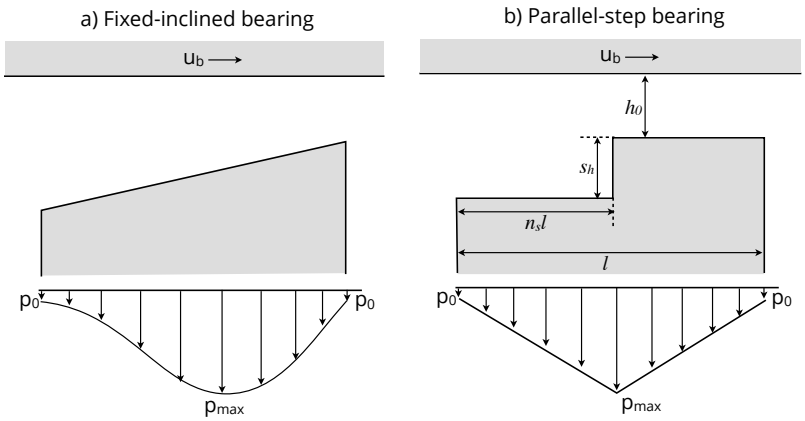


Figure 3.1.Comparison of pressure profiles of fixed inclined bearing (a) and parallel-step bearing (b)

3.2. Selected surface structures

The left plot in Figure 3.2 illustrates the slipper geometry, with the most important parameters the pocket diameter d_{inG}, the sealing land outer diameter d_{outG} and the gap height h. To further investigate the influence of surface structures on the hydrostatic and hydrodynamic pressure build-up, two representative surface designs (see Figure 3.2) from literature were selected for detailed analysis:

- **Design a):** A step structure several micrometers in height, resulting in a reduced gap at the inner land of the slipper surface. This concept is derived from classical step slider bearing designs and, compared to the other structures, offers significant advantages in terms of manufacturing simplicity.
- **Design b):** A parabolic surface profile, which mimics the typical run-in wear pattern of the slipper surface. This profile arises due to thermal and pressure-induced deformation, leading to abrasion at the inner radius, and tilting-induced wear at the outer radius. Darbani optimized this design using the CASPAR simulation environment [16]. However, no manufacturing or experimental validation has been reported to date.

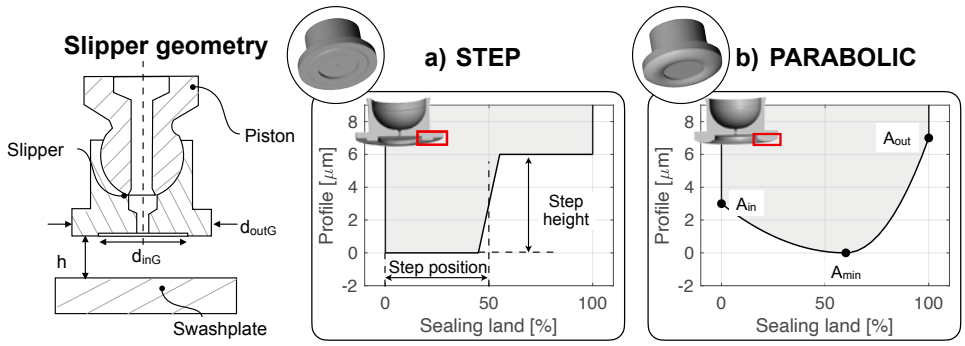


Figure 3.2. Slipper geometry and design spaces of two different surface structures

These surface structures were chosen to represent a spectrum of complexity and manufacturability, as well as to capture different mechanisms of pressure generation and

wear compensation. In the following, their impact on the pressure build-up in the lubricating gap will be systematically analyzed and compared.

3.3. Methodology

In the following, the influence of the design parameters on the hydrostatic and hydrodynamic lift force is investigated. In order to take into account the later dimensions of the test pump at this point and at the same time remaining company-neutral, the outer diameter d_{outG} of the slipper is set to 29.6 mm, which is an average value of industrial slipper outer diameters with a 21 mm piston. To directly see the influence of the surface structure, the inner radius is calculated such that a balance of 100 % is achieved. A gap height of h=20 µm during high pressure stroke is set to calculate the hydrostatic and hydrodynamic pressure build-up. To quantify the hydrostatic forces and hydrodynamic moments based on the principles of the Reynolds equation, simplified analytical methods are employed. Bergada detailly derives the calculation of the lift force of a multi-land slipper design in [22]. His approach is transferable to micro-structured sealing lands, as it discretizes the radial gap height *h(r)*, using the following formula to calculate the lift force F_{fz}:

$$F_{fz} = p_{\text{in}} \cdot \pi \cdot \left[\frac{\sum_{i=1}^{i=n} \frac{1}{h_i^3} \cdot \frac{r_{i+1}^2 - r_i^2}{2}}{\sum_{i=1}^{i=n} \frac{1}{h_i^3} \cdot \ln\left(\frac{r_{k+1}}{r_k}\right)} - r_1^2 \right] \tag{3.3}$$

This calculation accounts for the modified pressure profile caused by the structure. It is important to distinguish this from a conventional balance calculation based solely on the fixed pocket and sealing land diameters, which would not capture the influence of the meso-structure.

For the analysis of hydrodynamic effects, a simplified 1D analytical model, as implemented by Knoll is used [23]. This approach solves a one-dimensional form of the Reynolds equation (Equation (3.4), assuming the slipper sealing land is an infinitely wide bearing. This reduces the problem to the direction of motion (x_G) and allows for rapid evaluation of the pressure profile *p(x)* generated by the Couette term.

$$\frac{\partial}{\partial x_G}\left(h^3 \frac{\partial p}{\partial x_G}\right) = 6u\mu \frac{\partial h}{\partial x_G} \tag{3.4}$$

The resulting 1D pressure profile p(x) is then integrated across the sealing land width to calculate the hydrodynamic tilting moment M_y around the y_G-axis, providing a direct measure of the structure's tendency to generate a stabilizing or destabilizing pitch:

$$M_y = \int p(x) \cdot x \, dx. \tag{3.5}$$

4. RESULTS

This chapter presents the influence of key geometric parameters on the hydrostatic balance and the hydrodynamic tilting moments first for the step and second for the parabolic design. The findings for the step structure are further compared with results from a comprehensive simulation to contextualize the analytical approach.

4.1. a) Step structure

4.1.1. Influence on hydrostatic pressure and balance

The initial investigation focuses on the examination of the impact of the step design parameters, specifically the step height and position, on the hydrostatic. Figure 4.1 (left) shows, as a dashed black line, the typical logarithmic pressure drop across the sealing land for a flat, unstructured slipper with a balance of 100 % (pocket diameter of 14.4 mm) at a pocket pressure of 100 bar and a gap height of 20 µm (design *a3*). The colored lines represent three example pressure distributions for a step located in the middle of the sealing land, with varying step heights. Designs *a1* and *a2* feature a step with a reduced gap height at the inner radius, elevated by 6 µm and 3 µm, respectively. In contrast, design *a4* has a 3 µm-high step at the outer radius, reducing the gap height there. A smaller gap at the inner radius causes a more rapid pressure drop in the inner region, resulting in a lower lift force and thus a reduced balance. Conversely, a reduced gap at the outer radius (*a4*) leads to a slower pressure drop in that area, increasing the hydrostatic force and balance. The right-hand plot in Figure 4.1 shows the resulting balance as a function of step height and step position, with the four designs *a1–a4* indicated.

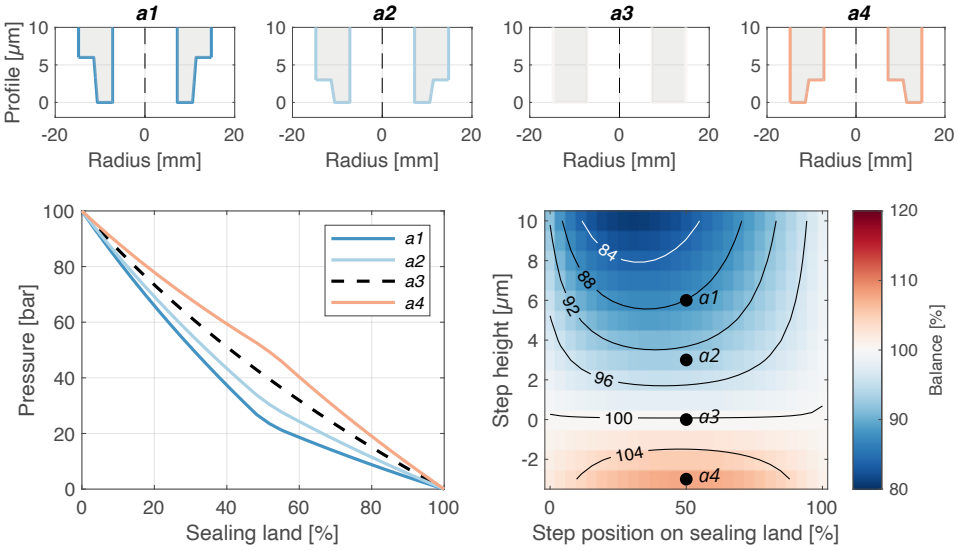


Figure 4.1. Sealing land Pressure drop over sealing land for different surface structured designs (*a1, a2, a4*) in comparison to an unstructured slipper *a3* (left) and balance for different step heights and step positions (right)

Figure 4.2 left shows that the gap height for an unstructured flat slipper (design *a3*) has no impact on the balance, while it strongly changes the balance for the structured slippers (*a1, a2* and *a4*) especially for smaller gap heights below 10 µm. It is particularly noticeable that the designs *a1* and *a2* with a smaller gap at the inner ring have less lift for smaller gap heights, dropping below 80 % for a 3 µm gap height. Design *a4* on the other side exhibits a balance of over 140 % for a 3 µm gap height providing extra lift in case of critical mixed friction conditions. All designs have a pocket radius of 7.2 mm.

Figure 4.2 right illustrates the balance for different pocket diameters ranging from 7-8 mm and gap heights from 3-30 µm for the structured design *a1*. While the pocket diameter has a linear impact on the balance value, the impact of the gap height on the balance value is significantly higher at lower gap heights.

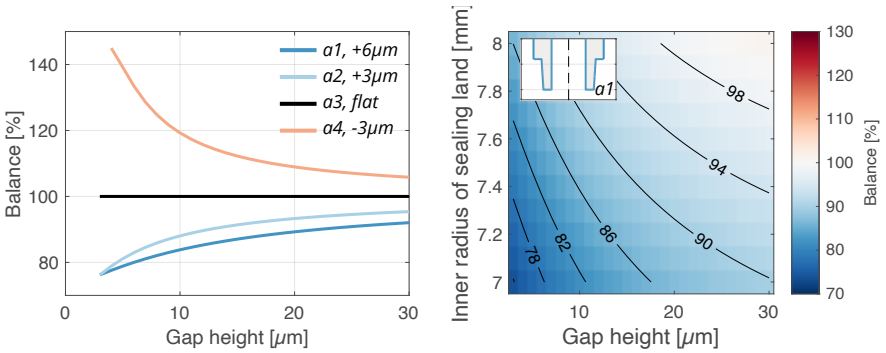


Figure 4.2. Influence of gap height on balance for different designs *a1*-*a4* (left) and balance for different pocket diameter and gap height combinations for design *a1* (right)

4.1.2. Influence on hydrodynamic lift and tilting moment

Figure 4.3 (top) shows the radial cross-sectional profiles of the three designs *a1, a2*, and *a4*. The lower part of the figure presents a qualitative representation of the resulting hydrodynamic pressure distribution, calculated at a rotational speed of 500 rpm and an external pressure of 0 bar without any slipper tilt. The direction of movement *u* is defined to the right. The triangular hydrodynamic pressure build-up caused by the step corresponds to that of a parallel-step bearing, as shown in Figure 3.1.

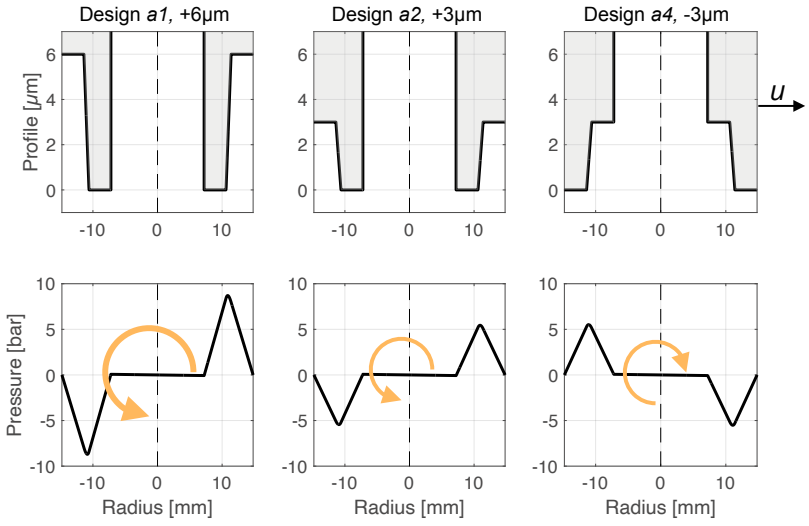


Figure 4.3. Radial cross-sectional profiles of designs *a1*, *a2*, and *a4* (top) and corresponding qualitative hydrodynamic pressure distribution at 500 rpm (bottom)

A diverging gap in the direction of motion leads to a positive pressure build-up, whereas a converging gap results in a negative pressure. While negative pressures cannot occur in reality and are merely the result of a mathematical solution under the given boundary conditions, the simplified example nonetheless illustrates the effect induced by the step. In a physical system, a liquid cannot sustain significant tension, and the pressure will not drop below the fluid's cavitation pressure. A smaller gap at the inner radius (as for *a1* and *a2*) causes an uneven hydrodynamic pressure distribution due to the wedge effect in the flow direction, resulting into a positive tilting moment, causing the slipper to pitch upward and align itself in a way that promotes further hydrodynamic pressure build-up. In contrast, a step in the opposite direction as for design *a4* induces a negative tilting moment, which may cause the slipper to "sink" or lose its load-carrying capacity.

Figure 4.4 illustrates the tilting moments generated by the step structure and calculated with a 20 µm gap height and a dynamic viscosity of $\eta = 0.0705$ Pas. Positioning the step near the inner radius results in a positive tilting moment, which acts to pitch the slipper upward and enhance hydrodynamic support. In contrast, a step near the outer radius creates a negative, potentially destabilizing moment. This fundamental behavior aligns with the principles first established by Lord Rayleigh for ideal parallel-step bearings, which predict an optimal configuration for maximizing lift (refer to formula (3.2)) [20].

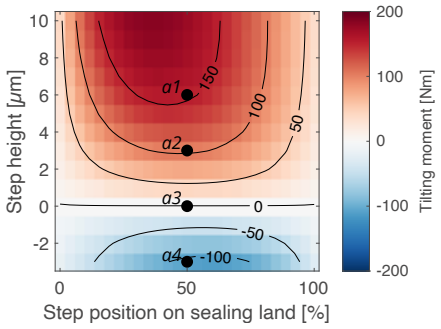


Figure 4.4. Tilting moment generated by hydrodynamic pressure for various step heights and step positions on the sealing land

The results in Figure 4.4 show that for the slipper geometry in this study, the maximum positive tilting moment is found for a step position of approximately 40 % of the sealing land width. This is in the same region as, though slightly outward from, the theoretical optimum of ~30 % predicted by Rayleigh's theory. Similarly, the optimal step height of around 10 µm (for a 20 µm gap) is in the same order of magnitude as the classical prediction, confirming the validity of the general principle.

4.2. b) Parabolic structure

4.2.1. Influence on hydrostatic pressure and balance

The parabolic design is showing a typical wear profile of the slipper sealing land. Starting with the influence of the parabolic design parameters A_{in}, A_{out} and A_{min} on the hydrostatic behavior, Figure 4.5 (left) shows the pressure profiles across the sealing land. The reference design *b3*, without any structural modification, is shown as a dashed line. The colored curves

illustrate four modified designs: *b1* and *b2* feature a parabolic recess at the outer radius with depths of 10 µm and 5 µm respectively, that can for example be caused by slipper tilting. *b4* and *b5* have a recess at the inner radius with the same depths and can be produced by convex deformation of the slipper surface. As already observed in the step design, material removal at the outer radius (*b1, b2*) causes a reduction in pressure compared to the unprofiled reference, whereas a recess at the inner radius (*b4, b5*) leads to an increase in pressure. This behavior corresponds to a decrease or increase in the hydrostatic balance of the slipper, as shown in the right plot of Figure 4.5. Here, the design space covers also quadratic designs with a recess at both, the outer and inner radius, with a curve minimum at A_{min}=50 %.

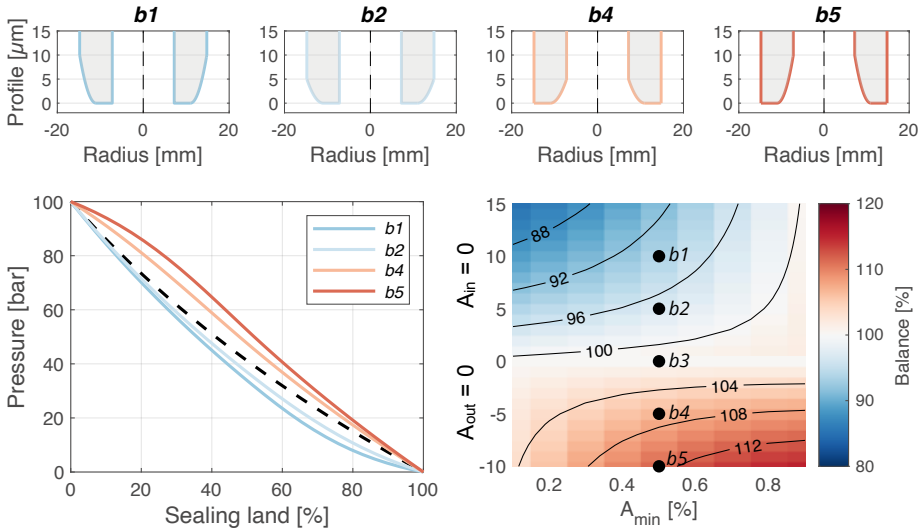


Figure 4.5. Hydrostatic pressure drop over the sealing land for different parabolic designs *b1-b5* (left) and hydrostatic balance map for different combinations of the parameters A_{in} and A_{out}, with different curve minimums A_{min} (right)

Further balance maps with different curve minimums at 20, 50 and 80 % are presented in Figure 4.6. A shift of the curve minimum to the inner radius reduces the balance while a shift to the outer radius increases the balance, aligning with the trends of the step design.

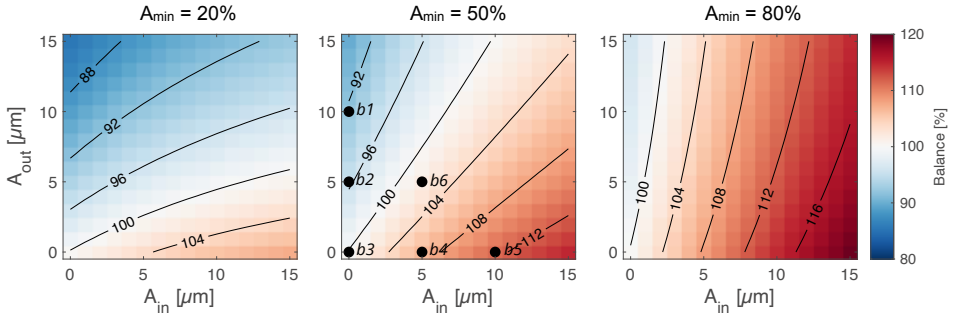


Figure 4.6. Influence of quadratic curve minimum A_{min} on balance for different combinations of A_{in} and A_{out}

4.2.2. Influence on hydrodynamic lift and tilting moment

In Figure 4.7 on the top, the radial cross-sectional profiles of the three designs *b2*, *b6* and *b4* are drawn in more detail. Below, the corresponding qualitative hydrodynamic pressure distributions are calculated at 20 µm gap height, 500 rpm and no slipper tilt. Design *b2* and *b4* produce the same hydrodynamic pressure however mirrored such that wear at the outer radius results in a positive tilting moment, while wear at the inner radius produces a negative tilting moment. A recess at the inner and outer radius (*b6*) varies the hydrodynamic pressure according to the positive and negatives edges of the profile. In total, a slightly positive tilting moment result.

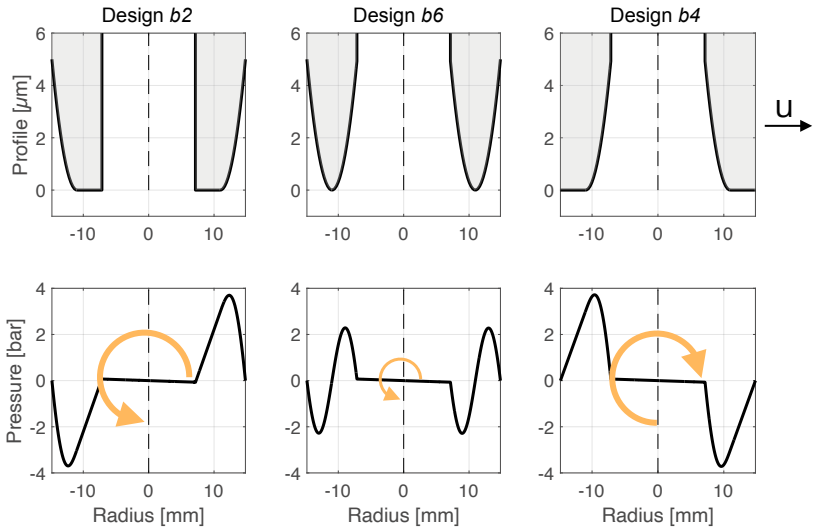


Figure 4.7. Radial cross-sectional profiles of designs *b2*, *b6* and *b4* (top); corresponding qualitative hydrodynamic pressure distribution at 500 rpm (bottom)

Figure 4.8 summarizes the tilting moments of a wide range of combinations of the parabolic design parameters A_{in} and A_{out} for three different values of A_{min}. The more at the inner radius, the curve minimum exists, the more the recess at the outer radius dominates resulting into a broader area of a positive tilting moment (compare the large red area in the left plot with the large blue area in the right plot).

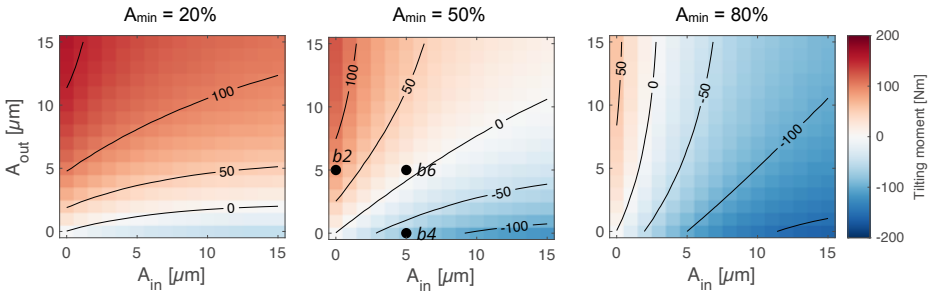


Figure 4.8. Tilting moment generated by hydrodynamic pressure for different quadratic curve minimums A_{min} over A_{in} and A_{out}

5. COMPARISON WITH SIMPLIFIED SIMULATION RESULTS

To validate the fundamental trends observed in the simplified analytical model, a comparison is made with the comprehensive FSTI CASPAR simulation. This step is necessary to confirm that the basic physical effects predicted by the static, analytical approach hold true within a more complex simulation that seeks a system's force and moment equilibrium. However, a direct quantitative validation presents a methodological challenge. The analytical approach calculates forces for a predefined, fixed gap height, producing maps of force balances versus geometry. In contrast, the FSTI simulation is designed to find the system's equilibrium state. It iteratively adjusts the gap height and slipper tilt until the forces and moments precisely balance the external loads (e.g., the piston force). Consequently, the simulation's primary output is the final equilibrium gap height for a given operating point, not a map of forces at various non-equilibrium gap heights.

5.1. Simulation setup and parameter selection

To create a qualitative comparison despite this fundamental difference, several steps were taken. First, the FSTI simulation was simplified to more closely match the assumptions of the analytical model by deactivating squeeze effects and deformations. Second, a key challenge was identifying simulation parameters that yielded stable, converged results with physically meaningful gap heights for all designs. Many initial simulations failed to converge. Therefore, only a smaller selection of the step designs *a1-a4* is chosen for the comparison. Runs with a smaller pocket radius (7.2 mm, as used in the analytical study) resulted in predicted gap heights that were too small, particularly for design *a1*. This approached a mixed-friction regime that the simulation cannot accurately depict. Consequently, a larger pocket radius of 8.2 mm was selected. At an operating condition of 100 bar and 500 rpm, this configuration produced reasonable and comparable equilibrium gap heights.

Figure 5.1 (left) illustrates the slipper coordinate system, in which the direction of movement is defined by x_G and the velocity by u. The high- and low-pressure regions are marked in red and blue and the angle of rotation is defined by ϕ, which ranges from 0 to 360°.

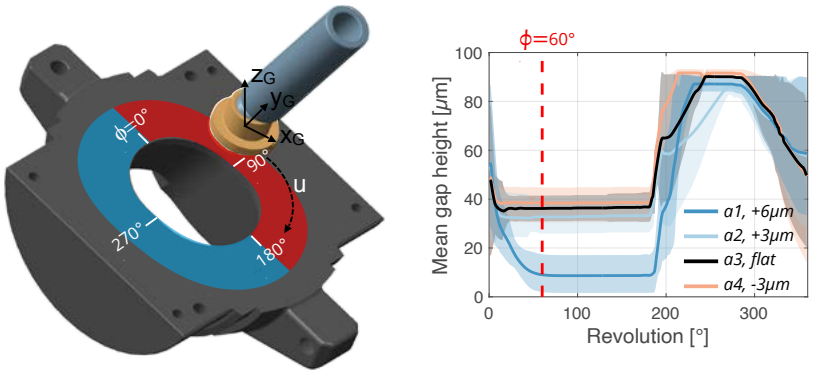


Figure 5.1. Mean gap height over one pump revolution for all designs *a1-a4* with shaded regions indicate min-max range (right)

Figure 5.1 (right) presents the resulting mean gap heights for designs *a1-a4* over one pump revolution, with shaded regions indicating the spread between minimum and maximum values. As predicted by the analytical balance calculations, design *a1*, which has the lowest hydrostatic balance, also exhibits the lowest minimum gap heights in the simulation. The maximum value of the curves during the low-pressure phase is due to the slipper hold-down mechanism, which limits the maximum gap height to 90 µm. The red dotted line indicates a rotation angle of 60°, at which point gap heights and pressure fields are evaluated further.

The next two paragraphs analyze the influence of the meso-structures on the hydrostatic and hydrodynamic pressure build-up in isolation. As the outputs of the FSTI simulation comprise the results of the total Reynolds equation, it is necessary to separate the hydrostatic and hydrodynamic components of the pressure field from the final simulation outputs. To achieve this, a post-processing script was used [23], which implements the methodology for pressure component separation proposed by Ernst and Vacca [18]. This method first uses the results from the full fluid-structure interaction simulation (total pressure p_{tot} and equilibrium gap height). It then resolves the Reynolds equation using only the hydrostatic (Poiseuille) term, treating the final gap height and fluid properties as known constants. The resulting pressure field is the pure hydrostatic component (p_s), and the hydrodynamic component (p_d) is subsequently found by subtraction: $p_d = p_{tot} - p_s$ [18].

Figure 5.2 illustrates the results of the pressure separation for design *a2* at 60° during the high-pressure phase. The gap height distribution (left) reveals the slipper's backward tilt, which directly influences the pressure fields. The hydrostatic component (middle) exhibits the expected radial symmetry, whereas the hydrodynamic component (right) reflects the wedge effect created by the tilt: regions of positive pressure (yellow) form where the gap converges, and sub-ambient pressure (blue) develops at the location of smallest gap height. It should be noted that the negative pressure values are a mathematical result of the separation methodology. In reality, the fluid cannot sustain such tensions and would cavitate; however, this does not affect the validity of the qualitative trends. The red dotted line in b) marks the path analyzed in Figure 5.3.

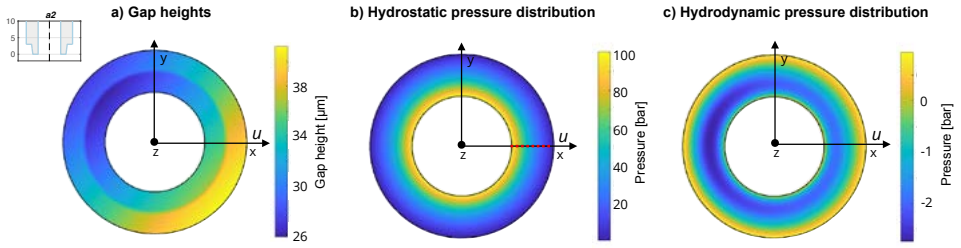


Figure 5.2. Gap height (left), hydrostatic (middle) and hydrodynamic (right) pressure fields for design *a2* at 60°; the red dotted line marks the path analyzed in Figure 5.3.

5.2. Influence on hydrostatic pressure and balance

Figure 5.3 compares the simulated (left) and analytically calculated (right) pressure drop across the leading sealing land for all four designs at $\phi = 60°$. Unlike in chapter 4, the analytical model employs the same pocket diameter of 8.2 mm as the simulation and uses the different mean gap heights at $\phi = 60°$ from Figure 5.1 as inputs. The resulting pressure profiles are qualitatively and apart from minor differences also quantitatively comparable.

Those minor deviations may arise from the pressure component separation methodology [18].

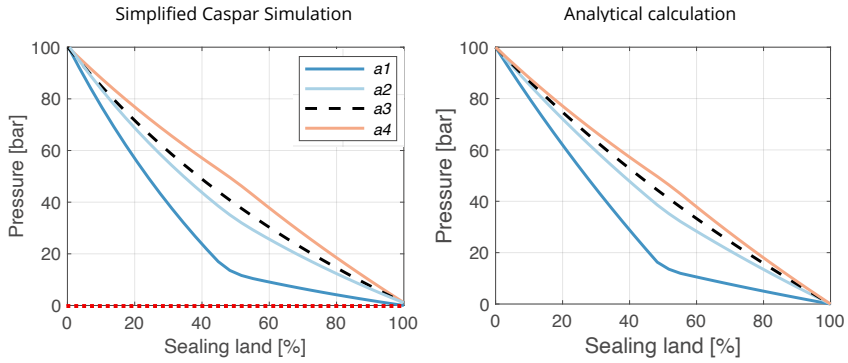


Figure 5.3. Pressure drop over the leading sealing land for all designs *a1-a4* at 60° pump revolution from simulation (left) and analytically calculated (right)

5.3. Influence on hydrodynamic lift and tilting moment

Figure 5.4 displays the respective hydrodynamic pressure fields for designs *a1-a4*, with the resulting tilt axis marked as a dashed line. The opposite tilting moment directions—negative for the outer step (*a4*) versus positive for the inner steps (*a1-a3*)—are clearly visible in the pressure field orientations relative to slipper velocity u. Among designs *a2-a4*, which share similar gap heights, the trend is solely geometry-driven: increasing inner step height leads to a larger positive tilting moment and corresponding tilt axis rotation, aligning with the analytical findings in chapter 4. Design *a1* exhibits a notably stronger hydrodynamic effect due to its much smaller gap height, which amplifies the hydrodynamic lift beyond what the step geometry alone would produce.

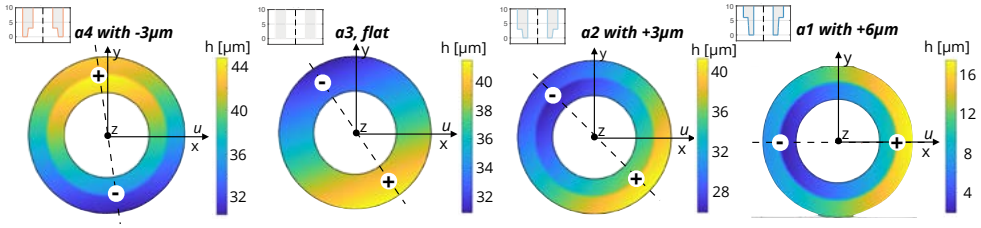


Figure 5.4. Gap height distributions for different step designs *a1-a4* with marked tilt axis at 30° pump revolution during high pressure phase

Figure 5.5 shows the result of this separation for design *a2* at 30° of the pump revolution (high pressure regime), now at an elevated operating pressure of 200 bar (compared to 100 bar in the previous analysis) while maintaining the same speed of 500 rpm. The higher operating pressure was selected because it produces smaller equilibrium gap heights, which in turn amplifies the hydrodynamic pressure generation, making the effect more clearly visible for analysis. The plot on the left displays the isolated hydrodynamic pressure field across the entire slipper land. It clearly shows a region of positive pressure build-up (yellow) and a region of sub-ambient pressure (blue), which is characteristic of the wedge effect. The plot on the right shows the pressure profile along the central x-axis, which is qualitatively

similar to the analytical results for a parallel-step bearing (refer to Figure 2.2), confirming that the step is the primary driver of this hydrodynamic effect.

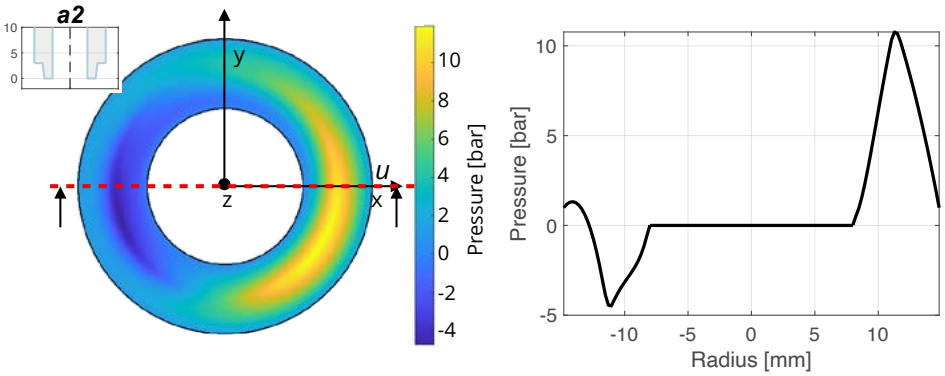


Figure 5.5. Hydrodynamic pressure field at $\phi = 30°$ for design *a2* (left) and hydrodynamic pressure distribution through x-axis (right)

6. DISCUSSION AND OUTLOOK

This study systematically analyzes the influence of meso-structures on the pressure-generating mechanisms in a slipper-swashplate interface, revealing a fundamental trade-off between hydrostatic and hydrodynamic performance. The analytical models and simulation results consistently demonstrate two opposing principles. A reduced gap at the inner side on the sealing land decreases the hydrostatic balance but generates a beneficial positive tilting moment in the direction of motion that promotes hydrodynamic lift. A reduced gap at the outer side of the sealing land by contrast, increases the hydrostatic balance but at the cost of a potentially destabilizing negative tilting moment.

Of particular significance is the observation that the influence of these structures is not constant. The effect on both, hydrostatic balance as well as hydrodynamic lift, becomes more pronounced at smaller gap heights. This non-linear behavior means that the structure's impact is strongest precisely when the slipper operates in critical mixed-friction regimes. This provides a fundamental explanation for the improved reliability of structured surfaces observed in previous experimental work findings provide a clear framework for application-specific design, moving beyond a "one-size-fits-all" approach. For applications dominated by low-speed or start-stop operation where hydrodynamic stability is paramount, a design that guarantees a positive tilting moment is preferable. The associated reduction in hydrostatic balance can be compensated for by adjusting the pocket diameter.

The primary limitation of this analytical study is the neglect of thermo-elastic deformations. As noted in the literature and observed in the FSTI simulations, pressure and temperature-induced deformations can create their own relevant micro-geometry, which may be convex or concave depending on the design and operating point. This deformation can either enhance or counteract the effect of the manufactured structure and has been taken into account for the optimization study in [1].

Beyond the fundamental design trade-offs, the analytical maps presented in this study also provide practical guidance for manufacturing. The contour lines indicate that certain parameters are more tolerant to manufacturing variations than others. For instance, the step position shows relatively flat gradients in the mid-range of the sealing land (approximately 30–60%), meaning that moderate positional deviations have limited impact on the hydrostatic balance. In contrast, the step height exhibits steeper gradients and thus requires tighter manufacturing tolerances to achieve the target performance. This insight is particularly valuable for industrial implementation, as it allows manufacturers to focus precision efforts on the most sensitive parameters while relaxing tolerances where the design is inherently robust.

The findings of this study, particularly the trade-off between hydrostatic balance and hydrodynamic tilting moment, and the non-linear sensitivity of meso-structures at small gap heights, are governed by thin-film lubrication physics and are therefore expected to be transferable to other tribological interfaces in displacement pumps. This includes the valve plate-cylinder block and piston-cylinder bore interfaces in axial piston pumps, as well as the axial thrust bearings in external gear pumps.

ACKNOWLEDGEMENTS

The authors would like to thank Meike Ernst and Andrea Vacca for their foundational methodology on separating hydrostatic and hydrodynamic pressure fields [18], [24]. We are particularly grateful to Meike Ernst for generously providing the original implementation scripts. Access to this sophisticated code greatly facilitated our analysis and saved significant development effort. Special thanks are also extended to Simon Knoll for his diligent work in adapting these scripts for the specific analyses in this study. His development of the 1D analytical model for calculating tilting moments, which formed a significant part of his diploma thesis [23], is also gratefully acknowledged.

REFERENCES

[1] S. Horn, R. Ivantysyn, and J. Weber, “Tribo-optimized lubricating interfaces in hydrostatic pumps with surface shaped slippers,” *IFK2022*, 2022.

[2] R. Ivantysyn, S. Horn, and J. Weber, “Design of a Lead-Free Slipper Bearing for Low Speed Axial Piston Pump Applications,” *IEEE GFPS 2022*, 2022.

[3] S. Horn, R. Ivantysyn, and J. Weber, “Validated efficiency improvements by implementation of structures on the slipper surface of an axial piston pump,” *IEEE GFPS 2022*, 2022.

[4] S. Horn, R. Ivantysyn, and T. Schmidt, “Tribological properties of different slipper designs of an axial piston pump,” *Ifk2024*, no. in mm, pp. 1–18, 2024.

[5] S. Horn, T. Schmidt, and P. Kajewski, “Measuring friction of different slipper designs using a hydrostatic tribometer test-rig with pressure pulsation,” in *Fachtagung Gesellschaft für Tribologie*, 2025, pp. 1–12.

[6] C. J. Hooke and Y. P. Kakoullis, “The effects of non-flatness on the performance of slippers in axial piston pumps,” *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 197, no. 4, pp. 239–247, 1983, doi: 10.1243/PIME_PROC_1983_197_104_02.

[7] M. Borghi, E. Specchia, and B. Zardin, “Numerical analysis of the dynamic behaviour of axial piston pumps and motors slipper bearings,” *SAE Int. J. Passeng. Cars - Mech. Syst.*, vol. 2, no. 1, pp. 1285–1302, 2009, doi: 10.4271/2009-01-1820.

[8] N. D. Manring, R. E. Johnson, and H. P. Cherukuri, “The impact of linear deformations on stationary hydrostatic thrust bearings,” *J. Tribol.*, vol. 124, no. 4, pp. 874–877, 2002, doi: 10.1115/1.1482118.

[9] S. Hashemi, H. Friedrich, L. Bobach, and D. Bartel, “Validation of a thermal elastohydrodynamic multibody dynamics model of the slipper pad by friction force measurement in the axial piston pump,” *Tribol. Int.*, vol. 115, no. May, pp. 319–337, 2017, doi: 10.1016/j.triboint.2017.05.013.

[10] J. Ivantysyn, “Sliding bearing with self-adjusted load bearing capacity,” 5983781, 1999.

[11] M. Ivantysynova and J. Baker, “Power loss in the lubricating gap between cylinder block and valve plate of swash plate type axial piston machines,” *Int. J. Fluid Power*, vol. 10, no. 2, pp. 29–43, 2009, doi: 10.1080/14399776.2009.10780976.

[12] Reddy; Ivantysyn; Rauschenberger; Weber, “Advanced Micro-Surfacing : Tribological Optimization for Variable Speed Hydraulic Drives,” 2025.

[13] J. Baker and M. Ivantysynova, “Axial sliding bearing and method of reducing power losses therof,” WO 2009/146060 A3, 2009.

[14] S. Ye, H. Tang, Y. Ren, and J. Xiang, “Study on the load-carrying capacity of surface textured slipper bearing of axial piston pump,” *Appl. Math. Model.*, vol. 77, pp. 554–584, 2020, doi: 10.1016/j.apm.2019.07.058.

[15] Z. Wang, L. Sun, B. Han, X. Wang, and Z. Ge, “Study on the Thermohydrodynamic Friction Characteristics of Surface-Textured Valve Plate of Axial Piston Pumps,” *Micromachines*, vol. 13, no. 11, 2022, doi: 10.3390/mi13111891.

[16] A. A. Darbani, L. Shang, J. R. Beale, and M. Ivantysynova, “Slipper surface geometry optimization of the slipper/swashplate interface of swashplate-type axial piston machines,” *Int. J. Fluid Power*, vol. 20, no. 2, pp. 245–270, 2019, doi: 10.13052/ijfp1439-9776.2025.

[17] H. Mo, Y. Hu, and S. Quan, “Thermo-Hydrodynamic Lubrication Analysis of Slipper Pair Considering Wear Profile,” *Lubricants*, vol. 11, no. 5, pp. 1–14, 2023, doi: 10.3390/lubricants11050190.

[18] M. Ernst and A. Vacca, “Hydrostatic vs. hydrodynamic components of fluid pressure in the tribological interfaces of axial piston machines,” *Tribol. Int.*, vol. 157, no. November 2020, p. 106878, 2021, doi: 10.1016/j.triboint.2021.106878.

[19] O. Reynolds, “On the Theory of Lubrication and its Application to Mr. Beauchamp Tower’s Experiments including an Experimental Determination of the Viscosity of Olive Oil.,” *Philos. Trans. R. Soc. London*, vol. 177, pp. 157–234, 1886.

[20] B. J. Hamrock and S. R. Schmid, *Fundamental of Fluid Film Lubrication*. 1991.

[21] Lord Rayleigh, “Notes on the Theory of Lubrication,” *Philosophical Magazine*, vol. 35, pp. 1–12, 1918.

[22] J. M. Bergada and J. Watton, “Axial piston pump slipper balance with multiple lands,” *ASME Int. Mech. Eng. Congr. Expo. Proc.*, pp. 69–73, 2002, doi: 10.1115/IMECE2002-39338.

[23] S. Knoll, “Einfluss der Hydrodynamik auf die Gleitschuhperformance verschiedener meso-strukturierter Gleitschuhe,” TU Dresden, 2025.

[24] M. Ernst and A. Dissertation, “Enabling High-Pressure Operation With Water for the Piston-Cylinder Interface in Axial Piston Machines,” 2021.

Biographies



Svenja Horn received her diploma degree in Aircraft Construction from the Technical University of Dresden in 2018. Since then, she has been working as a research assistant at the Chair of Fluid-Mechatronic Systems, focusing on improving the efficiency and service life of axial piston pumps. Her research concentrates on tribological phenomena at the slipper–swashplate interface, which she investigates using simulations, a hydrostatic tribometer and pump measurements. Furthermore, she applies machine learning algorithms for pump condition monitoring and lifetime prediction.

Effects of contamination on the wear process of piston/cylinder interface in axial piston machines considering three-body abrasive wear model

Liangyu Song[1], Fei Lyu[1,2]*, Junhui Zhang [1], Bing Xu [1]

[1] *The State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, China*

[2] *Institute for Fluid Power Drives and Systems, RWTH Aachen University. Campus-Boulevard 30, 52074 Aachen, Germany*

** Corresponding author: Tel.: (49)1739447126; E-mail address: fei.lyu@ifas.rwth-aachen.de*

Abstract.

Contamination level is a critical indicator in hydraulic systems. The piston/cylinder interface (PCI) serves vital functions in lubrication and load-bearing within axial piston pumps. Under contaminated conditions, PCI performance degrades rapidly, potentially resulting in complete pump failure. A wear model incorporating multiple wear mechanisms at the PCI was developed, along with a predictive method specifically accounting for three-body abrasive wear. An isolated PCI test bench that can accurately control the contamination degree was designed for verification. Both simulation and experimental results demonstrate that contaminated abrasive particles alter wear distribution and significantly accelerate the wear process.

Keywords. Contamination level, piston/cylinder interface, three-body abrasive wear, abrasive particle, wear process prediction.

1. INTRODUCTION

Axial piston machines are widely used as hydraulic power sources in industrial operation systems, offering higher power density in high-pressure applications. The piston/cylinder interface (PCI) is a critical tribological interface in axial piston machines, responsible for lubrication and load support. During operation, this interface undergoes increased wear caused by various mechanisms, including adhesive wear, two-body abrasive wear, and three-body abrasive wear induced by the rolling action of entrained wear particles within the contact interface. These complex wear mechanisms can degrade the oil suction and discharge performance and eventually lead to system failure.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Analyzing the wear process of the piston-cylinder interface (PCI) is crucial for effective predictive maintenance, accurately assessing its service life and reliability, and providing key insights for design optimization aimed at wear resistance and anti-wear properties. During operation, the surface contour of the piston/cylinder pair degenerates over time, accompanied by complex oil film pressure and thickness boundaries. As a result, the wear state of the interface is highly dynamic and coupled with oil film characteristics. Fei Lyu et al. conducted numerical simulations on two-body abrasive and adhesive wear processes of the piston-cylinder interface (PCI) [1,2]. Yu Lin et al. established a dynamic model based on the equivalent load method and proposed a novel wear prediction approach for the PCI by integrating the Archard wear model [3]. Lars Brinkschulte et al. proposed a simulation model, based on the Archard and Fleischer approaches, to predict wear in the piston-cylinder contact [4]. Roman Ivantysyn et al. combined numerical simulation with experimental measurements to analyze the run-in wear behavior of the slipper-swash plate interface in axial piston pumps, demonstrating the stabilization of the lubricating gap and wear profile during early operation [5]. Zunling Du et al. established a mixed lubrication model for the slipper pair within an elasto-hydrodynamic lubrication (EHL) framework, integrating two-body abrasive and Archard adhesive wear theories to formulate an eccentric wear prediction method [6]. Svenja Horn et al. investigated the tribological performance of different slipper designs in axial piston pumps, showing that tailored surface structures can effectively enhance hydrodynamic pressure generation and improve wear condition under mixed lubrication conditions [7]. Xuguang Li et al. proposed a wear prediction model for the cylinder block/valve plate interface, considering the multi-scale evolution of surface topography, including both macroscopic profile wear and microscopic asperity deformation [8].

However, the impact of high oil contamination levels was not considered in these simulations. High contamination levels can cause severe wear on the friction pairs, especially under high-pressure and high-speed PCI conditions, which are more sensitive to contamination levels [9]. Compared with other friction pairs such as the slipper-swash plate interface and the cylinder block/valve plate interface, the PCI shows a stronger dependence of its frictional behavior on clearance design, resulting in increased sensitivity to contaminant particles [10]. The distribution of abrasive particle size and concentration is related to the level of contamination, such as the standard NAS1638 for oil contamination level [11]. Three-body abrasive wear, which is caused by abrasive particles entering the friction pairs, is significantly influenced by the size, hardness, and concentration of the abrasive particles. Therefore, studying three-body abrasive wear of PCI under high contamination levels is crucial for better predicting the wear process.

There is extensive basic research on three-body wear of tribo-pairs. Under high oil pollution conditions, three-body abrasive wear accounts for a large proportion [12], which affects the service life and performance of friction pairs. Different types and sizes of contaminant particles in oil can lead to different wear results of friction pairs [13]. Some studies have explored the effects of materials, textured surfaces, and abrasive particle sizes on the wear process of bearings [14,15]. Nejc Novak et al. studied the effect of oil contaminants on accelerated gear wear [16]. The performance of lubricants and coatings can also affect three-body abrasive wear [17,18]. However, the extreme operating conditions of PCI lead to complex boundary conditions that form wedge-shaped oil films and multiple wear mechanisms acting simultaneously. These studies mainly analyze local specific wear

through wear experiments, and there is no way to directly analyze the distributed lubrication state and multiple complex wear mechanisms in PCI. Therefore, it is necessary to study the coupling of wear and oil film in PCI after considering three-body wear.

This paper proposes a method for predicting wear process in the PCI that incorporates three-body abrasive wear caused by contaminated oil. Section 2 outlines the simulation of lubrication calculation and wear cycle. Section 3 describes the experimental setup for testing a single PCI under different contamination conditions. Section 4 compares simulated and experimental results. Section V concludes the study.

2. WEAR PROCESS SIMULATION

The development of a PCI wear model begins with a dynamic analysis to calculate lubrication boundary conditions such as oil film pressure and thickness. Following this, a distributed wear sub-model performs the wear calculation. The resulting updated contour is then coupled with the oil film, enabling the prediction of the wear process.

2.1. Dynamic Analysis and Lubrication Calculation of PCI

Figure 2.1 shows the general configuration of the axial piston pump, the main external forces applied on the piston are shown in Figure 2.2. The support force of the swash plate F_{SS} and the pressure force F_{P} cause the piston to reciprocate within the cylinder. F_{SSy} is the lateral component of F_{SS}. The friction force of the slipper F_{FS} is generated as the slipper slides on the swash plate. The resultant force of F_{SSy} and F_{FS} is F_{L}. F_{FC} is the friction force between the cylinder and the piston, F_{C} is the centrifugal force of the piston, F_{I} is the inertial force of the piston, and F_G is the gravitational force of the piston. Due to the significant lateral force F_{L}, the piston tilts within the cylinder bore. This radial motion leads to a wedge-shaped oil film being squeezed. When the oil film becomes very thin, solid contact pressure arises. The resulting reaction force that balances the external forces is generated through a combination of the oil film's squeezing effect, dynamic pressure buildup, and solid contact pressure.

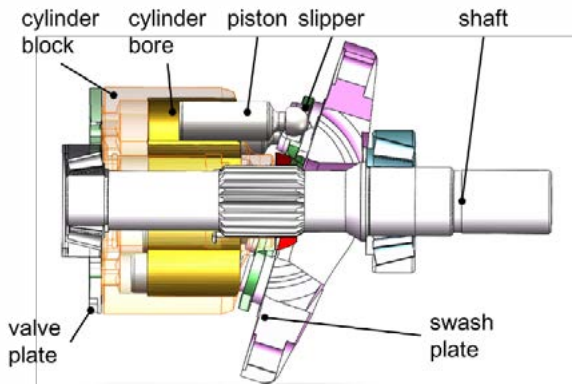


Figure 2.1. Axial piston pump basic structure

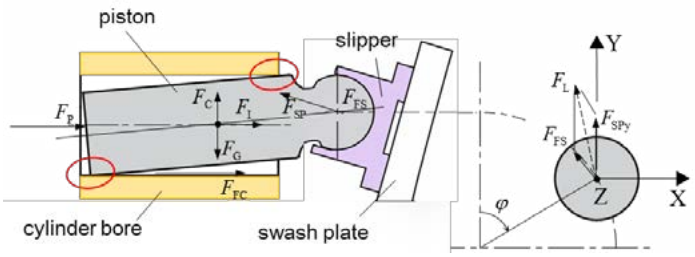


Figure 2.2. PCI dynamic analysis

The oil film thickness between the PCI is related to the posture of the piston in the cylinder bore. When the surface wear occurs, the wear depth $w(l,\theta)$ must be considered to achieve more accurate oil film thickness. Figure 2.3 shows the titling state of the piston in the cylinder bore exaggeratedly. The posture of the piston can be represented by the eccentricities of two sections relative to the cylinder bore. The oil film thickness can be defined as the clearance between the piston and the cylinder bore in the radial direction, and it is given by:

$$h(l,\theta) =$$

$$(D_{\mathrm{C}} - D_{\mathrm{P}})/2 + w(l,\theta) - [e_1 - (e_1 - e_2)l/L_{\mathrm{C}}]cos\theta - [e_3 - (e_3 - e_4)l/L_{\mathrm{C}}]sin\theta \qquad (2.1)$$

where e_1, e_2 and e_3, e_4 are the eccentricities in the X direction and Y direction of the two sections respectively. L_{C} is the length of the cylinder bore, D_{C} and D_{P} are the diameter of the cylinder bore and the piston. The pressure field is divided into $a \times c$ grids and can be expanded to a plane for intuitive as shown in Figure 2.3.

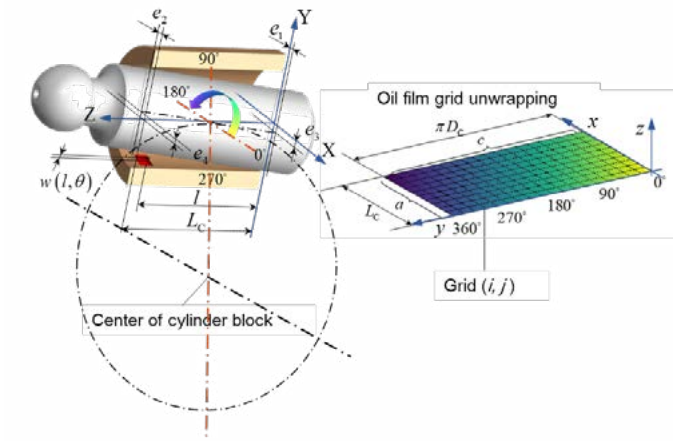


Figure 2.3. Titling state of the piston and expansion of the oil film

The minimum contact oil film thickness hm can be determined by the surface roughness of the friction pair is set to estimate whether the contact occurs. The definition of $h_{\min}$ [19] is shown in Eq. (2.1), where R_{qp} and R_{qc} are the root mean square roughness of the piston and the cylinder bore respectively:

$$h_{\min} = 3\sqrt{{R_{\mathrm{qp}}}^2 + {R_{\mathrm{qc}}}^2} \tag{2.2}$$

The hydrodynamic pressure of the oil film p can be represented by using Reynolds equation:

$$\frac{\partial}{\partial x}\left(\frac{\partial p}{\partial x}\frac{h^3}{\mu}\right) + \frac{\partial}{\partial z}\left(\frac{\partial p}{\partial z}\frac{h^3}{\mu}\right) = 6\left(\omega_{\mathrm{P}}\frac{D_{\mathrm{C}}}{2}\frac{\partial h}{\partial x} + u\frac{\partial h}{\partial z} + 2\frac{\partial h}{\partial t}\right) \tag{2.3}$$

where μ denotes the oil viscosity, and ω_{P} is the spinning speed of piston, u is the sliding velocity of the piston surface. To ensure numerical stability, the oil film thickness h is set to $h_{\min}$ when the calculated result of Eq. (2.2) is less than $h_{\min}$. In Eq. (2.3), the first two terms on the right-hand side represent the dynamic pressure effect generated by the wedge-shaped oil film, while the third term corresponds to the squeezing effect due to micro-scale radial motion of the piston. The Reynolds equation is discretized using the finite volume method, and the Cyclic Tridiagonal Matrix Algorithm is employed to solve the resulting pressure field of the PCI.

The pressure distribution is obtained by solving the discretized Reynolds equation, thereby determining the oil film reaction force acting on the piston. The Newton-Raphson method is then applied to iteratively solve the piston force balance equation.

In Figure 2.4, k is the rotated angle of the cylinder with the resolution of 1°. $\Delta\dot{e}$ is the variation of Δe in each iteration, adjusted automatically via the Newton-Raphson method. P, P_{cs} refers to dynamic oil film pressure and contact pressure between piston and cylinder block. F_{out}, F_{out}, F_{cs} refers to external force, dynamic oil film force and contact force between piston and cylinder block. A convergence threshold $F_{\mathrm{b}} = 0.1$ N is set for the force balance residual. The Newton-Raphson iteration terminates when the condition $\|f(\dot{e}_k)\| \leq F_{\mathrm{b}}$ is satisfied. The simulation is considered converged and lubrication parameters are output when periodic phenomenon is exhibited.

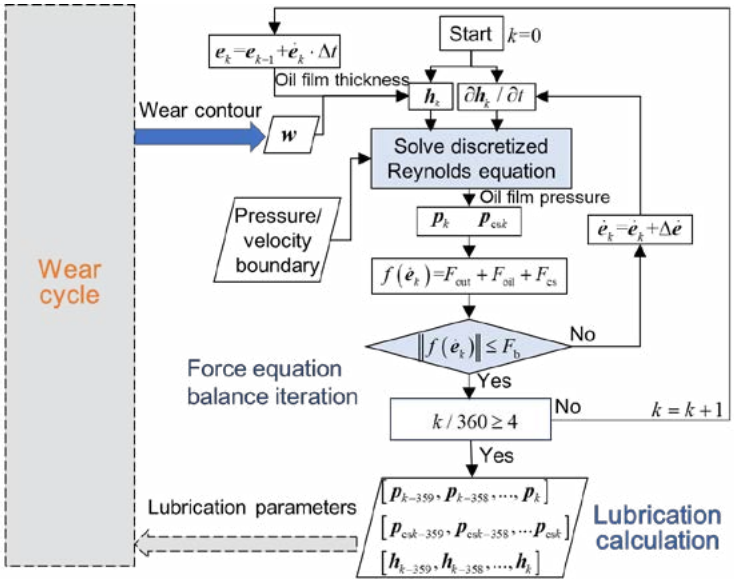


Figure 2.4. Schematic diagram of lubrication calculation

2.2. Wear Cycle

As shown in Fig. 2.5, inputs including the oil film thickness $h(i,j)$, pressure $p(i,j)$, and solid contact pressure $p_{cs}(i,j)$ field for each grid cell from the lubrication calculation model, as well as the contaminant particles size distribution d_a measured in the contaminated oil, are introduced into the wear cycle model. Depending on the local oil film thickness boundaries, each computational grid may undergo different wear mechanisms-including adhesive wear, two-body wear, and three-body abrasive wear-leading to distinct wear depths. And the wear contour is linearly superimposed with the revolutions until the wear depth exceeds a threshold value. Once the threshold is exceeded, the wear curve is fed into the oil film model to update the contours of both the piston and cylinder. The updated wear curve is also output, and the corresponding operating time can be calculated from the pump revolutions and speed. This completes the simulation of the time-varying wear process of PCI.

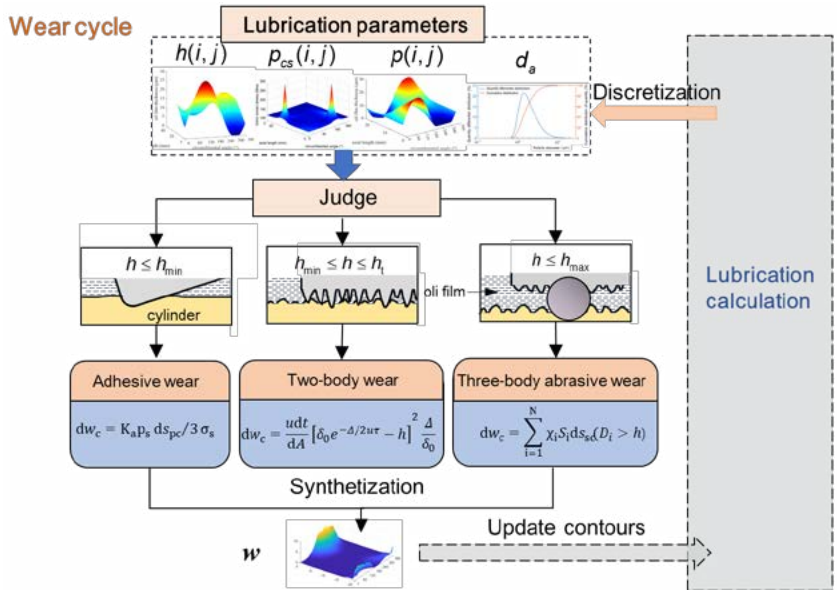


Figure 2.5. Wear cycle model

When the oil film cannot provide sufficient counterforce to balance the external load, the piston and cylinder are squeezed to slide against each other. Under high solid contact pressure, rupture of the oil film occurs, resulting in adhesive wear. The wear depth of the cylinder bore can be calculated using the equation below.

$$\begin{cases} 0 < h < h_{\min} \\ \mathrm{d}w = K_{\mathrm{a}} p_{\mathrm{s}} d_{\mathrm{pc}} / 3\sigma_{\mathrm{S}} \end{cases} \tag{2.4}$$

where d_{pc} is the displacement of the piston relative to the cylinder, σ_{S} is the compressive yield limit, K_{a} is the wear coefficient, p_{s} is the solid contact pressure.

When the oil film thickness is less than h_{t} which characterizes the height of the roughness peak after viscoelastic deformation but remains comparatively larger than the critical thickness $h_{\min}$ for adhesive wear, two-body wear may initiate. Specifically, two-

body wear occurs when the deformed roughness peak exceeds the oil film thickness. The wear depth of the cylinder bore is shown below [27].

$$\begin{cases} h_{\min} < h < h_{\mathrm{t}} = \delta_0 \mathrm{e}^{-\Delta/2u\tau} \\ \mathrm{d}w_{\mathrm{c}} = \frac{u\mathrm{d}t}{\mathrm{d}A}[\delta_0 \mathrm{e}^{-\frac{\Delta}{2u\tau}} - h]\frac{\Delta}{\delta_0} \end{cases} \tag{2.5}$$

where t is the sliding time, and A is the wear area which is equal to the grid area of oil film thickness field, δ_0 is the height of roughness peak, Δ is the wavelength of roughness peak, τ is the delay time of the piston material.

As shown in the Figure 2.6, the real-time oil film contour formed by the reciprocating motion of the piston in the cylinder (as described in Section 2.1) exhibits a wedge-shaped oil film. The contamination level is used to determine the volume concentration within a particle size range. Laser analysis of the contaminated oil sample reveals the proportion of specific particle sizes, yielding the size and concentration distribution of the abrasive particles. When the particle size is smaller than the maximum oil film thickness, there is a certain probability χ_i that the particle will enter the PCI and cause wear [20]. Because the fluid is considered a homogeneous suspension, the probability of a particle entering the interface is assumed to be proportional to the volume of oil in the gap and the volume concentration of particles of different sizes. The influence of abrasive particles on adhesive wear extension is neglected in the simulation [21]. Wear primarily occurs in areas where the particle diameter exceeds the local oil film thickness, resulting in indentation and material removal [22]. In areas where the particle diameter is smaller than the oil film thickness, almost no wear occurs.

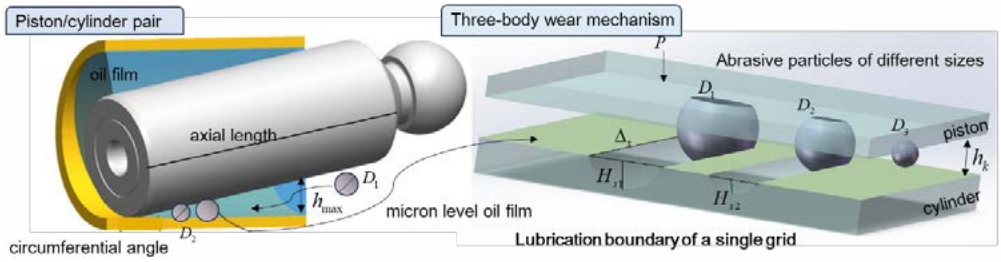


Figure 2.6. PCI three-body wear mechanism

Assuming the abrasive particles are spherical, for each simulation grid cell, particles of different sizes embedded into the cylinder surface will induce wear depths such as H_{S2} [23], which are inversely proportional to the hardness of both the cylinder and piston surfaces. When the particle diameter $D_i > h$, the wear depth on the cylinder can be calculated using Eq. (2.6). Since the size of the computational grid cell is much larger than the abrasive particle diameter, particles of different diameters $(D_1, D_2, \cdots)$ will cause wear in grid cells where the oil film thickness is less than the particle diameter. The cumulative wear volume per unit two-body distance is computed by summing the contributions from particles of various diameters. However, the displacement caused by abrasive particles on the cylinder surface is different from adhesive wear. According to the empirical formula that surface friction is inversely proportional to hardness [24], the sliding distance d_{sc} of the abrasive particles relative to the cylinder surface is calculated by torque balance and rolling kinematics, as shown in Eq. (2.7) and Figure 2.6(b). It is worth noting that different test

conditions can produce two-body grooving or three-body rolling abrasive wear mechanisms [25]. For high-hardness abrasives such as quartz sand, when $D/h > 2$, particles may be embedded in the softer cylinder surface [26]. The sliding distance of the embedded particles relative to the piston surface becomes d_{pc}, which significantly increases wear and causes severe directional grooving on the harder surface. However, in this study, stainless steel powder with a hardness comparable to that of the cylinder material was used. The possibility of these particles adhering to the cylinder surface and causing scratches is very low.

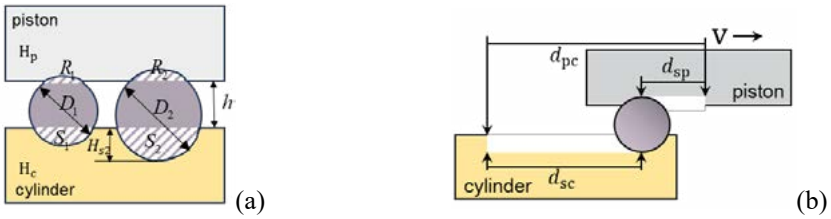


Figure 2.7. Three-body abrasive model

$$H_{S2} = \frac{\mathrm{H_p}}{\mathrm{H_p}+\mathrm{H_c}}(D_i - h) \tag{2.6}$$

$$d_{\mathrm{sc}} = \frac{\mathrm{H_p}}{\mathrm{H_p}+\mathrm{H_c}} d_{\mathrm{pc}} \tag{2.7}$$

Where $\mathrm{H_C}$ and $\mathrm{H_P}$ are the hardness of the cylinder and piston, d_{pc}, d_{sc} and d_{sp} are the sliding distance relatively in Figure 2.7(b).

The wear volume of the cylinder and piston surface meshes is calculated by Eq. (2.8). Given that the calculated three-body wear volume of the piston is approximately 20% of the cylinder's and considering the piston's spinning motion (which promotes uniform axial wear distribution), the experimental analysis does not focus on the piston's wear contour. To optimize wear calculations, the wear contour is linearly superimposed with the revolutions until the wear depth exceeds a threshold value. Once the threshold is exceeded, the wear curve is fed into the oil film model to update the contours of both the piston and cylinder. The updated wear curve is also output, and the corresponding operating time can be calculated from the pump revolutions and speed. This completes the simulation of the time-varying wear process of PCI.

$$\begin{cases} \mathrm{d}w_{\mathrm{c}} = \sum_{i=1}^{\mathrm{N}} \chi_i S_i ds_{\mathrm{sc}}(D_i > h) \\ \mathrm{d}w_{\mathrm{p}} = \sum_{i=1}^{\mathrm{N}} \chi_i R_i ds_{\mathrm{sp}}(D_i > h) \end{cases} \tag{2.8}$$

Where $\mathrm{d}w_{\mathrm{c}}$ and $\mathrm{d}w_{\mathrm{p}}$ are the displacement of the abrasive particles relative to the cylinder and piston. χ_i is the probability of different abrasive particles participating in wear in this grid. S_i and R_i are the areas of the semicircular wear regions in Figure 2.7(a). D_i represents the set of abrasive particle diameters that can cause wear on the grid (i, j) with oil film thickness h.

3. EXPERIMENTAL INVESTIGATION

The isolated PCI test bench that can control the contamination level is designed to carry out the control test under NAS1638 8 and 12. The following is the detailed test method.

3.1. Test Bench Design

To experimentally validate the three-body abrasive wear simulation results of PCI across varying contamination levels, an isolated PCI test bench was built, capable of precisely and directly controlling the contamination levels in the PCI and isolating it from the influence of other friction pairs. In Figure 3.1, add external contamination and detect the contamination level of PCI through an online particle counter (9). Then switch the filter (2,12) through the directional valve (1,13) to control contamination level, when the contamination level gradually increases and exceeds the rated value during operation. The overflow valve (6) maintains the high-pressure input of PCI, and the driving swash plate speed is controlled by the driver (17). Actual picture of the isolated PCI test bench is shown in Figure 3.2.

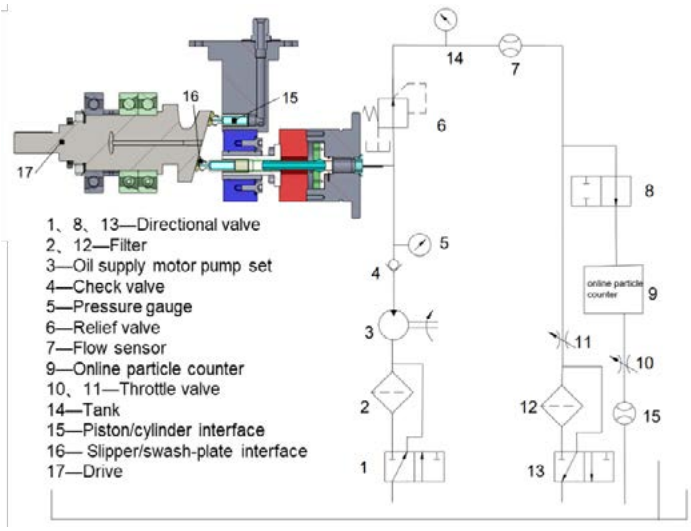


Figure 3.1. Hydraulic schematic diagram of the isolated PCI test bench

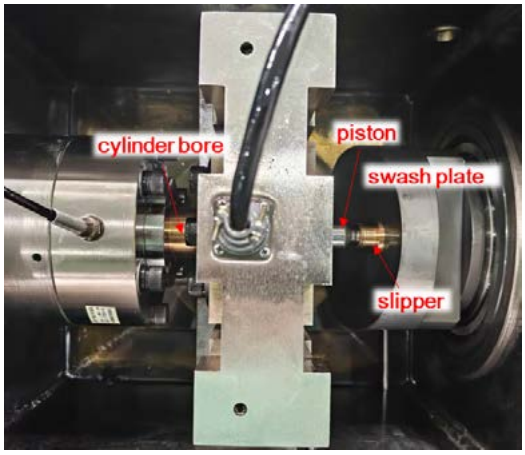


Figure 3.2. Physical photo of the isolated PCI test bench

3.2. Test Samples and Abrasive Particles Preparation

In Figure 3.3, the piston sample (a) and cylinder sample (b) used in this test are made of 38CrMoAl and manganese brass (CuZn37Mn3Al2PbSi) respectively, with a surface roughness of Ra 0.2μm after processing. The red mark is the effective length in contact with the piston.



(a) (b)

Figure 3.3. Test samples

During the test run, 300 mL of oil with additional contaminated abrasive particles was taken, which was the same oil as that used in the PCI operating environment. The abrasive particles added to the oil are mixed with a mass ratio of 30% for <3 μm, 40% for 5 μm, and 30% for 6.5 μm to be closer to the actual contaminated oil abrasive particle distribution. An ultrasonic crusher was employed to ensure optimal dispersion of the particles. Subsequently, the contaminated oil sample underwent wet particle size analysis utilizing a laser particle sizer (Model LS-609, Zhuhai OMEC Instruments Co., Ltd., China). The particle size distribution measurement results are shown in Figure 3.4, and the proportion of each particle size can be obtained. Based on the aforementioned data, the differential percentage by count for specific particle size ranges (e.g., 5-15 μm, corresponding to the first interval of contamination level) can be further derived, as shown in Figure 3.5. Thus, based on the concentration of wear particles within the measured interval range (5-15 μm) detected by the contamination sensor, the proportion of each particle size required to participate in the wear cycle (Figure 2.5) is obtained, and the range is 0-30 μm.

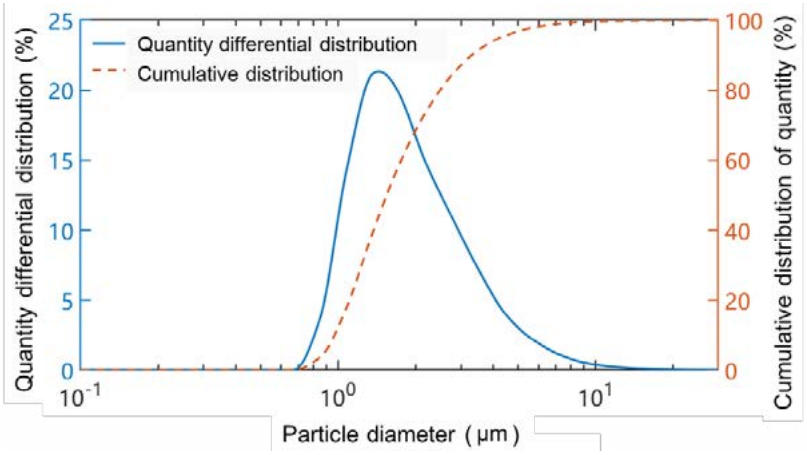


Figure 3.4. Particle size distribution

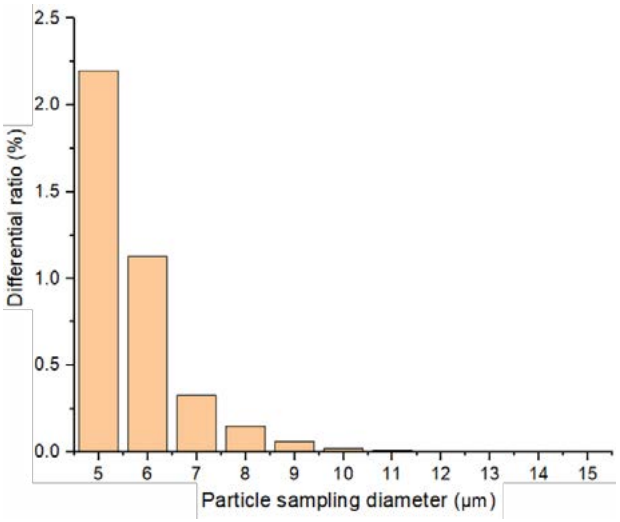


Figure 3.5. Different proportions of sampled abrasive particles

3.3. Experimental Procedures

The operating conditions were maintained at a pressure of 14 MPa and a rotational speed of 1500 rpm. A predetermined concentration of ultrasonically dispersed stainless steel contaminant particles was introduced into the 50 L oil tank. The oil supply motor pump was subsequently activated. After the hydraulic circuit and piston chamber were fully primed with oil, the system pressure was elevated to 14 MPa, corresponding to the preset cracking pressure of the relief valve. A minor volume of excess oil was discharged through a particle counter sensor integrated at the relief valve. Contaminant particles were supplemented in real time based on continuous monitoring until the oil contamination level stabilized. The drive motor was then engaged, and the rotational speed was increased to 1500 rpm to initiate the test. Whenever the contamination level exceeded the predetermined threshold, diversion filtration was activated to prevent particulate accumulation and clogging. Upon completion of the test, a recirculating filtration process was carried out for system flushing and cleaning.

The test was divided into two groups: one with clean oil without added particles, and the other with contaminated oil with a contamination level of 12, containing stainless steel abrasive particles. The two groups were run for 0.5 hours, 1.5 hours, 3 hours, and 5.5 hours, respectively. Stop at the set time, remove the sample and measure the inside of the cylinder using a three-coordinate instrument (Leitz PMM-C). The wear amount is calculated by subtracting the fitted contour before and after wear, thereby obtaining the wear process curve and contour change.

By monitoring the oil contamination level in real time with a particle counter sensor, the curve of the change in the number and volume of abrasive particles (5-15 μm) in the test with abrasive particles is shown in Fig. 14. It can be seen that the contamination level fluctuates within the range of 12. The small fluctuations observed in the experiment indicate the stability and reliability of the contamination control.

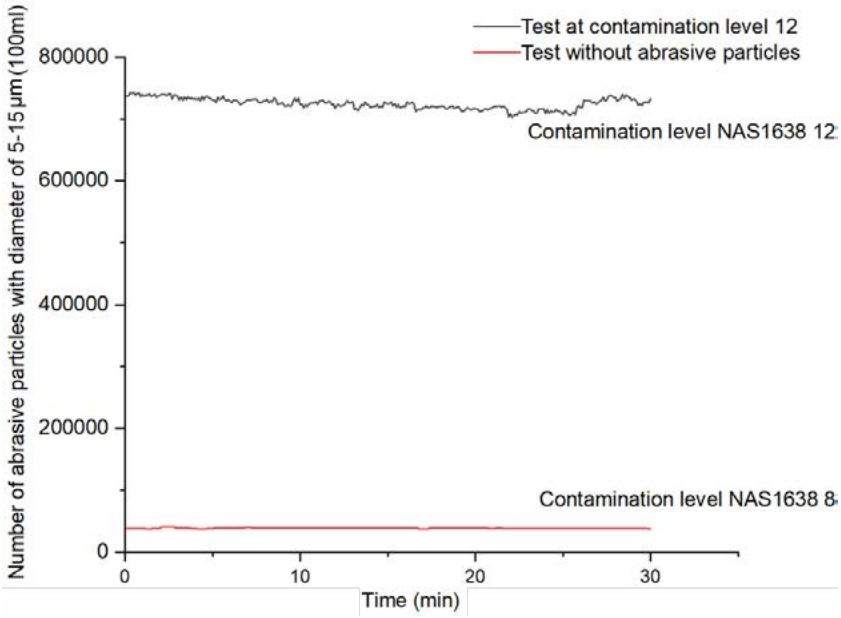
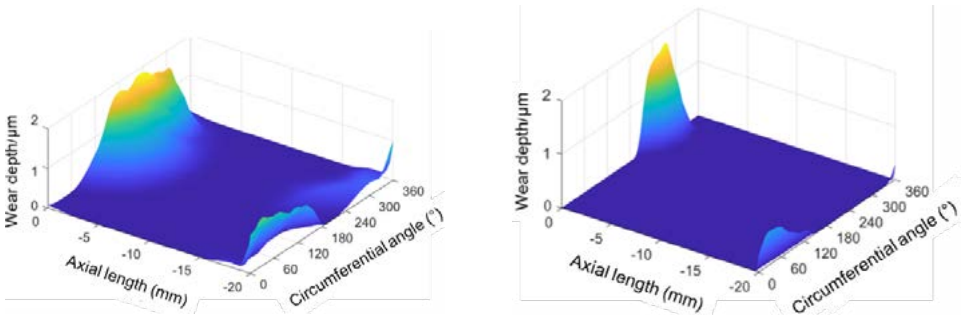


Figure 3.6 Fluctuations in contamination levels during the test

4. RESULTS AND ANALYSIS

Figures 4.2(a) and (b) present the simulated wear contours of the cylinder surface with and without three-body abrasive wear, respectively. The three-body wear model in Figure 4.2(a) reveals a wear distribution extending 90°-360° around the cylinder port, with a depth of 0-7μm. Notably, severe wear is concentrated between 180° and 300° near the port. Additionally, moderate wear extends across the cylinder bottom from 0°-180° and 240°-360°. In contrast, the model without three-body wear in Figure 4.2(b) exhibits a significantly reduced wear zone, confined to 135°-340° at the port with a depth of 0-2.5μm and a distinct wear peak. The area of severe wear is limited, with only minor wear observed at the cylinder bottom.



(a) with three-body abrasive wear (b) without three-body abrasive wear

Figure 4.2. Simulated wear contour of the cylinder surface

Figure 4.4 shows the contour of cylinder after wear test at contamination level 12, while Figure 4.3 shows the contour of cylinder after wear without added abrasive particles. The wear contour is calculated by subtracting the run time after 0.5 hours from the wear time after 5.5 hours.

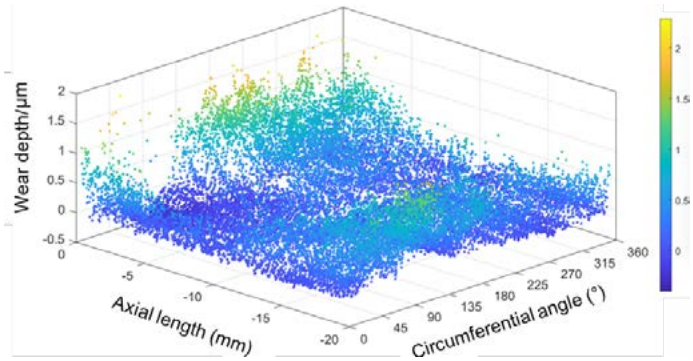


Figure 4.3. Contour of cylinder after wear test at contamination level 12

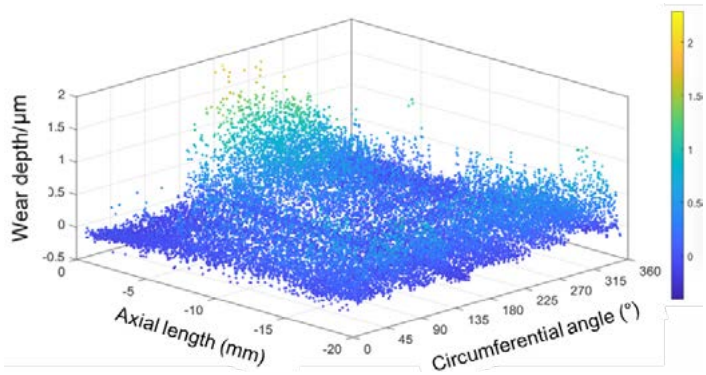


Figure 4.4. Contour of cylinder after wear without abrasive particles

Comparing the simulation results above, Figure 4.3 shows a similar wear range distribution to (a), with wear distribution at 90°-360° and 0-7μm in the cylinder port area, a larger wear distribution at 135°-290°, and a corresponding wear distribution at the bottom of the cylinder, only within a reasonable range of deviation. However, it is somewhat different from Figure 4.2(b), indicating that when the contamination level is high, the model including three-body wear is more accurate for simulation and has a more reasonable wear range.

Figure 4.4 shows the wear contour under the condition of no contamination. The distribution range of the larger wear is smaller, and the wear distribution is similar to the results of the model without the three-body abrasive particle addition, indicating that the wear process can be simulated without the three-body wear model under the condition of no contamination.

5. CONCLUSION

This paper introduces a predictive method for PCI wear that includes three-body abrasive wear due to oil contamination. Comparisons between simulated and experimental surface contours after operation under normal and highly contaminated (NAS1638 12) conditions reveal that, although the maximum wear depth is not significantly affected in the initial hours, the severe wear area and the overall worn area are greatly expanded. This accelerated wear process can deteriorate lubrication, potentially leading to premature failure. The model demonstrates a superior ability to predict both the wear extent and volume under high contamination, highlighting its significance for studying tribo-pairs in contaminated environments.

6. REFERENCES

[1] Lyu, F., Zhang, J., et al. Research on wear prediction of piston/cylinder pair in axial piston pumps. *Wear*, 2020, 456-457: 203338.

[2] Zhang, J., Lyu, F., et al. Simulation and experimental investigation on low wear rate surface contour of piston/cylinder pair in an axial piston pump. *Tribology International*, 2021, 162: 107127.

[3] Yu, L., Wang, H., et al. A novel wear prediction method and wear characteristic analysis of piston/cylinder pair in axial piston pump. Wear, 2024, 550-551, 205402.

[4] Brinkschulte, L., Mutter, J., et al. An approach to wear simulation of hydrostatic drives to improve the availability of mobile machines [Paper presentation]. 11th International Fluid Power Conference (11. IFK), Aachen, Germany, 2018.

[5] Ivantysyn, R., Shorbagy, A., et al. Investigation of the wear behavior of the slipper in an axial piston pump by means of simulation and measurement [Paper presentation]. 12th International Fluid Power Conference (12. IFK), Dresden, Germany, 2020.

[6] Du, Z., Zhang, Y., et al. Research on Prediction Methods for Eccentric Wear of the Slipper Pair in Axial Piston Pumps and Lubrication Performance Analysis. *ASME Journal of Tribology*, 2024, 146(9), 092201.

[7] Horn, S., Ivantysyn, R., et al. Tribological properties of different slipper designs of an axial piston pump [Paper presentation]. 14th International Fluid Power Conference (14. IFK), Germany, 2024.

[8] Li, X., Zhang, J., et al. Multi-scale topography wear mechanism of the valve plate in axial piston pumps during running-in process: A case study. Engineering Failure Analysis, Volume 182, Part B, 110058, 2025.

[9] Aerospace Industries of America. *NAS 1638: Cleanliness Requirements of Parts Used in Hydraulic Systems*. Washington D.C., USA, 2001.

[10] Han, L., Zhao, X. An abrasive wear model between valve plate and cylinder block in axial piston pumps. Book Chapter CSAA/IET International Conference on Aircraft Utility Systems, 2018.

[11] Junhui Zhang, Yining Shen, et al. Tolerance design guideline for piston/cylinder interface of electro-hydrostatic actuator (EHA) pumps based on a Thermal-Fluid-Structure model. Tribology International, 2024, 191, 109208.

[12] Woldman, M. An experimental and theoretical investigation into three-body abrasive wear [PhD Thesis]. University of Twente, 2014.

[13] Koulocheris, D., Stathis, A., et al. Experimental study of the impact of grease particle contaminants on wear and fatigue life of ball bearings. *Engineering Failure Analysis*, 2014, 39: 164-180.
[14] Litwin, W., Kropp, S. Sliding bearings with sintered bronze bush lubricated by contaminated water with solid particles: Theoretical and experimental studies. *Wear*, 2023, 532-533: 205070.
[15] Zhao, X., Zhang, Y. Analysis of the tribological and dynamic performance of textured bearings under contaminated conditions. *Tribology International*, 2023, 187: 108732.
[16] Novak, N., Trajkovski, A., et al. Wear of hydraulic pump with real particles and medium test dust. *Wear*, 2023, 532-533: 205101.
[17] Qi, J., Liu, H., et al. Influences of added sand-dust particles on the tribological performance of graphite-like coating under solid-liquid lubrication. *Tribology International*, 2014, 71: 69-81.
[18] Qi, J., Wang, L., et al. The tribological performance of DLC-based coating under the solid-liquid lubrication system with sand-dust particles. *Wear*, 2013, 297(1-2): 972-985.
[19] Wen, S., Huang, P. *Principles of Tribology*. Wiley Online Library, 2012.
[20] Wang, L., Zhou, X., et al. Effect of sediment particle size on the friction and wear performance of NBR/copper rings under contaminated water lubrication conditions. *Wear*, 2025, 571: 205832.
[21] Baydoun S, Arnaud P, et al. Modelling adhesive wear extension in fretting interfaces: an advection-dispersion-reaction contact oxygenation approach. *Tribology International*, 2020, 151: 106490.
[22] Kusano, Y., Hutchings, I.M., et al. Modelling the entrainment and motion of particles in a gap: Application to abrasive wear. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 2003, 217(6): 427-433.
[23] Adachi, K., Hutchings, I.M. Wear-mode mapping for the micro-scale abrasion test. *Wear*, 2003, 255(1-6): 23-29.
[24] Moore, A.J.W., Tegart, W.J.M. Relation between friction and hardness. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 1952, 212(1110): 452-458.
[25] Trezona, R.I., Allsopp, D.N., et al. Transitions between two-body and three-body abrasive wear: influence of test conditions in the microscale abrasive wear test. *Wear*, 1999, 225-229(Part 1): 205-214.
[26] Williams, J.A., Hyncica, A.M. Abrasive wear in lubricated contacts. *Journal of Physics D: Applied Physics*, 1992, 25(3A): A81-A90.

Biographies



Liangyu Song is a doctoral student in mechanical engineering hydraulics at Zhejiang University. His research interests focus on tribological analysis and predictive maintenance of hydrostatic pumps.



Fei Lyu received the Ph.D. degree in mechatronics engineering from Zhejiang University, Hangzhou, China, in 2022. He is currently a Post-Doctoral Researcher with the State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University and is conducting research exchanges at RWTH Aachen University in Germany. His research interests focus on tribological analysis and predictive maintenance of hydrostatic pumps and motors.



Junhui Zhang received the Ph.D. degree in mechatronics engineering from Zhejiang University, Hangzhou, China, in 2012.He is currently an Associate Professor with the State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University. He has authored or coauthored more than 60 articles indexed by SCI and applied for more than 30 National Invention Patents which are granted. He is supported by the National Science Fund for Excellent Young Scholars. His research interests include high-speed hydraulic pumps/motors and hydraulic robots.



Bing Xu received the Ph.D. degree in fluid power transmission and control from Zhejiang University, Hangzhou, China, in 2001. He is currently a Professor and a Doctoral Tutor with the Institute of Mechatronic Control Engineering and the Director of the State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University. He has authored or coauthored more than 200 journals and conference papers and authorized 49 patents. Dr. Xu is a Chair Professor of the Yangtze River Scholars Program and a Science and Technology Innovation Leader of the Ten Thousand Talent Program.

Hydraulic Oil Temperature in Bridge Circuit and Hydraulic Bent Axis Piston as Load

Komsan Ngamkham ,Kittapas Homraruen, Jeerattikul Kaharn, Kantayut Trebooniti

Authors affiliation, komsan.ng@rmutsb.ac.th

Abstract.

The aim of this research is to study the behavior of and controlling hydraulic oil temperature in bridge circuit with through path bent axis piston motor as load in the test bench. The testing of hydraulic motor efficiency was conducted in accordance with the test circuit for motor unit of ISO 4409. Measurement of inlet flow rate, drainage flow rate, output torque and test fluid temperature were all taken at full rotational speed range of the motor. The load of this circuit used for testing is the hydraulic bent axis motor 160 cc/rev connected with torque sensor 0-10 kNm. The testing of the hydraulic motor started when the load was driven and the piston moved hydraulic oil to flow and pass through the bridge circuit. The valves were checked and circulated into circuit by the proportional pressure control valve. The result showed that a high hydraulic oil flow rate of 160 l/min and high pressure of 150 bar rapidly increased the temperature up to 80 degrees Celsius. The hydraulic oil circulated in this circuit was later separated from the box valve and put into an oil cooler to reduce the temperature. Increasing the oil temperature depends significantly on the variability of the rate of oil flow and pressure. The fan at the cooler driven by a high-speed low torque hydraulic motor was controlled by a proportional relief valve and decentralized controller. The temperature of the hydraulic oil during test should be 33- 60 degrees Celsius while the speed of the hydraulic motor is regulated by input signal from sensor sent to the controller. It was found that the hydraulic bent axis motor can be used to function as load for hydraulic motor testing because the cooling efficiency of the oil greatly regulates the temperature, reduces the degree of the hydraulic bent axis motor wearing and instead increases the lifespan of the oil.

Keywords. Hydraulic oil temperature, bridge circuit, heating generation, charge system, flushing, cooling.

1.INTRODUCTION

Hydraulic motor has been widely used for driving agriculture machinery including rice harvester sugar cane harvester and construction machinery. The transmission of these machinery is purely hydrostatic consisting of an open and close circuit. The repaired hydraulic motor should have high efficiency before being used. The test rig machine built

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

up from outsource can be operated for efficiency evaluating both hydraulic motor and hydraulic pump using the same driving power and reservoir. With regards to hydraulic pump testing, the system is driven by a close-loop three phase electric motor of 75 kW and an electric displacement control (EDC) close-loop pump of 75 cc/rev. The power drive of this machine can be used in testing hydraulic motor too Figure 1. The size of the reservoir is 400 litres. When testing the hydraulic motor, the hydraulic motor is installed in the test rig machine with its gear shaft connected with another gear shaft of the bent axis motor by chain coupling. When testing the dynamic efficiency, the repaired hydraulic motor has to rotate the load which is the bent axis hydraulic motor A2FM of 160 cc/rev. When it was driven it tend to be a pump as well as a load for testing. At port A and B of the pump, the hydraulic oil circulates through within the hydraulic bridge circuit (Graetz circuit) which consist of four cartridge check valves, proportional pressure relief valve (PRV) and charging system as seen in Figure 2.

At the start of testing the hydraulic motor, hydraulic oil ISO VG68 circulated from pump A2FM to block valve or hydraulic bridge circuit. There are two lines; the high pressure line (HPL) and low pressure line (LPL) with different pressure known as delta pressure (Δp). This circuit is similar to hydrostatic transmission whose major elements are pump (A2FM), motor (hydraulic bridge circuit or graetz circuit) and charge system.

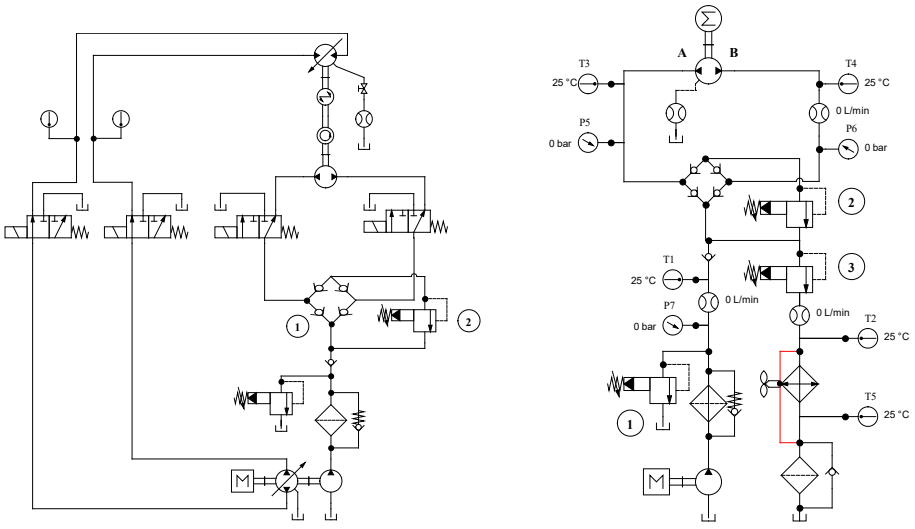
Four check valves together with proportional pressure relief valve are mounted in the block valve, as seen in Figure 2. There are two ports A and B and another port connected to the charge pump and proportional pressure relief valve. This can control the pressure in charge line which corresponds with the flow and circulation of hydraulic oil or speed of load pump. The flow rate of hydraulic oil depends on the speed of rotation of the pump load. When the speed of pump load rotation increases, the flow rate of hydraulic oil also increases, leading to high pressure at proportional pressure relief valve. Also, when the speed of pump load rotation decreases, the flow rate of hydraulic oil decreases, but there will be no corresponding effect on the pressure.

When the loop-line flow rate is high, a pressure drop develops across the pressure control valve, generating a pressure differential and causing hydraulic losses that are converted into heat. The load on the hydraulic motor being tested is increased by increasing the pressure at the proportional pressure control valve. In hydraulic motor efficiency testing, the presence of pressure differentials and high flow rates leads to energy losses that result in elevated heat generation. The hydraulic oil temperature during volumetric efficiency testing of hydraulic pumps and motors should be maintained within the range of 33–60 °C, in accordance with ISO 4409. The hydraulic oil temperature in the low-pressure line during hydraulic motor efficiency testing exceeds 80 °C. This leads to the deterioration of the anti-wear additive performance. As illustrated in the original circuit in Figure 2, the hydraulic oil in the test circuit undergoes no cooling or filtration prior to returning to the reservoir, and only a limited amount returns via the load pump drain line. The reduction of hydraulic oil temperature and the filtration of hydraulic oil are improvements that should be implemented in this test rig. As discussed previously, the load-generating circuit in this efficiency test exhibits characteristics of a closed-loop system and can be improved by incorporating a flushing line to control hydraulic oil temperature and contamination. The improvement analysis of this circuit will be performed by treating it as a closed-loop system, with consideration given to the high-pressure line, low-pressure line, charge pump

pressure and flow, flushing pressure and flow, and the cooling system. In the hydraulic bridge (Graetz) circuit, the flow through the pressure relief valve is unidirectional, and an additional flushing line on the low-pressure side should be connected downstream of the pressure relief valve. The criteria for adjusting the pressure and flow rate of the charge pump, as well as the flushing and cooling flows, are determined based on closed-loop circuit theory. Another proportional pressure relief valve dedicated to flushing is added to the low-pressure line downstream of the load-control proportional pressure relief valve, in order to control the pressure to match the charge pump pressure. Hydraulic oil temperature control within the specified range is achieved using a LabJack T7 Pro, which is employed to monitor key variables such as charge pressure, charge flow, flushing flow, and cooling temperature, and has been found to operate effectively.



Figure 1. Cross-sectional view of the repaired hydraulic motor installed in the test rig and connected as a load motor.



(a) (b)

Figure 2. (a) The original hydraulic circuit used in the hydraulic motor test rig (b) Modified hydraulic circuit incorporating flushing and cooling for the hydraulic motor test rig.

2. The temperature behaviours of the hydraulic oil in the load circuit used for hydraulic motor testing

In the investigation of temperature behaviours in the load-simulation system for a hydraulic motor, the system is treated as an integrated system comprising an A2FM hydraulic motor, a hydraulic bridge circuit, and a charge system. Measurements and data logging are carried out using the HMG 4000 data recorder and HYDAC pressure sensors and temperature sensor. Following testing conducted under ISO 4409 conditions to gain an understanding of temperature behaviours, improvements were then made to the circuit and control system and the hydraulic oil used in the test is ISO VG 68. To study the temperature behaviour caused by heat generation under variable pressure with constant flow rate and variable flow rate with constant pressure, the hydraulic system was operated in the configuration shown in Figure2. (a)and(b). The experimental results indicate that the thermal behaviour is governed by the fundamental energy balance equation

$$mc_p \frac{dT}{dt} = Q_{gen} - Q_{loss} \qquad (2.1)$$

During the experiments shown in Figures 4 and 5, the hydraulic system operated without active cooling, and the oil exchange with the reservoir was limited. Under these conditions, heat dissipation to the surroundings was relatively small, particularly during each constant operating stage. Therefore, the heat loss term Q_{loss} can be considered negligible compared to the generated heat, and the equation can be approximated

$$mc_p \frac{dT}{dt} \approx Q_{gen} \qquad (2.2)$$

The heat generation in a hydraulic system resulting from pressure losses can be expressed as follows

$$Q_{gen} = \Delta pQ \qquad (2.3)$$

The equation is applied to assess heat generation due to pressure losses arising from pump motor losses, throttling valve losses, and flushing relief losses that is employed to calculate the heat generation within the load-simulation loop of the test rig.

2.1 Hydraulic motor testing under no-load conditions.

In this test, a variable-displacement motor is employed as the load driver. Operating at 700 rpm, it produces a circuit flow rate of 130 L/min, with pressures of 27 bar in the high-pressure line and 6.6 bar in the low-pressure line, resulting in a pressure differential of approximately 20 bar. In the no-load test, the temperature rises linearly with a constant slope from 28 °C to 93.2 °C within 10 minutes at the low-pressure line, while the temperature at the high-pressure line reaches 90.7 °C, yielding a temperature difference of 2.5 °C. Although no load pressure is applied, the temperature increases to 93 °C due to the pressure differential between the inlet and outlet of the hydraulic bridge circuit, where the

associated losses are converted into heat. Heat generation results from energy losses caused by flow resistance in drilled orifices, pressure drops across two check valves and the pressure relief valve, as well as valve throttling.

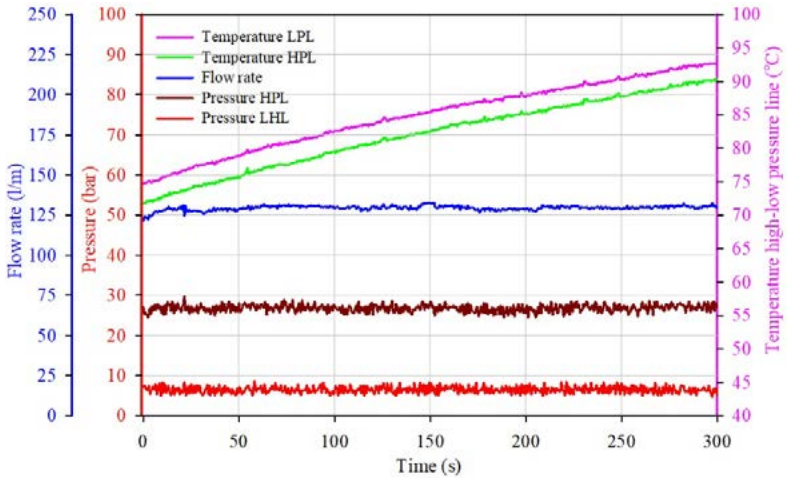


Figure 3. Oil temperature in the high-pressure and low-pressure lines at constant pressure and constant flow rate.

2.2 Hydraulic motor testing under constant pressure with increasing flow rate.

Hydraulic motor testing is performed under multiple operating conditions depending on the motor type and size, since different displacements result in different load capacities. Accordingly, testing is carried out under high-pressure, low-flow-rate and low-pressure, high-flow-rate conditions, and the temperature response due to heat generation under non-cooled conditions is presented in Figure 4. In the initial stage, the load condition was simulated at an operating pressure of 104 bar for 94 s. The pressure differential between the high-pressure and low-pressure lines was 86 bar, and the flow rate was 26 L/min, yielding a heat generation rate of 3.734 kW. In the subsequent stage, under a comparable operating pressure of 109 bar for 64 s, the pressure differential was 89 bar. With the flow rate increased to 42 L/min, the resulting heat generation was 6.242 kW. In the final stage of the graph, under an operating pressure of 101 bar for 10 s, a pressure differential of 81 bar and an increased flow rate of 62 L/min result in a heat generation of 8.386 kW. Despite the short duration, the high heat generation leads to a linear temperature rise with a pronounced slope. Although the pressure differential remains at a comparable level across the three stages, the increase in flow rate leads to a significant rise in heat generation from 3.73 to 8.39 kW, resulting in a linear temperature increase with a steeper slope, even over a short operating duration.

2.3 Hydraulic motor testing under constant flow rate with increasing pressure.

In this test, the initial flow rate was set to 47 L/min, while the hydraulic oil temperature was maintained in the range of 43.5–45.5 °C.

Table 2.1 Heat generation under constant pressure with increasing flow rate.

Stage	Δp HPL–LPL (bar)	Flow (L/min)	Time (s)	Heat generation (kW)
1	86	26	94	3.73
2	89	42	64	6.24
3	81	62	10	8.39

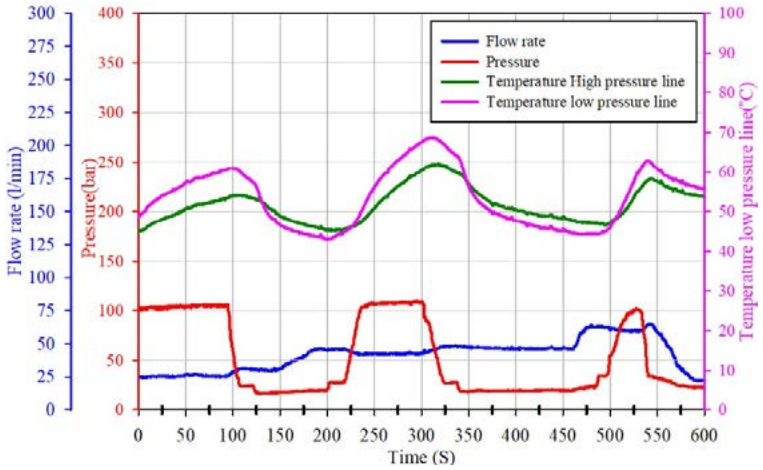


Figure 4. Oil temperature in the high-pressure and low-pressure lines under constant pressure with increasing flow rate.

During the first stage, the simulated load pressure was increased to 53.6 bar for 53 s, producing a pressure differential of 34.86 bar between the high-pressure and low-pressure lines and a corresponding heat generation of 2.74 kW. In the subsequent stage, the simulated load pressure was increased to 79 bar for 87 s. This produced a pressure differential of 61.3 bar between the high-pressure and low-pressure lines, resulting in a heat generation of 4.483 kW. In the final pressure stage, the simulated load pressure was increased to 105.5 bar for 88 s, producing a pressure differential of 88.4 bar between the high-pressure and low-pressure lines and a corresponding heat generation of 6.274 kW. During this period, the hydraulic oil temperature rose continuously to the range of 67.3–74.9 °C, as shown in Figure 5.

Table 2.2 Heat generate constant flow rate with varying pressure.

Stage	Δp HPL–LPL (bar)	Flow (L/min)	Time (s)	Heat generation (kW)
1	34.86	44.4	53	2.58
2	61.3	43.8	87	4.48
3	88.4	42.5	88	6.27

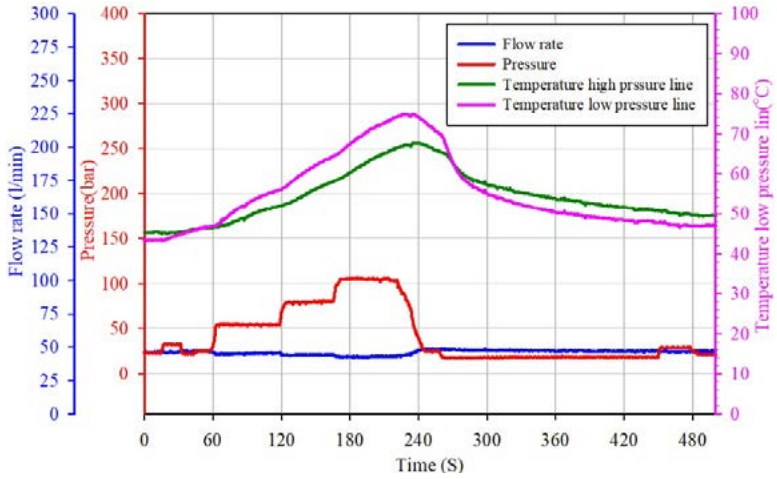


Figure 5. Oil temperature in the high-pressure and low-pressure lines at constant flow rate with varying pressure.

To study the temperature behaviour caused by heat generation under vary pressure with constant flow rate and vary flow rate with constant pressure, the hydraulic system was operated in the configuration shown in Figure 2(a). The experimental results indicate that the thermal behaviour is governed by the fundamental energy balance equation

2.4 Hydraulic motor testing under constant flow rate with increasing pressure.

As described in the preceding paragraph, the operating principle of the load-simulation circuit used for hydraulic motor testing is analogous to that of a hydrostatic closed-loop drive. The main distinction is that the hydraulic motor test circuit lacks a flushing system, whose function is to extract a portion of the high-temperature oil for cooling and filtration prior to its return to the reservoir or reintroduction into the loop. In response to the excessive hydraulic oil temperature in the loop circuit during testing, and based on the heat generation analysis revealing significant thermal generation over short operating durations, the test circuit was modified by incorporating a proportional pressure relief valve, an air cooler, a filtration unit, and a control system, as illustrated in Figure 2(b). The hydraulic oil from this section is returned to the reservoir**.** The modified circuit is installed downstream of the load proportional pressure relief valve on the low-pressure line of the loop.The parameter tuning is carried out in accordance with closed-loop circuit principles. Based on the original circuit configuration, the line equipped with proportional pressure relief valve No. 1 functions as the charge line supplying hydraulic oil to the loop. The charged oil flow rate is governed by the pressure setting of valve No. 1, whereas the drain flow from the load is negligible relative to the main loop flow. As a result, the charge pump is unable to supply sufficient oil to the loop. In accordance with closed-loop circuit theory, the charge flow rate should generally be maintained at approximately 20–30% of the main loop flow. This constitutes an additional cause of the heat generation observed in the previous graph.

To examine the charge flow supplied to the main loop, a test was performed as illustrated in Figure 6. Under a main loop flow rate of 28 L/min and a load pressure of 200 bar applied for 31 s, with the oil temperature rising from 33 °C to 57.3 °C and a charge

pressure of 17 bar, the charge flow into the loop was limited to 0.89 L/min, representing only 4.45% of the main loop flow. Based on the test results, once the operating time exceeds 150 s, and at a charge pressure of 17 bar, the charge pump supplies hydraulic oil to the main loop at a flow rate of 7.25 L/min. This leads to a convergence of the cooling temperatures in the high-pressure and low-pressure lines, which is advantageous because it reduces thermal accumulation within the system.

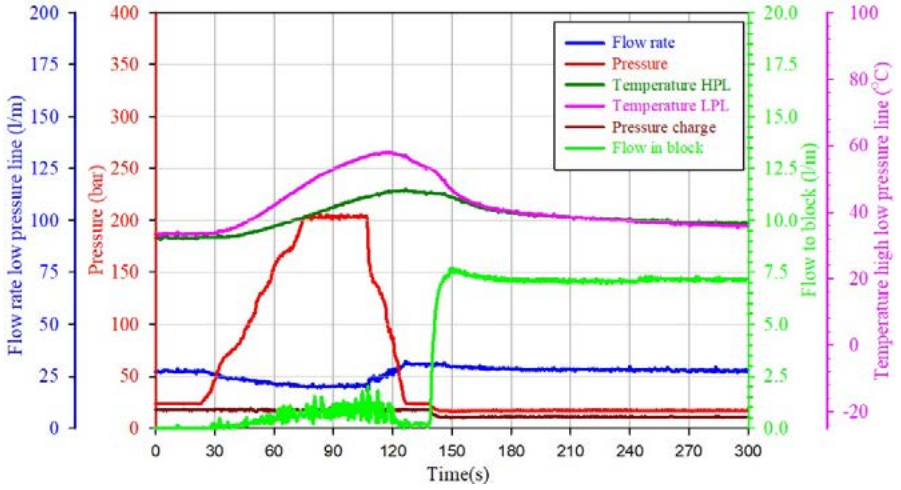


Figure 6. Measured charge flow rate of hydraulic oil supplied to the low-pressure line during testing.

From the study of heat generation in the load-simulation circuit, which is similar to a closed-loop system in mobile machinery, the pressure differential between the inlet and outlet of components leads to energy losses that are converted into heat. The heat generated in this system causes degradation of the hydraulic oil and subsequently leads to failure of hydraulic components. As a corrective measure, the circuit in Figure 2(a) should be modified to allow charge oil supply to the low-pressure line by installing a flushing line downstream of pressure relief valve No. 2, as illustrated in Figure 2(b). In the modified circuit, two proportional pressure relief valves are employed. Valve 1 is used to control the pressure of the hydraulic oil charged into the loop, whereas Valve 2 regulates the flushing line pressure, directing the oil through the cooler and filtration unit before returning it to the tank. In a closed-loop circuit, the charge pump pressure is adjusted to be slightly higher than the flushing line pressure. Both proportional pressure relief valves can be programmed and calculated based on Equation (2.1) to control the hydraulic oil temperature within the range of 33–60 °C under the following conditions.The equation used to control the proportional pressure relief valve.

$$\frac{\mathrm{dT}}{\mathrm{dt}} = \frac{\mathrm{Q_{gen}}}{mc_p} - \frac{hA}{mc_p}(T_{average} - T_{charge}) \qquad (2.4)$$

$$\frac{\mathrm{dT}}{\mathrm{dt}} = 0 \qquad (2.5)$$

$$\frac{\mathrm{dT}}{\mathrm{dt}} > 0 \qquad (2.6)$$

$$\frac{dT}{dt} < 0 \qquad (2.7)$$

According to the conditions specified in Equations (2.5)–(2.7), the program can regulate the oil temperature, as demonstrated by the temperature results presented in Figure 7.

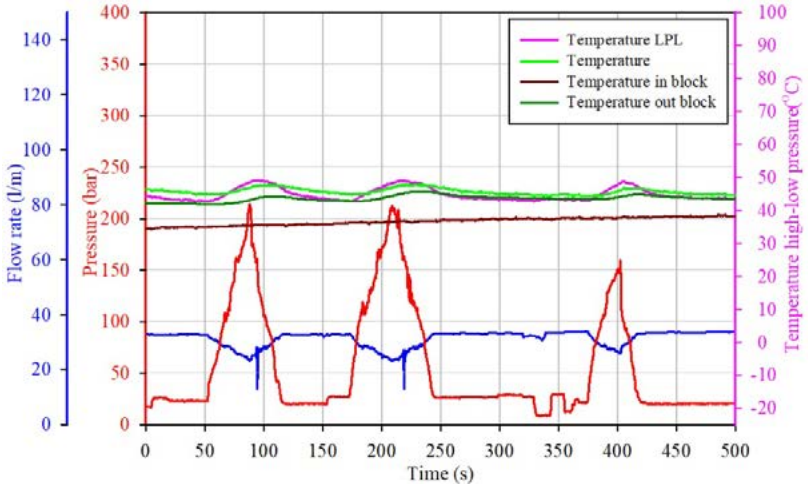


Figure 7. Temperature control achieved through the implementation of a flushing line and regulation of the charge pressure.

3. Conclusion

During no-load motor testing, it was observed that hydraulic oil still flows from the pump through the valve block, resulting in a continuous increase in temperature. This occurs because the high flow rate causes a significant pressure drop across the proportional pressure relief valve, leading to energy losses that are converted into heat.

Operating characteristics under high-pressure load and low-speed(low flow rate) conditions include a high pressure differential in the loop (high Δp) and a low flow rate (low Q). Under these conditions, heat generation is dominated by pressure-induced losses, resulting in a high heat generation per unit flow. Consequently, the loop temperature rises rapidly. High-pressure/low-speed operation therefore represents the worst thermal condition for a hydrostatic drive system.

Operating characteristics under high-speed and low-load conditions include a low pressure level in the loop and a high flow rate. In this regime, the pressure differential (Δp) is low, resulting in relatively small pressure-related losses. Internal leakage per unit time increases slightly, while viscous losses increase with speed. Overall, heat generation remains moderate. The high oil circulation rate enhances heat dissipation, leading to relatively stable loop temperatures with a small temperature difference between the loop and the tank. Flushing has a less pronounced effect compared to heavy-load conditions. In summary, high-speed/low-load operation represents a thermally stable condition.

The addition of flushing and programmed feedback control enables the mitigation of high hydraulic oil temperatures, particularly because hydraulic motor testing involves motors of various sizes and operating functions.

4. REFERENCES

[1] Kwon, H., Keller, N., Ivantysynova, M., Thermal Management of Open and Closed Circuit Hydraulic Hybrids – A Comparison Study,11th IFK, Aachen, Germany, March 19–21, 2018.

[2] Chao Wu, Cong Xu and etc, 'Heat analysis in constant-pressure hydraulic system base on energy analysis', 1st International global on renewable energy and development (IGRED)., 2017.

[3] Q. Zhang, X. Liu, Z. Tong, and Z. Cheng, "Research on heat generation law and cooling system performance of hydraulic system of combined machine tool," Energies, vol. 16, no. 21, Art. no. 7322, 2023.

[4] Minav, T., Papini, L., Pietola, M., A Thermal Analysis of Direct Driven Hydraulics, 10th International Fluid Power Conference (IFK), Dresden, Germany, 2016.

[5] Guo, X., Vacca, A., Advanced Design and Optimal Sizing of Hydrostatic Transmission Systems, Actuators, vol. 10, Art. no. 243, 2021.

[6] Carmen Debeleac, 'Simulation of the Working Cycle for a Hydraulic Press Equipped with a Flow Regulator Mounted in Graetz System', Hidraulica – Magazine of Hydraulics, Pneumatics, Tribology, Ecology, Sensorics, Mechatronics, no. 3, pp. 7–12, 2022.

[7] Fan, Q., Zhang, J., Li, R., Fan, T., Review of Research on Hydrostatic Transmission Systems and Control Strategies, Processes, vol. 13, Art. no. 317, 2025.

[8] Zhang, Q., Liu, X., Tong, Z., Cheng, Z., Research on Heat Generation Law and Cooling System Performance of Hydraulic System of Combined Machine Tool, Energies, vol. 16, art. no. 7322, 2023.

Biographies



Komsan Ngamkham received the bachelor's degree in industrial technology in material handling technology in 1998, the master's degree in mechanical engineering Srinakharinwirot University in 2009, and the doctor of engineering in Mechanical Engineering from The Sirindhon International Thai-German graduate school of engineering (TGGS) King Mongkut's University of Technology North Bangkok Thailand in 2015, respectively. He is currently working as an Assistant Professor at the Department of Mechanical Engineering, Faculty of Industrial Education, Rajamangala University of Technology Suvarnabhumi. His research areas include mobile efficiency.

High-Resolution Force Control in a Multi-Pressure Hydraulic System for Excavator Work Functions

Mohamed Allam[1], Mikko Huova[1], Kim Heybroek[2]

[1]Tampere University, Tampere, Finland, *mohamed.allam@tuni.fi*, *mikko.huova@tuni.fi*

[2]Volvo Construction Equipment, Eskilstuna, Sweden, *kim.heybroek@volvo.com*

Abstract

Conventional centralized throttling-based hydraulic systems in excavators are robust but inefficient, especially under partial loads and multi-actuator operation. To address the inefficiencies, this work presents a multi-pressure hydraulic system for a 20-ton excavator and the implementation of its control. The system employs four discrete pressure levels, maintained by accumulators and a displacement-controlled pump, and distributed to the Swing, Boom, Arm, and Bucket actuators via 2/2 proportional valves. This enables more efficient use of energy but requires the development of a controller to select the pressure level, meter the flow and maintain machine performance. A force-based control strategy is developed that can manage these tasks while remaining relatively simple for practical implementation. The proposed controller combines discrete pressure levels with continuous modulation yielding near-continuous force resolution. Simulation in MATLAB/Simulink demonstrates 67% reduction in energy consumption against the original machine's load sensing system, and a 32% reduction against a benchmark controller developed previously for the multi-pressure system. Results also showed accurate tracking and preserved actuator dynamics.

Keywords: Mobile Machines, Hybrid Systems, Multi-Pressure System, Excavators, Control, Energy Efficiency, Force Control, Novel System.

1. INTRODUCTION

Hydraulic systems are extensively employed in heavy machinery due to their high-power density and robust performance. However, traditional hydraulic architectures, particularly those relying on throttling-based flow control, often exhibit energy losses. These conventional systems lack regeneration capabilities, differential flow control, and independent metering, resulting in inefficiencies. Recent research has focused on enhancing the energy efficiency of hydraulic work functions through various methods, including hydraulic transformers [1], multi-chamber cylinders [2], displacement control [3], hybrid technologies [4], and electro-hydraulic actuators [5]. Despite demonstrating significant energy savings, none of these approaches has yet seen widespread industrial adoption. While these solutions hold promise for future mobile machinery applications, this study pursues a multi-pressure system (MPS) approach. In an MPS, the actuator is connected to a discrete number of pressure levels achieving several different control modes that can be increased tremendously when coupled with proportional valves which requires the development of an appropriate control algorithm. The controller needs to accomplish different tasks

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non-Commercial 4.0 International

simultaneously such as selection of pressure levels, modulation of proportional valves and management of mode switching.

Several authors have addressed the challenge of discrete multi-pressure or multi-chamber system control. For instance, Heybroek [6] applied model predictive control to balance force peaks and energy efficiency. Hansen proposed a force switching algorithm for wave energy converters [7]. Huova also presented controllers for a multi-pressure actuator [8] and also tackled a model predictive control for a MPS in [9]. Xu also presents an actuator controller for a Multi-pressure multi-chamber cylinder system, which handles the pressure trajectories during transients using a MPC [10]. Building upon the insights presented in these research, the force-based controller developed in this work integrates concepts from Linjama's [2] work on secondary control along Hansen's [11] effort in reducing pressure oscillations to yield a practical and implementable control solution for the MPS.

This paper presents the development and evaluation of a force-based control strategy aimed at maximizing the energy-saving potential of the MPS. The proposed controller uses velocity feedback to generate a reference force needed to accelerate the piston and dynamically selects the chamber pressure levels maintaining controllability and reducing losses. The algorithm generates valve commands for both inflow and outflow actuator valves to satisfy operator inputs. This control approach also effectively mitigates force peaks caused by the different pressure dynamics during a mode switch in individual actuator chambers, as detailed in the control algorithm section. The proposed controller is benchmarked against a previously developed concept and the original machine's system using simulations.

2. SYSTEM ARCHITECTURE

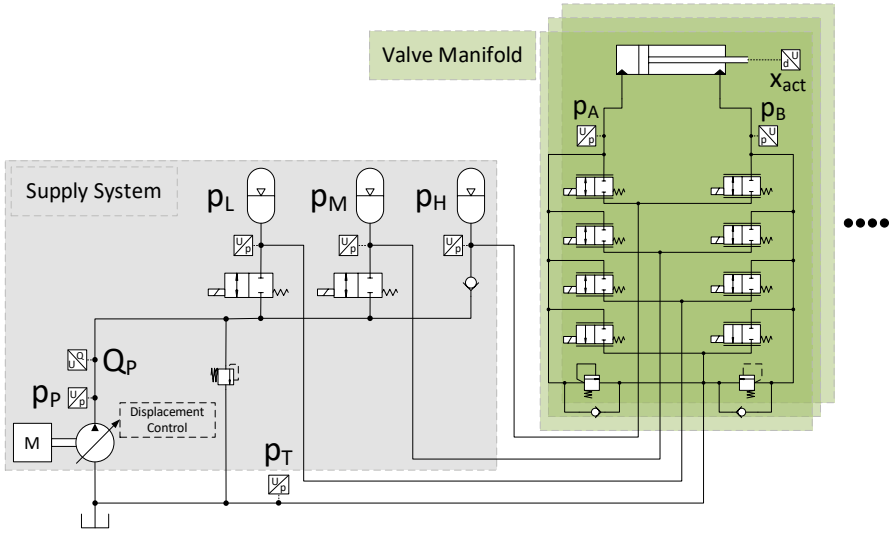


Figure 2.1: Simplified hydraulic circuit diagram of the Multi-Pressure System detailing the supply system and valve manifolds

The MPS illustrated in Figure 2.1. consists of three hydraulic accumulators operating at different pressure levels $[p_L, p_M, p_H]$ and a tank line p_T which establishes the system's multi-pressure supply lines. These lines are connected to the excavator's hydraulic work functions

[Swing, Boom, Arm, Bucket] via valve manifolds equipped with sets of 2/2 proportional bidirectional valves. Each valve modulates flow into and out of the actuator chambers by connecting to the appropriate pressure line and controlling flow metering in both piston and rod chambers. Pressure sensors on both actuator ports and a piston position sensor deliver feedback signals essential to the valve controller functionality. This feedback enables the control of valve commands to achieve desired actuator force outputs. The accumulator pressure levels are maintained between set thresholds by the excavator's hydraulic supply system, comprising of a diesel engine coupled to a variable-displacement piston pump. The pump output is regulated to charge the accumulators through dedicated charge valves.

3. CONTROL ALGORITHM

In the proposed MPS each actuator manifold contains four proportional bi-directional valves at the piston side and four at the rod side each connected to a pressure level. The role of the control algorithm is to choose the appropriate pressure level and actuate its corresponding valve according to the system requirements and load conditions. This is accomplished by using signals from the chamber $[p_A, p_B]$ and supply line pressures $[p_T, p_L, p_M, p_H]$, the cylinder position x as well as the joystick signal from the operator v_{ref} and the piston velocity v which can be estimated from the position sensor. The next section presents how the controller is formulated and how each decision is made, and the overall flow of the algorithm is shown in Figure 3.1.

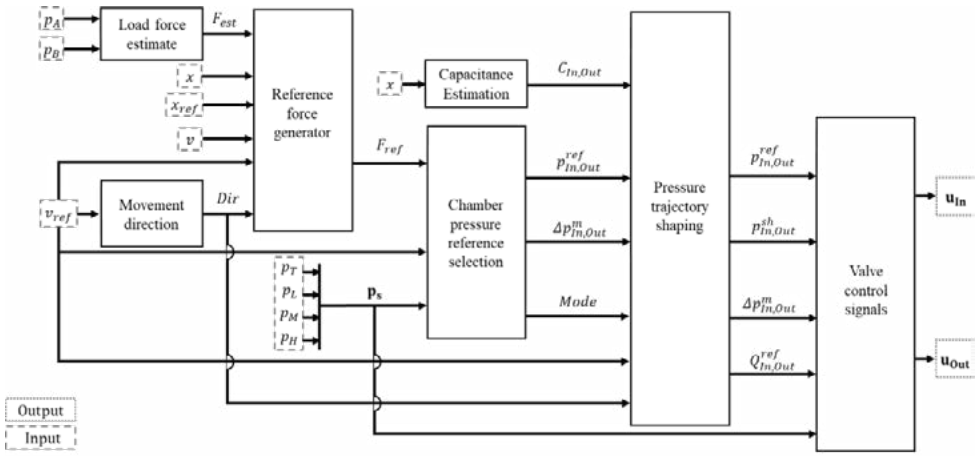


Figure 3.1: Force control algorithm - overall flow

The main idea is to generate a reference force F_{ref} based on the velocity error between the velocity reference v_{ref} (which could be e.g. a joystick command) and the measured velocity estimated from the piston position x (in case of a cylinder actuator). Following the force reference requires high resolution of achievable forces which can be accomplished by on/off-type control of the valves, for example, as demonstrated in Linjama [2]. However, when using two-chamber cylinders and four discrete pressure levels, the resolution obtained in this way is not sufficient for practical operation. To overcome this limitation, the proportional control valves employed are used to introduce a pressure differential in each of

the chambers, thereby enabling finer adjustment of the actuator forces and significantly improving the achievable force resolution.

The pressure selection chooses chamber pressure references for the inflow and outflow sides $p_{In,Out}^{ref}$ that match the generated reference force. The chamber reference selection is done through a cost function that minimizes force error, unnecessary variation of chamber pressures and power losses.

In the next step, the valve commands are selected such that the chamber pressure references are achieved. The references are generated in two different ways depending on the operating condition. During regular motion, when no pressure transient is needed i.e., during a pressure shift, it is kept as the selected pressure reference, whereas during a pressure shift where the selected lines are changed, it follows a predefined pressure trajectory $p_{In,Out}^{Sh}$. In both cases, the valve commands are determined based on the required flow rate and the calculated pressure differential, which accounts for both the supply pressure and the chamber pressure reference. The required flow rate is obtained from the piston velocity, the derivative of the chamber pressure reference, and the estimated hydraulic capacitance. This is done in order to avoid force peaks originating from different pressure dynamics in the piston and rod side chambers during a mode switch. The details of each step are illustrated in the next sections.

3.1. Movement direction

The first step is determining the movement direction of the actuator's piston. This is done using the velocity reference generated by the operator. This step is necessary to assign the variables of the controller to the inflow and outflow sides. Where the variable Dir is calculated using:

$$Dir = \begin{cases} 1 & If \begin{cases} v_{ref} \geq v_{start}\ OR \\ v_{ref} > v_{stop}\ AND\ Dir^{-1} = 1 \end{cases} \\ -1 & If \begin{cases} v_{ref} \leq -v_{start}\ OR \\ v_{ref} < -v_{stop}\ AND\ Dir^{-1} = 1 \end{cases} \\ 0 & else \end{cases} \tag{1}$$

The logic includes hysteresis to avoid repetitive initiation of motion, where v_{start} and v_{stop} are velocity thresholds to start and stop the movement and Dir^{-1} is the direction variable in the previous timestep. The parameters, inputs and outputs of the control algorithm are assigned based on the value of the direction variable to the inflow and outflow sides. The logic is used throughout the controller, and an example is shown for the effective areas of the chambers for the inflow A_{In} and the outflow A_{Out} where:

$$A_{In} = \begin{cases} A_A & Dir \geq 0 \\ A_B & Dir < 0 \end{cases}; A_{Out} = \begin{cases} A_B & Dir \geq 0 \\ A_A & Dir < 0 \end{cases} \tag{2}$$

where A_A and A_B are the chambers effective areas on the piston and rod sides.

3.2. Reference force generator

The reference force is generated using a PI – controller based on the velocity error. An outer position loop is added to the existing velocity controller to correct for position error where the corrected reference velocity v_{ref_x} is calculated using:

$$v_{ref_x} = \big(x_{ref}(t) - x(t)\big)K_{px} + v_{ref} \tag{3}$$

where x_{ref} is the position reference, K_{px} is the gain for position loop. The reference force F_{ref} is generated using:

$$F_{ref} = \big(v_{ref_x} - v\big)K_P + K_I \int \big(v_{ref_x} - v\big)dt \tag{4}$$

K_p and K_i are the proportional and integral gains respectively and v is the measured piston velocity. An important note, in operating intervals where no velocity command is provided by the operator, i.e. *Dir* equals zero, the reference force is gradually driven toward an estimated hydraulic force F_{est}, computed from the measured chamber pressures as:

$$F_{est} = p_A A_A - p_B A_B \tag{5}$$

The motivation for this is to prevent delays or wrong movement directions when valve actuation starts. By aligning the reference force with the estimated force during inactive periods, the system ensures a smooth transition between inactive and active phases of operation, thereby improving responsiveness and allowing for smoother operation.

3.3. Chamber pressures selection

The principle of pressure selection is to match the reference force by choosing appropriate chamber pressures from a set of virtual levels. Since there is a limited number of accumulators and they cannot provide arbitrary pressure values, intermediate pressure levels are virtually generated by linearly interpolating between the current accumulator pressures. These "virtual" levels represent pressures that can be realized in practice through throttling across the metering valves for the inflow and outflow sides. The available accumulator pressures $\mathbf{p_s}$, is defined as:

$$\mathbf{p_s} = [p_T, p_L, p_M, p_H] \tag{6}$$

These serve as the boundary values for constructing virtual multi-level buses along with the minimum and maximum system pressures. This produces a denser grid of chamber pressures for the inflow and outflow sides using the constructed vector $\mathbf{p_{virt}}$. The vector expands the resolution of the achievable force levels beyond the discrete accumulator values while preserving physical feasibility. A combination matrix $\mathbf{M}$ is formed based on the number of virtual levels n created where:

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & \dots & 1 & 2 & 2 & \cdots & n \\ 1 & 2 & \dots & n & 1 & 2 & \cdots & n \end{bmatrix} \tag{7}$$

where the matrix $\mathbf{M}$ has the size $2x n^2$. Pressure references are created using the vectors $\mathbf{p_{In}^{ref}}$, $\mathbf{p_{Out}^{ref}}$ and $\mathbf{p_{virt}}$:

$$\begin{aligned} \mathbf{p_{In}^{ref}} &= \mathbf{p_{virt}}\big(\mathbf{M}(1,:)\big) \\ \mathbf{p_{Out}^{ref}} &= \mathbf{p_{virt}}\big(\mathbf{M}(2,:)\big) \end{aligned} \tag{8}$$

For each candidate pair of inflow and outflow side pressures an achievable actuator force **F** is calculated where:

$$\mathbf{F} = \mathbf{p_{In}^{ref}} \mathrm{A_{in}} - \mathbf{p_{Out}^{ref}} \mathrm{A_{Out}} \tag{9}$$

where A_{in} and A_{Out} are the effective areas of the inflow and outflow sides dictated by the Dir variable.

To calculate the static power losses for the cost function each candidate is paired with an accumulator that can meter the flow in or out of the chamber depending on the direction of motion. For the inflow side an accumulator with higher or equal pressure to the virtual chamber pressure reference is chosen and for the outflow side an accumulator with lower or equal pressure is chosen. This creates the bus matrices $\mathbf{p_{In}^{Bus}}, \mathbf{p_{Out}^{Bus}}$. The two vectors can be used to calculate the throttle free force and consequently the static power loss $\mathbf{P_{loss}}$ where:

$$\mathbf{P_{loss}} = \left(\left(\mathbf{p_{In}^{Bus}} A_{In} - \mathbf{p_{Out}^{Bus}} A_{Out}\right) - \mathbf{F}\right) * |v_{ref}| \tag{10}$$

On the other hand, not all chamber pressures are hydraulically realizable. Therefore, a feasibility check is performed to eliminate the unfeasible pairs where:

$$\begin{aligned} \mathbf{feas} = \ & \mathbf{p_{In}^{Bus}} - \mathbf{p_{In}^{ref}} \geq \Delta p_{In}^{m} \text{ AND} \\ & \mathbf{p_{In}^{Bus}} > 0 \text{ AND} \\ & \mathbf{p_{Out}^{ref}} - \mathbf{p_{Out}^{Bus}} \geq \Delta p_{Out}^{m} \end{aligned} \tag{11}$$

Where Δp_{In}^{m} and Δp_{Out}^{m} are the minimum differential pressure across the valve to allow for metering and is calculated using the valve orifice equation:

$$\Delta p_{In,Out}^{m} = \left(\frac{A_{In,Out} v_{ref}}{K_{In,Out}^{V}}\right)^2 \tag{12}$$

where $K_{In,Out}^{V}$ is the valves flow coefficient calculated using the nominal flow rate and pressure difference.

A cost function **J** is formulated to choose the reference chamber pressures for the inflow and outflow where:

$$\begin{aligned} \mathbf{J} = \left|\mathbf{F} - F_{ref}\right| + W_{prs}\left(\left|p_{In}^{ref^{-}} - \mathbf{p_{In}^{ref}}\right| A_{In} - \left|p_{Out}^{ref^{-}} - \mathbf{p_{Out}^{ref}}\right| A_{Out}\right) \\ + W_{loss}\mathbf{P_{loss}} + W_{dp}\mathbf{\Delta p_{eq}} \end{aligned} \tag{13}$$

where W_{prs}, W_{loss} and W_{dp} are weighing terms for the static losses, the magnitude of the chamber pressure change and for equalizing the pressure differential at the inflow and outflow sides, $p_{In}^{ref^{-}}$ and $p_{Out}^{ref^{-}}$ are the selected chamber pressure references at the previous time step.

The final term in the cost function attempts to ***equally*** distribute the needed overall pressure differential to both the inflow and outflow sides, this helps in reducing the valve activity and also avoids having high pressure difference in the inflow side and low-pressure difference on the outflow side. The term $\mathbf{\Delta p_{eq}}$ is calculated using:

$$\mathbf{\Delta p_{eq}} = \left|\left(\mathbf{p_{In}^{Bus}} - \mathbf{p_{In}^{ref}}\right) - \left(\mathbf{p_{Out}^{ref}} - \mathbf{p_{Out}^{Bus}}\right)\right| \tag{14}$$

The infeasible combinations are set to infinity and the chosen chamber pressure references for the inflow p_{In}^{ref} and outflow p_{In}^{Out} sides are chosen as:

$$p_{In,Out}^{ref} = \mathbf{p}_{\mathbf{In,Out}}^{\mathbf{ref}}[k] \tag{15}$$

where k is the index of minimizing the cost function. Each selected chamber pressure has a corresponding chosen accumulator that is combined to the $Mode$ variable where:

$$Mode = \left[\mathbf{p}_{\mathbf{In}}^{\mathbf{Bus}}(k)\ \mathbf{p}_{\mathbf{Out}}^{\mathbf{Bus}}(k)\right] \tag{16}$$

The chosen chamber pressure references $\left[p_{In}^{ref}, p_{Out}^{ref}\right]$, and M$ode$ are input to the trajectory shaping algorithm which is explained in the next section.

3.4. Pressure trajectory shaping

Chamber pressure references can change in two ways, which determines how valve commands are generated. For small changes within the chamber that do not require switching accumulators; flow is displacement induced, and valve commands are computed using the calculated pressure differential (supply and chamber reference) and the valve orifice equation. Larger changes, which require one or both chambers to switch accumulators, actuate multiple valves simultaneously can produce a "force peak". Uneven pressure build-up owed to different hydraulic capacitance results in sudden acceleration or deceleration of the piston due to the force peak. To prevent this, valve commands are shaped to make the chamber pressures follow a predefined pressure trajectory. The implemented algorithm is derived from the work in [11].

The idea is to generate a flow rate reference Q_{ref} and consequently a valve opening u_v that corresponds to the pre-defined pressure trajectory. The trajectory used in this work is a third order polynomial which is calculated from:

$$p_{In,Out}^{sh} = -\frac{2\Delta p_{In,Out}}{\tau_{Sw}^3}t^3 + \frac{3\Delta \mathrm{p}_{In,Out}}{\tau_{Sw}^2}t^2 + p_{In,Out}^0 \tag{17}$$

where $\Delta p_{In,Out}$ can be defined as:

$$\Delta p_{In,Out} = p_{In,Out}^{ref} - p_{In,Out}^0 \tag{18}$$

and p_0 is the chamber pressure at the beginning of the shift, t is the current time step and τ is the time dedicated for the pressure switch to be completed.

To calculate the flow rate reference needed to complete the pressure switch for both sides, the pressure build-up equation can be used after re-arranging for Q_{In}^{ref} where:

$$\begin{aligned} Q_{In}^{ref} &= A_{In}v_{ref} + C_{In}\left(-\frac{6\Delta p_{In}}{\tau_{Sw}^3}t^2 + \frac{6\Delta p_{In}}{\tau_{Sw}^3}t\right) \\ Q_{Out}^{ref} &= -A_{Out}v_{ref} + C_{Out}\left(-\frac{6\Delta p_{Out}}{\tau_{Sw}^3}t^2 + \frac{6\Delta p_{Out}}{\tau_{Sw}^3}t\right) \end{aligned} \tag{19}$$

where C_{In} and C_{Out} are the hydraulic capacitances estimated from the piston position [12].

It is important to note that the pressure shaping algorithm is only active at times where there is a mode switch occurring dictated by the variable $Mode$ changing values. When no mode switching is occurring the valve commands are calculated using the displacement flow reference which is the first part of equation (19) and $p_{In,Out}^{sh}$ are set as $p_{In,Out}^{ref}$.

3.5. Valve actuation logic

The valve reference generation for the inflow and outflow sides is formulated such that the reference flow rates $Q_{In}^{ref}, Q_{Out}^{ref}$ are distributed to the accumulators that are hydraulically feasible. For the inflow side, the accumulator is considered feasible If the pressure differential to p_{In}^{sh} is higher than or equal to Δp_{In}^{m}. The feasible accumulators are sorted by their proximity to p_{In}^{ref} at each time step. The maximum achievable flow from one accumulator is calculated from:

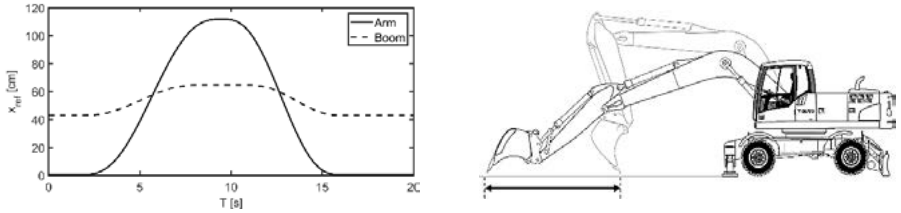
$$Q_{In,Out}^{max} = \begin{cases} K_{In}^{V}\sqrt{\mathbf{p_s}(j) - p_{In}^{sh}} & \mathbf{p_s}(j) - p_{In}^{sh} \geq \Delta p_{In}^{m} \\ -K_{Out}^{V}\sqrt{p_{Out}^{sh} - \mathbf{p_s}(j)} & p_{Out}^{sh} - \mathbf{p_s}(j) \geq \Delta p_{Out}^{m} \end{cases}$$

where j is the index of the feasible accumulator after sorting. The valve command is then calculated using:

$$u_{In,Out}(j) = \frac{Q_{In,Out}^{ref}}{Q_{In,Out}^{max}} \tag{20}$$

and the non-feasible indices are set to zero. If the valve is not fully opened this means that the flow reference is satisfied. However, if the opening is more than one this means that another valve needs to be opened to meet the requirement. This is done by setting the current valve to fully open and the new flow reference to the difference between current valve's maximum flow and the original flow required and calculating the opening using the new reference. The outflow side is calculated in a similar fashion with the main difference being that the p_{Out}^{sh} must be higher than accumulator pressure by Δp_{Out}^{m}. Finally, the inflow and outflow valve commands $\mathbf{u_{In}}$, $\mathbf{u_{Out}}$ are routed to the piston and rods sides using the Dir variable.

4. SIMULATION AND RESULTS



(a) (b)

Figure 4.1: (a) Reference trajectories for the boom and arm actuators (b) Motion of the machine's components relative to the commands

The simulation uses a 20s air grading cycle from the JCMAS standard which is shown in Figure 4.1, where the boom and arm cylinders are extended and retracted simultaneously representing a leveling operation. The cycle was adopted for this simulation study as it is well established in literature and provides a suitable basis for evaluating system performance under multi-actuator operation and high flow demands [13]. It is also based on measurements performed with the original base-line machine performing the same trajectories. The detailed simulation model, cycle description and the previously developed

controller, used here as a benchmark, are presented in [14]. In this work, the focus is on assessing the proposed force-based controller relative to that benchmark as well as the original machine's system which is also detailed in the cited publication. The main parameters used in the simulation model are listed in Table 4.1 while the controller parameters are illustrated in Table 4.2.

Table 4.1: Main parameters for the simulation

Machine mass	21160	kg
Accumulator pressure	[0, 106, 213, 320]	bar
Valve coefficient K_V	$1.088\text{x}10^{-5}$	$\text{m}^3/(\text{s}.\text{Pa}^{1/2})$
Valve delay	30	ms
Valve time constant	10	ms

Table 4.2: Control algorithm parameters

K_P	$[2\text{x}10^5, 1\text{x}10^6]$ *	$\text{N}/(\text{m}/\text{sec})$
K_I	$2.5\text{x}10^6$	$1/\text{sec}$
K_{px}	2	$1/\text{sec}$
v_{start}	$2\text{x}10^{-3}$	m/sec
v_{stop}	$1\text{x}10^{-3}$	m/sec
n	49	ul
W_{loss}	$[0.65, 4]$*	W
W_{prs}	$[0.01, 0.02]$*	ul
W_{dp}	$2.5\text{x}10^{-3}$	Pa
τ_{Sw}	90	ms

Arm, Boom

4.1. Performance

Figure 4.2 and Figure 4.3 show the simulation results for the boom and arm cylinders during the simulated cycle. The first two plots show the position and velocity tracking. The second diagram shows the multi-pressure lines, chamber pressure references, shaped chamber pressures and finally the simulated chamber pressures. Next diagram illustrates the forces including the reference force F_{ref}, selected force (representing the selected pressure references $p^{ref}_{In,Out}$), shaped force (representing the shaped pressure references $p^{sh}_{In,Out}$) and finally the boom force calculated from the chamber areas and pressures. Finally, the valve commands for the piston (A) and rod (B) sides are shown where the colors match those of the pressure lines in the third plot.

Figure 4.2 illustrates the dynamic behavior of the boom actuator, highlighting the performance of the control strategy and MPS. The piston velocity and position closely follow their reference trajectories, indicating that the controller tracks the commanded inputs well. Minor deviations in velocity tracking are observed during both extension and retraction phases attributable to arm motion conditions. However, these do not significantly affect the position tracking of the boom, which remains accurate throughout.

The tracking performance is enabled by the regulation of chamber pressure references to follow the reference force, as shown in the pressure subplot. At the start of motion, when the piston velocity is low, the controller connects piston side chamber to the medium-pressure line and the rod side chamber to the low-pressure line. This configuration allows

the tracking of the desired reference force through throttling on both sides. As the piston velocity increases along with associated power losses the controller transitions piston side chamber to the low-pressure line and the rod side chamber to the tank-line. This mode is maintained throughout the extension phase as it balances between force tracking and static losses. As the piston approaches its target position and decelerates, the controller reduces the output force as needed by increasing the pressure in the rod side chamber while simultaneously decreasing the pressure in piston side chamber. This allows close tracking of the reference force that is decreasing.

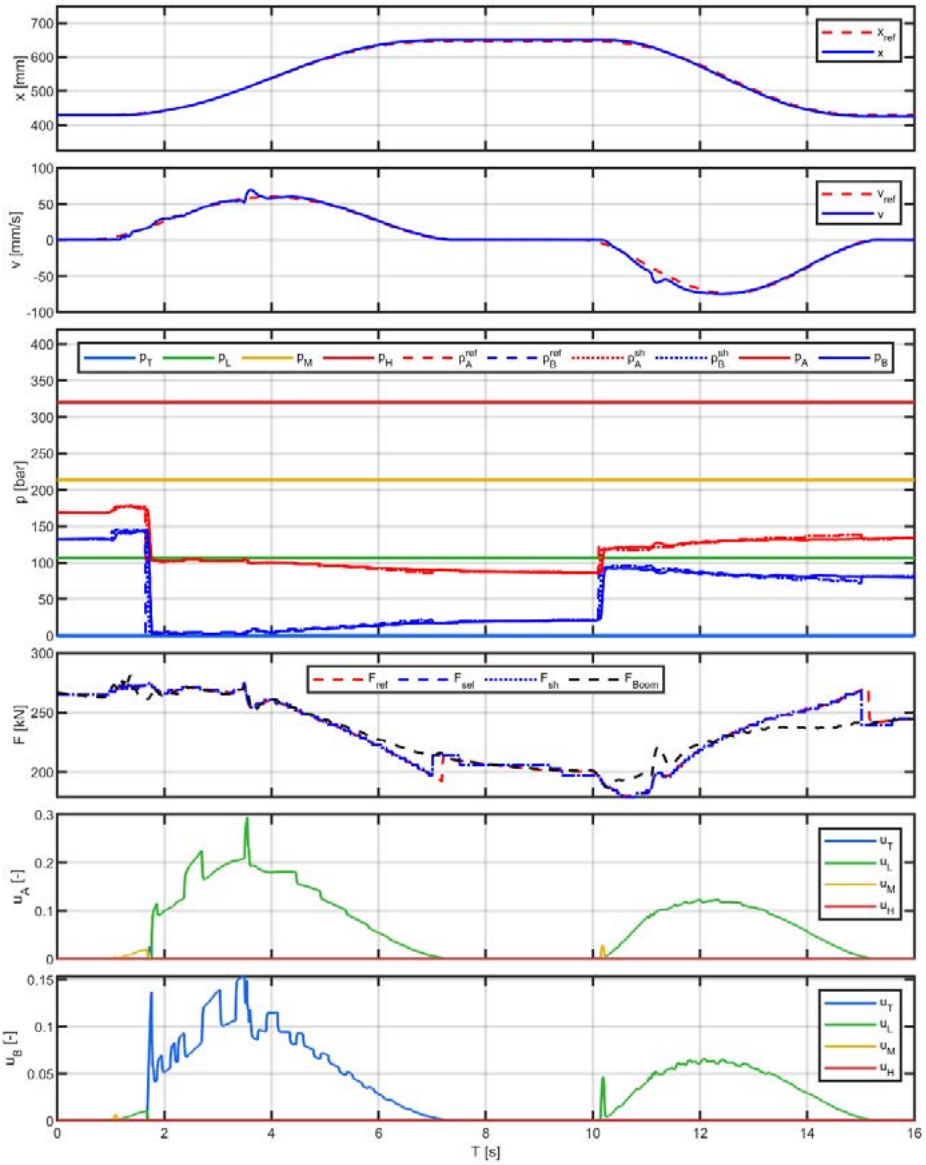


Figure 4.2: Simulation results for the control algorithm of the boom actuator

During retraction, both chambers are connected to the low-pressure line, implementing a differential flow strategy. When the reference force increases, the controller responds by raising the pressure in the piston side chamber and reducing the pressure in the rod side chamber, thereby maintaining effective velocity tracking while reducing the losses. There is an instance where the controller operates at near minimum pressure differential for both chambers at the start of retraction motion contributing to fewer losses and is a core strength of the proposed logic. Additionally, it can be seen that the chamber pressure references change in a stepwise manner at mode switches, but the actual pressures follow the trajectory shaping pressure reference which is detailed more in the upcoming sections.

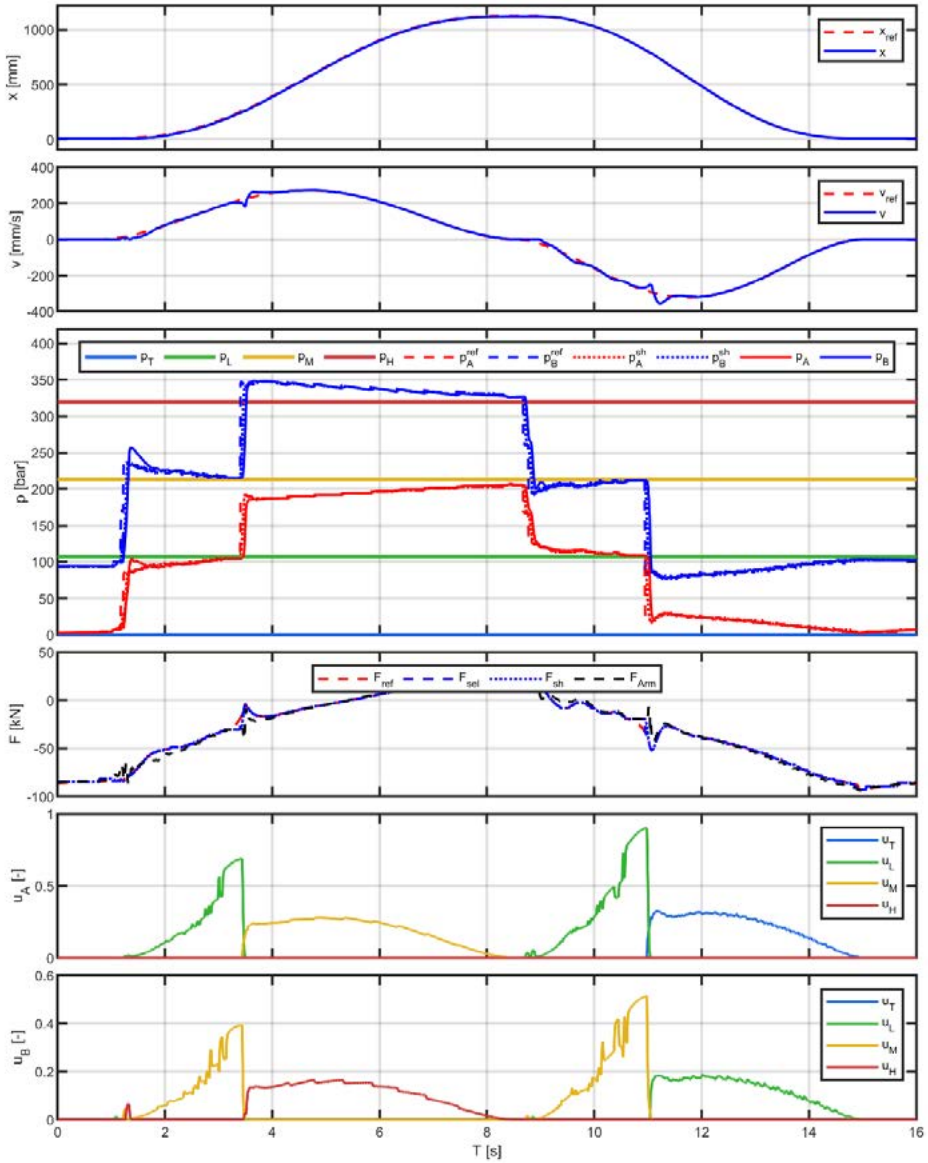


Figure 4.3: Simulation results for the control algorithm of the arm actuator

The arm actuator shows similar results in terms of precise velocity and position tracking with the similar observation of smaller deviations in velocity that are not noticed in position response during motion. The arm actuator chamber pressure references show more activity than the boom actuator mainly due to the force distribution resulting from the available pressure levels and effective areas.

At the beginning, the piston side chamber is connected to low-pressure line while the rod side chamber is connected to medium-pressure line achieving the required negative reference force. Similarly to the boom actuator, the reference force is matched by increasing the pressure in the piston side chamber and reducing pressure in the rod side chamber. The deviation in velocity seen at around 3.5 seconds is due to the algorithm selecting a mode with negative force error that enables reduction of power losses, which is another key strength of the proposed controller in terms of energy savings. As the force error increases the algorithm shifts to reduce the error and introduces throttling as a consequence where it now chooses medium-line for piston side and high-line for rod side. This mode is kept for the rest of the duration of extension. The reference force is again matched by manipulating the chamber pressure references for both edges.

During retraction, the controller initially selects low-line for the piston side chamber and high-line for the rod side chamber. As the losses increase quickly the chambers are switched to a lower pressure reference using the same line for the piston side chamber and switching to medium-line for the rod side chamber, which yields lower losses. The mode remains feasible and highly efficient as it has near minimum pressure differential. The force reference is again matched by regulating both chamber pressures until just before 11 seconds, where it becomes infeasible. At this instance, some force error is allowed by the controller to reduce losses before switching to a different mode which can be seen in the velocity tracking (piston slowing down then speeding back up). To reduce the force error, the control switches to a different mode that can match the reference at the cost of power losses. The chosen mode is tank for the piston side chamber and low for the rod side chamber. Similarly to the end of the extension stroke the mode is preserved and the force reference is matched by increasing the pressure in the rod side chamber and reducing the pressure in the piston side chamber.

The stepwise nature of some of the valve command signals on both sides of both actuators is attributed to the discrete pressure reference selection algorithm. While this approach simplifies implementation, it introduces a higher utilization of the valves. A finer and more consistent valve command profile could be achieved by increasing the number of virtual pressure levels available n. This would enhance control resolution, but at the cost of increased computational load. An optimization study for the number of virtual levels along with other parameters are left for future studies.

4.2. Pressure trajectory shaping

The pressure shaping algorithm is illustrated in Figure 4.4 where the top figures show the pressures in both chambers and how they evolve during the switching period $[t_0 \rightarrow \tau_{sw}]$ and their corresponding flow rate references in the bottom plots.

The figure highlights the pressure shift occurring in the boom actuator just before 2 seconds during extension. For piston side chamber the pressure reference shifts from medium to low.

It can be seen that the pressure reference p_A^{ref} changes instantly while the chamber pressure steadily follows the generated shaped pressure reference shown in dotted. This is accomplished by modifying the flow rate reference using the third order polynomial defining the needed flow to complete the shift. Initially, the flow rate reference is due to the displacement flow needed by the piston motion, at the instant where the chamber reference shifts the flow reference is modified to shape the chamber pressure. The reference is followed by sequential valve control starting with closing the medium-line valve to reduce the flow rate and pressure as needed, then opening the tank valve to supply the required negative flow (flow out of the chamber) which reduces the pressure further. At the end of the shift the low line valve is opened to complete the shift and continue supplying the displacement flow.

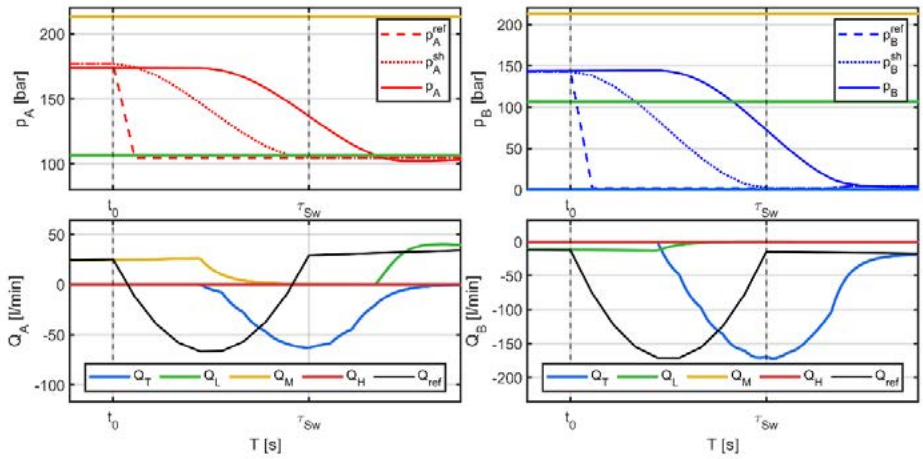


Figure 4.4: An excerpt from the results highlighting the pressure shaping algorithm

On the other hand, at the rod side chamber (figure to the right) similar behavior is observed with the exception that the flow reference doesn't change signs during the shift. The shaping can be accomplished by closing the low-pressure line valve and then opening the tank-line valve. More details on the reasoning behind flow direction changes can be found in Hansen's work [11]. The delay between the flow rate reference and the achieved one is attributable to the valve dynamics. Finally, it is important to note that the switching time τ_{sw} is set to 90 ms and the sampling rate is 10 ms which explains the steps in the flow rate signal.

4.3. Energy savings

A key strength point of the proposed algorithm is that it enables metering of the flow with minimum required differential pressure accomplished by manipulating the chamber pressure references. This is reflected in the power consumption of both actuators where there are several instances of the controller operating at minimal losses. Figure 4.5 shows the input and out powers of the proposed controller against the cited benchmark control [14]. The difference between proposed and benchmark can be seen in the boom actuator input power. The proposed controller initially matches the benchmark at the start of motion but as the piston speed increases it switches to a minimum differential pressure mode which critically minimizes the power consumption. The power losses slightly increase at the end of the extension stroke as some throttling is introduced to match the reference force. However, the

proposed controller outperforms even when throttling. During retraction both controllers exhibit similar performance as they both use an efficient differential mode.

Similar behavior is seen in the arm actuator, soon after the motion starts the proposed controller switches to a near minimum differential pressure mode reducing the power losses significantly. This can be observed during the acceleration phase of the extension stroke. Due to the reference force increasing the mode becomes infeasible and a compromise in power losses must be made, both controllers exhibit similar losses. As the piston retracts the proposed controller again chooses a highly efficient mode with minimal losses while the benchmark performance is poorer. Finally, both controllers show similar behavior during the rest of the retraction phase.

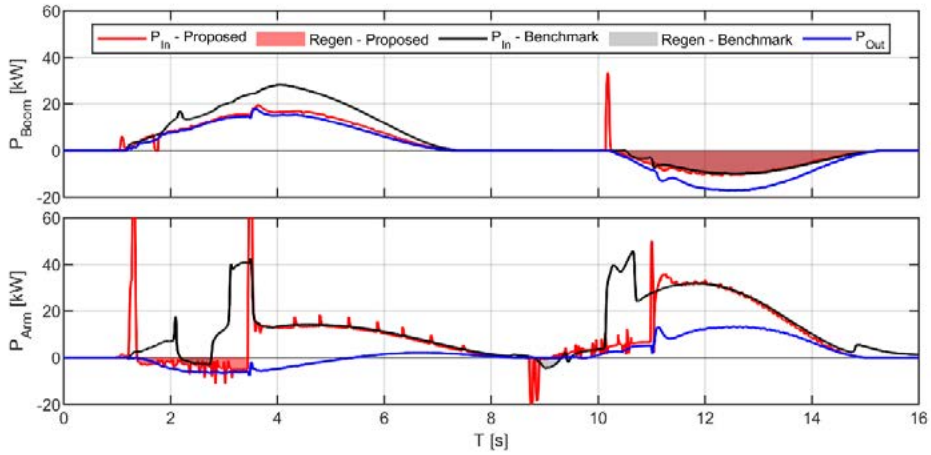


Figure 4.5: Power consumption of the proposed control algorithm against benchmark control

One key difference can be seen during the durations of pressure switching where the proposed controller concentrates the switch in a short time τ_{sw} while the benchmark one takes longer contributing to higher losses. Another advantage of the proposed controller is that it can allow for slight negative force error which allows for more utilization of efficient modes. Although this reflects in the piston velocity where it decelerates and accelerates, the deviation is small and unnoticeable.

Figure 4.6 shows the input energy of the proposed control against the original machine's load sensing system and the benchmark control. The results demonstrate the effectiveness of the proposed MPS system and control algorithms in reducing power losses compared to the conventional load sensing (LS) system. The LS baseline exhibited total power consumption of 518 kJ. By introducing the MPS system with the benchmark controller, the losses were reduced to 250 kJ, corresponding to a 52% reduction. Further improvement was achieved with the proposed controller, where the energy consumption decreased to 176 kJ, representing a 67% reduction relative to the original LS system and an additional 32% reduction compared to the benchmark controller. These results highlight the significant efficiency gains enabled by the MPS architecture and the proposed control strategy.

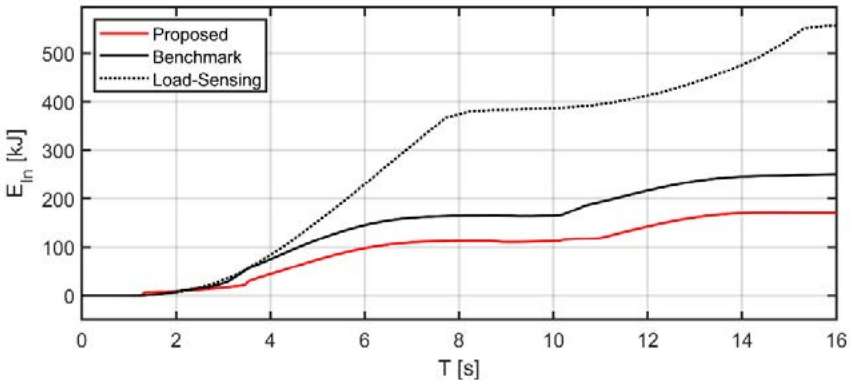


Figure 4.6: Input energy of proposed control algorithm against benchmark control

While these findings demonstrate the potential of the proposed control, it is important to note the limitations of the air grading cycle used for evaluation. The cycle does not fully represent real operating conditions. In particular, it may exaggerate the energy savings when compared to an LS system, which inherently shows higher compensation losses in no-load situations. Furthermore, the inclusion of energy recovery can distort the efficiency profile, suggesting greater savings than might actually occur under practical working conditions. Therefore, the results should be interpreted as an indication of relative system improvements rather than a direct prediction of real-world energy savings. The MPS along with the proposed controller are planned for future experimental work to provide more concrete energy saving estimates under real-world operating conditions.

5. Conclusion and Future Work

This paper presented a novel control strategy for a Multi-Pressure System (MPS) for an excavator's work functions. Simulation results for boom and arm actuators demonstrated accurate tracking of position, velocity, and force references, while significantly reducing energy consumption. The proposed controller dynamically selects chamber pressures to minimize throttling losses and optimize efficiency. The proposed controller ability to operate at near-minimum pressure differential and tolerate slight force error allows for higher energy savings. The proposed method achieved up to 67% reduction in energy consumption compared to conventional Load Sensing (LS) systems and 32% compared to an MPS benchmark controller, highlighting its potential for improving the machine's overall efficiency.

6. References

[1] P. Achten, "Fundamentals of Hydraulic Transformers," in *14th International Fluid Power Conference*, Dresden, Germany: River Publishers, 2024, pp. 796–815. doi: 10.13052/rp-9788770042222C63.

[2] M. Linjama, H.-P. Vihtanen, A. Sipola, and M. Vilenius, "Secondary controlled multi-chamber hydraulic cylinder," in *The 11th Scandinavian International Conference on*

Fluid Power SICFP'09, Linköping, Sweden, June 2-4 2009, Linköping, Sweden, 2009. Accessed: June 07, 2024.

[3] J. Zimmerman, E. Busquets, and M. Ivantysynova, "40% Fuel Savings by Displacement Control Leads to Lower Working Temperatures - A Simulation Study and Measurements," in *Proceedings of the 52nd National Conference on Fluid Power*, Las Vegas, USA, Mar. 2011.

[4] Vukovic, "STEAM : a holistic approach to designing excavator systems," in *9th International Fluid Power Conference (9th IFK)*, 2014. Accessed: Mar. 19, 2024.

[5] D. Fassbender, C. Brach, and T. Minav, "Experimental Investigations of Partially Valve-, Partially Displacement-Controlled Electrified Telehandler Implements," *Actuators*, vol. 12, Jan. 2023, doi: 10.3390/act12020050.

[6] K. Heybroek and J. Sjöberg, "Model Predictive Control of a Hydraulic Multichamber Actuator: A Feasibility Study," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 3, pp. 1393–1403, 2018, doi: 10.1109/TMECH.2018.2823695.

[7] R. Hansen, T. Andersen, and H. Pedersen, "Analysis of discrete pressure level systems for Wave Energy Converters," in *International Conference on Fluid Power and Mechatronics (FPM), 2011*, 2011, p. 558. doi: 10.1109/FPM.2011.6045825.

[8] M. Huova, M. Linjama, P. Kahra, J. Elonen, P. O. Box, and F. Oy, "New Implementation Of Digital Hydraulic Multi-Pressure Actuator," in *Workshop on Digital Fluid Power*, 2019.

[9] M. Huova and M. Linjama, "Energy efficient throttling control of a multi-pressure system using a genetic algorithm and model predictive control," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 236, no. 2, pp. 406–417, 2022, doi: 10.1177/09596518211027713.

[10] Z. Xu, M. Bertolin, A. Vacca, and J. Nilsson, "A Hydraulic Architecture Based on Multi-common Pressure Rail Principle Using Multi-chamber Cylinders for Excavators," in *Advancements in Fluid Power Technology: Sustainability, Electrification, and Digitalization*, L. Ericson and P. Krus, Eds., Cham: Springer Nature Switzerland, 2025, pp. 339–356. doi: 10.1007/978-3-031-84505-5_22.

[11] A. H. Hansen and H. C. Pedersen, "Reducing pressure oscillations in discrete fluid power systems," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 230, no. 10, pp. 1093–1105, Nov. 2016, doi: 10.1177/0959651815625015.

[12] M. Linjama and M. Huova, "Model-based force and position tracking control of a multi-pressure hydraulic cylinder," 2018, doi: 10.1177/0959651817754035.

[13] *JCMAS - Earthmoving machinery - Energy consumption test method - Hydraulic excavator*, 2010.

[14] M. Allam and M. Huova, "Energy Efficiency and Control of a Multi-pressure System - An Excavator Study," *International Journal of Fluid Power*.

Biographies



Mohamed Allam received a MSc degree at Ain Shams University, Egypt, in 2020. He is currently working as a doctoral researcher at Tampere University, Tampere Finland. He is pursuing a Ph.D. degree in Automation Science and Engineering. Mohamed is currently working in Tampere University's Innovative Hydraulics and Automation Lab. Mohamed's work is mainly in energy efficient machines, hybrid technologies, powertrains modelling and simulation.



Mikko Huova received a DSc degree at Tampere University of Technology, Finland, in 2015. The topic of his thesis is related to energy-efficient digital hydraulic systems, and he is continuing the work on this subject as a postdoctoral researcher at Tampere University. The areas of interest include energy efficient actuator systems, control design, modelling and simulation.



Kim Heybroek was born in Vastervik, Sweden, in 1981. He received the M.Sc. degree in mechanical engineering from Linkoping University (LiU), Linkoping, Sweden, in 2006. In 2008, he joined Volvo Construction Equipment in Eskilstuna, where he is currently working as a Research Engineer and has a specialist role in the field of hydraulics. In 2017, he received the Ph.D. degree in hydraulics at the Department of Fluid Power and Mechatronic Systems (FluMeS) at LiU.

Research on the Performance of a Heavy-Duty Robotic Arm Featuring Pump-Valve Coordination and Independent Metering Control with Hydraulic-Pneumatic Energy Storage

Zhixin Yan[1,2], Long Quan*[1,2], Lei Ge[1,2], Yunxiao Hao[1,2], Shufei Qiao[1,2], He Liu[1,2], Shuofeng Zhang[1,2], Yuxin Yang[1,2]

[1] *Key Lab of Advanced Transducers and Intelligent Control System of Ministry of Education and Shanxi Province, Taiyuan University of Technology, Taiyuan 030024, China*

[2] *College of Mechanical Engineering, Taiyuan University of Technology, Taiyuan 030024, China*

E-mail: yanzx163@163.com

Abstract.

Hydraulic systems dominate heavy-duty robotic arm applications due to their high power density and inherent impact resistance. However, conventional architectures suffer from severe throttling losses at valve ports and considerable dissipation of gravitational potential energy, resulting in overall system efficiencies of less than 30 % - a key contributor to the high energy consumption of heavy-duty equipment. To overcome these limitations, this paper proposes a novel hydraulic system integrating a three-chamber hydraulic cylinder with hydraulic-pneumatic energy storage, incorporating pump-valve coordination and independent metering control for heavy-duty robotic arms. In the proposed system, the energy-storage chamber of the three-chamber hydraulic cylinder is directly connected to a hydraulic accumulator. A multi-stage pressure-adjustable valve is designed to adaptively regulate energy storage under varying working conditions, enabling efficient recovery and reuse of gravitational energy. Pre-charging the hydraulic accumulator to counterbalance the arm's self-weight improves dynamic response and significantly reduces flow demand. The two driving chambers of the three-chamber hydraulic cylinder adopt independent metering control, combined with a pump-valve coordinated control strategy dominated by velocity feedforward and regulated by position feedback. This approach minimizes throttling losses and mitigates the reverse nonlinear charging characteristics of hydraulic accumulators. This strategy also overcomes traditional constraints on actuator chamber area imposed by conventional valve configurations. To fully exploit the synergy between independent metering and hydraulic-pneumatic energy storage, a newly designed independent metering valve manifold is introduced, offering low pressure loss, rapid response, and high precision and reliability. A 38-ton hydraulic excavator boom system was selected as a representative case study. Firstly, energy flow analysis was conducted. Subsequently, a multidisciplinary

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

co-simulation model was developed. Then, system parameters and control strategies were optimized based on the machine's operational characteristics. Finally, a prototype was built to evaluate performance and energy efficiency. Results demonstrate that, compared with existing solutions under identical conditions, the proposed system reduces flow demand by approximately 38 %, peak power by around 48 %, and achieves a gravitational energy recovery efficiency of nearly 70 %, leading to overall energy savings of up to 40 %. Additionally, the system exhibits excellent tracking accuracy, rapid dynamic response, and stable operation with minimal pressure and velocity fluctuations.

Keywords. Heavy-duty robotic arm, pump-valve coordinated control, independent metering control, hydraulic-pneumatic energy storage, velocity-position compound control.

1. INTRODUCTION

Heavy-duty robotic arms are typically powered by centralized hydraulic systems with valve-controlled power distribution, which provide advantages such as high power density and strong impact resistance. However, relevant studies have shown that the overall energy efficiency of such systems is often below 30 % [1], becoming a critical bottleneck for the sustainable and energy-efficient development of heavy equipment. The main reasons are twofold: on the one hand, traditional four-edge valve spools exhibit mechanical coupling between the inlet and outlet ports, leading to simultaneous throttling and severe energy losses. Moreover, only one control variable can be regulated at a given time, resulting in low degrees of freedom and poor controllability. On the other hand, during the lowering process of heavy-duty robotic arms, large amounts of gravitational potential energy are dissipated as heat through throttling orifices, not only wasting energy but also requiring additional cooling devices, which increase system complexity and failure risk.

Independent metering control offers an effective means to decouple actuator chamber regulation, thereby improving both energy efficiency and control performance. Backé pioneered the concept by developing a hydraulic cylinder system controlled by four cartridge valves [2]. Subsequently, Jansson and Palmberg addressed the problem of energy dissipation in valve-controlled systems by employing four Valvistor valves with pressure compensation and flow regeneration to reduce losses [3]. Hu and Zhang further introduced a proportional valve in parallel at the pump outlet based on a four-spool configuration, enabling multiple neutral control functions [4]. Lyu, Chen and Yao proposed a pump-valve coordinated scheme, where the pump supplied the main flow and the valve ensured accuracy, reducing energy consumption by about 50 % [5]. Zhong et al. employed high-speed switching valves as the pilot stage of a dual-spool electro-hydraulic proportional valve, improving dynamic response by 21 % and control accuracy by 16 % [6].

In terms of control strategies, Meyer, Weiler and Murrenhoff developed a discrete logic control method for independent metering control systems, dynamically adjusting piston-rod side pressure to enhance stiffness and efficiency [7]. Lu and Yao proposed a multi-level adaptive robust control strategy utilizing five cartridge valves and an accumulator, resulting in 30–40 % energy savings [8]. Xu et al. combined pump-valve control strategies to optimize both flow and pressure regulation, significantly reducing energy consumption while improving performance [9]. Shi et al. introduced a mode-switching velocity–position compound control strategy based on pump-valve coordination and independent metering control, achieving smooth operation and high positioning accuracy under low energy

consumption [10]. Li et al. integrated adaptive robust control with online parameter estimation to dynamically compensate for fluid compressibility and valve flow deviations, thereby improving control accuracy and stability [11].

Energy recovery from heavy-duty robotic arms primarily involves hydraulic, electrical, and mechanical approaches, which are also crucial for enhancing system efficiency. In hydraulic methods, Tong et al. proposed coupling high- and low-pressure accumulators with pump-motor torque, reducing energy consumption by 28–35 % [12]. Yang et al. developed a system using a three-chamber accumulator, increasing energy storage density by 43 % [13]. Tan et al. combined auxiliary and intensifying cylinders for energy recovery, achieving an efficiency of 58 % [14]. Beyond academic research, several manufacturers, including Liebherr (Germany), Sennebogen (Germany), Mantsinen (Sweden), Changlin Group (China), and Sunward Intelligent Equipment Group (China), have implemented mass compensation solutions based on energy-storage cylinders combined with accumulators. Reported applications indicate that such solutions can reduce the overall energy consumption of the machine by more than 30 % [15]. In electrical methods, Li, Wu and Cao controlled boom energy recovery based on excavator operation and battery state, improving recovery efficiency [16]. Lin et al. used proportional valves to control hydraulic motors, obtaining 35 % recovery efficiency with enhanced dynamic performance [17]. Do and Ahn employed hydraulic motors with planetary gearboxes, achieving 73 % regeneration efficiency and extending battery life by 16 % [18]. In mechanical methods, Latas and Stojek integrated a variable-inertia flywheel with a variable-displacement pump/motor in a novel hydromechanical accumulator, enabling multi-mode storage and improving energy density [19]. Li and Zhao combined flywheels, pump/motors, and regenerative valves, achieving 62 % recovery efficiency [20]. Cronk, Van and Strohmaier hybridized hydraulic accumulators with flywheels, reaching a conversion efficiency of 78 % and realizing dual-domain storage of hydraulic and kinetic energy [21].

Despite these significant advances, several challenges remain. independent metering control systems improve efficiency and control performance but involve complex multi-variable coordination, which limits their widespread adoption. Meanwhile, energy recovery techniques often involve multiple conversion stages, high system complexity, and trade-offs between efficiency and cost-effectiveness. Independent metering control and hydraulic-pneumatic energy storage offer complementary advantages in enhancing energy efficiency and control performance, yet their deep integration remains a challenge.

To address this gap, this study proposes a novel pump-valve coordinated and independent metering control system for a three-chamber hydraulic cylinder with hydraulic-pneumatic energy storage in heavy-duty robotic arms. In this system, the energy-storage chamber is directly connected to a hydraulic accumulator, together with a multi-level pressure-adjustable control valve. Through hydraulic-pneumatic balance, the system efficiently stores and utilizes gravitational potential energy under varying operating conditions. The two driving chambers of the cylinder adopt independent metering control principles, combined with a pump-valve coordinated strategy integrating velocity feedforward and position feedback, enabling smooth switching between operating modes to achieve high power density, high performance, and high energy efficiency. Based on theoretical analysis, a multidisciplinary co-simulation and experimental system was developed to investigate the

operating characteristics and energy efficiency of the proposed heavy-duty robotic arm system.

2. SYSTEM PRINCIPLE

The principle of the proposed pump-valve coordinated and independent metering control system with hydraulic-pneumatic energy storage for heavy-duty robotic arms is illustrated in **Figure 2.1**. The system consists of an electrohydraulic power source, an independent metering valve manifold, a three-chamber hydraulic cylinder, a heavy-duty robotic arm, a hydraulic accumulator, and a multi-stage pressure-adjustable valve. The three-chamber hydraulic cylinder includes one energy-storage chamber and two driving chambers, with the two driving chambers designed to have equal effective areas.

The energy-storage chamber of the three-chamber cylinder is directly connected to the hydraulic accumulator via the multi-stage pressure-adjustable valve. Through hydraulic-pneumatic balancing, the self-weight of the robotic arm is compensated, allowing for the efficient storage and utilization of gravitational potential energy [14]. Depending on operating conditions, the energy-storage chamber can also be connected to the electrohydraulic power source or the tank via the adjustable valve to satisfy oil replenishment or pressure release requirements. The two driving chambers are connected to the independent metering valve manifold, enabling fully decoupled control [3]. Furthermore, the manifold is optimized to integrate flow regeneration, allowing direct interconnection between the two driving chambers of the cylinder.

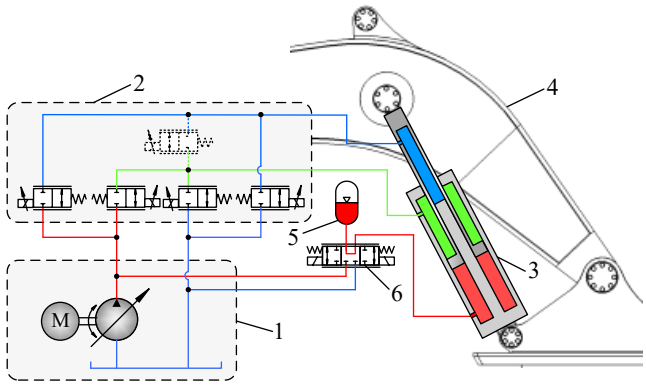


Figure 2.1. Principle of the pump-valve coordinated and independent metering control system with hydraulic-pneumatic energy storage for heavy-duty robotic arms.
1-Electrohydraulic power source,2-Independent metering valve manifold,3-Three-chamber hydraulic cylinder,4-Heavy-duty robotic arm,5-Hydraulic accumulator,6-Multi-stage pressure-adjustable valve.

According to load and operating conditions, the system can switch among different modes by adjusting motor speed, pump displacement, and valve openings [5]. This enables pump-valve coordinated control, independent metering control, energy recovery, and flow regeneration.

2.1. Resistive Extension Mode

As shown in **Figure 2.2**, the resistive extension mode operates in the first quadrant. In this mode, the inlet valve, outlet valve, and pressure-adjustable valve are fully open. The two driving chambers of the three-chamber cylinder are connected to the hydraulic pump and the tank, respectively, while the energy-storage chamber is connected to the accumulator. The cylinder is therefore driven jointly by the hydraulic pump and the accumulator to extend against the external load. The cylinder velocity is controlled through motor speed and pump displacement, with the system operating in pump control mode. This significantly reduces throttling losses while lowering overall flow demand.

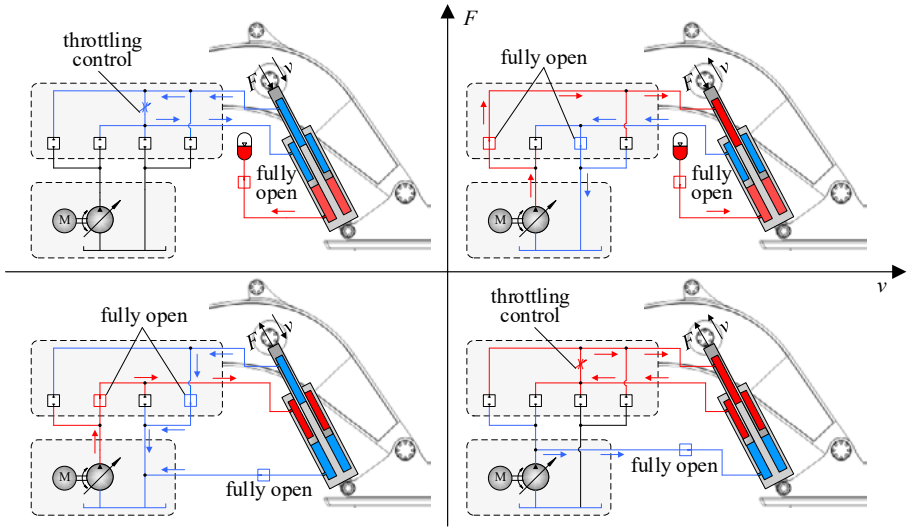


Figure 2.2. Operating modes of the pump-valve coordinated and independent metering control system with hydraulic-pneumatic energy storage for heavy-duty robotic arms.

2.2. Overrunning Retraction Mode

As shown in **Figure 2.2**, the overrunning retraction mode operates in the second quadrant. In this mode, the pressure-adjustable valve is fully open, and the two driving chambers of the cylinder are interconnected through the flow regeneration valve, while the energy-storage chamber is connected to the accumulator. In this condition, the gravitational potential energy of the robotic arm is directly recovered as hydraulic energy and stored in the accumulator. The cylinder velocity is regulated by throttling through the flow regeneration valve, which also mitigates the nonlinear back-pressure characteristic of the accumulator during charging. The system therefore operates in valve control mode with both energy recovery and flow regeneration.

2.3. Resistive Retraction Mode

As shown in **Figure 2.2**, the resistive retraction mode operates in the third quadrant. Similar to the resistive extension mode, the inlet valve, outlet valve, and pressure-adjustable valve are fully open. The two driving chambers are connected to the hydraulic pump and tank, respectively, while the energy-storage chamber is connected to the tank. In this case, the hydraulic pump drives the cylinder to retract against the load. The cylinder velocity is

controlled by adjusting motor speed and pump displacement, and the system operates in pump control mode.

2.4. Overrunning Extension Mode

As shown in **Figure 2.2**, the overrunning extension mode operates in the fourth quadrant. Similar to the overrunning retraction mode, the pressure-adjustable valve is fully open. The two driving chambers are interconnected through the flow regeneration valve, while the energy-storage chamber is connected to the pump. In this case, the hydraulic pump replenishes oil into the energy-storage chamber to prevent cavitation and oscillations. The cylinder velocity is regulated through throttling by the flow regeneration valve, and the system operates in valve control mode.

3. SYSTEM ANALYSIS

A theoretical framework was developed for the proposed system, combining energy-flow analysis with mathematical modeling to support control strategy design.

3.1. Energy Flow Analysis

The energy flow of the pump-valve coordinated and valve-orifice independent control system with hydraulic-pneumatic energy storage for heavy-duty robotic arms is illustrated in **Figure 3.1**. Pump-valve coordinated control reduces throttling losses by switching between pump-control and valve-control modes. In addition, by coordinating motor speed with pump displacement, the pump supply matches actuator demand, minimizing overflow losses. Independent metering control enables independent regulation of pressure and flow in each chamber. This decouples the oil supply and return paths, eliminates simultaneous throttling, and enhances system controllability while reducing losses. Furthermore, return oil is either recovered by the hydraulic accumulator or directly regenerated to another actuator chamber, thereby realizing energy recovery and circulation [8]. Unlike conventional systems, where potential energy is dissipated as heat, the proposed system converts these losses into reusable energy, improving overall efficiency.

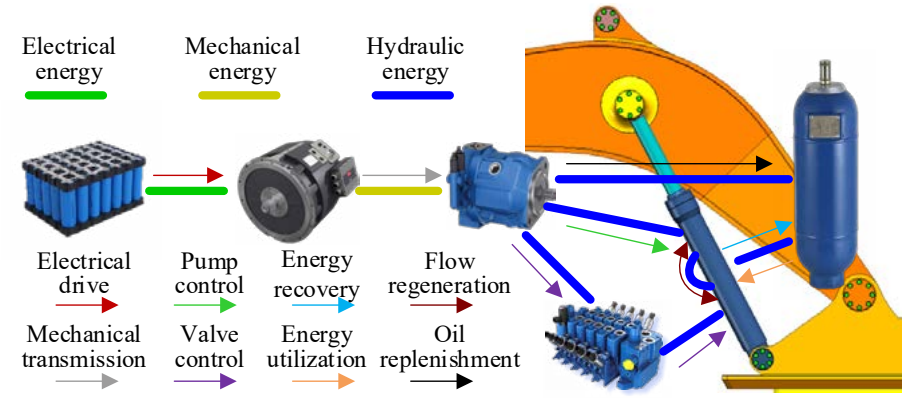


Figure 3.1. Energy flow of the pump-valve coordinated and independent metering control system with hydraulic-pneumatic energy storage for heavy-duty robotic arms.

3.2. System Modeling

Mathematical models are created for the main components and links. These models provide the basis for simulating system performance and evaluating energy efficiency.

The discharge power of the battery is:

$$\mathrm{P_b} = \frac{\mathrm{T_e}\omega_\mathrm{e}}{\eta_\mathrm{e}} \tag{3.1}$$

where P_b is the battery discharge power, T_e is the electromotor torque, ω_e is the electromotor speed, and η_e is the electromotor efficiency.

The torque balance of the electromotor and hydraulic pump is:

$$J_{ep}\dot{\omega}_e + B_{ep}\omega_e = T_e - T_p \tag{3.2}$$

where J_{ep} is the total rotational inertia of the electromotor and pump, B_{ep} is the total viscous damping coefficient, and T_p is the pump torque.

The pump torque is:

$$T_p = \frac{D_p p_p}{\eta_p} \tag{3.3}$$

where D_p is the pump displacement, p_p is the pump pressure, and η_p is the pump efficiency.

The pump flow rate is:

$$Q_p = D_p\omega_e - C_{lp}(p_p - p_t) \tag{3.4}$$

where Q_p is the pump flow, C_{lp} is the pump leakage coefficient, and p_t is the tank pressure.

The valve flow equation is:

$$Q_{v,j} = C_d w_{v,j} x_{v,j}\sqrt{\frac{2\Delta p_{v,j}}{\rho}} \cdot \mathrm{sgn}(\Delta p_{v,j}) \tag{3.5}$$

where $Q_{v,j}$ is the flow through the valve orifice, C_d is the flow coefficient, $w_{v,j}$ is the orifice area gradient, $x_{v,j}$ is the valve spool displacement, $\Delta p_{v,j}$ is the pressure drop across the valve, and ρ is the oil density.

The continuity equation of the three-chamber hydraulic cylinder is:

$$\dot{p}_j = \frac{\beta_e}{V_j}(Q_{v,j} - A_j\dot{x}_c - C_{lc}\Delta p_c) \tag{3.6}$$

where p_j is the chamber pressure, β_e is the effective bulk modulus, V_j is the chamber volume, A_j is the piston area, x_c is the piston displacement, C_{lc} is the cylinder leakage coefficient, and Δp_c is the chamber pressure differential.

The force balance of the three-chamber hydraulic cylinder is:

$$m_{eq}\ddot{x}_c + B_c\dot{x}_c + K_c x_c = A_1 p_1 - A_2 p_2 + A_3 p_3 - F_L \tag{3.7}$$

where m_{eq} is the equivalent mass of piston rod and load, B_c is the viscous damping, K_c is the stiffness, A_1, A_2 are the driving chamber areas with pressures p_1, p_2, A_3 is the energy-storage chamber area with pressure p_3, and F_L is the external load.

The accumulator pressure dynamics are:

$$\dot{p}_3 = \frac{\beta_a}{V_a} Q_a \tag{3.8}$$

where β_a is the effective bulk modulus of the accumulator, V_a is the effective volume, and Q_a is the accumulator flow.

The energy stored in the accumulator is:

$$E_a = \frac{p_y V_y - p_x V_x}{\gamma - 1} \tag{3.9}$$

where E_a is the energy stored in the hydraulic accumulator, p_x, p_y are the pressures of the accumulator, V_x, V_y are the volumes of the accumulator, and γ is the adiabatic index.

4. CONTROL STRATEGY

To further improve system performance and efficiency, a pump-valve coordinated control strategy with velocity feedforward and position feedback regulation is adopted [9]. Velocity feedforward serves as the primary control loop, generating real-time commands for motor speed, pump displacement, and valve opening based on the desired velocity profile. This enables rapid response to target motions. Position feedback is introduced as a compensatory loop: the deviation between desired and actual position is used to correct the feedforward commands, thereby suppressing tracking errors caused by nonlinear factors such as leakage, friction, and fluid compressibility [11].

The motor speed feedforward signal u_{ef} is:

$$u_{ef} = \frac{v_e \cdot A}{D_p \omega_{emax}} \tag{4.1}$$

where A is the piston area of the driving chamber, and ω_{emax} is the maximum motor speed.

The pump displacement feedforward signal u_{pf} is:

$$u_{pf} = \frac{v_e \cdot A}{D_{pmax} \omega_e} \tag{4.2}$$

where D_{pmax} is the maximum pump displacement.

The valve opening feedforward signal u_{vf} is:

$$u_{vf} = \frac{v_e \cdot A}{Q_n} \sqrt{\frac{\Delta p_n}{\Delta p_{v,j}}} \tag{4.3}$$

where Q_n is the rated valve flow at pressure drop Δp_n.

The position feedback signal u_b is:

$$u_b = K_p \left(\int_{t_0}^{t_1} v_e \mathrm{d}t - x_c \right) + K_i \int_{t_0}^{t_1} \left(\int_{t_0}^{t_1} v_e \mathrm{d}t - x_c \right) \mathrm{d}t + K_d \frac{\mathrm{d}\left(\int_{t_0}^{t_1} v_e \mathrm{d}t - x_c \right)}{\mathrm{d}t} \tag{4.4}$$

where K_p, K_i, and K_d are proportional, integral, and derivative gains, respectively.

The overall control input is:

$$U = \left[u_{ef} + u_b, u_{pf} + u_b, u_{vf} + u_b \right] \tag{4.5}$$

This strategy leverages velocity feedforward for rapid response and position feedback for nonlinear compensation. Combined with pump-valve coordinated and energy recovery mechanisms, it enables high-precision control and energy-efficient operation of the heavy-duty robotic arm hydraulic system [21].

5. Simulation Study and Preliminary Experimental Validation

The boom system of a 38-ton hydraulic excavator was selected as a representative case for investigation. A multidisciplinary co-simulation model was developed in SimulationX, incorporating the mechanical, electrical, hydraulic, and control subsystems [10]. To further validate the operating characteristics and energy-saving performance, an experimental platform was constructed on a 38-ton excavator, as shown in **Figure 5.1**.

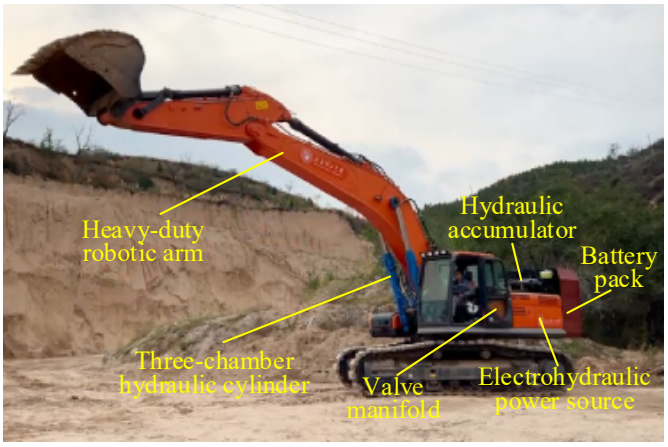


Figure 5.1. Experimental platform of the 38-ton hydraulic excavator.

The displacement and velocity curves of the boom cylinder are illustrated in **Figure 5.2**, where the maximum extension velocity was set to 200 mm/s and the maximum retraction velocity to 400 mm/s. The boom lifted from 2 to 7.5 s, remained stationary from 7.5 to 9 s, and descended from 9 to 12.5 s. As shown in **Figure 5.2**, both the original valve-controlled two-chamber cylinder system and the proposed pump–valve coordinated and independent metering controlled three-chamber cylinder system were able to track the desired displacement with overall smooth velocity profiles. During the initial lifting phase, the energy-storage chamber of the new system was directly connected to the hydraulic accumulator, thereby counterbalancing the boom's gravitational load and overcoming the inertial resistance. With the assistance of the accumulator, the proposed system exhibited faster dynamic response compared to the conventional system. However, the nonlinear charging/discharging characteristics of the accumulator introduced minor velocity fluctuations.

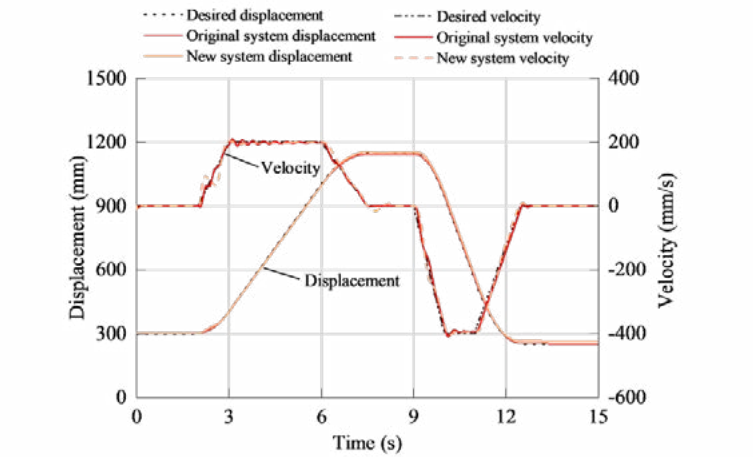


Figure 5.2. Displacement and velocity curves of the boom cylinder.

Figure 5.3 presents the pump output flow and accumulator pressure curves of both systems. For the conventional system, the required pump flow reached 424 L/min during boom lifting and 432 L/min during lowering. In contrast, the proposed system required only 173 L/min during lifting and 357 L/min during lowering. In the lifting phase, the accumulator discharged into the energy-storage chamber of the three-chamber cylinder, with its pressure gradually decreasing as stored energy was directly utilized. During lowering, the energy-storage chamber charged the accumulator, causing its pressure to increase as gravitational potential energy was converted into hydraulic energy and stored for subsequent reuse. Compared with the conventional two-chamber system, the smaller effective areas of the two driving chambers in the proposed three-chamber cylinder allowed a significant portion of the required flow to be supplied by the accumulator, thereby reducing the pump output flow by approximately 38 %.

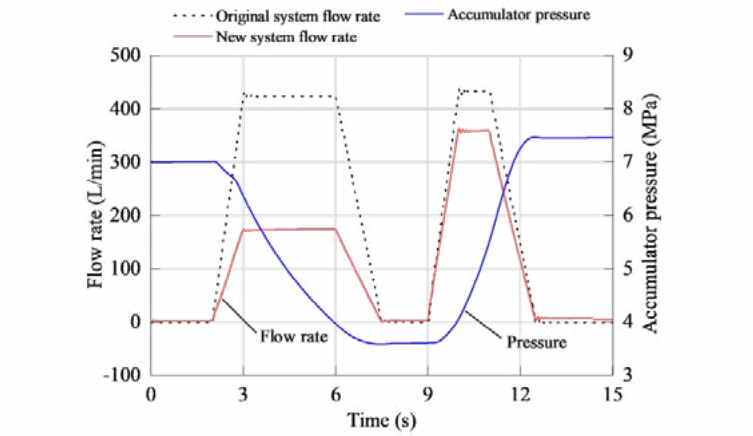


Figure 5.3. Pump output flow and accumulator pressure curves.

Figure 5.4 shows the system power and energy profiles. In the conventional system, the peak power reached 104 kW during lifting and 224 kW during lowering, with a total energy

consumption of 821 kJ. In the proposed system, the corresponding peak powers were reduced to 54 kW and 175 kW, with a total energy consumption of 479 kJ. By integrating chamber-independent valve control and the three-chamber cylinder, the proposed system effectively reduced throttling losses and directly recovered gravitational potential energy. As a result, the peak power demand was reduced by approximately 48 %, the gravitational energy recovery rate reached about 70 %, and the overall energy savings were around 40 %.

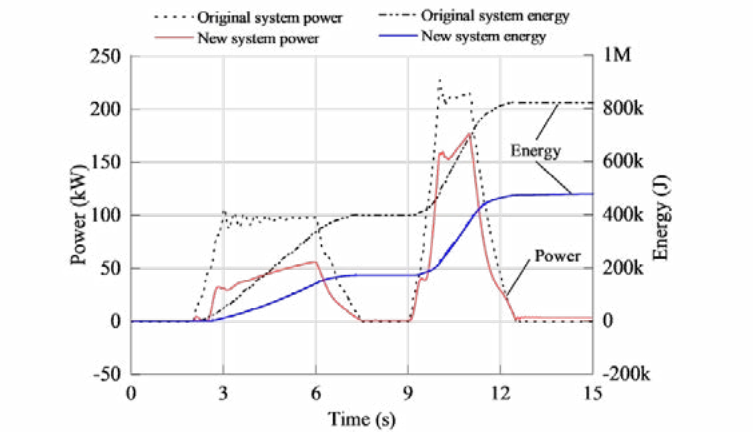


Figure 5.4. System power and energy curves.

6. CONCLUSIONS

A pump–valve coordinated and independent metering control three-chamber hydraulic cylinder system with hydraulic–pneumatic energy storage, together with a velocity feedforward–position feedback regulated pump–valve coordination strategy, has been proposed for heavy-duty robotic arms. Taking the boom system of a 38-ton hydraulic excavator as a representative case, the main conclusions are summarized as follows:

(1) The operating principle of the proposed system was analyzed. The system integrates pump–valve coordination and independent metering control with energy recovery and flow regeneration. It is capable of switching among different operating modes according to working conditions, thereby achieving high power density, superior performance, and improved energy efficiency.

(2) The energy flow of the system was examined, and a comprehensive mathematical model was established. A velocity feedforward–position feedback pump–valve coordination strategy was developed, enabling fast dynamic response through velocity feedforward and nonlinear error compensation through position feedback. Coupled with energy recovery, this strategy further enhances both control performance and system efficiency.

(3) Simulation and experimental validation demonstrated that, under identical working conditions compared to conventional systems, the proposed system reduced pump flow demand by approximately 38 %, peak power by 48 %, and achieved a gravitational energy recovery rate of about 70 %, resulting in overall energy savings of approximately 40 %. Moreover, the system exhibited excellent tracking accuracy and rapid dynamic response without significant pressure or velocity fluctuations.

This study provides a high-performance, high power density, and energy-efficient drive and transmission solution for heavy-duty equipment, offering significant potential for practical implementation and scalability in energy-saving, cost-reduction, and carbon-reduction applications.

7. ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFB3406601) and the National Natural Science Foundation of China (Grant No. 52505068).

8. NOMENCLATURE

A	piston area of the driving chamber	m^2
A_j	piston area	m^2
A_1, A_2	driving chamber areas	m^2
A_3	energy-storage chamber area	m^2
B_c	viscous damping	N·s/m
B_{ep}	total viscous damping coefficient	N·m/(rad/s)
C_d	flow coefficient	-
C_{lc}	cylinder leakage coefficient	$m^3/(s·Pa)$
C_{lp}	pump leakage coefficient	$m^3/(s·Pa)$
D_p	pump displacement	m^3/rad
D_{pmax}	maximum pump displacement	m^3/rad
E_a	energy stored in the hydraulic accumulator	J
F_L	external load	N
J_{ep}	total rotational inertia of the electromotor and pump	$kg·m^2$
K_c	stiffness	N/m
K_d	derivative gains	-
K_i	integral gains	-
K_p	proportional gains	-
m_{eq}	equivalent mass of piston rod and load	kg
P_b	battery discharge power	W
p_j	chamber pressure	Pa
p_p	pump pressure	Pa
p_t	tank pressure	Pa

p_x, p_y	pressures of the accumulator	Pa
p_1, p_2	driving chamber pressures	Pa
p_3	energy-storage chamber pressure	Pa
Q_a	accumulator flow	m^3/s
Q_n	rated valve flow	m^3/s
Q_p	pump flow	m^3/s
$Q_{v,j}$	flow through the valve orifice	m^3/s
T_e	electromotor torque	N·m
T_p	pump torque	N·m
u_b	position feedback signal	-
u_{ef}	motor speed feedforward signal	-
u_{pf}	pump displacement feedforward signal	-
u_{vf}	valve opening feedforward signal	-
V_a	effective volume	m^3
V_j	chamber volume	m^3
V_x, V_y	volumes of the accumulator	m^3
x_c	piston displacement	m
$x_{v,j}$	valve spool displacement	m
β_a	effective bulk modulus of the accumulator	Pa
β_e	effective bulk modulus	Pa
γ	adiabatic index	-
Δp_c	chamber pressure differential	Pa
Δp_n	pressure drop	Pa
$\Delta p_{v,j}$	pressure drop across the valve	Pa
η_e	electromotor efficiency	-
η_p	pump efficiency	-
ρ	oil density	kg/m^3
ω_e	electromotor speed	rad/s
ω_{emax}	maximum motor speed	rad/s
$w_{v,j}$	orifice area gradient	m^2/m

9. REFERENCES

[1] Love L J, Lanke E, Alles P. 'Estimating the impact (energy, emissions and economics) of the US fluid power industry'. Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2012. ORNL/TM-2011/14

[2] BACKÉ W. 'Design systematics and performance of cartridge valve controls'. International Conference on Fluid Power, pp. 1-48, Tampere, Finland, March 24-26, 1987.

[3] Jansson A, Palmberg J.O. 'Separate controls of meter-in and meter-out orifi ces in mobile hyraulic systems'. SAE transactions, pp.377-383, 1990. DOI: 10.4271/901583

[4] Hu H, Zhang Q. 'Multi–function realization using an integrated programmable e/h control valve'. Applied engineering in agriculture, 19(3), p.283, 2003. DOI: 10.13031/2013.13660.

[5] Lyu L, Chen Z, Yao B. 'Advanced valves and pump coordinated hydraulic control design to simultaneously achieve high accuracy and high efficiency'. IEEE Transactions on Control Systems Technology, 29(1), pp.236-248, 2020. DOI: 10.1109/TCST.2020.2974180

[6] Zhong Q, Xu E, Jia T, Yang H, Zhang B, Li Y. 'Dynamic performance and control accuracy of a novel proportional valve with a switching technology-controlled pilot stage.' Journal of Zhejiang University-SCIENCE A, 23(4), pp. 272-285, 2022. DOI: 10.1631/jzus.A2100463.

[7] Meyer C, Weiler D, Murrenhoff H. 'Optimal nominal pressurization generation: A novel idea for a discrete logic control using an independent metering valve configuration'. Dynamic Systems and Control Conference, Vol. 54761, pp. 351-354, January, 2011. DOI: 10.1115/DSCC2011-5947.

[8] Lu L, Yao B. 'Energy-saving adaptive robust control of a hydraulic manipulator using five cartridge valves with an accumulator'. IEEE Transactions on Industrial Electronics, 61(12), pp.7046-7054, 2014. DOI: 10.1109/TIE.2014.2314054.

[9] Xu B, Ding R, Zhang J, Cheng M, Sun T. 'Pump/valves coordinate control of the independent metering system for mobile machinery'. Automation in Construction, 57, pp.98-111, 2015. DOI: 10.1016/j.autcon.2015.04.012.

[10] Shi J, Quan L, Zhang X, Xiong X. 'Electro-hydraulic velocity and position control based on independent metering valve control in mobile construction equipment'. Automation in Construction, 94, pp.73-84, 2018. DOI: 10.1016/ j.autcon.2018.06.005.

[11] Li C, Ding R, Cheng M, Chen Z, Yao B. 'Accurate motion control of an independent metering actuator with adaptive robust compensation of uncertainties in pressure dynamics'. IEEE/ASME Transactions on Mechatronics, 29(5), pp.3877-3889, 2024. DOI: 10.1109/TMECH.2024.3360236.

[12] Tong Z, Jiang Y, Tong S, Zhang Q, Wu J. 'Hybrid drivetrain with dual energy regeneration and collaborative control of driving and lifting for construction machinery'. Automation in Construction, 150, p.104806, 2023. DOI: 10.1016/j.autcon.2023.104806.

[13] Yang C, Zhou L, Wang J, Xu T, Yang C, Ye G. 'Research on energy saving system of hydraulic excavator based on three-chamber accumulator'. Journal of Energy Storage, 72, p.108571, 2023. DOI: 10.1016/j.est.2023.108571.

[14] Tan L, He X, Xiao G, Jiang M, Yuan Y. 'Design and energy analysis of novel hydraulic regenerative potential energy systems'. Energy, 249, p.123780, 2022. DOI: 10.1016/j.energy.2022.123780.
[15] Xia L, Quan L, Yang J, Zhao B. 'Research on the method of hydraulic-gas energy storage balancinghydraulic excavator boom weight'. Journal of Mechanical Engineering, 54(20), pp.197-205, 2018. DOI: 10.3901/JME.2018.20.197.
[16] Li W, Wu B, Cao B. 'Control strategy of a novel energy recovery system for parallel hybrid hydraulic excavator'. Advances in Mechanical Engineering, 7(10), p.1687814015613769, 2015. DOI: 10.1177/1687814015613769.
[17] Lin T, Chen Q, Ren H, Huang W, Chen Q, Fu S. 'Boom energy recovery system with auxiliary throttle based on hybrid excavator'. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 231(22), pp.4250-4262, 2017. DOI: 10.1177/0954406216664546.
[18] Do T.C, Ahn K.K. 'Hybrid powertrain with dual energy regeneration for boom cylinder movement in a hydraulic excavator'. Automation in Construction, 171, p.105974, 2025. DOI: 10.1016/j.autcon.2025.105974.
[19] Latas W, Stojek J. 'A new type of hydrokinetic accumulator and its simulation in hydraulic lift with energy recovery system'. Energy, 153, pp.836-848, 2018. DOI: 10.1016/j.energy.2018.04.040.
[20] Li J, Zhao J. 'Energy recovery for hybrid hydraulic excavators: flywheel-ba sed solutions'. Automation in construction, 125, p.103648, 2021. DOI: 10.10 16/j.autcon.2021.103648.
[21] Cronk P, Van de Ven J, Strohmaier K. 'Design optimization, construction, and testing of a hydraulic flywheel accumulator'. Journal of Energy Storage, 44, p.103281, 2021. DOI: 10.1016/j.est.2021.103281.

Biographies



Zhixin Yan received the B.S. degree in Mechanical Design, Manufacturing, and Automation from Taiyuan University of Technology, China, in 2021. He is currently pursuing a Ph.D. degree in Mechanical Engineering at Taiyuan University of Technology. His research interests include energy-saving technologies and intelligent control of heavy-duty equipment.



Long Quan received the B.S. degree in Mechanical Engineering from Taiyuan University of Technology, China, in 1981, the M.S. degree in Fluid Power Transmission and Control from Zhejiang University, China, in 1988, and the Ph.D. degree in Fluid Power Transmission and Control from Xi'an Jiaotong University, China, in 1993. From 1998 to 1999 and from 2003 to 2004, he was a visiting scholar at the Institute of Fluid Power and Control, Dresden University of Technology, Germany. He is currently a Professor with the College of Mechanical Engineering, Taiyuan University of Technology. His research interests include fluid power transmission and control, energy-saving and intelligent control of electromechanical equipment, and electro-hydraulic control for construction and heavy machinery.



Lei Ge received the B.S. degree in Mechanical Design, Manufacturing, and Automation in 2011, the M.S. degree in Mechanical Engineering in 2014, and the Ph.D. degree in Mechanical Engineering in 2018, all from Taiyuan University of Technology, China. He is currently a Research Fellow with the College of Mechanical Engineering, Taiyuan University of Technology. His research interests include energy-saving control of engineering equipment, electro-hydraulic servo power sources, electro-hydraulic control components and system integration, and digital twin technology for electromechanical systems.



Yunxiao Hao received the Ph.D. degree in mechanical engineering from Taiyuan University of Technology, Shanxi, China, in 2021. He has been a Postdoctoral Researcher with the Taiyuan University of Technology since 2021. His current research interests include electro-hydraulic control systems, energy-saving technology of mobile machinery, and the design of hydraulic elements.



Shufei Qiao received the B.S. degree in mechanical engineering from Shenyang Chemical Engineering University, China, in 2015, and the Ph.D. degree in mechanical engineering from Taiyuan University of Technology, Shanxi, China, in 2024. From 2015 to 2017, he was with State Grid for hydraulic work. His research interests include the hydraulic system of excavators and intelligent control.



He Liu received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Taiyuan University of Technology, Shanxi, China, in 2017, 2020, and 2025, respectively. His current interests include fluid transmission and control, electro-hydraulic proportional control, and the design of hydraulic components.



Shuofeng Zhang received the B.S. degree in Mechanical Design, Manufacturing, and Automation from Taiyuan University of Technology, China, in 2021. He is currently pursuing a Ph.D. degree in Mechanical Engineering at Taiyuan University of Technology. His research interests include energy-saving control of hydraulic systems.



Yuxin Yang received the B.S. degree in Mechanical Design, Manufacturing, and Automation from Sichuan University, China, in 2020. He is currently pursuing a Ph.D. degree in Mechanical Engineering at Taiyuan University of Technology. His research interests include electro-hydraulic control systems for construction machinery.

Operator effect on the performance of a wheeled excavator

Bernhard Sender*, Johannes Sprink, Katharina Schmitz

RWTH Aachen University, Institute for Fluid Power Drives and Systems (ifas), Campus-Boulevard 30, 52074 Aachen, Germany

**Corresponding Author. E-mail: bernhard.sender@ifas.rwth-aachen.de*

Abstract.

This paper investigates how human operators influence the performance of a wheeled excavator. A 16t wheeled excavator with a load-sensing hydraulic system was instrumented with sensors enabling kinematic and hydraulic measurements. Two professional and five non-professional operators executed a standardized 90° dig-and-dump-operation under controlled operating conditions. From the recorded data, three indicators were derived: mean cycle time, simultaneity level of actuator operation and mean hydraulic pump power. The correlation between these indicators was analysed across multiple repetitions performed by each operator. The results highlight the influence of operator behaviour on machine performance.

Keywords. Excavator, mobile machines, mobile hydraulics, operator, duty cycle

1. INTRODUCTION

Improving the energy efficiency of off-highway mobile machinery is a key lever for reducing fuel consumption, operating costs, and emissions in construction and mining. Hydraulic excavators are among the most energy-intensive machine types. While extensive research is conducted on improving component and system efficiency (optimized hydraulic architectures, energy recovery, advanced control strategies etc.), the influence of the human operator on the energetic performance of the machine has received comparatively less attention. Modern excavators offer increasing degrees of automation and assistance, making it important to understand which aspects of operator behaviour most strongly affect performance, and how they interact with the hydraulic system.

This gap is addressed by experimentally studying the influence of different operators on the performance of a wheeled excavator during a standardized load cycle. However, the scope of this work is limited to a 90° dig-and-dump-operation.

2. STATE OF THE ART

While many authors agree that the operator is one of the largest influences on machine performance, rigorous quantitative studies are still relatively scarce, especially for excavators. This section briefly summarizes a selection of existing research on this topic.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Kunze et al. [1] analysed 153 measurements of 7 different operators, ranging from novice to expert level, performing trenching operations with a 20t wheeled excavator. From the gathered data and video analysis they computed the mean cycle time and simultaneity level ("[…] the number of hydraulic cylinders being manipulated through the joystick axes at the same time, averaged over a closed working cycle." [1]). The exact signal-processing steps are not described in detail. They found that cycle time correlates negatively with simultaneity level and used k-means clustering to separate operators into groups (Figure 2.1), that correspond well to their skill level.

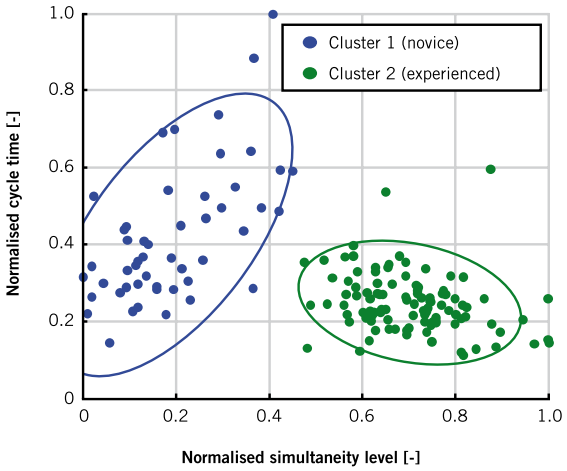


Figure 2.1. Clustering excavator operators by skill level [1]

Mieth et al. [2] extended this approach to 285 trenching cycles and included mean cycle power as an additional feature. Their results indicate that more skilled operators not only achieve higher simultaneity and shorter cycle times, but also higher hydraulic power usage.

Extensive research by *Filla et al.* [3–7] and *Frank et al.* [8, 9] addressed operator interaction with wheel loaders, ranging from development of operator models [3–5], to studies on machine operability (i.e. the ease with which the operator can perform the task) by using psychophysiological measurements (heart rate, finger temperature etc.) [5–7] and operator influence on machine performance [8, 9]. *Frank et al.* [8, 9] performed different loading cycles with 73 operators and measured productivity (moved material per time , t/h), and efficiency (moved material per fuel consumption, t/l). Their results show that for predefined duty cycles, productivity can increase by up to eight times, and fuel efficiency by up to three times for highly skilled operators. One interesting finding from machine operability research is that design changes in components or control logic can unintentionally increase operator workload, worsen bucket handling and reduce fuel efficiency if they conflict with established operator strategies [5].

Sturm [10] used measurements from three excavator operators to validate a developed operator model, but did not systematically analyse the operator influence itself.

Inderelst et al. [11] studied various influences on a 21t crawler excavator, including hydraulic system principle (LS vs. NFC), engine speed, operator (four different operators), and ground conditions. They found that the spread in fuel consumption and digging

performance caused by different operators is comparable in magnitude to changes achieved by hardware or parameter modifications.

Fecke [12] used fleet management telemetry data from nearly 12000 machines (excavators and wheel loaders). His work focusses on deriving standardized duty cycles and reducing operator variability in efficiency measurements.

3. METHODOLOGY

The objective of this study is to quantify the influence of the human operator on the performance of a wheeled excavator. The used methodology is similar to *Kunze et all* [1] and *Mieth et al.* [2].

To this end, a Volvo EW160D wheeled excavator was instrumented with an extensive sensor array and operated by drivers with different experience levels under controlled test conditions.

In detail, the methodology comprises four main elements:

- Test machine and instrumentation: description of the excavator and the additional measurement equipment installed on the hydraulic and mechanical subsystems.
- Test operation and operators: definition of the task and the participating operators.
- Derivation of performance indicators: computation of used hydraulic pump power, simultaneity level of actuator operation, and mean cycle time from the recorded data.
- Assessment of limitations: discussion of constraints and uncertainties of the experimental design and data evaluation.

The following sections describe each element in detail.

3.1. Test Machine

The experiments were conducted using a Volvo EW160D, a 16t wheeled excavator equipped with a load-sensing hydraulic system. The machine remained in its original configuration; no major modifications were made to the drivetrain or the control system. Only measurement equipment was added to enable data acquisition.

3.2. Data Gathering

The excavator was equipped with a range of sensors to capture the relevant hydraulic and kinematic quantities during operation. There is additional measurement equipment installed (for example, optical sensors), these were used for research outside the scope of this work and therefore aren't described here.

3.2.1. Kinematic Measurements

Figure 3.1 schematically depicts the placement of the wire-based linear position sensors on the boom, stick, bucket, and auxiliary boom cylinders. These sensors (S1–S4) measure the extension and retraction of the respective cylinders, enabling the calculation of the linear actuator velocities via numerical differentiation.

In addition, Inertial Measurement Units (IMUs) were mounted on the upper carriage and on the individual components of the excavator arm. These IMUs provide the rotational motion

of the upper carriage and the arm links as well as the inclination, allowing reconstruction of the excavator arm kinematics.

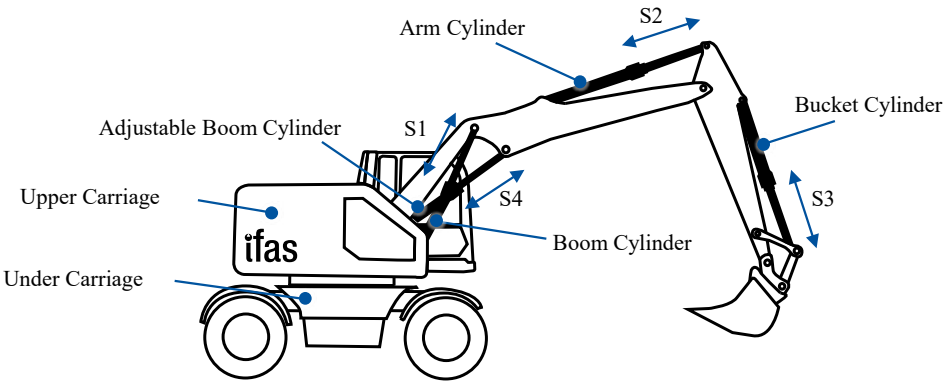


Figure 3.1. Measuring technology mounted on the test machine to record movement

3.2.2. Hydraulic Measurements

Figure 3.2 presents a simplified hydraulic circuit schematic, showing the locations of the sensors integrated into the hydraulic system. Numerous pressure sensors were installed at different points in the hydraulic circuit to measure:

- Actuator chamber pressures (excluding dozer shield and outrigger cylinders)
- Pressures at the main control valve (for the linear actuators)
- Load-sensing pressure p_{LS}
- Supply: Pressure of the main (p_{pump}) and secondary pumps (supply pilot pressure, braking and steering)
- Tank pressure
- Actuator pilot pressures

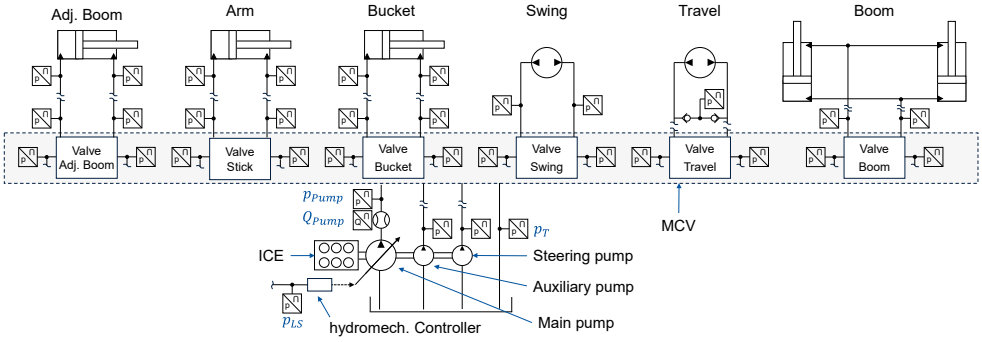


Figure 3.2. Measuring technology mounted to the hydraulic drivetrain

A turbine-type volume flow sensor was installed in the main pump outlet line to measure the pump flow. Note, that this introduces an additional pressure loss in the supply line, that is not present in series machines.

3.3. Performed Operation

Wheeled excavators are versatile machines typically used for digging, trenching, and grading operations [13]. Several defined operations were performed with the test machine during the measurement campaign; however, this paper focuses exclusively on an elevated 90° dig-and-dump-operation.

Figure 3.3 depicts the executed 90° dig-and-dump-operation. Each cycle consists, in simplified form, of:

1. Positioning the bucket at the excavation area
2. Filling the bucket with material
3. Swinging the upper carriage by approximately 90°
4. Dumping the material at the designated dump location
5. Returning to the initial position to start the next cycle

For this operation, the machine was positioned on top of a hill, approximately three meters above ground level. The material was loaded at ground level and dumped at excavator level. All operators performed repeated cycles under comparable conditions.

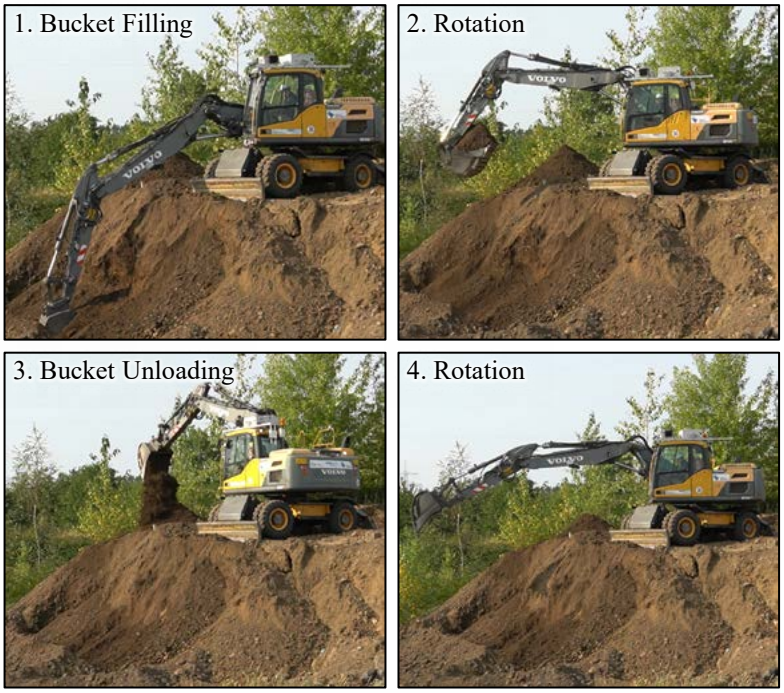


Figure 3.3. Example of performed 90° dig-and-dump-operation

The test series included two operator groups:

- Two professional operators (P1, P2), with substantial prior experience in operating excavators in daily construction work

- Five non-professional operators (NP1–NP5), with varying amounts of experience, ranging from intermediate to novice level

All operators executed the same predefined dig-and-dump-operation using the same machine and test setup (for example positioning), so that differences in performance can be attributed primarily to operator behaviour.

3.4. Calculated System Variables

From the recorded sensor data, several indicators were derived to quantify operator performance:

- Mean Pump power (compare *Mieth et al.* [2])
- Simultaneity level of actuator operation (compare *Kunze et al.* [1])
- Mean cycle time of the 90° dig-and-dump-operation (compare *Kunze et al.* [1])

The following subsections describe each indicator in detail.

3.4.1. Pump Power

Figure 3.5 shows the load-sensing pressure, pump pressure, flow and power for an excerpt of one of the measurements.

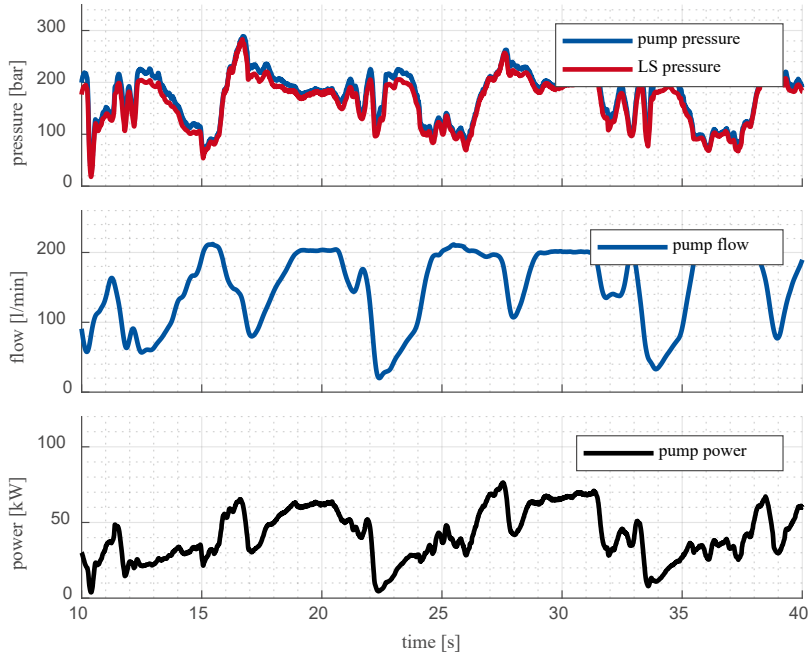


Figure 3.5. LS pressure, pump pressure, flow and power

The pump flow Q_{Pump} can be calculated from recorded impulses of the turbine-type volume flow sensor using a corresponding calibration curve. The instantaneous hydraulic power of

the Pump P_{Pump} is calculated from the pump flow Q_{Pump} and pressure p_{pump} (equation 3.1).

$$P_{Pump} = Q_{Pump} \cdot p_{pump} \qquad (3.1)$$

During the measurement campaign, the pump pressure sensor proved to be unreliable, meaning that for some of the performed measurement the pump pressure value was not available. Instead, the pump pressure is calculated according to equation (3.2),

$$p_{Pump} = p_{LS} + \Delta p_{ls} \qquad (3.2)$$

where p_{LS} is the load-sensing pressure (maximum supply pressure of all active actuators) and Δp_{ls} is the load-sensing pressure difference. In equation (3.2) it is assumed, that Δp_{ls} is a constant offset, however in reality this is not true, as figure 3.6 shows (measurement with intact pump pressure sensor)

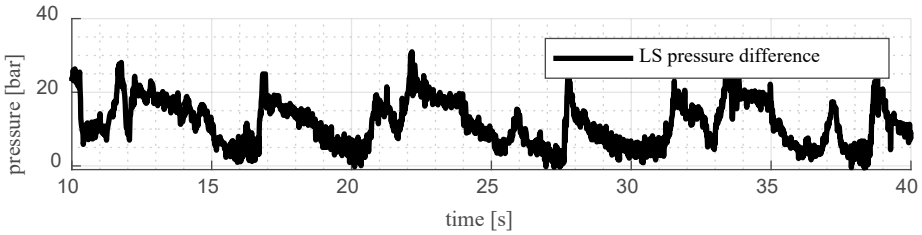


Figure 3.6. LS pressure difference

For the pump pressure calculation, a constant value derived from averaging was used. This introduces a small error in the pump power calculation. A more sophisticated approach would be evaluating the dependency of Δp_{ls} on pump pressure and flow, fit a function to the data und use this as the pump pressure estimation for the affected measurements, but this is outside the scope of this work.

With this approach focussing on pump power, potential operator influences on the total machine energy consumption or fuel consumption that arise from different engine operating points cannot be identified. The analysis is limited to the hydraulic subsystem. Specifically, the operator influence on the hydraulic power demand is captured, while effects related to engine and drivetrain efficiency lie outside the scope of this work.

The mean pump power $P_{Pump,mean}$ can be calculated by averaging P_{Pump} across operation duration.

3.4.2. Simultaneity Level

The simultaneity level, described by *Kunze et al.* [1] quantifies the mean number of actuators that are active in parallel during operation. Since operating several actuators simultaneously requires more coordination, a higher simultaneity level is typically associated with higher operator skill [1]. Within this work, the simultaneity is calculated as follows:

First a binary logic signal (filter) is generated for each actuator i, indicating whether this actuator is actively commanded at time t_n.

Figure 3.7. shows the measured stroke and pilot pressures for the bucket cylinder and the calculated cylinder velocity and activity filter signal for an excerpt of one of the measurements.

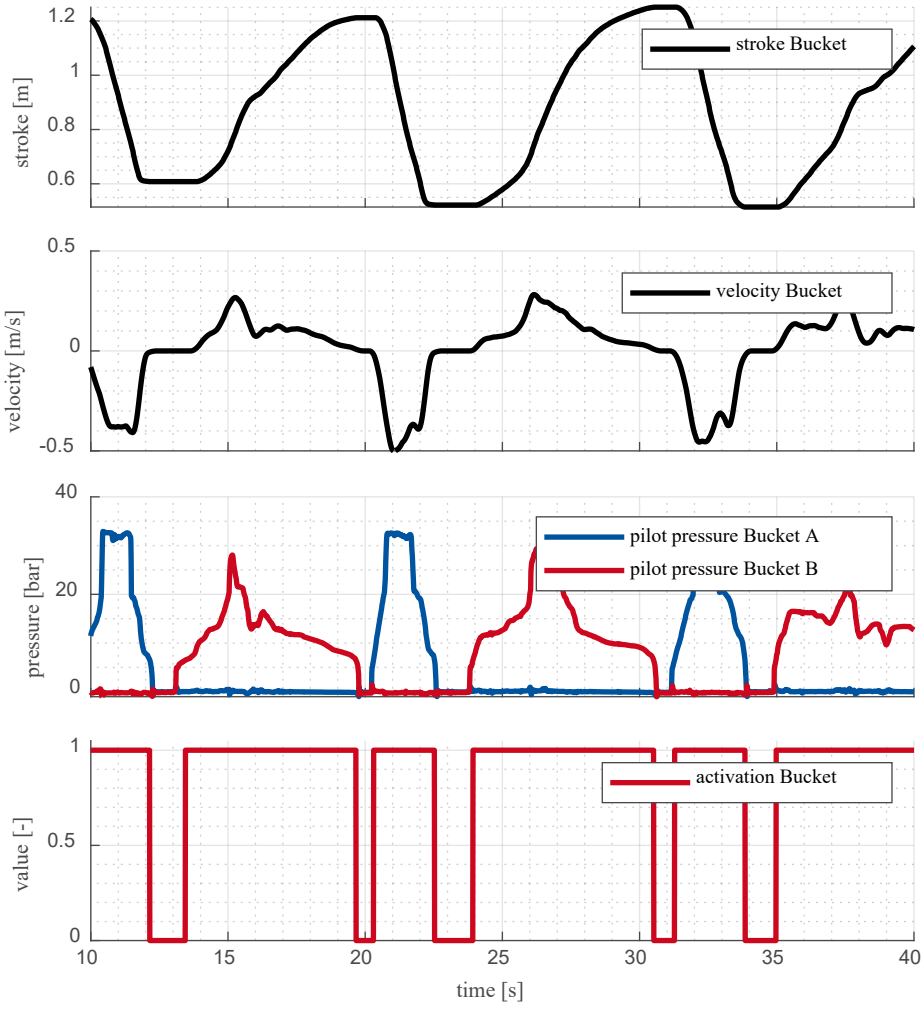


Figure 3.7. Bucket stroke, velocity, pilot pressure and activity filter signal

A straightforward approach would be to decide based on non-zero actuator (angular) velocity. However, numerical differentiation of the measured cylinder strokes introduces noise into the velocity signals. Filtering these velocities can in turn introduce phase delays, which complicates accurate detection of the onset and end of actuator movement.

To avoid this issue, the method used here relies on the pilot pressures of the hydraulic joysticks. The machine is equipped with hydraulic joysticks which pilot pressures actuate the spools of the main control valves. Thus, the pilot pressures $p_{\mathrm{pilotA},i}$ and $p_{\mathrm{pilotB},i}$ for each actuator i can be used as an indicator of operator command.

For each actuator i, the logic filter filter$_i[t_n]$is defined by equation 3.3 as

$$filter_i[t_n] = \begin{cases} 1\,, if\ p_{pilotA,i}[t_n] \geq p_{tresh}\ or\ p_{pilotB,i}[t_n] \geq p_{tresh} \\ 0\,, otherwise \end{cases} \quad (3.3)$$

where $p_{\text{thresh}}$is a pressure threshold determined experimentally. Note that only the main actuators were considered: Boom, Adjustable Boom, Arm, Bucket, Swing, Travel (not used during the evaluated operation). Auxiliary functions like the dozer blade and outriggers were not considered.

The simultaneity level is then obtained from the individual filters of all considered actuators by addition at each time step and then averaging over the entire observation period:

$$\text{simultaneity level} = \text{mean}(\textstyle\sum_i filter_i) \quad (3.4)$$

This yields the average number of actuators that are commanded simultaneously during the task.

3.4.3. Cycle Time

Cycle time is used in this study as a proxy for operator productivity. A more precise metric would be measuring the productivity directly in t/h, obtained by weighing the moved material and relating it to the working time. However, such measurements are beyond the scope and resources of this work.

All test operations were recorded on video. The number of completed buckets (i.e., cycles) was determined manually by reviewing the footage. For each operator and test run, the mean cycle time $t_{cycle,mean}$ was calculated with equation (3.5) as

$$t_{cycle,mean} = \frac{t_{eff}}{N_{buckets}} \quad (3.5)$$

where t_{eff} is the total effective working time and N_{buckets} is the number of counted buckets. Only successful cycles with a clearly full bucket were counted. This choice intentionally penalizes poor bucket filling, as cycles with low bucket fill would reduce the achievable t/h productivity and are therefore not treated as equivalent to full cycles. Non-productive pauses within were excluded from the total effective working time by cutting these intervals during the timing evaluation. In this way, the resulting mean cycle time reflects the net time required by each operator to perform one productive dig-and-dump-operation under the given test conditions.

3.5. Limitations

The methodology employed in this study is subject to several limitations and potential sources of uncertainty that must be considered when interpreting the results. First, the influence of operator fatigue and the sequencing of the different tests was not systematically controlled or analysed. Possible learning effects, warm-up effects, or performance degradation over time may therefore affect the measured performance, but investigating these effects in detail would require a substantially more extensive experimental design and is beyond the scope of this work.

Second, the characteristics of the excavated material were not captured. Neither the mass of the excavated material nor other properties like its moisture content was measured. The material was not uniform; it consisted mostly of soil with varying amounts of rubble.

Variations in density, cohesion, and moisture may influence digging resistance and thus the required hydraulic power and achievable productivity. Consequently, some portion of the variability in the results may be attributable to changing material conditions rather than purely to operator behaviour.

Third, the number of repetitions per operator and the total number of participating operators is relatively low. This limited sample size restricts the statistical robustness of the findings and makes it difficult to fully average out day-to-day performance fluctuations of individual operators. The results should therefore be interpreted as indicative trends rather than as statistically rigorous population estimates.

In addition, the determination of cycle times and bucket counting was carried out manually by reviewing video recordings. This process is inherently prone to human error. Although care was taken to ensure a consistent procedure, small inaccuracies cannot be ruled out.

Moreover, the chosen performance indicators simultaneity level and cycle time primarily describe the operator's handling of the machine during the active working phases. As previous work by the authors has shown [14]–[16], the machines spent a substantial proportion of the shift on typical road construction sites in idle, and the time share of different operations may fluctuate heavily from day to day. Consequently, depending on the type of construction site, the derived indicators may have limited explanatory power with respect to the operator's overall daily productivity. The results are likely more representative for quarry-like sites or other applications with relatively constant, repetitive operations than for urban road construction sites with highly variable tasks.

The dig-and-dump-operation used in the tests represents a simplified, standardized duty cycle that does not fully reflect the variability and irregularity of real-world operation. Instead, the focus and strength of the present methodology lies in comparing relative differences between operators under controlled, reproducible conditions, with particular emphasis on the hydraulic subsystem.

4. RESULTS

This chapter presents the results for the 90° dig-and-dump-operation. For each operator, the three calculated indicators mean cycle time, simultaneity level and mean pump power were evaluated. Pairwise relationships between these indicators are analysed using scatter plots (Figures 4.1–4.3) and the Pearson correlation coefficient r, which measures the linear association between two variables [17].

Figure 4.1 shows the mean cycle time and simultaneity level. A strong, negative correlation is observed ($r = -0.86$), measurements performed by operators with higher simultaneity levels achieve shorter cycle times. The two professional operators (P1, P2) lie close together, combining the lowest cycle times with the highest simultaneity levels. The non-professional operators (NP1–NP5) are more widely scattered, with generally lower simultaneity and longer cycles. NP1, who has the most experience among the non-professionals, approaches the region occupied by P1 and P2. This supports the interpretation of simultaneity level as an indicator of operator skill. It also confirms, that this effect found during trenching operation in [1, 2] appear to apply to 90° dig-and-dump-operation as well.

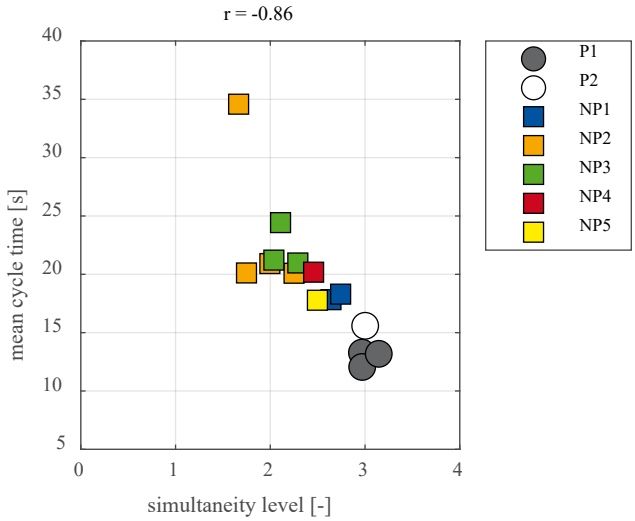


Figure 4.1. mean Cycle time and simultaneity level

Figure 4.2 plots mean cycle time against mean pump power. Here, a moderate negative correlation is found ($r = -0.49$). Overall, faster operation comes with a higher mean pump power, but the relationship is not strong. Further analysis can be conducted. This result aligns with results found during from trenching operations reported in [2]. An approach could be to analyse the net used power and system efficiency to find how effective the machines power is used by differently skilled operators.

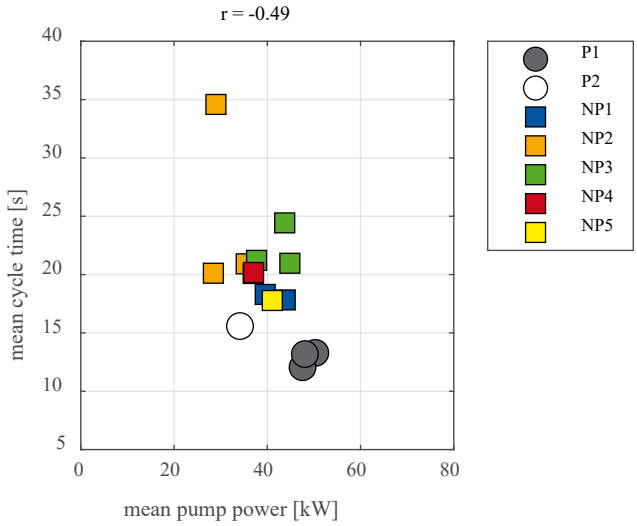


Figure 4.2. mean Cycle time and mean Pump Power

Figure 4.3 shows simultaneity level as a function of mean pump power. In this case, a moderate positive correlation is obtained ($r = 0.7$). Operators with higher simultaneity

levels tend to exhibit higher mean pump power, which is consistent with the expectation that more frequent parallel actuator operation increases hydraulic power demand. This result corresponds to a similar trend found during trenching operation in [2].

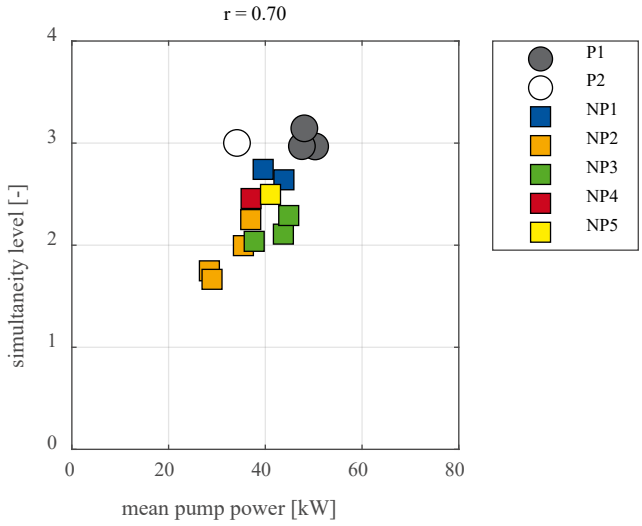


Figure 4.3. Simultaneity level and mean pump power

In summary, simultaneity level correlates strongly with cycle time and with mean pump power, whereas mean pump power and cycle time are only moderately related in this task.

5. Conclusion and Outlook

This work showed that operator behaviour has a clear and measurable influence on the performance of a wheeled excavator during a standardized 90° dig-and-dump-operation. Using an instrumented Volvo EW160D and a mixed group of professional and non-professional operators, it was possible to relate objective indicators such as cycle time, pump power and a pilot-pressure-based simultaneity level to differences in how the machine is handled. In particular, the simultaneity level proved to be a useful proxy for operator skill, capturing how effectively actuators are coordinated.

Building on these findings, future work should move from power-based to energy-based evaluation by relating input energy at the pump to useful work at the actuators, enabling a more direct assessment of hydraulic efficiency under different operators and load cycles. It is also promising to extend the analysis to the additional recorded operations to understand how robust the identified relationships are across varying tasks. In the longer term, the indicators could be used in operator training or assistance systems and help improve both productivity and energy efficiency of mobile hydraulic machines in real-world applications.

6. References

[1] Kunze, G., Mieth, S., Voigt, S. 'Bedienereinfluss auf Leistungszyklen mobiler Arbeitsmaschinen', ATZoffhighway, Vol. 4, Nr. 1, pp. 70–79, 2011. DOI: 10.1365/s35746-011-0010-2.

[2] Mieth, S., Voigt, S., Kunze, G. 'Typical performance cycles of mobile machinery taking into account the operator influence', 8th International Fluid Power Conference, 2012.

[3] Filla, R., Ericsson, A., Palmberg, J.-O. 'Dynamic Simulation of Construction Machinery: Towards an Operator Model', 2005. DOI: 10.48550/arXiv.cs/0503087.

[4] Filla, R. 'A Methodology for Modeling the Influence of Construction Machinery Operators on Productivity and Fuel Consumption', 2009. DOI: 10.1007/978-3-642-02809-0_65.

[5] Filla, R. 'Quantifying Operability of Working Machines', Dissertation, Linköping University, Linköping, 2011. DOI: 10.13140/RG.2.1.4702.0004.

[6] Filla, R. 'Study of a method for assessing operability of working machines in physical and virtual testing', International Journal of Vehicle Systems Modelling and Testing, Vol. 7, Nr. 3, pp. 209, 2012. DOI: 10.1504/IJVSMT.2012.048939.

[7] Filla, R., Olsson, E. M. G. et al. 'A Case Study on Quantifying the Workload of Working Machine Operators by Means of Psychophysiological Measurements', Proceedings from the 13th Scandinavian International Conference on Fluid Power, June 3-5, 2013, Linköping, Sweden, pp. 293–304, 2013. DOI: 10.3384/ecp1392a29.

[8] Frank, B., Skogh, L. et al. 'On Increasing Fuel Efficiency by Operator Assistance Systems in a Wheel Loader', Proceedings from the 2012 International Conference on Advanced Vehicle Technologies and Integration, 2012. DOI: 10.13140/RG.2.1.3129.1362.

[9] Frank Bobbie, Skogh Lennart, Alaküla Mats: 'On wheel loader fuel efficiency difference due to operator behaviour distribution', 2012.

[10] Sturm, C. 'Bewertung der Energieeffizienz von Antriebssystemen mobiler Arbeitsmaschinen am Beispiel Bagger', KIT Scientific Publishing, 2015. DOI: 10.5445/KSP/1000047809.

[11] Inderelst, M., Weidner, F. et al. 'Quantification of Energy Saving Influencers in a 21t Excavator Hydraulic System – A Holistic Investigation?', Fluid power networks: proceedings: 19th - 21th March 2018: 11th International Fluid Power Conference Vol.1, 2018. DOI: 10.18154/RWTH-2018-224628.

[12] Fecke, M. 'Bewertung der Energieeffizienz von Baumaschinen mithilfe einer praxisnahen Lastzyklusentwicklung für einen In-Situ-Test', Dissertation, Bergische Universität Wuppertal, Wuppertal, 2018.

[13] Vukovic, M. 'Hydraulic hybrid systems for excavators', Dissertation, RWTH Aachen, Aachen, 2017.

[14] Schmitz, K., Deuster, S. et al. 'Recent Developments on Mobile Machine Research – Electrification, Connectivity, Sustainable Pressure Media', 2026. DOI: 10.1007/978-3-032-07392-1_24.

[15] Sprink, J., Sender, B., Schmitz, K. 'Digital Twin-Based Classification of Hydraulic Excavator Duty Cycles in Road Construction', 2025. DOI: 10.1007/978-3-031-84505-5_7.

[16] Sprink, J., Sender, B., Schmitz, K. 'Echtzeitauswertung von Messdaten zur Erstellung eines Zyklusprofils von Baggern im Straßenbau', 13. Kolloquium Mobilhydraulik: Karlsruhe, 8./9. Oktober 2024, 2024.

[17] Puhani, J. 'Statistik - Einführung mit praktischen Beispielen', 14. Aufl., Springer Fachmedien Wiesbaden, Wiesbaden, 2025. DOI: 10.1007/978-3-658-48352-4.

Biographies



Bernhard Sender received the bachelor's degree in mechanical engineering and the master's degree in product development from RWTH Aachen University. Since 2023, he is working as a research associate at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. His research focusses on hydraulic systems in mobile machines.



Johannes Sprink studied mechanical engineering at RWTH Aachen University. Since 2021, he has been a research associate at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. Johannes is the group leader of the Mobile Machines research group, which focuses on fluid-mechatronic and information technology innovations at the machine level. He investigates greenhouse gas emissions in the fluid power industry; his research also includes analysing operating cycles of hydraulic excavators.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Optimal preliminary design of electro-mechanical actuators over drive-cycle for mobile machinery

Róger Mateus Sehnem, Samuel Kärnell and Liselott Ericson

Linköping University, roger.mateus.sehnem@liu.se

Abstract.

Electromechanical actuators (EMAs) are increasingly being considered as replacements for hydraulic systems in mobile machinery due to electrification trends and efficiency requirements. Conventional EMA design approaches select representative steady-state operating points from the drive cycle, potentially yielding suboptimal designs when systems must operate over wide dynamic ranges. This paper explores the benefits of incorporating full drive-cycle dynamics into the preliminary design process using a bilevel optimization framework. A simplified hypothetical crane drive cycle demonstrates the approach by comparing dynamic optimization against the more conventional static multi-design-point optimization. Despite a higher computational cost per optimization round, the dynamic approach consistently outperformed the static approach, achieving not only copper losses that are four times lower than using the static optimization approach, but also smaller diameter, and superior trajectory tracking while respecting voltage and current limits. It is shown that the performance advantages persist when evaluated on drive cycles outside the optimized one. While the simplified models (neglecting iron losses, thermal limits, and saturation effects) limit direct applicability to real designs, the results demonstrate significant potential benefits of drive-cycle-aware optimization for preliminary EMA design in mobile machinery applications.

Keywords. Optimization, EMA, PMSM, Mobile machines, Optimal control.

1. Introduction

With the current global trend towards electrification and stricter regulations on emissions, mobile machinery is being pushed towards being electrically driven. Since current battery technology has lower power density than traditional fuel-driven solutions, the efficiency of the actuators is now becoming a much more relevant design goal. This is one of the reasons why electromechanical actuators (EMAs) are now being considered as, to some extent, replacements for hydraulics in mobile machinery.

Designing an EMA can be framed as a complex optimization problem that needs to consider several physical domains. Optimality in this context can often be framed as maximizing the power density of the actuator and its efficiency while ensuring that the relevant constraints are met or not exceeded, e.g., maximum current and desired power. Most systems in engineering can be precisely described by nonlinear differential equations. Dealing with differential equations in preliminary design is, however, difficult, and it is a common practice to simplify the design process by considering only static or steady-state behavior. To this end, the designer needs to

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

choose a design point that is representative of the system's operation range. In cases where the system operates in a wide range of conditions, as in EMAs for mobile machinery, this choice is by no means trivial, and this approach can lead to suboptimal designs.

In reality, the designer is either attempting to optimize the performance of the system over the entire drive cycle on which the EMA will operate or a significant reference drive cycle exemplified for a lorry crane application in [1]. In any case, for drive cycles that require significant transient behavior of the system, it is reasonable to expect that considering the full dynamics of the system can lead to better designs than using steady state approaches. The main goal of this paper is to show that this is indeed the case.

1.1. *Previous Work*

Considering more of the drive cycle in design has been proposed by several authors in different ways. For instance, [2] proposed a convex model fitted on high-fidelity simulation data to allow efficient convex optimization and global optimum results. The approach considers a simplified dynamical model, thus attempting to be more representative of the system behavior. Another approach is to somehow aggregate several different steady state operating points into some that are then used to design the electrical machine, this is the approach of [3], where the energy center of gravity (ECG) of the torque-speed plot is used to create 8 clusters, which are then weighted according to their energy usage in the drive cycle and then have their centers used as steady state design points. Another example of this approach is [4], which uses a multi-objective optimization approach to optimize the design of a PMSM considering multiple operating points and uses simple formulas to estimate the transient losses in the optimization problem. However, all those methods have strong simplifying assumptions about the true dynamics of the system.

The main challenge in the attempt to consider the full dynamics of the system in the design process is that the output of the system depends on the control input, and if one chooses a suboptimal control strategy, the performance of the system will be underestimated. The problem of finding an optimal control strategy for a given system is known as an optimal control problem (OCP), which is known to be numerically difficult to solve in the nonlinear case [5, 6]. Considering the optimization of the system parameters together with the control inputs is thus expected to be an even more sensitive and hard problem.

However, when one considers linear systems with quadratic cost functionals in the state deviations and control inputs, known as the Linear Quadratic Regulator (LQR) problem [5, 6], the problem has an analytical solution. Although the LQR problem is designed for linear systems, it can be used to design controllers for nonlinear systems with the usage of some form of linearization technique, giving a controller that in practice works well for a wide range of operating conditions in many applications.

1.2. *Contribution*

To address the conjoint optimization of an EMA and the solution of the OCP in a sensible way, in this work, a two-step optimization process is used. The lower problem is a standard OCP, which, for a given system, finds the optimal control inputs that minimize a relevant cost functional. The upper problem then chooses the system parameters to minimize a cost function that penalizes energy loss, tracking error and the size of the EMA, which is evaluated by solving the lower OCP. Dividing the optimization as described is known as bilevel optimization and has been used in several applications [7]. In this way, it is possible to directly consider the full dynamics of the system in the design process. To simplify the OCP problem, a standard field-oriented control with feedback linearization and P controllers for the currents and velocity and a PI controller for the position are used. The controller gains are chosen using the LQR method on an error system, allowing the efficient solution of the internal OCP for a given set of system parameters.

As a proof of concept, this paper explores the potential benefits of considering the full dynamics of the system in the preliminary design of an EMA for a simplified hypothetical loader crane drive cycle. A standard PMSM model from [8] with parameters calculated from physical dimensions is considered, together with very simple models for the gears and ball screw. The approach is then compared to using only some relevant design points considering only the steady-state dynamics of the system, analogous to the approach of [4]. The PMSM model used in this work neglects several effects that are present in real machines, such as saturation, thermal limits, eddy currents, hysteresis losses and non-sinusoidal magnetic field distributions. The same can be said about the mechanical components, their models are overly simplified to be useful in the design of a real EMA. However, the models are considered to be sufficient for the purpose of this work, which is to explore the potential benefits of considering the full dynamics of the system in the design of an EMA for a hypothetical drive cycle.

The results show that the EMA optimized considering the full system dynamics dissipates four times less energy to follow the design trajectory, all while being able to respect the electrical constraints and having the desired geometry.

2. Electro-mechanical Actuator Model

The EMA model consists of a PMSM, a gearbox and a ball screw and is shown schematically in Figure 2.1. In the following subsections, the model of each component is described in detail.

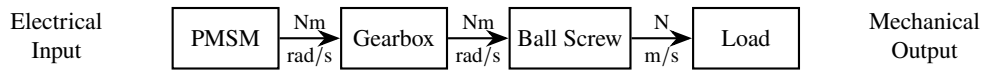


Figure 2.1 Schematic representation of the EMA system considered. The PMSM drives a ball screw through a gearbox, which then moves a load linearly. The electrical subsystem is represented on the left, while the mechanical subsystem is represented on the right. The arrows indicate the power flows in the system.

2.1. *Permanent Magnet Synchronous Machine*

The PMSM model is traditionally composed of electro-mechanical constants that are usually obtained from experimental data or finite element analysis. In this work, however, the PMSM parameters are estimated from simplified physical considerations of the motor geometry and magnetic field distribution. This not only allows a more direct connection between the physical dimensions of the motor and its performance but also has very low computational complexity. The model follows the developments in [8] closely, with the additional parametrization for the torque constant K_m and stator resistance R_S from physical dimensions, which are not directly given in the reference. The PMSM model assumes a sinusoidal magnetic field distribution, no saturation, eddy current or hysteresis losses and is given by the following set of differential equations in the stator reference frame [8]:

$$\begin{aligned} \frac{d\theta_r}{dt} &= \omega, \\ J\frac{d\omega}{dt} &= K_m\left(-i_{Sa}\sin(n_p\theta_r) + i_{Sb}\cos(n_p\theta_r)\right) - f(\omega) - \tau_L, \\ L_S\frac{di_{Sa}}{dt} &= -R_S i_{Sa} + K_m\sin(n_p\theta_r)\omega + u_{Sa}, \\ L_S\frac{di_{Sb}}{dt} &= -R_S i_{Sb} - K_m\cos(n_p\theta_r)\omega + u_{Sb}. \end{aligned} \tag{2.1}$$

where θ_r is the rotor position, ω is the rotor angular velocity, i_{Sa} and i_{Sb} are the stator currents in phases a and b, u_{Sa} and u_{Sb} are the stator voltages in phases a and b, J is the rotor inertia,

K_m is the motor torque constant, n_p is the number of pole pairs. L_S is the stator inductance, R_S is the stator resistance, $f(\omega)$ is a friction model and τ_L is the load torque.

The derivation of L_S closely follows the developments in [8]. Both L_S and K_m assume a full-pinched winding, but the model can be adapted to short-pinched or long-pinched windings by adding the winding factor κ_w to the equations for L_S and K_m.

2.1.1. *Stator Inductance from Physical Dimensions*

For the parameters calculation, assume that the magnetic field distribution in the air gap is approximated by a sinusoidal distribution for each phase. Following [8], the total magnetic field in the air gap due to the stator currents is then given by

$$\boldsymbol{B}_s(i_{Sa}, i_{Sb}, r, \theta) = \boldsymbol{B}_{Sa} + \boldsymbol{B}_{Sb} = \frac{\mu_0 N_s r_r}{2g} \frac{1}{r} \left(i_{Sa} cos(\theta) + i_{Sb} \sin(\theta)\right) \boldsymbol{e}_r, \tag{2.2}$$

where μ_0 is the permeability of free space, N_s is the number of turns per phase, r_r is the rotor radius, g is the air gap length, r is the radial position in the air gap, θ is the angular position in the air gap, and $\boldsymbol{e}_r$ is the radial unit vector. The magnets in the rotor of the PMSM are distributed such that the magnetic field in the air gap generated by the rotor is also sinusoidal, thus

$$\boldsymbol{B}_R = B_m \frac{r_r}{r} \cos(\theta - \theta_r) \boldsymbol{e}_r, \tag{2.3}$$

where B_m is the peak magnetic field in the air gap generated by the rotor.

The total flux linkage of the stator describes how much magnetic flux flows through the stator windings and is a useful quantity to calculate K_m and L_S. It is defined as

$$\lambda_{Sa}(\boldsymbol{i}_s, \theta_R) = {}_s\lambda_{Sa}(\boldsymbol{i}_s, \theta_r) + {}_r\lambda_{Sa}(\theta_r) = \int_{\mathbb{L}_{Sa}} \boldsymbol{B}_s d\boldsymbol{S} + \int_{\mathbb{L}_{Sa}} \boldsymbol{B}_R d\boldsymbol{S}, \tag{2.4}$$

where $\boldsymbol{i}_s = \begin{bmatrix} i_{Sa} & i_{Sb} \end{bmatrix}^T$ and $\mathbb{L}_{Sa}$ is a set of surfaces that envelops all stator windings of phase a. To calculate ${}_s\lambda_{Sa}$, one calculates the flux in a single loop of the stator and then integrates over all loops. If the given loop is at an angle θ from the a phase of the stator, then its other side is, considering full-pinched windings, at $\theta - \pi$. Thus, the flux in the cylindrical surface is given, for the a phase of the stator, by

$${}_s\phi_{Sa}(\boldsymbol{i}_s, \theta) = \int_{\theta-\pi}^{\theta} \int_0^{l_1} \boldsymbol{B}_S(\boldsymbol{i}_s, \theta)|_{r_s} \boldsymbol{e_r} \cdot (r_s d\theta dz \boldsymbol{e_r}) = \frac{\mu_0 r_r l_1 N_s}{g} \left(i_{Sa} \sin(\theta) - i_{Sb} \cos(\theta)\right). \tag{2.5}$$

The flux linkage ${}_s\lambda_S$ between θ and $\theta + d\theta$ is then given by multiplying by turns density distribution, as

$$d\,{}_s\lambda_{Sa} = {}_s\phi_{Sa}(\boldsymbol{i}_s, r, \theta) \frac{N_s}{2} \sin(\theta) d\theta = \frac{\mu_0 r_r l_1 N_s}{g} \left(i_{Sa} \sin(\theta) - i_{Sb} \cos(\theta)\right) \frac{N_s}{2} \sin(\theta) d\theta, \tag{2.6}$$

where the $\frac{N_s}{2} \sin(\theta)$ term comes from the winding distribution in the stator. Integrating Equation (2.6) from 0 to π gives the total flux linkage in phase a of the stator due to the stator magnetic field, which is given by

$${}_s\lambda_{Sa}(\boldsymbol{i}_s) = \frac{\mu_0 r_r l_1 \pi N_s^2}{4g} i_{Sa} = \frac{\mu_0 l_1 l_2 \pi N_s^2}{8g} i_{Sa} \coloneqq L_S i_{Sa}, \tag{2.7}$$

where $l_2 = 2r_r$ is the stator diameter. Equation (2.7) defines the stator inductance L_S. The flux linkage in phase b of the stator can be calculated in a similar way, giving the same result for L_S.

2.1.2. *Torque Constant from Physical Dimensions*

To get the rotor flux linkage, it is important to account for the leakage κ in the magnetic field. Following [8], the rotor flux linkage at the stator side is given by modifying Equation (2.3) to

$$\boldsymbol{B}_R|_{r_s} = \kappa B_m \frac{r_r}{r_s} \cos(\theta - \theta_r)\boldsymbol{e}_r, \tag{2.8}$$

where r_s is the stator radius.

To calculate ${}_r\lambda_{Sa}$, the same approach as in the stator flux linkage is used. The flux in a single loop of the stator at angle θ is given by

$${}_r\phi_{Ra} = \int_{\theta-\pi}^{\theta}\int_0^{l_1} \kappa B_m \frac{r_r}{r_s}\cos(\theta - \theta_r) r_s dr dz = 2\kappa B_m l_1 r_r \sin(\theta - \theta_r). \tag{2.9}$$

With that is possible to write the rotor flux linkage differential, using the sinusoidal winding for the stator phase a, as

$$d\,{}_r\lambda_{Sa} = \phi_{Ra}\frac{N_S}{2} sin(\theta) d\theta = \kappa B_m l_1 r_r N_S \sin(\theta - \theta_r) sin(\theta) d\theta. \tag{2.10}$$

Integrating Equation (2.10) from 0 to π gives

$${}_r\lambda_{Sa} = \int_0^{\pi} \kappa B_m l_1 r_r N_S \sin(\theta - \theta_r) sin(\theta) d\theta = \kappa B_m l_1 r_r N_S \frac{\pi}{2}\cos(\theta_r) \coloneqq K_m \cos(\theta_r), \tag{2.11}$$

which defines the torque constant K_m.

2.1.3. *Stator Resistance from Physical Dimensions*

The stator resistance is greatly defined by the length of the wire used in the winding, which greatly depends on how it is wound. Figure 2.2 shows how one loop is wound in the stator.

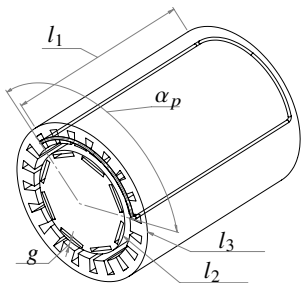


Figure 2.2 3D view of the PMSM with the stator and rotor, showing the stator length l_1, stator outer diameter l_3, stator inner diameter l_2, air gap g and winding with pitch angle α_p.

Figure 2.2 shows a single loop of winding. This winding is made with a wire of diameter ϕ_w, and it is of a distributed winding type with pitch angle α_p. Assuming no skin and proximity effects, the resistance R_S of the stator winding can be estimated from the total length l_s of wire used in the stator, its cross section S_s and the electrical resistivity of the wire $\rho_{e,w}$ using

$$R_S = \frac{\rho_{e,w} l_s}{S_s}. \tag{2.12}$$

In a distributed winding stator, the total length of wire l_s is given by the arc made by the angle α_p, which is assumed to be at the stator's inner diameter l_3, together with the length of wire needed to traverse the stator's length l_1. Thus, the total length of wire is given by

$$l_s = (l_3\alpha_p + 2l_1) N_S n_p n_f, \tag{2.13}$$

where n_f is the number of phases, l_3 is given by

$$l_3 = l_2 + 2g. \tag{2.14}$$

Thus, the resistance of the stator winding is fully defined by the physical dimensions of the motor as

$$R_S = \frac{(l_3\alpha_p + 2l_1)N_S n_p n_f \rho_{e,w}}{S_s}. \tag{2.15}$$

2.2. *Gearbox and Ball Screw*

The gearbox and ball screw mechanisms are modeled in a very simplified way, assuming a constant efficiency η_g and η_{bs} for the gearbox and ball screw, respectively. Each of the mechanisms is modeled as a simple torque and speed transformation, given by

$$\begin{aligned} \tau_{out} &= \eta_g G \tau_{in}, \\ \omega_{out} &= \frac{\omega_{in}}{G}, \end{aligned} \tag{2.16}$$

for the gearbox, where G is the gear ratio, τ_{in} and ω_{in} are the input torque and speed, respectively, and τ_{out} and ω_{out} are the output torque and speed, respectively. The ball screw is modeled in a similar way, given by

$$\begin{aligned} F_{out} &= \eta_{bs} \frac{2\pi}{p} \tau_{in}, \\ v_{out} &= \frac{p}{2\pi} \omega_{in}, \end{aligned} \tag{2.17}$$

where p is the pitch of the ball screw, τ_{in} and ω_{in} are the input torque and speed, respectively, and F_{out} and v_{out} are the output force and speed, respectively.

3. CONTROLLER

To be able to evaluate the performance of the EMA over the drive cycle, a control strategy that allows the EMA to follow the desired trajectory needs to be defined. This can be achieved by following [8] and using a standard field-oriented control with feedback linearization, P controllers for the currents and velocity and a PI controller for the position. The controller gains are in turn chosen using the LQR method, allowing the efficient computation of the optimal controller gains for a given set of system parameters.

3.1. *Field-oriented Control*

The field-oriented control (FOC) is a widely used control strategy for PMSMs, which allows decoupling of the torque and flux control. The FOC is based on the transformation of the stator currents and voltages from the stationary reference frame to the rotating reference frame, which is aligned with the rotor magnetic field, known as the d-q reference frame.

To transform Equation (2.1) to the d-q reference frame, one simply rotates the stator currents and voltages by the electrical angle $n_p\theta$ using the counter-clockwise rotation matrix [8]. Applying the transformation to the PMSM model given by Equation (2.1), resulting in the

following equations in the d-q reference frame

$$\begin{aligned} L_S \frac{di_d}{dt} &= -R_S i_d + n_p L_S \omega i_q + u_d, \\ L_S \frac{di_q}{dt} &= -R_S i_q - n_p L_S \omega i_d - K_m \omega + u_q, \\ J \frac{d\omega}{dt} &= K_m i_q - f(\omega) - \tau_L, \\ \frac{d\theta}{dt} &= \omega, \end{aligned} \tag{3.1}$$

where i_d and i_q are the d-axis and q-axis stator currents, respectively, and u_d and u_q are the d-axis and q-axis stator voltages, respectively.

For a given reference trajectory $r(t) = \begin{bmatrix} \theta_{\text{ref}}(t) & \omega_{\text{ref}}(t) & \alpha_{\text{ref}}(t) \end{bmatrix}^T$, it is possible to calculate the reference state and control inputs trajectories needed to follow the reference trajectory. By considering its steady-state condition, it is possible to obtain the $i_{d,\text{ref}}$ that maximizes the torque, which in turn allows the obtention of $i_{q,\text{ref}}$ from ω_{ref} and α_{ref}. With those, the reference voltages needed to follow the $r(t)$ directly follow.

With the reference system dynamics and control inputs, it is possible to create error dynamics by subtracting the reference dynamics from the system dynamics in Equation (3.1). This results in a nonlinear system that can be linearized using the following feedback linearization [8]

$$\begin{aligned} u_d &= -n_p L_S \omega i_q + u_{d,\text{ref}} + n_p L_S \omega_{\text{ref}} i_{q,\text{ref}} - L_S v_d, \\ u_q &= n_p L_S \omega i_d + u_{q,\text{ref}} - n_p L_S \omega_{\text{ref}} i_{d,\text{ref}} - L_S v_q, \end{aligned} \tag{3.2}$$

where v_d and v_q are the new control inputs for the error dynamics and the variables with subscript *ref* are the reference values. Using Equations (3.2) in the error dynamics results in a linear system that is amenable to standard linear control techniques.

To eliminate the steady-state error, an integrator state for the position error is introduced, $e_z = \int e_\theta dt$, which gives the augmented error state vector $\boldsymbol{e} = \begin{bmatrix} e_{i_d} & e_{i_q} & e_\omega & e_\theta & e_z \end{bmatrix}^T$. The augmented error dynamics is then given by

$$\frac{d\boldsymbol{e}}{dt} = \begin{bmatrix} -\frac{R_S}{L_S} & 0 & 0 & 0 & 0 \\ 0 & -\frac{R_S}{L_S} & -\frac{K_m}{L_S} & 0 & 0 \\ 0 & \frac{K_m}{J} & -\frac{f'(\omega)}{J} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \boldsymbol{e} + \begin{bmatrix} \frac{1}{L_S} & 0 \\ 0 & \frac{1}{L_S} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{v}, \tag{3.3}$$

where $\boldsymbol{v} = \begin{bmatrix} v_d & v_q \end{bmatrix}^T$, for which a state feedback controller can be designed using the LQR method.

3.2. *Linear Quadratic Regulator Controller Design*

The setup of the LQR problem consists of finding the control input $\boldsymbol{u}(t)$ that minimizes the quadratic cost functional

$$J = \int_0^\infty \left(\boldsymbol{e}(t)^T Q \boldsymbol{e}(t) + \boldsymbol{u}(t)^T R \boldsymbol{u}(t) \right) dt, \tag{3.4}$$

where Q is a positive semi-definite matrix and R is a positive definite matrix. For a system given by

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}(t) + B\boldsymbol{u}(t), \tag{3.5}$$

the LQR problem consists of finding the control input $\boldsymbol{u}(t)$ that minimizes the cost functional of Equation (3.4), subject to the system dynamics of Equation (3.5).

This optimal control problem (OCP), for the linear systems, has an analytical solution given by the state feedback law, that is, a function $\boldsymbol{u}(x(t))$ that minimizes the cost functional of Equation (3.4), given by

$$\boldsymbol{u}(t) = -R^{-1}B^T S(t)x(t), \tag{3.6}$$

where S is the solution of the matrix Riccati equation. To simplify the notation, the Kalman gain matrix is defined as

$$K(t) = R^{-1}B^T S(t). \tag{3.7}$$

Since the matrix $K(t)$ saturates to a constant matrix K as $t \to \infty$ [6], it is then possible to use what is known as the suboptimal control law, given by

$$\boldsymbol{u}(t) = -K(\infty)x(t). \tag{3.8}$$

To calculate the Kalman gain matrix $K(\infty)$, one solves the algebraic Riccati equation, given by

$$0 = A^T S + SA - SBR^{-1}B^T S + Q, \tag{3.9}$$

yielding the steady-state solution $S(\infty)$, which can be used to calculate the steady-state Kalman gain matrix $K(\infty)$ using Equation (3.7).

3.3. *Controller Summary*

By defining state and an input penalizing matrices Q and R, it is possible to calculate the optimal state feedback gain matrix $K(\infty)$ by simply solving the algebraic Riccati Equation (3.9). This can be done using standard numerical methods optimized for this task and available in scientific computing libraries in Python [9], turning the lower level problem of finding the controller gains into a computationally efficient task.

The advantages of using the LQR method in this context are that solving the OCP is computationally efficient and that the resulting controller has good robustness properties [6], which is important since the same decision matrices Q and R are used for all the different EMA design attempts made by the optimizer.

It is important to note that the controller itself is based on a linearized model of the EMA, thus, it is not necessarily optimal for the nonlinear system. However, it is a good starting point and it is also representative of the controllers used in practice.

4. Optimization Problem Formulation

With the model and controller defined, the optimization problem can be formulated. The goal of the optimization problem is to find the design variables that minimize the cost function, while satisfying the constraints. The design variables are the physical dimensions of the EMA, which in turn define the parameters of the model. In the next subsections, the cost function of the upper problem is motivated and defined, the constraints are defined and the design variables are presented.

4.1. *Cost Function*

To examine the usefulness of considering the whole drive cycle in the optimization problem, the optimization of an EMA for a crane application is considered. This kind of application is nowadays dominated by hydraulic actuators, which have a high power density and are able to deliver high forces. To compete with hydraulic actuators, the designed EMA needs to have similar characteristics. There are, however, some differences between the two technologies that can be more easily accommodated in a crane application. One of those differences is related to

the power density, it is probably much easier to accommodate a higher length than a larger diameter, thus, the power density might not be as critical as the diameter power density. This is one of the considerations made in the cost function, instead of trying to minimize the volume of the EMA, we try to minimize its diameter. Moreover, whatever the designed EMA is, it should be able to closely follow the desired load trajectory while respecting the constraints, the major ones being the maximum voltage and current that can be supplied to the EMA's motor.

One of the major motivators of substituting hydraulic actuators with EMAs is the potential increase in system efficiency. Thus, the cost function should also consider the efficiency of the EMA over the drive cycle. This is done by considering the total dissipated energy, which should be minimized. An important technical aspect to consider is that the optimizer might suggest designs that are not able to follow the desired trajectory and or are unstable. Those designs should be penalized in the cost function to prevent the optimizer from suggesting EMAs that are close to those designs, and thus, wasting computational resources evaluating those designs. This is done by penalizing the cost function with the squared residuals $\boldsymbol{r}(t)$ of the solver used to simulate the EMA dynamics over the drive cycle.

The above considerations motivate the following cost function for the dynamic optimization problem

$$J_{\text{dyn}}(\boldsymbol{x}(t)) = w_1 \int_0^T \frac{e_\theta(t)^2}{\max(|\theta_{\text{ref}}(t)|)} dt + w_2 l_2 + w_3 \frac{1}{E_{\text{ref}}} \int_0^T P_{\text{loss}}(t) dt + w_4 \int_0^T \|\boldsymbol{r}(t)\|^2 dt, \tag{4.1}$$

where the first term penalizes the position tracking error, the second term penalizes the diameter of the EMA, the third term penalizes the total energy loss over the drive cycle, and the fourth term penalizes the residuals of the EMA dynamics. The weights w_1, w_2, w_3 and w_4 are used to balance the different terms in the cost function, and E_{ref} is given by

$$E_{\text{ref}} = \int_0^T |F_{\text{ref}}(t) v_{\text{ref}}(t)| dt. \tag{4.2}$$

The other normalization terms in Equation (4.1) are used to make the different terms in the cost function have similar magnitudes.

For the static optimization approach, the cost function is similar to Equation (4.1), but instead of integrating over the entire drive cycle, only a few significant design points are considered in the drive cycle. Thus, the cost function for the static approach is given by

$$J_{\text{stat}}(\boldsymbol{x}) = w_2 l_2 + w_3 \frac{1}{E_{\text{ref}}} \sum_{k=1}^{N_{dp}} P_{\text{loss},k} \Delta t_k \tag{4.3}$$

where N_{dp} is the number of design points considered, Δt_k is the time spent at design point k, $P_{\text{loss},k}$ is the power loss at design point k.

4.2. *Constraints*

One of the main constraints in designing an EMA is the maximum voltage and current that can be supplied to its motor. In real applications, another important constraint is the thermal limits of the motor, which are usually the limiting factor in the motor's power density. However, the main focus of this work is to show the potential of considering the entire drive cycle in the optimization problem, and thus, it suffices to consider only the electrical constraints, given by

$$\begin{aligned} u_{\text{mag}}(t) &= \sqrt{u_d(t)^2 + u_q(t)^2} \leq u_{\max}, \\ i_{\text{mag}}(t) &= \sqrt{i_d(t)^2 + i_q(t)^2} \leq i_{\max}, \end{aligned} \tag{4.4}$$

where $u_{\max}$ and $i_{\max}$ are the maximum voltage and current magnitude that can be supplied to the motor, respectively.

The constraints of Equation (4.4) need to be satisfied at each time instant. To enforce those constraints in the dynamic optimization approach, the penalty method is used, adding the following penalty terms to the cost function

$$G_{\text{dyn}}(\boldsymbol{x}(t)) = w_5 \max\left(0, \max(u_{\text{mag}}(t)) - u_{\max}\right)^2 + w_5 \max\left(0, \max(i_{\text{mag}}(t)) - i_{\max}\right)^2, \quad (4.5)$$

where w_5 is the penalty weight.

The static approach needs to consider the constraints of Equation (4.4) at each design point k, but to account for the steady-state dynamics, there is an additional constraint that needs to be satisfied, the residual of the steady-state dynamics $\boldsymbol{r}_k$, that is, the right hand side of Equation (3.1) with the derivatives zeroed out and evaluated at design point k. Since the steady-state, optimization-driven design needs to satisfy a set of constraints at every operating point, the problem is much more sensitive, and it was therefore also necessary to penalize non-integer values of the number of pole pairs n_p and the number of turns per phase N_S. The constraints are enforced, in a similar way as in the dynamic case, that is, by adding the following penalty terms to the cost function

$$G_{\text{stat}}(\boldsymbol{x}) = w_6 \sum_{k=1}^{N_{dp}} \sin(\pi n_p)^2 + \sin(\pi N_S)^2 + w_7 \sum_{k=1}^{N_{dp}} \boldsymbol{r}_k^T \boldsymbol{r}_k + G_{\text{dyn}}(\boldsymbol{x}_k), \quad (4.6)$$

4.3. *Design Variables*

The design variables are the physical dimensions of the EMA, which in turn define the parameters of the model. The design variables are given, together with their respective bounds, in Table 4.1.

Table 4.1 Design variables and their respective lower and upper bounds, $\boldsymbol{l_b}$ and $\boldsymbol{u_b}$, respectively.

Design variable	Description	$\boldsymbol{l_b}$	$\boldsymbol{u_b}$	Dynamic	Static
n_p	Number of pole pairs	1	6	Yes	Yes
l_2	Stator diameter [m]	0.01	0.3	Yes	Yes
l_1	Stator length [m]	0.1	1	Yes	Yes
N_S	Number of turns per phase	1	40	Yes	Yes
g	Air gap [m]	0.001	0.01	Yes	Yes
α	Winding angle [rad]	$\frac{\pi}{20}$	π	Yes	Yes
g_{r_1}	First gear ratio	1	6	Yes	Yes
g_{r_2}	Second gear ratio	1	6	Yes	Yes
p	Ball screw pitch [m]	0.01	1	Yes	Yes
u_{q_k}	Voltage at design point k[V]	$-u_{\max}$	$u_{\max}$	No	Yes
i_{q_k}	Current at design point k[A]	$-i_{\max}$	$i_{\max}$	No	Yes

The chosen design variables allow one to fully define the EMA model, at least in the simplified form presented in this work. One should note that the design variables in the static approach include the voltages and currents at the design points, since those are needed to satisfy the steady-state dynamics at those points.

4.4. *Overview and Setup*

The dynamic and static optimization approaches were presented, an overview of how they were implemented is shown in Figure 4.1.

Both the static and dynamical optimization approaches were implemented using the same optimizer, Scipy's Differential Evolution [10], based on [11, 12]. A population size of 14 individuals was used, with a maximum of 1000 generations. The chosen weights are given in Table 4.2.

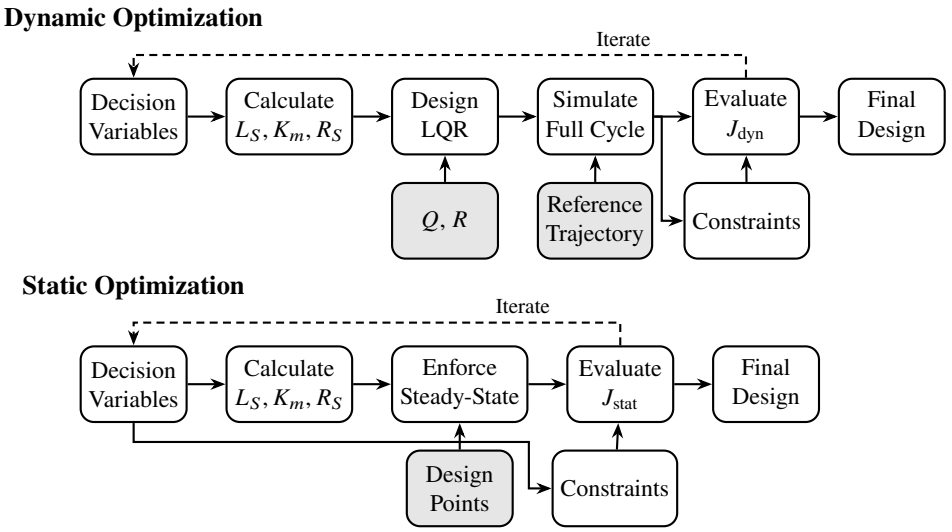


Figure 4.1 Schematic representation of the dynamic and static optimization approaches.

Table 4.2 Weights used in the cost functions and constraints.

Weight	w_1	w_2	w_3	w_4	w_5	w_6	w_7
Value Dynamic	100	$1 \cdot 10^{-3}$	$1 \cdot 10^{6}$	1	$1 \cdot 10^{3}$	N/A	N/A
Value Static	N/A	$1 \cdot 10^{-3}$	$1 \cdot 10^{3}$	N/A	$1 \cdot 10^{5}$	$1 \cdot 10^{10}$	$1 \cdot 10^{5}$

As can be seen in Table 4.2, the weights used in the dynamic approach are different from those used in the static approach. This is due to the different nature of the two approaches, which makes it necessary to tune the weights for each approach separately. The weights were tuned, however, so that the constraints are satisfied in both approaches. It should be noted, however, that the weights were tuned using a simple heuristic approach, and no systematic exploration of the Pareto front was conducted. Thus, the results could be improved further by choosing where in the Pareto front one wants to be, and then tuning the weights accordingly, allowing for a better trade-off between the different objectives.

For the static approach, four design points were considered: maximum torque and speed; half the maximum speed and maximum torque; half the maximum torque and maximum speed point; and one that has an additional torque term to account for the acceleration, where the acceleration was chosen as the maximum acceleration in the drive cycle together with the maximum torque and speed. Those points were chosen after several attempts; they are not only somewhat representative of the drive cycle but also give the best results found in terms of the dynamical performance of the designed EMA.

The chosen drive cycle is shown as the reference position and velocity in Figure 5.1. The drive cycle emulates a simplified crane operation, where a load of $1 \cdot 10^4$ N is lifted to a height of 10 cm, held there for 1 s, and then lowered back to the initial position, where a resting time of 1 s is applied. The load is zero whenever the position is zero, simulating a load being lifted from the ground. The total duration of the drive cycle in the dynamic optimization is 5 s, although it is simulated for 11 s in all figures to show more of the performance. This drive cycle was chosen because it has a wide dynamic range, as is typical for crane applications.

For the dynamical optimization approach, the controller matrices Q and R were chosen as $Q = \text{diag}(5 \cdot 10^1, 5 \cdot 10^1, 1 \cdot 10^6, 1 \cdot 10^2, 1 \cdot 10^4)$ and $R = \text{diag}(1, 1)$, where diag is the diagonal matrix. Those matrices were chosen as a trade-off between performance and control effort.

5. RESULTS

The optimization results shown are selected as the best design found in several attempts. The best design is chosen first by considering the constraints satisfaction, and then by the cost function value. The optimization for the dynamic approach took around 2.94 hours to complete, while the static approach took around 5.1 minutes to complete in the same machine, a standard desktop computer. At this point it must be emphasized that all the solutions found in the dynamic approach were respecting the constraints and had around the same characteristics. In contrast, in the static approach, the constraints are sometimes not satisfied and, even when they are, the characteristics of the solutions vary significantly, which makes it necessary to perform several attempts to find a good design.

The best found designs for both approaches are shown in Figure 5.1, where the position and velocity tracking are shown together.

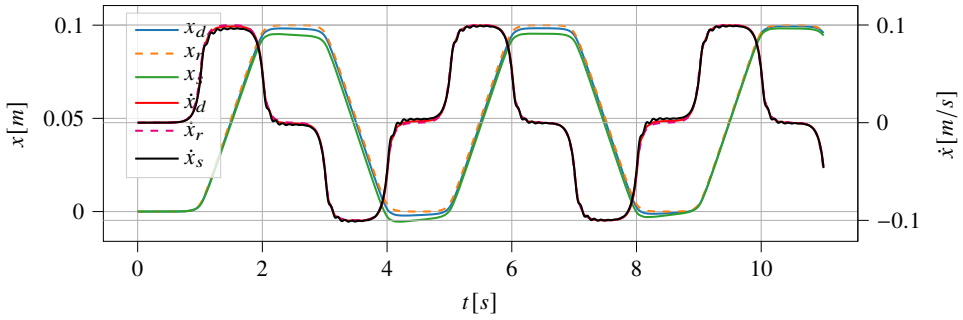


Figure 5.1 Position x and velocity $\dot{x}$ tracking over the drive cycle for both optimization approaches. Within the figure, the reference trajectory is shown with subscript r, the designed EMA using the dynamic approach with subscript d and the designed EMA using the static approach with subscript s.

The designed EMA using the dynamic approach was able to closely follow the desired trajectory, with very small position and velocity errors. The designed EMA using the static approach was also able to follow the desired trajectory, although not as closely. It is important to note that both designs are simulated using the same LQR controller structure with design matrices Q and R defined as in the dynamic optimization approach, yielding the best gain matrix K for each design and allowing consistent comparison. However, it is possible to obtain better tracking performance for the static design by changing the design matrices, although this comes at the cost of increased control effort.

One very important design objective is to choose an EMA that respects the electrical constraints while being able to deliver the needed tracking performance. The current and voltage magnitudes over the drive cycle for both designs are shown in Figure 5.2, where the solid lines represent the designed EMA using the dynamic approach, while the dashed lines represent the designed EMA using the static approach.

As can be seen in Figure 5.2, the designed EMA using the dynamic approach is able to respect the electrical constraints, represented as $u_{\max}$ for the voltage and $i_{\max}$ for the current. $u_{m,d}$ hits the constraint at all points of maximum positive acceleration, while the current $i_{m,d}$ does not hit the constraint at any point. On the other hand, the designed EMA using the static approach is not close to violating any electrical constraints, but the voltage magnitude $u_{m,s}$ is very low compared to the maximum allowed voltage, at the cost of having a higher current magnitude $i_{m,s}$, not ideal for efficiency.

It is important to evaluate the efficiency of the designed EMAs over the drive cycle. The total energy losses for both designs depend on the losses in the PMSM, for which the model only considers the copper losses, and the losses in the ball screw and gears, which are the same regardless of the resulting EMA design since they were defined as having simple constant

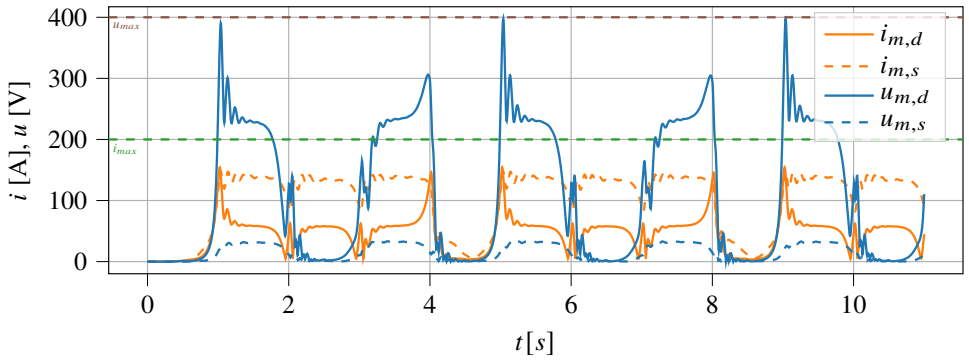


Figure 5.2 Current and voltage magnitudes over the drive cycle for both optimization approaches.

efficiencies. Thus, the only relevant comparison is the copper losses in the PMSM. The total copper losses to complete the trajectory shown in Figure 5.1 are 58 J for the dynamic design and 257 J for the static design, showing a difference in copper losses of approximately a factor of four between the two designs.

A natural question that arises is how the designed EMAs perform outside the drive cycle used in the optimization. This is reasonable to ask since the real world application is very rich in terms of possible trajectories that the EMA might need to follow. To evaluate this, both designs were simulated over a different drive cycle, one that both designs were not optimized for. The new drive cycle is shown as the reference position and velocity in Figure 5.3. This drive cycle is essentially the same as in the previous case, but with half the time to go from the initial height to the final height, thus having twice the velocity.

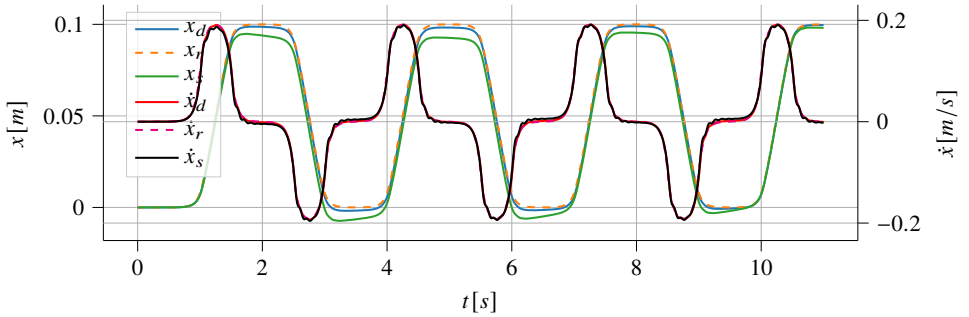


Figure 5.3 Position x and velocity $\dot{x}$ tracking over a different than optimized drive cycle for both optimization approaches. Within the figure, the reference trajectory is shown with subscript r, the designed EMA using the dynamic approach with subscript d and the designed EMA using the static approach with subscript s.

As can be seen in Figure 5.3, both designs have similar tracking performance over the new drive cycle, with the static design now performing even worse at tracking the reference. The currents and voltages over this new drive cycle are, as expected, higher than in the previous case, and for both designs it is exceeded at several points. One could simulate the system with restrictions on the maximum current and voltages, then, instead of the trajectory being followed at the cost of the constraints, the constraints would be kept at the cost of tracking. The former was chosen since it is much easier to implement and gives essentially the same conclusion but from a different perspective. The maximum $u_{m,d} = 1359.99$ V and $i_{m,d} = 274.09$ A for the dynamic design, while for the static design $u_{m,s} = 141.32$ V and $i_{m,s} = 197.47$ A. The total copper losses to complete this new trajectory are 131 J for the dynamic design and 261 J for

the static design. Although the dynamic design became relatively worse in percentage terms, its total energy loss was much lower than that of the static design.

Another major aspect to be considered in the design of EMAs is their diameter, which was considered in the cost function. The diameter, together with the other physical decision variables of both designs, is shown in Table 5.1.

Table 5.1 Design variables of the best found designs for both optimization approaches.

Design variable	n_p	l_2 [m]	l_1 [m]	N_S	g [mm]	α [rad]	g_{r_1}	g_{r_2}	p [mm]
Dynamic	1	0.1	0.99	40	1.32	0.449	1.906	2.18	38.89
Static	1	0.103	0.97	40	9.47	1.53	2.36	4.86	252.23

Table 5.1 shows that the dynamic design has a slightly smaller diameter than the static design, at the cost of having a bigger length. This means that the dynamic design has not only better performance in terms of tracking and efficiency, but also in terms of size. It should also be noted that the dynamic design not only converged to a solution with smaller air gap g, but also to a solution where the operational velocity of the PMSM is much higher, a result of the different gear ratios and ball screw pitch. Both aspects are important in the design of high-performance PMSMs.

In contrast, the static design converged to a relatively large air gap, close to the upper bound of the admissible range. Since the PMSM is being optimized for efficiency, one would expect a solution closer to the minimum allowable g, as observed in the dynamic design. The reason for this behavior is not entirely clear, however, one possible explanation is that the static optimization problem has a larger number of local minima, increasing the likelihood that the optimizer becomes trapped in such regions.

Another important aspect of the results is that l_1, l_2, N_S and n_p were found by the dynamic design approach to be optimal at their bounds, showing that one could obtain better designs if the maximum number of turns and the maximum length were increased. The diameter l_2 was considered in the cost function directly, thus, the fact that the optimizer found the optimal value to be at the bound of l_2 suggests that more efficient designs could be obtained if one were to allow softer restrictions regarding the diameter or by choosing a smaller w_2.

6. CONCLUSIONS

This paper compared two preliminary EMA design strategies for a mobile machinery drive cycle: a dynamic approach that evaluates each candidate design by simulating the full trajectory with a feedback-linearized, LQR-tuned controller, and a static multi–design-point approach that enforces steady-state equations at a few representative operating points. The main finding of this work is not only to reiterate the importance of considering the entire drive cycle in the optimization problem, but also to show that it is possible to do so at a reasonable computational cost by using simplified models and efficient optimal control techniques, allowing the problem to be solved without solving the full OCP, but rather a simpler LQR problem that can be solved efficiently. As seen, although the computational cost of the dynamic approach is significantly higher for one optimization round, the static approach required several attempts to find a marginally acceptable design, making the dynamic approach more efficient in practice.

As discussed, the designs in Table 5.1 are not valid for the actual design of real EMAs, since the model purposely neglects several important aspects that are usually the limiting factors, such as thermal limits, saturation effects and more complete loss models. Moreover, the considered drive cycle is oversimplified and choosing a representative one was not addressed. The main goal of this work is, however, to show the potential of considering the entire drive cycle in the optimization problem without neglecting the dynamics of the EMA, a picture that was clearly shown by the results, where every designed aspect considered was significantly better for the dynamic approach.

The significantly lower losses of the dynamic approach result from the model considering only copper losses: by respecting electrical constraints while tracking trajectories, it finds designs that require lower currents. While more complete models with iron losses would reduce the relative loss difference between approaches, the dynamic approach is expected to remain superior due to its consideration of full EMA dynamics over the drive cycle.

7. Future Work

Future work should address four main areas: incorporating more complete EMA models including thermal limits, saturation effects, and iron losses to enable realistic preliminary designs; validating the simplified optimization-oriented models; comparing the LQR-based control approach against full OCP solutions to assess the optimality gap in this system; and the pareto front of the different designs, that is, how the different weights in the cost function affect the resulting design. This would allow for a better understanding of the trade-offs involved in the design of EMAs for mobile machinery applications. Those improvements would allow much better preliminary design of EMAs, reducing the number of refinements needed using more complete finite element models.

References

[1] A. Rankka, A. Dell'Amico, and P. Krus, 'Drive cycle generation for a hydraulic loader crane', in The Sixteenth Scandinavian International Conference on Fluid Power, May 22-24, 2019, Tampere, Finland, Tampere University of Technology, 2019, pp. 42–53.

[2] O. Borsboom, M. Salazar, and T. Hofman, 'Electric Motor Design Optimization:A Convex Surrogate Modeling Approach', en, IFAC-PapersOnLine, vol. 55, no. 24, pp. 373–378, 2022, ISSN: 24058963. DOI: 10.1016/j.ifacol.2022.10.312.

[3] S. Pastellides, S. Gerber, R.-J. Wang, and M. Kamper, 'Evaluation of Drive Cycle-Based Traction Motor Design Strategies Using Gradient Optimisation', en, Energies, vol. 15, no. 3, p. 1095, Jan. 2022, Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1996-1073. DOI: 10.3390/en15031095.

[4] E. Roshandel, A. Mahmoudi, W. Soong, and S. Kahourzade, 'Optimal Design of Induction Motors Over Driving Cycles for Electric Vehicles', IEEE Transactions on Vehicular Technology, pp. 1–15, Jul. 2023. DOI: 10.1109/TVT.2023.3292901.

[5] A. E. Bryson, 'Applied Optimal Control: Optimization, Estimation and Control'. New York: Routledge, May 2018, ISBN: 978-1-315-13766-7. DOI: 10.1201/9781315137667.

[6] F. L. Lewis, D. Vrabie, and V. L. Syrmos, 'Optimal Control', English. Hoboken: Wiley, 2012, ISBN: 978-0-470-63349-6.

[7] B. Colson, P. Marcotte, and G. Savard, 'An overview of bilevel optimization', en, Annals of Operations Research, vol. 153, no. 1, pp. 235–256, Sep. 2007, ISSN: 1572-9338. DOI: 10.1007/s10479-007-0176-2.

[8] J. Chiasson, 'Modeling and High Performance Control of Electric Machines', en. John Wiley & Sons, May 2005, Google-Books-ID: cq6RPPsOyt8C, ISBN: 978-0-471-72234-2.

[9] 'Solve_continuous_are — SciPy v1.16.2 Manual'.

[10] 'Differential_evolution — SciPy v1.16.2 Manual'.

[11] M. Wormington, C. Panaccione, K. M. Matney, and D. K. Bowen, 'Characterization of Structures from X-Ray Scattering Data Using Genetic Algorithms', Philosophical Transactions: Mathematical, Physical and Engineering Sciences, vol. 357, no. 1761, pp. 2827–2848, 1999, Publisher: The Royal Society, ISSN: 1364-503X.

[12] J. Qiang and C. Mitchell, 'A Unified Differential Evolution Algorithm for Global Optimization', IEEE Transactions on Evolutionary Computation, Jun. 2014.

Biographies



Róger Mateus Sehnem. He received his Bachelor in Aerospace Engineering (UFSM) in 2021 and his M.Sc in Electrical Engineering (UFRGS) with emphasis on data driven control in 2023. He currently works as a PhD student with electromechanical actuators with focus on dynamics, control and optimization. His research interests include control, system modelling, optimal control, machine learning, and optimization.



Samuel Kärnell. He received his Ph.D. in Hydraulics at Linköping University (LiU), Sweden, in 2022. He currently works as a Assistant Professor at Fluid and Mechatronic Systems at LiU. His research interests include pump and motor design, electro-hydraulic systems, electrification of mobile machinery, and karaoke.



Liselott Ericson. She received her Ph.D. in Hydraulics at Linköping University (LiU), Sweden, in 2012. She currently works as a professor at Fluid and Mechatronic Systems at LiU. Her research interests include pump and motor design, electro-hydraulic systems, modelling and simulation, and optimisation.

Universal and Highly Efficient Speed Control of the Hydraulic Motor in a Wire Saw Unit for a Hybrid Constant Flow and Constant Pressure System

Daniel Kriegl, Jörg Edler, Max Überbacher, Michael Pasterk

Daniel Kriegl, *IFT TU-Graz / Schwing GmbH, dkriegl@schwing.at*

Jörg Edler, *IFT TU-Graz, joerg.edler@tugraz.at*

Max Überbacher, *TU-Graz, max.ueberbacher@tugraz.at*

Michael Pasterk, *Schwing-GmbH, mpasterk@schwing. at*

Abstract.

The control of hydraulic motors supplied via long hydraulic lines presents significant challenges, particularly for high-speed operation and loads with high rotational inertia. This study investigates a hybrid hydraulic wire saw attachment equipped with a fixed displacement motor and supplied by two alternative power sources: a constant-flow system driven by a variable-speed internal combustion engine and a constant-pressure system powered by an electric motor with a variable displacement pump. The proposed hybrid configuration combines operational flexibility and mobility with low-noise and low-emission electric operation, while allowing independent operation from either power source. The key contribution of this work is a dual-mode speed control strategy that switches from conventional valve-based flow control during transient phases to direct, valve-less speed control under steady-state conditions. This approach enables energy savings of approximately 13–42% during the steady cutting phase, which accounts for the majority of operating time. Maintaining stable motor speed under varying control modes remains the primary control challenge.

Keywords. Mobile Application, Hybrid machine, Direct Flow control, Construction and Mining Machinery

1. INTRODUCTION

The control of hydraulic motors in construction machinery is a well-established and widely studied topic. However, challenges arise when long hydraulic lines (>20 m) are involved, particularly in achieving precise control of hydraulic motors operating at high rotational speeds (>1000 RPM) and driving loads with substantial rotational inertia. The system analysed in this study is a hydraulic wire saw attachment with a fixed displacement motor implemented as a hybrid configuration, like shown in Figure 1. Hydraulic power is supplied

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

via two distinct sources: a constant-flow system driven by a variable-speed internal combustion engine, and a constant-pressure system powered by an electric motor coupled with a variable displacement pump. This configuration enables the combination of the mobility and flexibility of an internal combustion engine-powered machine with the low-noise and low-CO_2 emission operation of an electric motor. The saw unit can be operated independently by either system.

Zhongyi Quan [1] and Tianliang Lin [2] have proposed various control strategies for hydraulic motors, including multiple combinations of variable pumps and motors tailored to specific application requirements, with the goal of maximizing energy efficiency and recovery. A particular emphasis is placed on different forms of secondary control strategies. Also Berg and Ivantysynova [3] examined secondary control strategies for hydraulic motors utilizing variable displacement.

In contrast, the hybrid concept examined in this work requires a configuration capable of operating effectively with both constant-flow and constant-pressure supply systems with a fixed displacement motor. In this specific application, energy recovery is not deemed advantageous.



Figure 1 Wire saw attachment cutting, mounted on carrier vehicle including internal combustion engine and electric powerpack at the back

The novelty of the proposed concept lies in the ability to switch between two different speed control modes during operation. For the short, undefined setup and cutting initiation phase, speed control is handled conventionally via a flow control valve. Once steady-state conditions are achieved, the system transitions to direct, valve-less speed control—either by adjusting the internal combustion engine speed or through electric motor control using the swivel angle of the variable displacement pump. This enables energy savings of approximately 13–42% in the steady cutting phase, which typically accounts for over 90% of total operation time. The switching mechanism can be realized using a bypass valve to circumvent the flow control valve, a load-sensing (LS) pressure valve for targeted

pressurization of the compensator, or a control spool with a non-linear characteristic.
The main challenge is maintaining a constant motor speed in accordance with process requirements, where either the combustion engine speed or the pump swivel angle serves as the control variable.

2. SYSTEM OVERVIEW



Figure 2 Hydraulic setup of the saw motor

The hydraulic architecture of the system, shown in Figure 2, comprises a telehandler-mounted attachment equipped with a main valve block and a fixed displacement hydraulic motor, as well as two alternative power units: an internal combustion engine driving a gear pump and an electric motor coupled to a variable displacement axial piston pump. The attachment can be operated independently by either power unit, depending on the availability of an external electric power supply. The decision to combine a constant-pressure system for the electric drive with a constant-flow system for the internal combustion engine was made at an early stage of development. This dual-system approach was necessary because modifications to the existing hydraulic system of the telehandler were not permitted, while the electro-hydraulic power supply was required to be as compact and energy-efficient as possible—requirements that could not be met using a constant-flow system alone.

2.1. Hydraulic Plan

The main hydraulic block is composed of several standard sandwich-plate valves assembled to provide all functions required for operating the attachment. For the sake of clarity and reduced complexity, only the components relevant to motor control are shown in the schematic Figure **3**. The block incorporates an integrated pilot oil supply unit as well as a two-way safety valve.

For the experimental setup, two flow meters and pressure sensors were integrated into the hydraulic circuit: one upstream of the main valve block to measure the pump supply flow to the attachment, and a second upstream of the hydraulic motor. This instrumentation allows the determination of flow and pressure losses across the main valve block by comparison of the measured values.

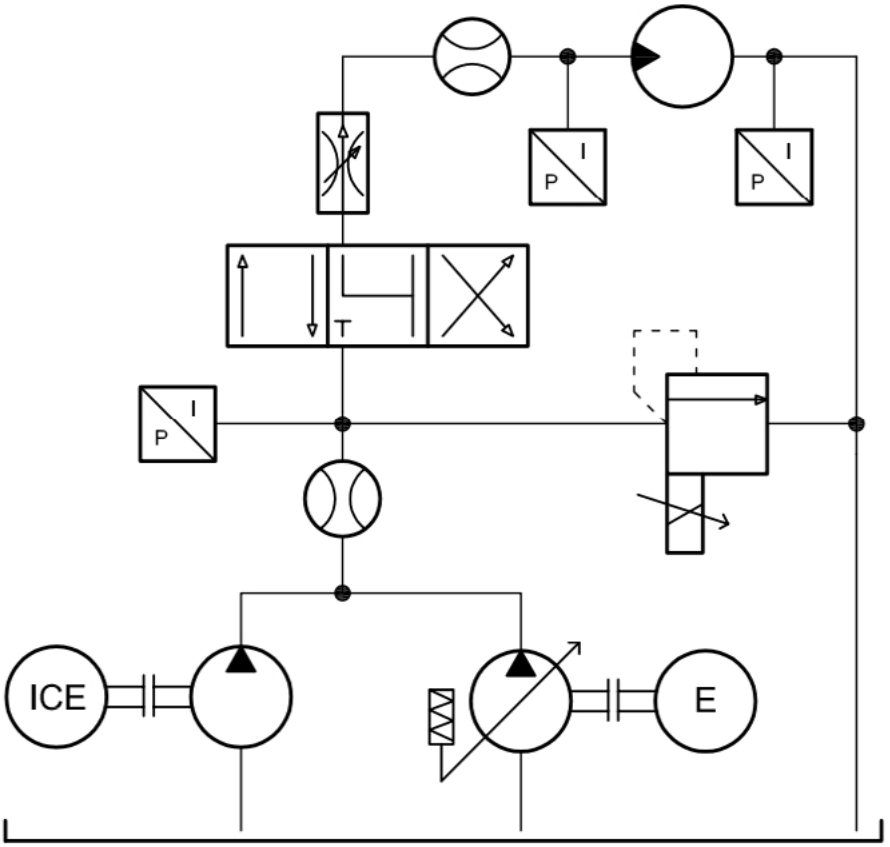


Figure 3 Hydraulic schematic (simplified)

2.2. Control Strategies

First, the hydraulic motor is controlled to reach the target rotational speed required to achieve the desired wire velocity. Second, the feed rate of the wire into the stone is adjusted to obtain the specified motor torque. These target values are predefined and are not the subject of this study. To ensure efficient operation, the hydraulic flow rate and pressure must be supplied in accordance with the required speed and torque demands, while avoiding excess energy provision.

2.2.1. Wire Speed

The wire speed is controlled using a 4/3-way valve equipped with a pressure compensator and a PID controller. This configuration exhibits stable behaviour, as the pressure compensator inherently adapts to pressure fluctuations. The PID controller primarily compensates for leakage and temperature-induced variations and contributes to the attenuation of oscillations. Oscillatory behaviour is a well-known challenge in the control of hydraulic motors driving large inertial loads in combination with long hydraulic hose lengths. During experimental testing with the proposed setup, oscillations were observed but without overshoot. Within the scope of these experiments, it was not possible to clearly distinguish whether the oscillations originated from the hydraulic circuit or from instabilities in the cutting process itself.

2.2.2. Hydraulic supply at constant pressure circuit

Hydraulic flow and pressure are supplied by a variable displacement axial piston pump driven by a three-phase electric motor. The pump is equipped with a pressure controller actuated by an electrically controlled valve, allowing the pump pressure level to be adjusted, while the flow rate is not actively regulated. As a result, the system operates as a pseudo load-sensing configuration. Due to the excessive length of the hydraulic hoses between the pump and the valve block, conventional transmission of a load-sensing signal via a hydraulic line is not feasible. However, under the relatively steady operating conditions encountered during cutting, load sensing is implemented electronically by adjusting the pump pressure based on measured pressures and predefined target values. The pump continuously supplies the flow required to maintain the desired pressure level at the attachment. The pressure difference between the pump outlet and the load pressure is regarded as a necessary and acceptable loss in this configuration.

To eliminate this pressure-related loss, the original pressure-controlled pump was replaced with a flow-controlled pump. The new pump is mechanically identical but is equipped with an electronic controller and a swashplate angle sensor, enabling the implementation of various control strategies purely through software. Load-sensing control can be realized using pressure feedback, while flow and power control are achieved by regulating the swashplate angle and motor speed. To minimize energy losses associated with Δp and to improve stability with respect to oscillations, motor speed is now controlled directly via the pump swashplate angle, while the 4/3-way valve is kept fully open. Consequently, pressure losses are limited to flow-dependent losses, with no additional losses introduced by control valves. Furthermore, the propensity for valve-induced oscillations is significantly reduced, as the valves operate in a fully open state.

2.2.3. Hydraulic supply at constant flow circuit

Hydraulic flow is supplied by a gear pump driven by the internal combustion engine of the telehandler. This pump provides hydraulic power for all machine functions, including boom actuation, stabilizers, and auxiliary attachments such as the wire saw. Consequently, the telehandler is equipped with a load-sensing hydraulic system in which a pressure compensator diverts excess flow to the tank at load pressure. In this configuration, only the pump flow rate and the engine speed can be adjusted. Owing to the operating principle of the circuit, auxiliary attachments must be controlled via the telehandler control valve to

ensure proper functionality. When the flow rate is reduced, the pressure compensator increases the system pressure to its maximum, resulting in unnecessary pressure losses between the actual load pressure and the maximum system pressure.

To mitigate this issue, a proportional pressure relief valve was integrated into the circuit, and the relationship between engine speed and pump flow rate was characterized. By setting the relief valve to a pressure level equal to the load pressure plus a defined Δp, the valve effectively acts as a pseudo pressure compensator. In this operating mode, the telehandler's pressure compensator remains closed, and excess flow is discharged at a lower pressure level. With appropriate adjustment of the engine speed, the amount of excess flow can be minimized, resulting in improved overall efficiency.

To further reduce these losses, an alternative control approach was investigated. This concept, identified inadvertently during testing, involves fully opening the 4/3-way valve while closing the pressure relief valve. Under these conditions, the hydraulic motor rotates at a speed directly determined by the pump flow rate. Provided that the internal combustion engine speed can be adjusted rapidly and accurately, the motor speed can be effectively controlled through engine speed regulation.

2.2.4. Additional hydraulic actuators in the circuit

In addition to the hydraulic motor, the pilot oil supply of the valve block, the wire tension system, and the feed cylinders operate simultaneously. Wire tension is maintained at a constant level by a hydraulic cylinder and therefore requires only the leakage flow of its load-sensing valve. The feed cylinders operate at a comparatively low velocity relative to the hydraulic motor, resulting in negligible hydraulic flow demand.

3. DESCRIPTION OF A STANDARD CUT

Cutting stone with a diamond wire is a comparatively slow process relative to other cutting methods and materials. Feed rates depend on several parameters, including stone type, diamond bead characteristics, cutting width, and wire speed. In the application considered here, feed rates typically range from 1 to 20 cm/min, with an average of approximately 4–6 cm/min. Cutting stones with heights between 1 and 3 m and widths between 2 and 5 m generally requires between 30 min and 3 h, with an average cutting duration of about 1 h.

The cutting process can be divided into three distinct phases, shown in Figure 4. The initial phase corresponds to the start of the cut, during which the wire is not yet fully engaged in the stone due to the rough surface and the absence of a fully developed cutting arc. During this phase, the motor torque is low and unstable until full engagement is achieved and a stable cutting arc is formed, which typically takes approximately 5 min.

The second phase constitutes the main cutting stage and represents the longest portion of the process. During this phase, the wire is continuously fed through the stone under a relatively constant high motor torque with only minor variations. This phase accounts for approximately 90% of the total cutting time.

The final phase occurs toward the end of the cut, when the wire is no longer actively fed because the guide wheels approach the ground. The wire is then used primarily to lower the

cutting arc as far as possible to ensure a clean edge before the stone is separated at the completion of the cut. During this phase, the motor torque decreases continuously to a low level, and the process concludes once only a minimal remaining section is left to be cut, typically lasting a few minutes.

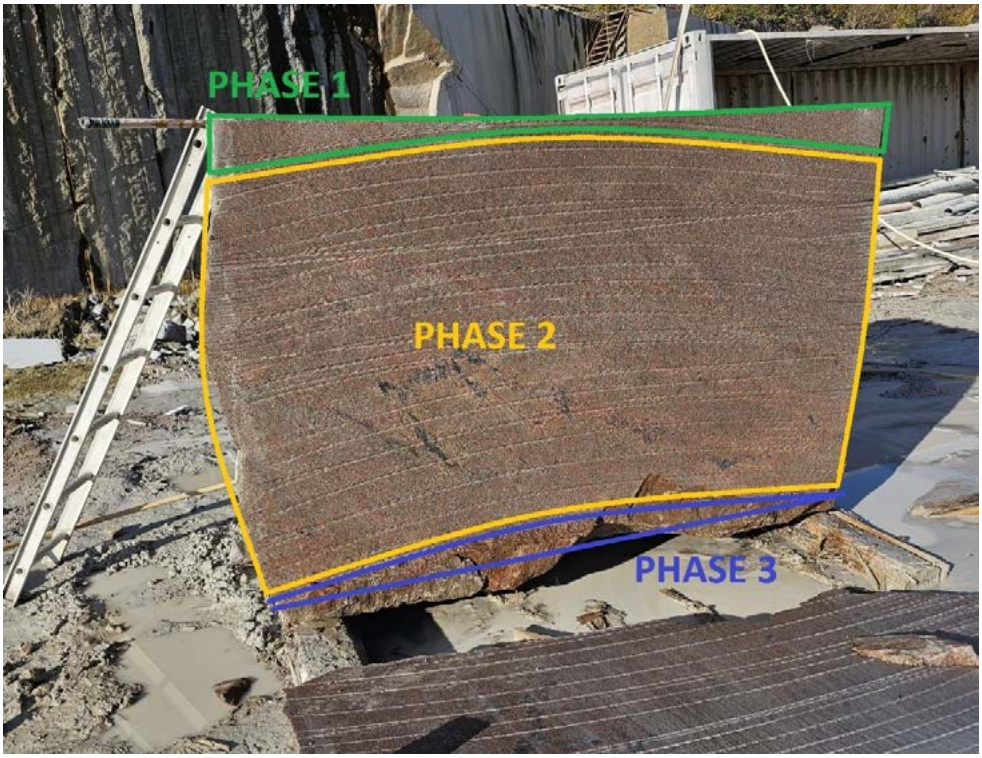


Figure 4 Stone block with the different cutting phases marked

4. TEST AND RESULTS

The experimental investigations were conducted using the same system configuration to ensure maximum comparability. All test cuts were performed on the same stone using the same diamond wire, and switching between the different control modes was carried out during the cutting process to minimize variations in operating conditions.

Due to the inherently irregular nature of the cutting process over time, a comparison based on instantaneous values is not meaningful. Instead, mean values over selected time intervals were used for the evaluation of the results. These intervals were chosen to be representative of typical cutting conditions and were kept relatively short to ensure comparable cutting states. Based on prior experience, these representative periods can be clearly identified within the measured data. The selected intervals are indicated by red and blue dotted rectangles in Figure 7 and Figure 8, and the corresponding mean values are summarized in Table 1. The blue rectangles denote measurements obtained using the proposed control mode, while the red rectangles represent measurements using the conventional control

mode. For improved comparability, all measured quantities were normalized using system-specific reference values p_{ref} and Q_{ref} according to:

$$p = \frac{p_n}{p_{ref}} \quad , \qquad Q = \frac{Q_n}{Q_{ref}} \qquad (4.1)$$

Figure 5 presents the data for the constant-pressure system operating under both control modes. The test sequence began with the proposed control mode, in which motor speed is regulated via the pump swashplate angle. After 175 s, the system was switched to the conventional valve-based control mode. The variable Q_{supply} denotes the flow rate supplied by the pump to the system, while Q_{motor} represents the flow rate delivered to the hydraulic motor. The difference between these values corresponds to pilot oil consumption and leakage flows discharged directly to the tank. The pressure p_{supply} is the main supply pressure measured upstream of the main valve block, p_{tank} is the motor return pressure, and p_{motor} is the load pressure measured at the motor inlet. The motor torque-related pressure p_{torque} is calculated as the difference between the motor load pressure and the return pressure.

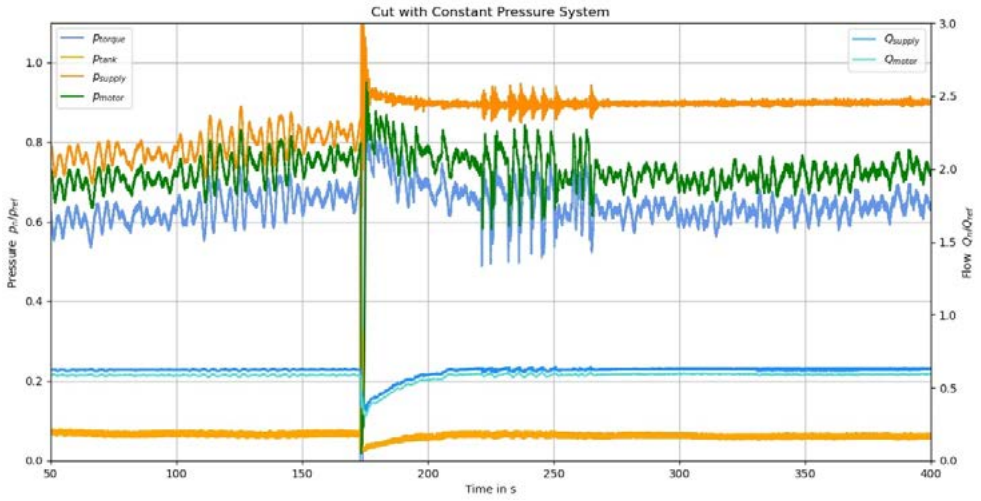


Figure 5 Test with constant pressure system

Figure 6 presents the data for the constant-flow system operating under both control modes. The test sequence began with the proposed control mode, in which motor speed is regulated via the internal combustion engines speed. After 260 s, the system was switched to the conventional valve-based control mode.

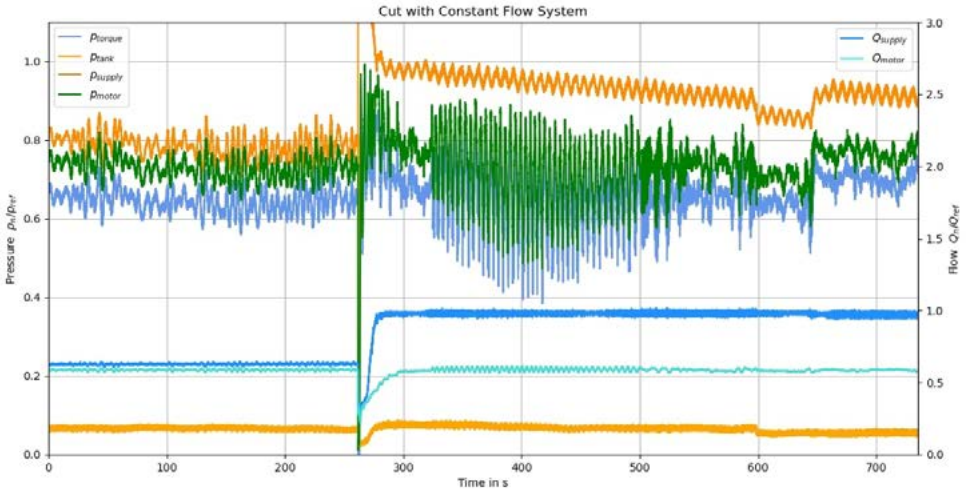


Figure 6 Test with constant flow system

4.1. Test with the constant pressure system

In the following figures Figure 7 and Figure 8, the green shaded areas represent the torque-related values, or equivalently the hydraulic power required for the cutting process. The yellow shaded areas indicate the pressure losses occurring in the valve block and the hydraulic hoses upstream of the motor. It is clearly evident that, after the mode switch, a significantly larger amount of hydraulic energy is dissipated across the pressure compensator. As expected, no significant difference in the flow rates can be observed, since the pump supplies only the flow required by the system, both before and after the change in control mode. The blue shaded areas represent the flow passing through the hydraulic motor.

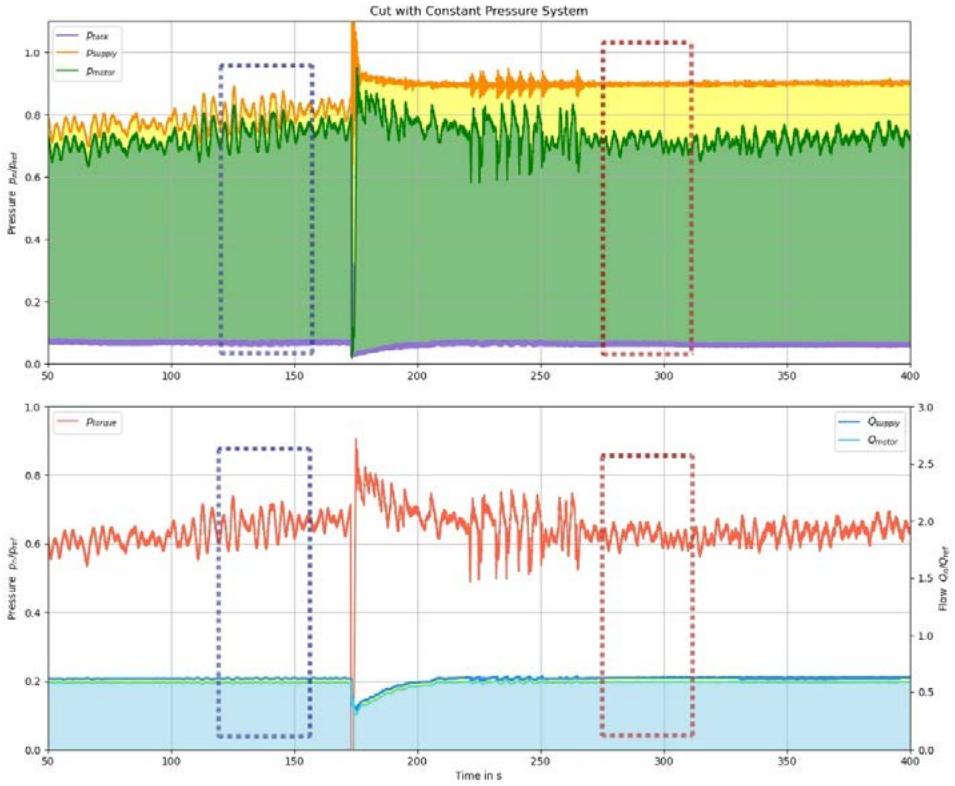


Figure 7 Result of the constant pressure system

4.2. Tests with the constant flow system

For the constant-flow system, the results are generally comparable, with the exception of the flow behaviour. The difference in flow is illustrated in Figure 8, where the yellow shaded area between the flow curves highlights the excess flow. This excess flow results in not only increased losses due to the oil discharged through the pressure relief valve, but also a substantially higher magnitude of loss than anticipated. During the experiments, it was observed that approximately 60% more flow than required was discharged, rather than the expected surplus of around 10%. This discrepancy was attributed to inaccuracies in the calibration of the engine-speed–to–flow characteristic, which therefore requires further refinement.

Because correct calibration could not be achieved at the time of testing, an additional efficiency calculation was performed within this study using the intended target values. This supplementary analysis illustrates the expected system performance under properly calibrated operating conditions.

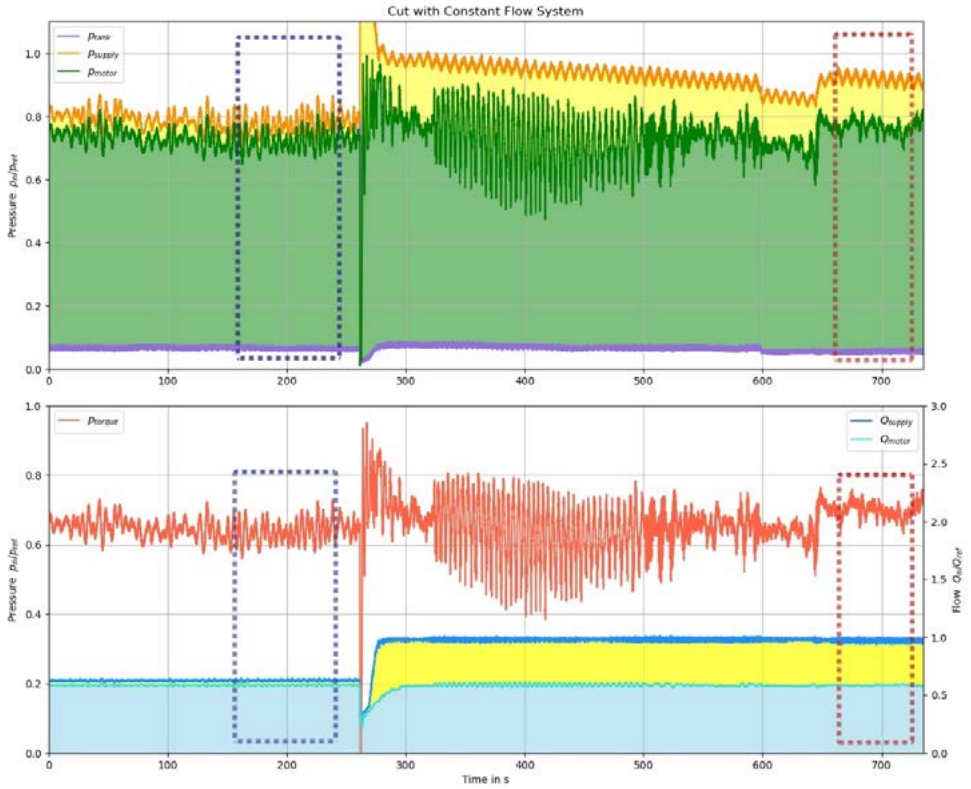


Figure 8 Result of the constant flow system

4.3. Results

To enable a meaningful comparison of the measured results, the converted hydraulic power was calculated based on pressure and flow rate measurements. This approach also allows a direct determination of potential energy savings. Furthermore, it facilitates comparison between different experiments, as all tests were conducted at a single operating point of the system.

4.3.1. Calculation of the results

To determine the total input power P_{in} the hydraulic flow rate and pressure measured upstream of the valve block are used:

$$P_{in} = Q_{supply} * p_{supply} \qquad (4.2)$$

The total cutting power at the motor P_{cut}, including mechanical losses, is calculated using the flow rate upstream of the motor and the pressure difference across the motor P_{torque}:

$$p_{torque} = p_{motor} - p_{tank} \qquad (4.3)$$

$$P_{cut} = Q_{motor} * p_{torque} \qquad (4.4)$$

The overall leakage flow of the valve block Q_{delta}, including the flow through the load-sensing pressure relief valve, is determined as the difference between the flow rates upstream and downstream of the valve block:

$$Q_{delta} = Q_{supply} - Q_{motor} \qquad (4.5)$$

Based on this, the power losses of the valve block $P_{valveblock}$ are calculated as:

$$P_{valveblock} = Q_{motor} * (p_{supply} - p_{motor}) + Q_{delta} * p_{supply} \qquad (4.6)$$

This formulation allows the influence of the compensator state (active or fully open) and the presence of flow through the pressure relief valve to be evaluated. For comparison between control modes, the difference in valve block losses is defined as:

$$P_{valveDiff} = P_{valveblock\ NewMode} - P_{valveblock\ Standard} \qquad (4.7)$$

The remaining losses, which include the return line and return filter losses, are calculated as

$$P_{lossOther} = P_{In} - P_{cut} - P_{valveblock} \qquad (4.8)$$

The efficiency improvement is quantified by the saving factor, defined as the ratio of the reduced valve block losses to the input power:

$$Saving = \frac{P_{valveDiff}}{P_{In}} \qquad (4.9)$$

Table 1 summarizes the calculated results for the comparison of control modes for both the constant-flow and constant-pressure systems. The third column presents the constant-flow system assuming an ideal flow rate under standard control conditions. The green-highlighted cells indicate improvements between control modes in terms of flow- and pressure-related losses, while the grey-highlighted cell denotes an ideally selected reference value.

	Constant Flow		Constant Pressure			Const. Flow improved	
	Standard	NewMode	Standard	NewMode		Standard	NewMode
p_{supply}	0,919	0,783	0,898	0,805		0,919	0,783
Q_{supply}	0,973	0,629	0,627	0,621		0,703	0,629
P_{In}	0,894	0,492	0,563	0,500		0,646	0,492
p_{torque}	0,700	0,642	0,625	0,651		0,700	0,642
Q_{motor}	0,586	0,588	0,592	0,584		0,586	0,588
P_{cut}	0,410	0,377	0,370	0,381		0,410	0,377
p_{motor}	0,767	0,725	0,712	0,742		0,767	0,725
Q_{delta}	0,387	0,041	0,035	0,037		0,117	0,041
$P_{valveblock}$	0,444	0,066	0,142	0,066		0,197	0,066
$P_{lossOther}$	0,039	0,049	0,051	0,053		0,039	0,049
$P_{valvediff}$	-0,378		$P_{valvediff}$	-0,076		$P_{valvediff}$	-0,131
$Saving$	-42,32%		$Saving$	-13,47%		$Saving$	-20,21%

Table 1 Test results

The results indicate that, for the investigated application, between 13% and 42% of the input energy can be saved by employing the proposed control mode. Even with an improved motor-speed–to–flow characteristic, the constant-flow system shows a potential energy saving of approximately 20% of the actual input power when operated with the new control strategy.

The Sankey diagrams provide an additional means of comparing energy losses. In these diagrams, all power flows are normalized with respect to the cutting power P_{cut}, which is used as the reference value. Consequently, the diagrams illustrate the differences in required input power for achieving the same cutting power.

Figure 9 illustrates that, in the constant-flow system under conventional control, approximately 108% of the cutting power is dissipated as losses to achieve 100% cutting power. When the new control mode is applied, these losses are reduced to approximately 17%, indicating a substantial reduction in dissipated energy and, correspondingly, in heat generation within the system.

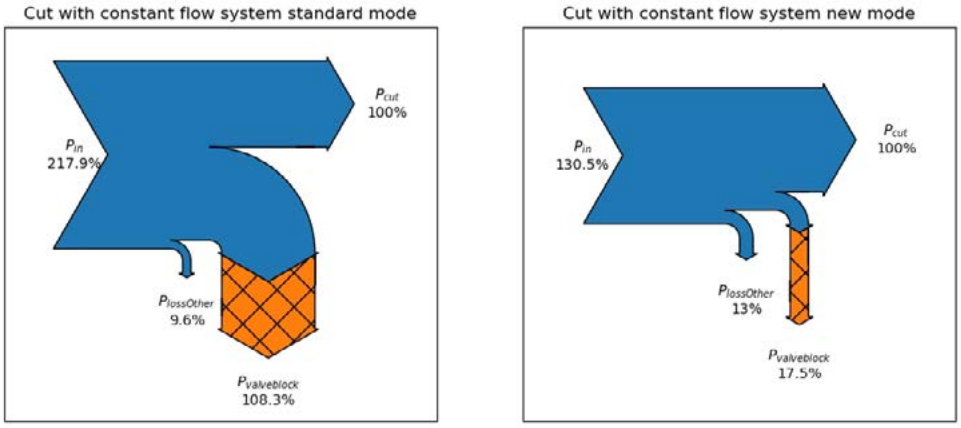


Figure 9 Sankey diagrams for constant flow system

Figure 10 further demonstrates that, for the constant-pressure system, a difference of approximately 20% in input power is required to achieve the same cutting power under conventional control. Both systems exhibit identical results when operated in the new control mode, which is expected, as the valve configurations and functions are identical in this case. In the proposed control mode, hydraulic flow is governed solely by the pumps, while system pressure adjusts according to the motor load.

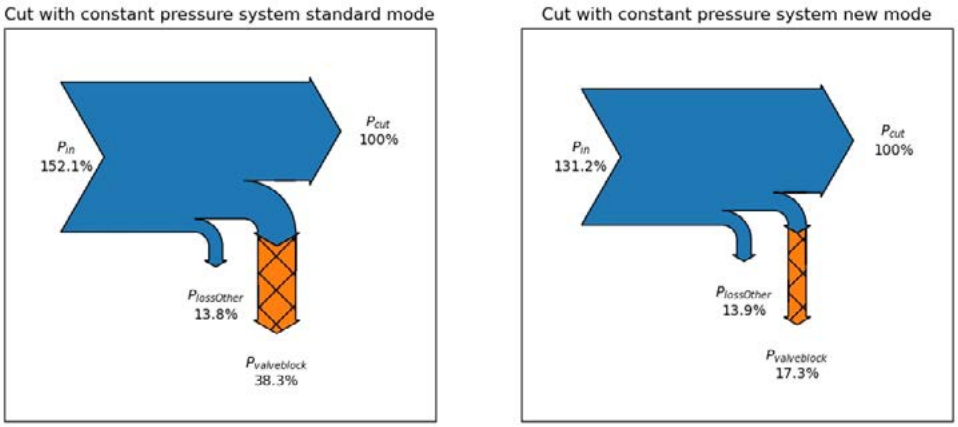


Figure 10 Sankey diagrams for constant pressure system

5. DISCUSSION AND CONCLUSION

The proportion of wasted energy observed in the constant-flow system is relatively high; however, the experiments were conducted on an entirely original system without additional modifications, apart from the integration of the new pump and measurement sensors. Further optimization of this configuration is therefore expected in future work. Nevertheless, even with improved adjustment, the results indicate a substantial potential for energy savings.

Additional improvement measures are conceivable for the investigated system. For example, the electrically driven piston pump could be replaced by a gear pump, and the main valve block could be equipped with a central pressure compensator, similar to an open-loop hydraulic system. Another potential approach would be the use of a variable displacement hydraulic motor capable of directly controlling its own rotational speed. However, the modifications presented in this study were deliberately kept to a minimum and can be implemented without major changes to existing machines. To further enhance system efficiency, other components such as hydraulic hoses and filters could also be reconsidered, as the experiments consistently indicated losses of approximately 13% attributable to these elements.

Further investigations at additional operating points would be beneficial, and the influence of oil temperature should be taken into account. Although the experiments were performed under nearly identical oil temperature conditions, explicitly incorporating temperature effects into the calculations would improve the accuracy of the results.

6. REFERENCES

[1] Zhongyi Q., Long Q., et al., Review of energy efficient direct pump controlled cylinder electro-hydraulic technology, in: Renewable and Sustainable Energy Reviews 35, pp. 336-346, 2014, DOI: http://dx.doi.org/10.1016/j.rser.2014.04.036

[2] Tianliang L., Qingfeng W., et al., Development of hybrid powered hydraulic construction machinery, in: Automation in Construction 19, pp. 11-19, 2010, DOI: http://dx.doi.org/10.1016/j.autcon.2009.09.005

[3] Berg H., Ivantysynova M, Robust position and speed control of variable displacement hydraulic motors, IFAC Robust Control Design, Budapest, Hungary. 1997

Biographies



Daniel Kriegl received a Master's degree in Mechanical Engineering from Graz University of Technology (TU Graz) in 2018. Since 2016, he has been employed at the Institute of Manufacturing Technology at TU Graz as well as at Schwing GmbH in St. Stefan, Carinthia. In 2018, he began doctoral studies with a research focus on hydraulic drives and control systems.

Digital Twins and AI in Fluid Power: Improving Systems by Connecting Real and Virtual Worlds

Katharina Schmitz and Malte Becker

RWTH Aachen University, Institute for Fluid Power Drives and Systems (ifas), Campus-Boulevard 30, 52074 Aachen, Germany

Abstract.

Fluid power is a widely used drive technology known for its durability, robustness, and cost efficiency, but it faces challenges related to energy efficiency and environmental impact. To address these challenges, the integration of software functions and digital frameworks is gaining importance for improving component performance, system integration, and product development. The integration of data and software within fluid power systems and components enables companies to increase their innovation potential, thereby establishing digital competence as a key determinant of competitive advantage. Consequently, fluid power is evolving into an interdisciplinary research field that combines mechanics, electronics, computer science, and data analytics. Current research focuses on the usage of digital twins across all phases of the product life cycle. As a result, numerous related concepts have emerged, such as digital frameworks and software functionalities that support component and system development while providing additional insights during operation, for example, through soft sensors that derive new variables from existing sensor data.
Despite these advances, the implementation and maturity of digital twin technologies vary significantly between domains. In the field of fluid power, in particular, the extent to which digital twins have been developed and integrated remains an open question. Therefore, this contribution examines the current state of digital twin development in fluid power through selected examples and identifies the key elements that enable interaction between the physical and virtual domains. The contribution shows that, while research already provides solutions addressing individual aspects of digital twins along the product life cycle, a fully integrated approach that connects these aspects is still lacking. These findings underscore the need for common standards and comprehensive frameworks to fully leverage the potential of digital twins.

Keywords. Digital Twin, Asset Administration Shell, Industry 4.0, Artificial Intelligence, Machine Learning, Soft Sensors, Digital Frameworks, Systems Engineering

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Fluid power is a key drive technology applied across almost all industry sectors. Its main advantages include high durability, robustness, and often cost efficiency in acquisition compared to alternative technologies. Combined with flexibility in system design, these properties enable fluid power to meet the specific requirements of a wide variety of applications. Today, fluid power technology forms a central pillar in heavy machinery, manufacturing systems, infrastructure, and energy production. It is also found in smaller applications, such as medical devices and power tools. [1]

Despite these advantages, the widespread use of fluid power presents notable challenges. Since many applications involve significant energy transfer, fluid power systems contribute substantially to industrial energy consumption and, consequently, to CO_2 emissions. An estimate indicates that the CO_2 emissions from hydraulic machinery account for approximately 5 % of global emissions [2]. For comparison, this corresponds to 2.3 of the global aviation emissions [2], underscoring the considerable potential for improving the efficiency and optimization of fluid power systems. In response, ongoing technological progress increasingly requires the integration of software functions and digital enhancements into products and processes. The effective use of data and information, combined with digital functionalities, is becoming essential for improving component and system performance, as it enables the development of intelligent and optimized systems that contribute to long-term emission reduction. Such advancements can be realized, for example, through the implementation of soft sensors [3, 4] or the digitalization of product development processes [5, 6]. As a result, fluid power is evolving into an increasingly interdisciplinary field, combining expertise from mechanics, electronics, computer science, and data analytics. Current research focuses on approaches such as soft sensors and digital frameworks that aim to enhance the system development process. In this context, the concept of the digital twin plays a central role, representing a fully digital counterpart of a fluid power system or of a component. Against this background, this contribution addresses the following research question: *What is the current development level of digital twins in fluid power?* To answer this question, the concept of digital twins is first defined. Subsequently, selected exemplary research findings related to digital twins in fluid power are presented, followed a brief assessment of the current state of development.

2. DEFINITION OF DIGITAL TWINS

The concept of the digital twin has emerged and is interpreted in various ways [7]. To provide a comprehensive understanding of digital twins, the terminology is first clarified and the interpretation used in this contribution is defined. The origin of the term “twin” can be traced back to NASA’s Apollo program, where engineers replicated the conditions of space missions on Earth [8]. NASA developed an identical physical replica of the spacecraft used in space to simulate mission scenarios on Earth [8]. To reduce the extensive resource requirements of such physical twins and enable their application in earlier stages of development, the idea of the digital twin was gradually introduced [9].

The expression “digital twin” was first introduced by Grieves in 2003, who characterized it as an ideal concept for product life cycle management, connecting physical components with their virtual counterparts through continuous information exchange [10]. In 2010, NASA described the digital twin as a multiphysics simulation model representing a technical system [11]. Since then, numerous definitions of the concept have been proposed, each one

highlighting specific aspects according to the perspective of the respective field or user, such as simulation, life cycle management, or data-driven services.

A general definition that serves as the foundation for this contribution is provided by [7], which describes digital twins as virtual representations of physical entities that are interconnected through real-time data exchange for various applications such as monitoring, design, optimization, maintenance, and remote access. The digital twin serves as a link between the physical and virtual domains. In this context, a digital twin can be understood as the virtual counterpart of a physical object. At the system level, this could include, for example, an excavator in mobile hydraulic applications, a press in stationary hydraulic systems, or a handling machine in pneumatic applications. At the component level, it may represent individual components such as a valve or a pump. Between the physical and virtual worlds, various tasks are carried out that correspond to different phases within the product life cycle (see **Figure 1**).

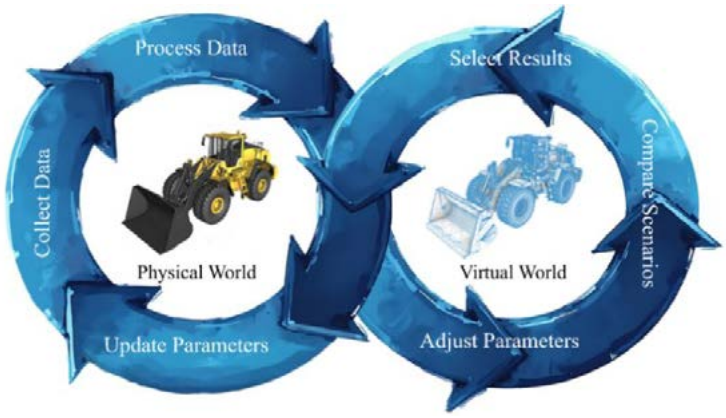


Figure 1: Connecting the virtual and physical worlds using digital twins.

With respect to the product life cycle, several stages can be identified that link the virtual representation to the physical asset. The digital twin facilitates the following steps:

- *Collect data*: Acquisition of data from various sources, such as field devices or sensors, followed by appropriate structuring and organization of the data.
- *Process data*: Transmission of the collected data between devices, specifically from the physical system to the digital twin, and preparation of the data for subsequent analysis. For instance, transferring parameters from the real application into simulation environments.
- *Adjust parameters*: Adaptation of parameters for specific applications, such as the calibration or parameterization of simulation models within the digital environment.
- *Compare scenarios*: Examination of different operating conditions using the digital twin, including parameter studies or evaluation of new configurations, e.g., assessing the performance of a new component.
- *Select results*: Identification and selection of suitable solutions based on scenario analysis, for example, determining an optimal set of parameters for controller settings or system components.

- *Update parameters*: Integration of the optimized parameters into the physical system, thereby closing the feedback loop between the digital and physical domains.

3. UNDERSTANDING THE RELEVANT COMPONENTS OF DIGITAL TWINS

Digital twins should be developed in parallel with product development and interact with the fluid power system during operation. To clearly define digital twins, their key aspects must be specified. This starts with defining data, information, and knowledge, as these constitute what digital twins exchange with real systems. A simplified product life cycle process for fluid power systems then illustrates where digital twins are created and applied. In the following, simulation models are addressed, as they often form the basis for digital twin modeling. Finally, the Asset Administration Shell (AAS) is introduced as an example of an interoperable interface for data exchange.

Data, information, knowledge – Clarification of terms often used synonymously

A continuous exchange of data between the digital twin and its corresponding physical object is essential. Therefore, it is important to establish a clear understanding of the concepts of data and information. However, in practice, the terms data, information, and knowledge are often not clearly differentiated and are frequently used interchangeably. Furthermore, a wide range of definitions can be found in the literature. To ensure conceptual clarity, this contribution adopts the framework proposed by North [12], as shown in **Figure 2**, which provides an overview of these terms along with examples from fluid power.

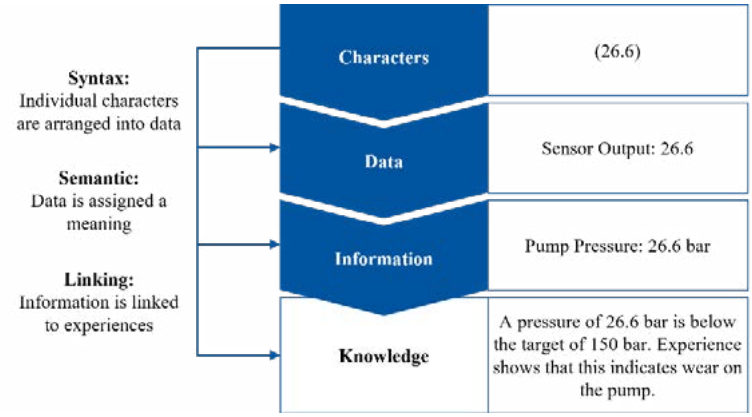


Figure 2: Different between characters, data, information, knowledge according to [12] extended with an example from fluid power.

According to North [12], data is generated through the application of syntactic rules to characters. It therefore represents arbitrary, non-interpreted sequences of symbols, such as a random string of numbers. In the context of fluid power, data can be understood as raw measurements obtained from a real application, for example a sensor output of 26.6. By assigning semantics, data is transformed into information. Information thus refers to data that is embedded within a meaningful context. For instance, interpreted data such as pressure signals from a specific pump (e.g., 26.6 bar). Knowledge, in turn, arises from linking

information with experience. For example, knowing that a certain system pressure is critical for the operation of a hydraulic system, or that a specific pressure value indicates possible pump wear.

To enable a successful exchange between the real system and its digital twin, a shared understanding is essential. The real system and the digital twin must both be able to interpret data in the same way.

Product life cycle – Developing and using digital twins

The product life cycle of fluid power systems, as illustrated in simplified form in **Figure 3**, can be divided into several phases. The division of the phases of the development process is adapted from [13]. The starting point is the requirement phase, where the requirements for the development task are defined and identified. After that, in the initial concept phase, the system layout is designed based on experience and system requirements, resulting in a schematic concept of the system layout. During the subsequent detail development phase, a simulation model is created by integrating component models within a suitable simulation environment. This approach aims to size the components appropriately and determine optimal parameter sets. After simulative optimization, the real system is commissioned. Although parameter transfer to the physical system is still predominantly performed manually, the simulation model continues to support the system during operation. Throughout the product life cycle, it is used for tasks such as efficiency enhancement, process monitoring, and the testing of planned reconfigurations. After the machine's operational phase, various considerations regarding its further use arise during the end-of-life phase. An explanation of various strategies for the further use of a product in context of circular economy can be found in [14].

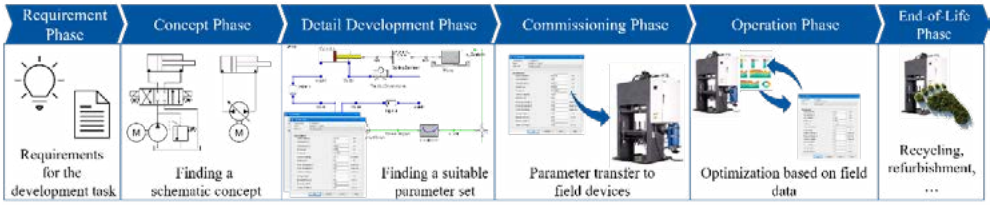


Figure 3: Simplified product life cycle phases.

Simulation models – Foundation for modeling digital twins

Simulation models form the foundation for modeling digital twins. Depending on the required level of accuracy, different types of simulation approaches are employed, as described in [1]. The simulation of fluid power components and systems is an essential tool in product development and optimization. It not only shortens development time but also provides access to parameters that are difficult or even impossible to measure experimentally. Once a model has been created and validated, extensive parameter studies can be performed efficiently and with minimal additional effort. Depending on the level of modeling detail, simulations can range from zero-dimensional to three-dimensional analyses. Furthermore, they can be classified as either static or dynamic, depending on whether time-dependent variations in state variables are considered or only stationary processes are examined. In concentrated (lumped)-parameter simulation, systems are modeled using discrete elements, and this approach is often used for fluid power system

simulation (see **Figure 4**). In contrast, distributed-parameter simulations enable spatial and temporal resolution of state variables and are often used to analyze specific geometries of fluid power components, such as a valve geometry.

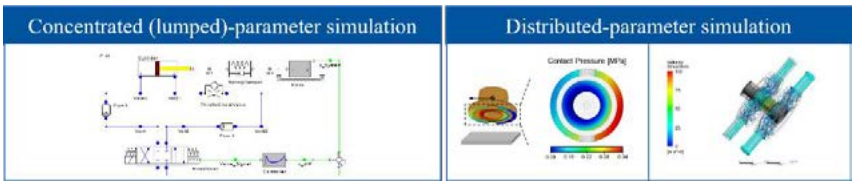


Figure 4: Exemplary representation of concentrated-parameter and distributed-parameter simulation according to [1].

AAS – Interoperable data and model interface

A possibility to transfer data and models through the digital twin is provided by the Asset Administration Shell (AAS). The AAS was developed in cooperation with the "Plattform Industrie 4.0" initiative [15] and the foundational framework of the AAS was published in DIN SPEC 91345 [16]. The AAS concept is becoming increasingly important in fluid power [17]. It serves as a digital representation of an asset, such as a hydraulic valve, enabling Industry 4.0-compliant interaction. It contains comprehensive descriptions of the asset and its associated services within a defined context throughout the entire product life cycle [16].

The AAS comprises submodels that describe an asset by means of SubmodelElements, for example properties or files [18]. Submodels can be standardized into templates to capture the relevant SubmodelElements needed for particular use cases (e.g., the submodel template "IDTA 02023 Carbon Footprint" to provide carbon-footprint-related data of an asset [19]) [18]. Both submodels and SubmodelElements can be annotated with semanticIds, which act as unique identifiers linking to semantic descriptions stored in standardized dictionaries (e.g., ECLASS [20] or IEC CDD [21]) [18]. These semantic references specify the exact meaning of the content contained within the annotated elements. The schematic concept of the AAS is shown in **Figure 5**.

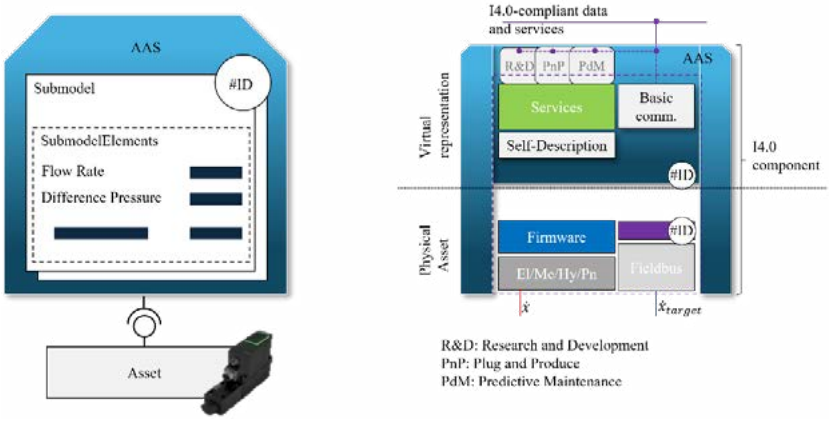


Figure 5: Schematic concept of AAS (left) and exemplary implementation of AAS in the application case of automated commissioning (right).

By providing an interoperable interface for product data and models, the AAS facilitates seamless collaboration among stakeholders, including component suppliers, system manufacturers, and operators, across the fluid power system development process [5]. By employing the AAS, data can thus be exchanged and utilized consistently across different systems and stakeholders.

4. CONNECTING THE VIRTUAL AND PHYSICAL WORLDS VIA DIGITAL TWINS

The following section presents relevant research findings relating to the process steps shown in **Figure 1.**

4.1. Collect data

Collecting and structuring data from various sources, such as field devices and sensors, is a crucial step in the development of digital twins. In addition to conventional hardware sensors, software-based solutions such as soft sensors play an increasingly important role. Soft sensors calculate or estimate process variables using mathematical models or data-driven algorithms instead of direct physical measurements. They are often used to determine variables that are difficult to measure from variables that are easy to measure. Moreover, in the field of hydraulic systems, current developments show a clear trend toward components with integrated intelligence and extended functionality. As a result, soft sensors are gaining importance as they offer a cost-effective way to obtain relevant process information while reducing the need for additional hardware.

One example of soft sensor application is the calculation of volume flow from pressure signals. Measuring the volume flow in fluid power systems continues to pose a major challenge. On the one hand, volume flow sensors are almost invasive sensors and therefore they have an influence on the flow. On the other hand, volume flow sensors are often very expensive, which is why fluid power systems are often designed without them. A virtual flow sensor enables the flow rate to be determined, which in turn makes it relatively easy to draw conclusions about the system, such as calculating hydraulic energy or, for example, detecting leaks at an early stage. Conventional sensors are often unable to detect highly unstable flow conditions due to their slow response times or their dependence on assumptions that only apply to steady, laminar flows, as described by Hagen-Poiseuille's law. By extending the applicability of measurements to turbulent and unstable flow regimes, soft sensors enable condition monitoring in fluid power systems. In addition, a soft sensor can also functionally improve the informative value of conventional sensors.

The development of a virtual volume flow sensor is shown in [3, 22], where the sensor was developed based on Hagen-Poiseuille's law to calculate the transient flow conditions of the volume flow. Theoretical models for transient, laminar, turbulent, and compressible flows were developed and validated by one-dimensional and three-dimensional simulations as well as experimental tests. The results showed that the soft sensor can measure transient volume flow and is well suited for practical industrial applications (see **Figure 6**).

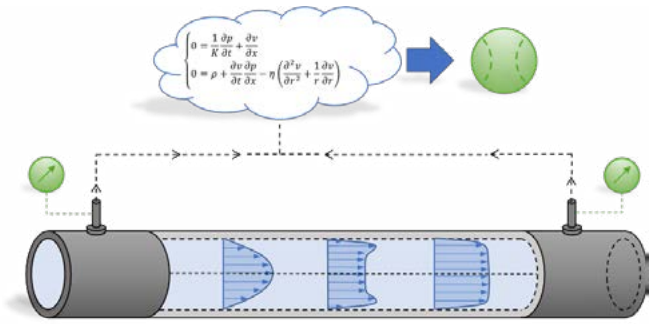


Figure 6: Schematic design of virtual volume flow sensor. The volume flow is calculated based on measured pressure signals.

Another example of the use of soft sensors is the flow calculation, based on the force of the electromagnetic valve actuator. The development of such a sensor is shown in [4] and the concept is illustrated in **Figure 7**. The approach aimed to develop a methodology for volume flow estimation on direct-actuated valves with solenoids for constant pressure systems. This approach provides a cost-effective, software-based sensor solution. By detailed modeling of the valve, inherent disturbances can be considered and thus the volume flow through the valve can be inferred from the electrical variables of the actuator as well as the armature position.

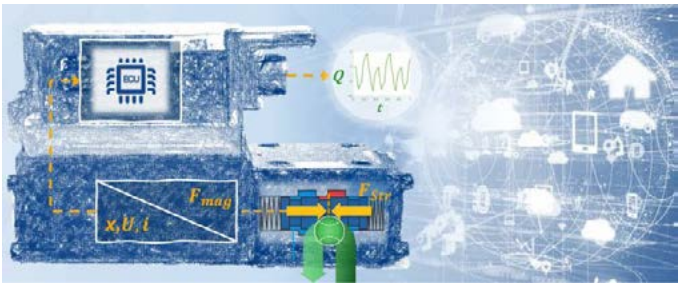


Figure 7: Schematic concept of virtual volume flow sensing, based on the force of the electromagnetic valve actuator.

In addition to soft sensors, there are other possibilities for data collection. One example is shown in [23], where simulation-based and measured data sets are used to train a condition monitoring algorithm based on machine learning. The approach is illustrated in **Figure 8** using the example of a hydraulic press. Machine learning is applied to identify patterns related to system failures (e.g., external and internal leakage). For this purpose, a database is built from both physical and virtual machines (simulation model). Subsequently, feature detection is performed, and machine learning is used to train the algorithm to recognize these features. In this research, the required amount of data and the combination of simulated and real data are also investigated. However, most algorithms are not designed to handle raw machine data directly and therefore require the data to be compressed and transformed into a smaller set of representative features. While several coding toolboxes aim to simplify feature generation through generic feature extraction methods, this work presents an enhanced procedure for automatically refining and selecting features. Moreover, an

approach for condition monitoring based on neural network is shown in [24]. The study shows that a feed-forward neural network can reliably identify different operating states of a pneumatic system based on vibration, pressure, and position data. It also becomes clear that even low-cost sensors enable reliable vibration-based condition monitoring, thereby supporting the industrial adoption of such maintenance strategies.

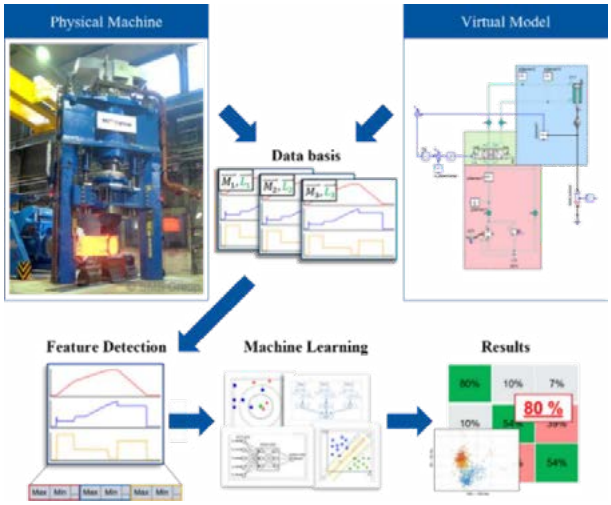


Figure 8: Collecting data from different databases to train a machine learning-based condition monitoring algorithm.

Another example of data collection for condition monitoring is given in [25]. This contribution proposes a cost-effective condition monitoring method for axial piston pumps using a single low-budget sensor capable of capturing multiple fault indicators. Data from various sensors operating under different conditions were used to detect faults such as valve plate cavitation, slipper wear, and axial clearance using machine learning algorithms.

4.2. Process data

Once the data has been collected, it is transferred between devices, specifically from the physical system to the digital twin, and prepared for analysis. For example, parameters can be transferred from the real application into simulation environments. The AAS provides an interoperable interface for data, models, and related information (see **Figure 9**). Consequently, within the development process, a variety of stakeholders such as component suppliers, system manufacturers, and operators can be seamlessly integrated. At the same time, heterogeneous data sources become accessible in a standardized and interoperable manner. By employing the AAS, data can thus be exchanged and utilized consistently across different systems and stakeholders. AAS is already being used in some industry-relevant use cases, an overview of which is given in [17]. In particular, within the domain of fluid power, several research activities are already being conducted in the context of simulation-based development and parameterization [5, 26, 27], and in the field of automated commissioning [28].

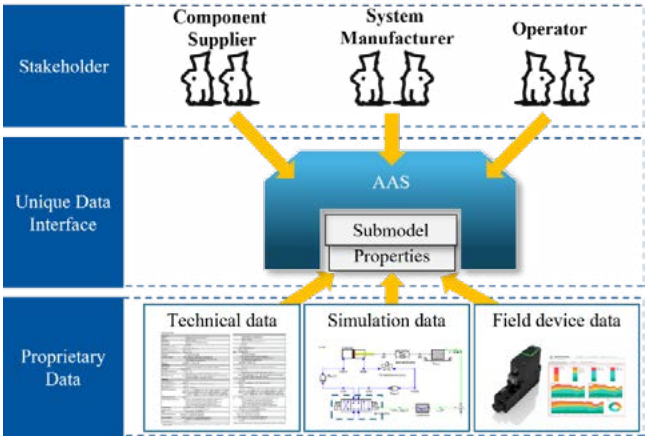


Figure 9: Integration of AAS into the product life cycle to enable a unique data interface.

4.3. Parameter adjustments

After the data is transferred, it is adapted for use in a virtual environment, for example for engineering applications. Adapting parameters to specific applications, such as parameterizing simulation models within a simulation environment, represents a key step in the development process. One example of parameter adjustment occurs during the detail development phase, as described in [5, 26] and illustrated in **Figure 10** and **Figure 11**.

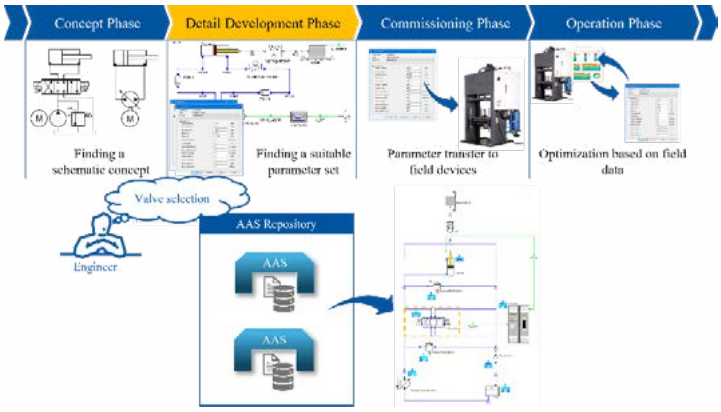


Figure 10: Exemplary use case for parameterization simulation models via AAS.

Engineers traditionally select components and set initial parameters manually by consulting various sources such as websites and data sheets, a process that is time-consuming and prone-to-errors. For example, an engineer may need to select a new valve and verify whether it has the correct size. This task usually involves clarifying the requirements and parameterization of the simulation with the relevant component data. This process can be made significantly easier using AAS.

The automatic component selection and parameterization of the simulation is described in **Figure 11** with more detail below.

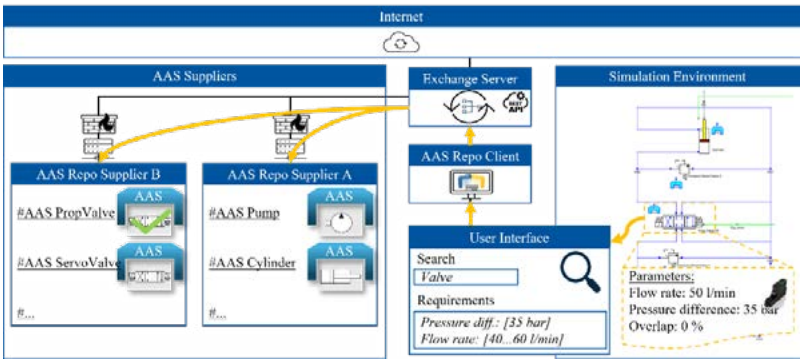


Figure 11: Parameter adjustment in the simulation-based development process.

A workflow is illustrated in which simulation data are adjusted based on parameters provided by the AAS. To enable this, a framework was developed that allows the parameterization of simulation models via the AAS. The framework is implemented using an HTTP REST API and considers manufacturer-specific safety requirements. The framework supports component selection by enabling searches in AAS Repositories provided by manufacturers to identify components that meet the desired requirements. The framework allows the definition of query criteria such as component type (e.g., valve or pump) and technical requirements (e.g., pressure difference and flow rate) specified as exact values or ranges. In the example considered in **Figure 11**, the engineer searches for a valve to ensure that the hydraulic press meets its functional requirements. Based on preliminary calculations, the engineer estimates the valve specifications, such as a nominal pressure difference of 35 bar and a flow rate of 40 to 60 l/min. Another example of adjusting parameters is shown in [29]. The contribution presents an approach for the identification, tuning, and verification of digital twins in hydraulic control systems. Another approach to parameter adjustments is presented in [30], which demonstrates a self-optimizing fuzzy PID controller for the independent control of the speeds of two non-identical electric motors. An example in the pneumatic area is shown in [31]. This contribution shows a concept for modeling and simulating the fluidic behavior of pneumatic vacuum generators. To increase the accuracy of the simulations, the approach uses instance-specific data rather than just object-specific data for the respective asset. The model and data are integrated into the digital twins of the pneumatic vacuum generators so that they can be combined with other components to represent digital twins of complete vacuum handling systems.

4.4. Scenario calculation and comparison

After parameterizing the engineering applications, such as simulation models, various scenarios are calculated and compared according to the engineering objectives. Finally, a suitable configuration is selected. Therefore, parameter studies are carried out for testing new configurations, for example evaluating the performance of a new component in a hydraulic system, finding suitable settings for a controller or testing different designs of slippers in hydraulic pumps. An example of scenario analysis is shown in [32] where displacement-controlled hydraulic systems are analyzed as an energy-efficient alternative to conventional valve-controlled systems in construction machinery. Another example is shown in [33], where different geometric designs of slippers for axial piston machines are

investigated. After the scenario comparison, suitable configurations are selected, such as an optimal set of parameters for a controller or the best design of a slipper. Therefore, tools are being researched that enable fast calculation of scenarios and allow efficient comparison of the results. Examples include [34, 35], in which a framework of fast computation of lubricated contacts using PINNs was developed. Conventional simulation tools are often time-consuming for extensive parameter studies and scenario comparison. Accurately representing all possible operating conditions requires many simulation cases, and in some applications, near real-time performance is even necessary for control tasks. Although extensive research has been conducted in this area, further efforts are particularly needed to accelerate calculations through AI-based approaches such as PINNs and neural networks, as well as in uncertainty management and cross-domain and cross-scale simulation.

The identified configurations and parameters are then made available again and can be used to adjust the real machine. At this stage, there is potential for an automatic transfer of parameters to the real application using tools such AAS. In **Figure 12** an overview of scenario calculation and comparison is shown. This process can be accelerated using AI-based simulation frameworks.

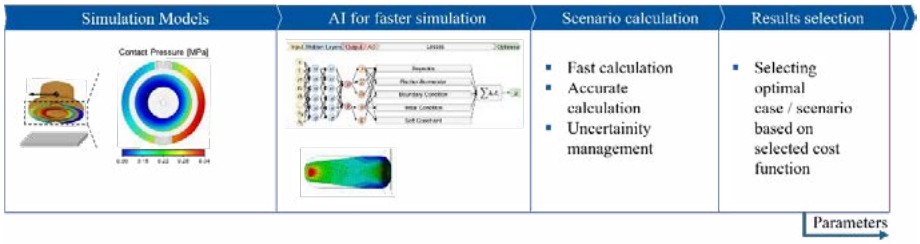


Figure 12: Scenario calculation and comparison with AI-based accelerated simulation.

Moreover, in [36] an optimization framework for hydraulic hybrid vehicles is presented. The paper introduces a simulation-based optimization approach for assessing a series hydraulic hybrid vehicle transmission, demonstrating that numerical optimization enables exploration of a wide design space and reveals trade-offs between component sizing and control strategies. This framework builds on the simulation-based optimization methodology originally introduced in [37]. Therefore, various scenarios defined by the design parameters of the hydraulic components are analyzed and systematically compared.

4.5. Update parameters

Feeding the optimized parameters back into the physical system closes the information loop between the digital and physical domains. By applying these updated parameters, the system operates more accurately and reliably, improving controllability, ensuring smoother performance, and enhancing overall output. At the same time, optimizing energy efficiency reduces energy consumption, resulting in a more sustainable and cost-effective operation. An example of updating parameters is presented in [38], where deep reinforcement learning (RL), an emerging branch of machine learning, addresses complex nonlinear optimal control problems through data-driven, trial-and-error learning. In this contribution, a deep RL controller is applied to a hydromechanical system, an inverted pendulum on a cart with a hydraulic drive, demonstrating a comprehensive framework that enables training in simulation to swing up and balance the pendulum. Another example is shown in [39], where

mobile off-highway machines are equipped with sensors that continuously record data such as position, inclination, power, and information about surrounding obstacles for analysis. Through sensor data fusion and modeling, this information can be enriched with geospatial data to generate semantic insights about the environment. As a result, a digital twin of the surroundings can be created, supporting geodata-based processes like digital construction planning or terrain analysis. Despite these research findings, much work remains to be done from both a scientific research perspective and an industrial standpoint. Transforming the available information into tangible benefits is a complex challenge, yet it is essential to fully realize the potential of digital twins. **Figure 13** summarizes the potentials of connecting data from real and virtual world, e.g., for the optimization of the machine controller.



Figure 13: Transferring parameters from the virtual to the real world.

5. CONCLUSION

Fluid power is a key technology applied across various industries but continues to face challenges in energy efficiency, environmental impact, and digital integration. Digital twins provide a promising approach to address these issues. However, the extent of existing research in this field remains unclear. Therefore, this contribution addressed the research question: *What is the current development level of digital twins in fluid power?* To answer this question, the term “digital twin” was first derived from scientific literature. Moreover, key aspects for understanding a digital twin were identified and explained, including, for example, simulation as a physical modeling of reality and the AAS as an example for a model and data carrier. Subsequently, this contribution presents relevant exemplary research findings from the field of digital twins in fluid power.

This contribution points out the significant progress made in recent years regarding concepts and solutions for various aspects of digital twins in fluid power. In addition, there are very advanced research activities related to collecting data with digital methods, for example using machine learning or the development of soft sensors. Moreover, there are research activities concerning process data and data adjustment, especially regarding the use of the AAS. Furthermore, in terms of scenario analysis and comparing scenarios, there are increasing activities focused on using PINNs to accelerate simulations and to enable an easy comparison of different scenarios. The identified research activities underline that there are many ongoing efforts in each sub-area of the digital twin process. Building on recent

advances, future research should focus on developing a unified and holistic framework for digital twins in fluid power.

6. REFERENCES

[1] K. Schmitz, *Fluidtechnik - Systeme und Komponenten,* 1st ed.: Shaker Verlag, 2021.

[2] L. Schmidt, K. Schmitz, T. Minav, and A. Senatore, "Estimating the Global Emissions by Fluid Power Machinery: Under review," *15th International Fluid Power Conference (IFK 2026)*, 2026.

[3] F. Brumand-Poor, T. Kotte, E. G. Pasquini, and K. Schmitz, "Signal Processing for Transient Flow Rate Determination: An Analytical Soft Sensor Using Two Pressure Signals," *Signals*, vol. 5, no. 4, pp. 812–840, 2024, doi: 10.3390/signals5040045.

[4] S. Hucko, H. Krampe, and K. Schmitz, "Evaluation of a Soft Sensor Concept for Indirect Flow Rate Estimation in Solenoid-Operated Spool Valves," *Actuators*, vol. 12, no. 4, p. 148, 2023, doi: 10.3390/act12040148.

[5] M. Becker, S. Heppner, R. Alt, T. Kleinert, and K. Schmitz, "Improving Fluid Power System Simulation Through an AAS-Based Simulation Framework," in *Proceedings of the 2023 ASME/BATH Symposium on Fluid Power and Motion Control*: ASME, 2023.

[6] D. Ritz, M. Becker, J. Weber, and K. Schmitz, "Accelerating Product Development: The Effort of Asset Administration Shells for aggregated Systems," in *The 19th Scandinavian International Conference on Fluid Power*, 2025.

[7] M. Singh, E. Fuenmayor, E. Hinchy, Y. Qiao, N. Murray, and D. Devine, "Digital Twin: Origin to Future," *ASI*, vol. 4, no. 2, p. 36, 2021, doi: 10.3390/asi4020036.

[8] S. Boschert and R. Rosen, "Digital Twin - The Simulation Aspect," *Mechatronic Futures*, pp. 59–74, 2016.

[9] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen, "About The Importance of Autonomy and Digital Twins for the Future of Manufacturing," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 567–572, 2015, doi: 10.1016/j.ifacol.2015.06.141.

[10] M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," in *Transdisciplinary perspectives on complex systems: New findings and approaches*, F.-J. Kahlen, S. Flumerfelt, and A. C. Alves, Eds., Cham: Springer, 2016, pp. 85–113.

[11] M. Shafto *et al.,* "Modeling, Simulation, Information Technology and Processing Roadmap: Technology Area 11," 2010.

[12] K. North, *Wissensorientierte Unternehmensführung: Wissensmanagement gestalten,* 6th ed. Wiesbaden: Springer, 2016.

[13] F. Engler, H. Baum, and R. von Dombrowski, "Development Environment for Fluid-Power-Mechatronic Systems," in *Proceedings of the 7th International Fluid Power Conference (IFK 2010)*, 2010.

[14] J. Kirchherr, D. Reike, and M. Hekkert, "Conceptualizing the circular economy: An analysis of 114 definitions," *Resources, Conservation and Recycling*, vol. 127, pp. 221–232, 2017, doi: 10.1016/j.resconrec.2017.09.005.

[15] S. Grüner *et al.,* "Integration von Module Type Package und Industrie 4.0 Verwaltungsschale," *Automation*, 2021, doi: 10.51202/9783181023921-421.

[16] *DIN SPEC 91345: Referenzarchitekturmodell Industrie 4.0 (RAMI4.0)*, DIN SPEC 91345:2016, DIN Deutsches Institut für Normung e. V., 2016.

[17] M. Becker, S. Karaoglu, F. Makansi, J. Radtke, and K. Schmitz, "Industry 4.0: Review of the State of the Art in Fluid Power Research and Industry," *International Journal of Fluid Power*, vol. 26, no. 04, pp. 573–624, 2025, doi: 10.13052/ijfp1439-9776.2643.

[18] Industrial Digital Twin Association, "Asset Administration Shell Specification - Part 1: Metamodel Schema: IDTA Number: 01001-3-0," 2023.

[19] Industrial Digital Twin Association, "IDTA 02023 Carbon Footprint: Submodel Template of the Asset Administration Shell," Version 1.0, 2025.

[20] ECLASS e.V. [Online]. Available: https://eclass.eu/ (accessed: Oct. 15 2025).

[21] International Electrotechnical Commission, *EC 61360-4 - IEC/SC 3D - Common Data Dictionary.* [Online]. Available: https://tc3.iec.ch/tc-activity/common-data-dictionary-cdd/ (accessed: Jul. 15 2025).

[22] F. Brumand-Poor, T. Kotte, E. Pasquini, F. Kratschun, J. Enking, and K. Schmitz, "Unsteady flow rate in transient, incompressible pipe flow," *ZAMM – Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 2024, doi: 10.1002/zamm.202300125.

[23] F. Makansi and K. Schmitz, "Feature Generation and Evaluation for Data-Based Condition Monitoring of a Hydraulic Press," in *Proceedings of the 13th International Fluid Power Conference (IFK 2022)*, K. Schmitz, Ed., 2022.

[24] M. Tiboni and C. Remino, "Condition Monitoring of Pneumatic Drive Systems Based on the AI Method Feed-Forward Backpropagation Neural Network," *Sensors (Basel, Switzerland)*, vol. 24, no. 6, 2024, doi: 10.3390/s24061783.

[25] S. Horn, S. Knoll, A. Shorbagy, M. Lenz, and J. Weber, "Economic Fault Detection for Axial Piston Pumps Under Variable Operating Conditions," *International Journal of Fluid Power*, vol. 2025, pp. 319–342, doi: 10.13052/ijfp1439-9776.2628.

[26] S. Heppner, T. Miny, T. Kleinert, M. Becker, K. Schmitz, and R. Alt, "Asset Administration Shells as Data Layer for Enabling Automated Simulation-based Engineering," in *2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2023, pp. 1–7.

[27] R. Alt *et al.,* "Seamless Integration of Device and Field Data into the System Simulation of a Hydraulic Servo-Press Using AAS," in *Proceedings of the 14th International Fluid Power Conference (IFK 2024)*, J. Weber, Ed., Aalborg: River Publishers, 2024, pp. 241–257.

[28] R. Alt, "Service-Oriented Architecture for Automated Commissioning of Fluid Power Systems," Dissertation, RWTH Aachen University, Aachen, 2023.

[29] G. Ćwikła and M. Szewczyk, "Station for Tuning and Testing Digital Twin in Hydraulic Actuator Control Systems under Programmable Load—Methodology, Design, and Tests," *Electronics*, vol. 13, no. 17, p. 3528, 2024, doi: 10.3390/electronics13173528.

[30] J. Heikkinen, T. Minav, and A. D. Stotckaia, "Self-tuning parameter fuzzy PID controller for autonomous differential drive mobile robot," in *2017 XX IEEE International Conference on Soft Computing and Measurements (SCM)*, Saint Petersburg, Russia, 2017, pp. 382–385.

[31] V. Stegmaier, W. Schaaf, N. Jazdi, and M. Weyrich, "Simulation Model for Digital Twins of Pneumatic Vacuum Ejectors," *Chem Eng & Technol*, vol. 46, no. 1, pp. 71–79, 2023, doi: 10.1002/ceat.202200358.

[32] P. Casoli, F. Scolari, T. Minav, and M. Rundo, "Comparative Energy Analysis of a Load Sensing System and a Zonal Hydraulics for a 9-Tonne Excavator," *Actuators*, vol. 9, no. 2, p. 39, 2020, doi: 10.3390/act9020039.

[33] F. Schlegel and K. Schmitz, "Geometric design of axial piston machine slipper-bearings made of high-performance plastics," *Forschung im Ingenieurwesen*, vol. 89, no. 1, 2025, doi: 10.1007/s10010-025-00813-2.

[34] F. Brumand-Poor, F. K. Azanledji, N. Plückhahn, F. Barlog, L. Boden, and K. Schmitz, "Extrapolation of cavitation and hydrodynamic pressure in lubricated contacts: a physics-informed neural network approach," *Adv. Model. and Simul. in Eng. Sci.*, vol. 12, no. 1, 2025, doi: 10.1186/s40323-025-00283-9.

[35] F. Brumand-Poor, M. Trautmann, and K. Schmitz, "Advancing deformation calculation: a physics-informed deep graph learning framework for hyperelastic materials," *Adv. Model. and Simul. in Eng. Sci.*, vol. 12, no. 1, 2025, doi: 10.1186/s40323-025-00304-7.

[36] K. Baer, L. Ericson, and P. Krus, "Framework for simulation-based simultaneous system optimization for a series hydraulic hybrid vehicle," *International Journal of Fluid Power*, pp. 1–13, 2018, doi: 10.1080/14399776.2018.1527122.

[37] P. Krus, "Simulation based optimisation for system design," *International Conference on engineering design*, 2003.

[38] F. Brumand-Poor, L. Kauderer, G. Matthiesen, and K. Schmitz, "Application of Deep Reinforcement Learning Control of an Inverted Hydraulic Pendulum," *TJFP*, 2023, doi: 10.13052/ijfp1439-9776.2429.

[39] Deuster, Sebastian, A. Bliznyuk, C. Haas, and K. Schmitz, "Methods of Environmental Information Acquisition by Sensor Data and Digital Twins," in *Proceedings of the 13th International Fluid Power Conference (IFK 2022)*, K. Schmitz, Ed., 2022.

Biographies



Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.



Malte Becker obtained his Master of Science degree in Mechanical Engineering from RWTH Aachen University. Since 2021, he has been working as a research associate at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. His research focuses on the Industrial Internet of Things (IIoT) and Systems Engineering. Since 2025, he has been serving as Chief Engineer at ifas.

Next Generation Valve Actuation Platform

Dr.-Ing. Dirk Wroblewski

Danfoss Power Solutions, Nordborgvej 81, DK-6430 Nordborg, Denmark

Email: dwroblewski@danfoss.com

Abstract.

Danfoss is going to develop a new generation of valve actuation platform for our PVG system. Featured with integrated sensors, enhanced connectivity, memory scalability, etc., it allows the creation of actuator variants customized to the different needs of the application systems. This paper will focus on the modular concept, and key features of this next generation actuation platform. One element of it is the voice coil driven pilot valve actuator. This is an important technology for highly advanced work function control tasks to create application performance driven by software design elements in favor of performance driven by mechanical design elements. This paper presents a design study with the goal to demonstrate the ability to design a cost-effective voice coil actuation for a PVG valve. The preliminary performance and key characteristics of the actuator will be discussed in this paper.

Keywords. pilot valve, voice coil, electro-hydraulic control valve, mobile machines.

1. INTRODUCTION

The quest to achieve seamless control of pressure and flow rate in fluid power systems has undergone major development in the past decades. Back in the 80's the industry shifted from mechanical/hydraulic controls to electro-hydraulic controlled proportional valves. These systems emerged as a cost-effective alternative to servo valves, enabling stepless control of pressure and flow rate via electrical signals.

To improve the performance of proportional valves, relative to the performance of servo valves, the manufacturers added linear variable displacement transducers (LVDTs) to proportional valves in order to sense the spool position. The output signal from the LVDT was used to calculate the difference between where the main spool was supposed to be and where it actually was, and that difference was passed on to modulate the pilot valve in an effort to position the main spool to achieve the desired output based on the input [1]. This simple, yet efficient closed loop control principle enabled machine builders and system integrators to make easy integration, i.e. plug 'n play of high performing control valves.

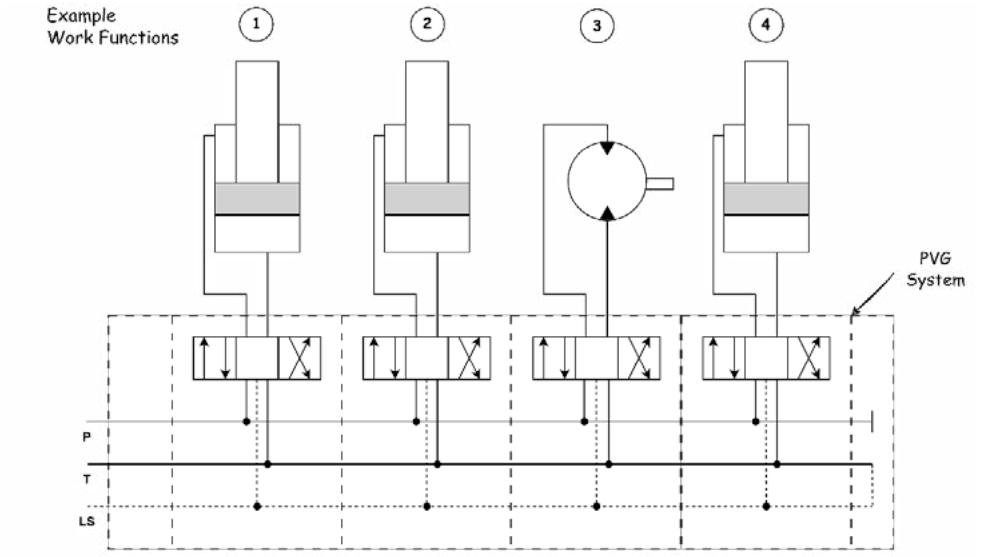
In 1988 Danfoss launched the PVG32 (Proportional Valve Group) series, a modular control valve system designed for mobile machinery. The system's concept was novel at the time

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Classified as Business

and offered a flexible, load-sensing directional control valve, designed for advanced electrohydraulic control. Over time, Danfoss and others have evolved electro-hydraulic valve actuation with microcontrollers, fault monitoring, integrated sensors, software-controlled options, and CAN bus communication to meet versatile requirements in a broadening application range for electro-hydraulic controlled proportional valves. Concurrently the PVG system has grown with adding new frame sizes, each demanding new variants or series of electro-hydraulic actuation.

Contrary to common assumption, the requirements for different work functions in many applications are far from uniform. The ability to configure a proportional control valve stack, serving versatile work functions, with valves and actuation technology tailored to the actual demand, provides scalable yet cost-efficient solutions to the machine builders to meet the individual value proposition of each work function. As a result, modular valve systems with flexible valve actuation technologies play an important role in helping machine builders to design a control system that is optimally customized to their applications, both in terms of cost and technology.



High Level Work Function Requirements	Proportional Control	Proportional Control	Proportional Control	ON-OFF Control
	Sensor controlled	Joystick. controlled	Joystick. controlled	Joystick. controlled
	Flow Rate: High	Flow Rate: Medium	Flow Rate: Medium	Flow Rate: Low
	Accuracy: High	Accuracy: High	Accuracy: Low	Accuracy: -
	Dynamics: High	Dynamics: Medium	Dynamics: Low	Dynamics: -
	Utilization: Always	Utilization: Often	Utilization: Rare	Utilization: Rare
	Productivity: High	Productivity: High	Productivity: Low	Productivity: Low

Figure 1. Work functions with basic requirements to illustrate the versatile value proposition of each work function

Classified as Business

2. MODULAR VALVE ACTUATION PLATFORM

As customer needs and technology evolve, so must the solution. Therefore, Danfoss is going to develop a new generation of valve actuation platform for our PVG system. It is a new hardware & software platform, fully modularized on electronic hardware and software domain, to support the prospective expansion of our proportional control valve portfolio. The new hardware & software platform addresses the most important needs of an increasingly growing variety of control valves.

Scalability: Product architects need to be able to easily add or remove components (e.g., drivers, sensors, actuators) without redesigning the entire system. Scalability in the design will support future upgrades and expansion.

Reusability: Modules (software blocks, hardware units) should be able to be reused across different product development projects, to reduce development time and cost.

Flexibility: The platform should support the customization of product variants. Flexibility in the design ensures the platform can support a wide range of applications from simple to highly complex applications.

Standardization promotes consistent design practices and interfaces and provides a common, well-known appearance to the system integrator / application engineer. Standardization enables easier integration with application systems. In short: If one knows a single product variant, one knows them all.

The core of the new platform will be a new microcontroller family with scalable memory to accommodate growing software, lockstep core technology and built-in self-tests (BIST) to increase fault coverage capabilities. Lockstep core technology is a mechanism where two processor cores run in perfect synchronization, executing the same instructions at the same time. Their outputs are continuously compared, and if a mismatch occurs, it signals a hardware fault. The external watchdog monitors, among other test points, the MCU's program sequencing ability with a question-response approach and disables power supply to the driver circuit via the safety switch. Secure Boot is a security feature designed to ensure that a device boots using only software that is trusted by the manufacturer. It helps protect against malware and unauthorized software running during the boot process. It supports compliance with security regulations (e.g. CRA).

A completely new software architecture will be developed, that exploits the microcontroller's capabilities but is still abstracted well enough to support the modular approach. The architecture will be designed with different peripheral and application specific layers. The layers will be vertically permeated by modular software components with well prescribed interfaces.

The microcontroller supports various analogue & digital input devices for application sensors, CAN Classic & CAN FD interfaces, plus Ethernet interface for future, data-hungry applications. On the actuation side the hardware & software platform primarily supports both solenoid and voice coil technology to control the associated pilot valves. But the electro-mechanical valve actuation technology can also be applied to this new platform to support the transition from mechanically actuated control valves to electrically actuated control valves.

Classified as Business

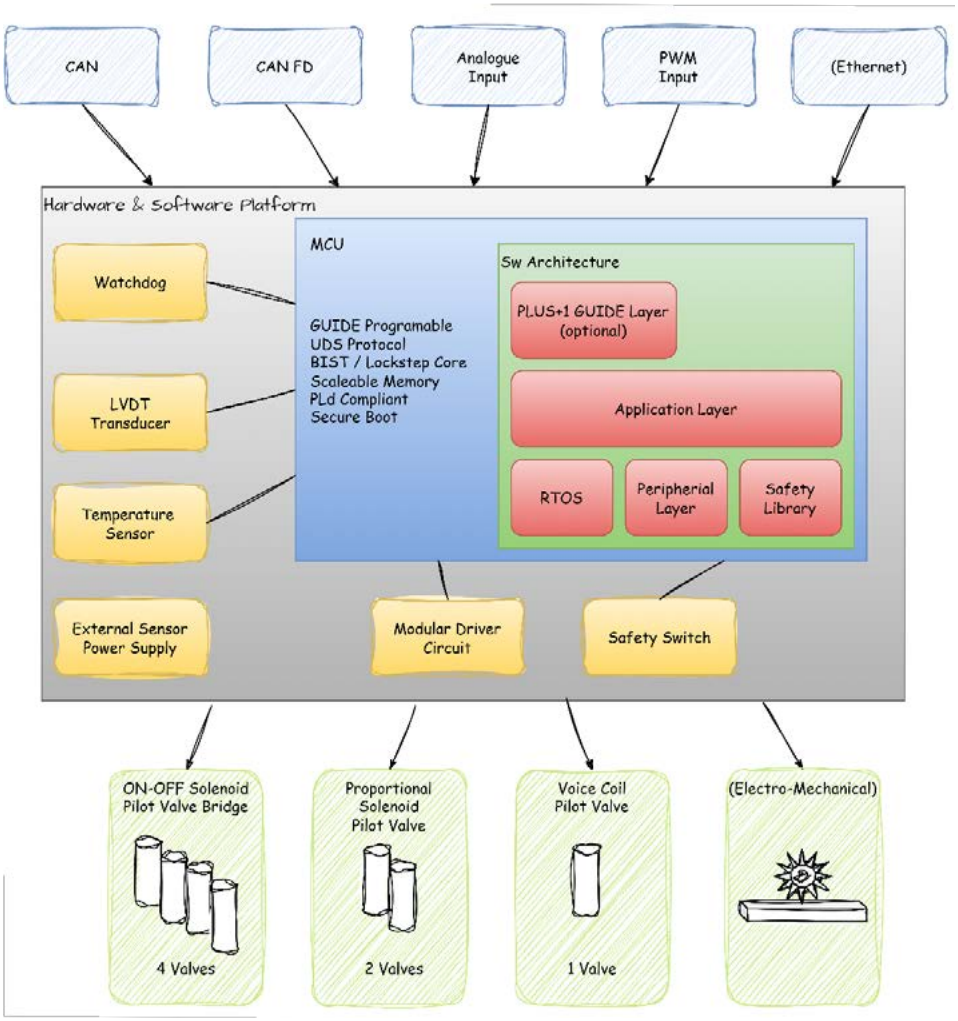


Figure 2. New PVG Valve Actuation Platform.

The platform will be designed to reach PL=d compliance (acc. ISO 13849) for application-level safety function **Safe Stop**. This means that whenever a safety-critical event occurs, i.e. overload situation, the boom movement must be safely stopped and an unexpected start-up from the rest position must be prevented. This platform supports these requirements by selecting highly reliable components and extensive diagnostics. The operation of the on-board electronics and the control valve is continuously monitored. If a failure is detected, the on-board electronics sends an error message on the CAN bus to the machine control unit and concurrently de-activates the solenoids via the integrated safety switch.

The interfaces and basic capabilities will be backwards compatible with current Danfoss PVE actuation series; however, we expect control performance enhancements due to increased microcontroller capabilities that would allow the implementation of sophisticated valve control algorithms. This new platform supports free combinations of input and actuation technology according to the upcoming demand and associated value proposition.

Classified as Business

For example, aside from conventional solenoid bridge technology with a classical CAN interface, Danfoss could also offer voice coil driven actuation with pressure or other application sensors. Or it could combine it with free programming space for customers.

All these technologies could be available in the same PVG valve system; individually tailored to the specific value and needs of the respective work function.

3. NEW FEATURES

The new platform provides a couple of new capabilities and features with the aspiration to be prepared for increasing integration, data exchange and control performance objectives.

Enhanced Control Valve Capability refers to offering new PVG features in addition to existing ones.

- Improvement on valve control through the implementation of advanced robust control techniques. In particular, the compensation of the inherent cold oil behaviour is a long-cherished desire.
- Integrated sensors for pressure, temperature and spool position can be utilized for detecting the work function's integrity, i.e. predictive maintenance, or they can be used to optimize the work function's control system.
- Realization of machine learning techniques to capture knowledge of highly skilled machine operators and transfer that knowledge into control logic to make it available to all machine users.

PLUS+1® GUIDE is a graphical software development environment from Danfoss, designed to simplify and accelerate the creation of control systems for mobile machinery. The tool is based on graphical software programming enabling users to develop their own individual work function application. When this development environment becomes part of the valve controller, customers will benefit from its close proximity to sensors and work functions in the development of their own intelligent solutions. This programable space is optional to the embedded application firmware.

CAN-FD (Controller Area Network – Flexible Data-rate) is an enhanced version of the original CAN protocol, designed to meet the increasing demands of modern automotive and industrial systems. It can handle larger and faster data exchanges, which are essential for modern, data-intensive applications. For instance, the controller software can be downloaded up to 5x faster than using classical CAN interface. CAN-FD nodes can coexist with classical CAN nodes, but classical CAN nodes cannot interpret CAN-FD frames.

Automotive **Ethernet** is a specialized version of Ethernet adapted for use in vehicles. It supports data rates from 10 Mbps to 10 Gbps, ideal for streaming large data frames, needed for advanced applications like autonomous driving.

Today it is not yet clear what kind of value Ethernet can offer to work functions. Eventually the move of centralized machine control functions to de-centralized machine control architectures might foster the use of Ethernet. A potential use case is the facilitation of many advanced, data dependent functions into a single valve system controller, servicing a whole valve stack. For example, this controller might process camera data as feedback signal for a given work function control task. When data bandwidth is no longer a constraint, many more options will become apparent.

Classified as Business

4. WHY VOICE COIL?

One element of the prospective valve actuation platform is the pilot valve with so called voice coil technology. Danfoss has already developed such a technology in conjunction with the development of CMA / PVX valves. These proportional control valves are known for their independent metering edge technology, their electronic load pressure compensation and superior controllability. They are well recognized by customers, but since they were developed independently, they do not benefit from the scaling and modularity of the PVG platform. On the other hand, the PVG platform does not benefit from the opportunities that the voice coil technology can offer.

The voice coil driven control valve achieves much higher dynamics, similar to servo valves, which enables them to react much more precisely and much faster to sensor signal induced valve commands. Features that were previously provided hydro-mechanically can now be provided by a software-enabled control function that uses sensor feedback and can be configured, tuned, and automatically adapted.

It is exactly the flexible control valve configuration by means of software that is appreciated by the customers. In the end it is the voice coil technology that enables significant gains in controllability/productivity beyond conventional pilot valves. For example, load pressure oscillations or the limitation of load pressure can be effectively controlled by voice coil driven control valves. **Figure 3** shows likewise the load pressure with / without active damping while lifting of a particular load. Actively dampening the load pressure oscillations contributes to stabilizing the work function acceleration, leads to decreased flow ramp-up time, to increased flow rate and operator comfort. Electronic load pressure limitation utilizes the full maximum load potential by precisely adjusting the port pressure by the control valve. All in all, these gains contribute to increasing the productivity of machines.

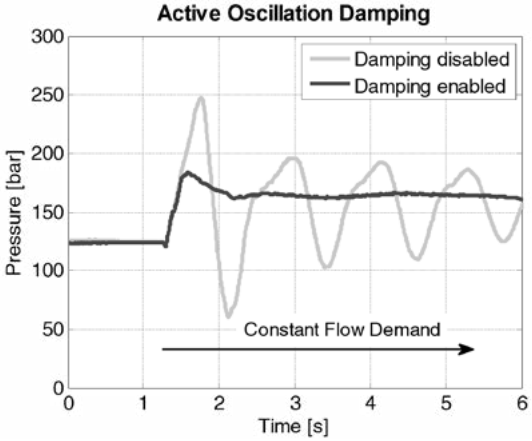


Figure 3. Effect of Active Oscillation Dampening. [2]

Voice coil driven valves also contribute to creating more energy efficient hydraulic systems. They provide the ability to switch the valve state very rapidly during energy regeneration or recuperation scenarios. These often require a fast redirection of the energy flow from one port to another, without interfering with the work function control.

Classified as Business

5. BASIC IDEA OF THE VOICE COIL

So, what is this "voice coil" all about? Voice coil principle is based on the Lorentz force. It is the force experienced by a moving charge in a magnetic or electric field. This principle is widely commercialized by electric motors but also in speakers that convert an electrical signal into a corresponding sound. Hence, the name voice coil. The magnetic component of the force is greatest when the direction of motion of the charge is perpendicular to the magnetic field lines, and zero when the charge moves along a field line. It always acts perpendicular to the direction of motion of the charge and to the magnetic field lines.

When a wire of length L is exposed to a magnetic field and carries a steady electric current I, each of the moving charges in the wire experiences the force. For a straight, stationary wire in a uniform magnetic field, this force F is given by

$$F = I \cdot L \times B \qquad (1)$$

where B is the magnetic flux density. The force is proportional to the coil wire length, the flux density and the coil current. The force direction will change when reversing the coil current. **Figure 4** illustrates the principle.

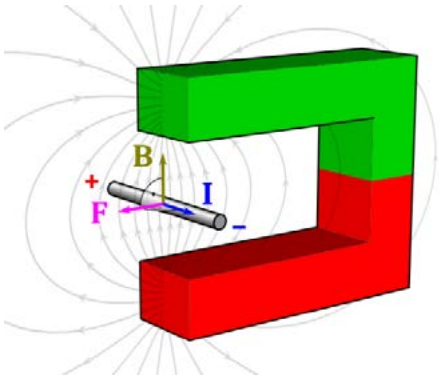


Figure 4. Lorentz force on a current-carrying wire exposed to a horseshoe magnetic field. [3]

This principle can be utilized by attaching a pilot valve to a moving coil and exposing the coil to the field of a permanent magnet. Due to the ability to reverse the coil force by reversing the current direction, the coil and pilot valve can move in both directions from a center position. A single 4/3 pilot valve can be designed to control a single main spool.

In contrast, traditional solenoids generate force in a single direction only, regardless of the current direction. Here, the solenoid force is generated by decreasing the magnetic circuit's reluctance (i.e. increasing inductance). Increasing the inductance happens when the solenoid's exerting force element, the armature, minimizes the air gap in the circuit. This movement happens independent of the coil's magnetic field polarity. Thus, these solenoid valves operate unidirectionally per given coil. However, in addition to the productive magnetic field created in the axial direction, a secondary (and unproductive) force is also created on the armature in the radial direction. This causes the armature to move laterally and stick to the solenoid body, introducing friction.

Classified as Business

Although the physics between armature and body is understood and minimized by surface coating means, this influence still generates significant friction and produces significant hysteresis in the solenoid valve behaviour. This inherent hysteresis decreases control valve resolution and reduces control valve dynamics. Both effects put proportional solenoid valves to a technologically inferior position against voice coil driven valves.

6. DESIGN STUDY

Despite the known performance advantage that voice-coil technology provides to hydraulic valves, one limitation has historically been a relatively higher cost. In this paper we reveal a design study, where we have built an experimental prototype with the goal to demonstrate the ability to design cost-effective PVG actuation utilizing the voice coil principle. See **Figure 5** and **Figure 6**.



Figure 5. Design study for voice coil driven PVG valve actuator.

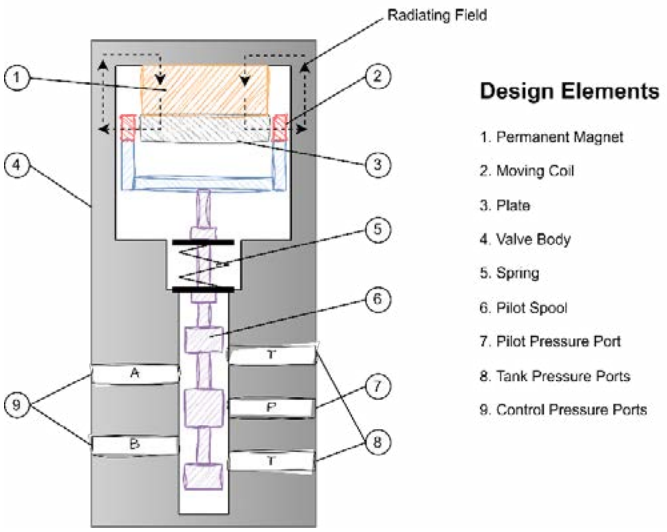


Figure 6. Sketch of the voice coil driven pilot valve. The moving coil is placed in the air gap where the magnetic flux density is highest.

Classified as Business

This proof-of concept consists of a permanent magnet (1), a moving coil (2), a valve body (4) and the pilot valve (5, 6, 7, 8, 9). The permanent magnet is assembled with a plate (3) to guide the magnetic field. The magnetic field radiates out of the plate, runs through the coil and enters the valve body. From there it travels through the valve body back to the permanent magnet.

The moving coil is connected to the pilot valve's spool. It is obvious that backlash is not allowed. The coil runs "wet", exploiting the oil bath to function as a heat sink. The coil wires are guided through a valve body opening (not displayed) to the ambience. There they are connected to the on-board electronics.

Further, the friction level of the moving parts must be low and predictable. The low clearance between spool and valve housing enables unambiguous defined bearing for the spool/moving coil assembly. The axial position of the spool/moving coil assembly is defined by a pre-loaded spring. To yield a high dynamic it is important to design the moving elements as light weighted as possible. The valve body offers hydraulic connection to pilot pressure, tank and two control pressure ports. Both control pressure ports are connected to either control chamber of the PVG main spool.

By applying a current to the coil, the spool moves out of the center position to connect pilot pressure with one of the control pressure ports. While the one control pressure port is connected to pilot pressure, the other one is connected to tank. And vice versa for reverted coil current.

As the pilot pressure travels into one of the PVG's main spool chamber, the main spool will move away from its neutral position. Once the main spool approaches its target position the logic reduces the coil current in order to reduce the pilot valve opening from pilot pressure to control port. Reducing the valve opening leads to adjusting the control chamber pressure so it will match the main spool spring force at particular main spool position. The pilot valve spool takes a position where all acting forces, such as coil force, spring force, friction, etc., are in equilibrium. Thus, holding a permanent PVG main spool position to meter a steady fluid flow to the work function requires the application of a permanent coil current.

When the coil is de-energized the pilot spool moves under the influence of the spring force to the center position at which the control pressure ports are connected to tank, and the PVG's main spool will return to its neutral position.

The law of the Lorentz force provides guidance in sizing the design elements to optimize the coil force to given space constraints. To identify the individual design element's sensitivity to the coil force, we have conducted numerous finite element simulations to calculate the coil force for a given design iteration. These simulations confirmed the underlying basic principles which have been discussed before.

The typical coil force characteristic is depicted in **Figure 7**. The force curve is fairly constant, where the force itself depends primarily on the coil current and much less on the coil position. In contrast, the armature force of proportional solenoids depends both on current and armature position. This is very important for the control understanding of any electro-hydraulic valve. Because the electric current induced force, whether voice coil or solenoid, is the control signal used to induce a dedicated reaction based on the control logic input. To achieve the desired closed loop control performance of the PVG valve the actuator

Classified as Business

force must be capable of precisely controlling the piloting power. Controlling the piloting power means the actuator force must find equilibrium with other force components in the pilot valve, like spring force, flow force and friction forces. For perfect controllability it is necessary that small current changes lead to small and immediate force changes, and that alternating current adjustments are not obscured by hysteresis effects.

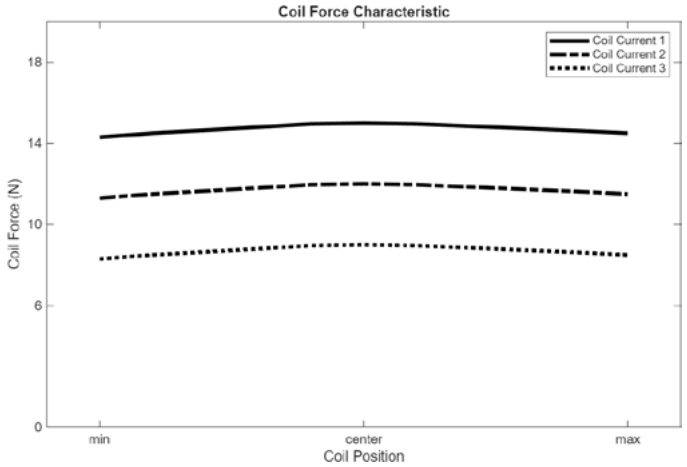


Figure 7. Simplified typical coil force characteristics for various coil currents.

Voice coils have a distinct advantage over solenoids because their force characteristics—being bi-directional, low-hysteresis, and proportional to the coil current—allow for quick and precise adjustments to find the next pilot valve position.

7. TEST RESULTS

The presented voice coil feasibility study was mocked together with a LVDT transducer to control a PVG32 valve. Functional tests carried out in the lab showed encouraging results. The voice coil valve was operated at Δ13 bar pilot pressure. The oil and valve components are heated at 50 °C to represent typical operational oil viscosity conditions. The focus for the lab test has been on tracing the spool position based on given commands.

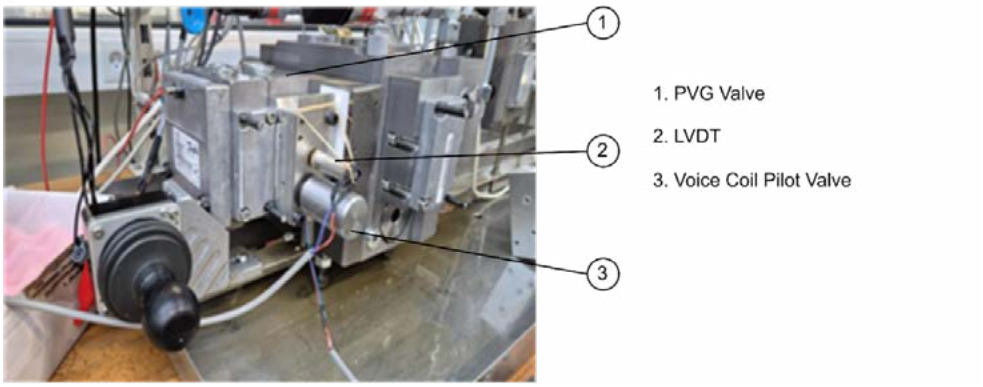


Figure 8. Proof-of-concept demonstration in the lab.

Classified as Business

Figure 9 shows the steady state characteristics of spool positioning. It plots the actual spool position over the demanded spool position, showing how much the spool position deviates from the command for almost steady command. The spool command undergoes a sweep from neutral to max stroke and back to neutral within 25 sec.

Figure 10 shows the step response characteristic. It plots the demanded and actual spool position as functions of time, showing how quickly and accurately the spool position gets adjusted to the dynamically changing command value. The response time is evaluated at 95% of the reached steady-state spool position.

Figure 11 shows the frequency response characteristic, describing how robustly the control valve reacts to alternating setpoint commands. The frequency response is characterized by two subplots: The magnitude plot, showing the amplitude attenuation, and the phase delay plot, showing the phase delay of the spool position in regard to its setpoint. A negative magnitude indicates the spool position oscillation is reduced, i.e. dampened, and the phase delay increases by increasing setpoint alternation frequencies. The set point alternation is configured as a sinusoidal signal with a bias of 50% nominal spool stroke and an amplitude of 25% nominal spool stroke.

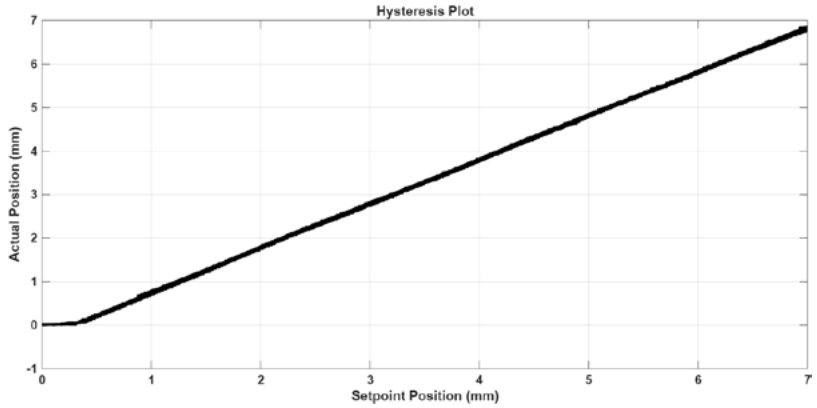


Figure 9. Steady state characteristic for the closed loop main spool position control.

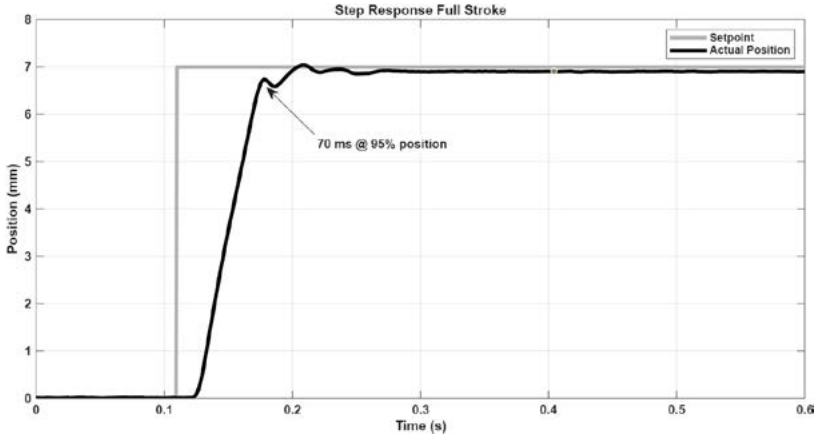


Figure 10. Step response characteristic for the closed loop main spool position control.

Classified as Business

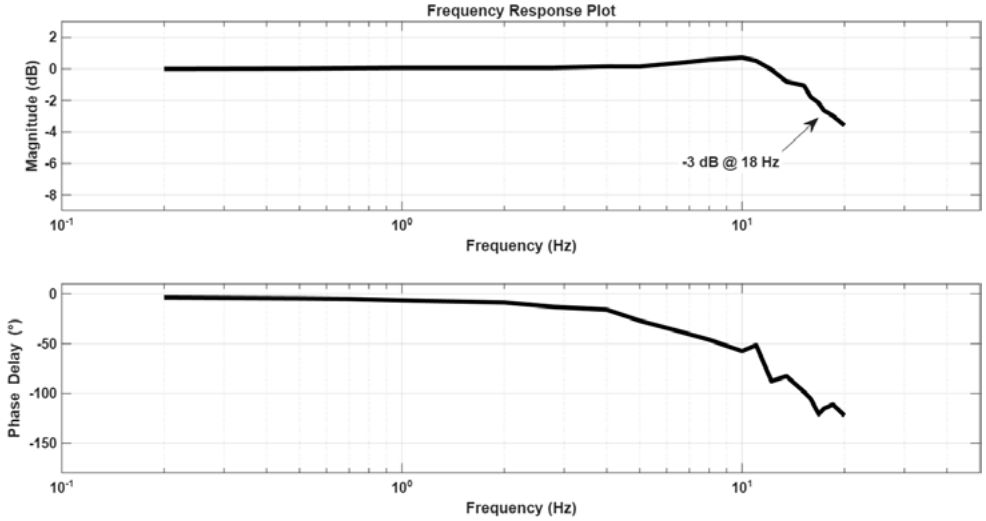


Figure 11. Frequency response characteristic for the closed loop main spool position control.

As can be seen in the figures, the PVG valve displays almost zero hysteresis and almost zero noise for the spool position control. The test results above reveal a deadband around neutral position which is attributed to the imperfect LVDT and control calibration. Besides this, the graph proves that the valve does very well in converting small and slowly changing incremental setpoint changes. The full-stroke step response of 70 ms, with little oscillation or overshoot, is noteworthy. This dynamic capability is underpinned by an impressive frequency response of approximately 18 Hz.

These test results of the feasibility study confirm the anticipated superior spool controllability of the voice coil technology. On a particular note, the frequency response data suggest that such a control valve is well-suited to be integrated into sensor- and software-controlled work functions.

8. SUMMARY

Danfoss is developing a new modular hardware and software platform for its PVG valve system to support future expansion and increasing variety in proportional control valves. The hardware and software platform emphasizes:

- **Scalability**: Easy addition/removal of components without full redesign.
- **Reusability**: Shared modules across projects to reduce cost and time.
- **Flexibility**: Simplified customization for diverse applications.
- **Standardization**: Consistent design and interfaces for easier integration.

Pilot valves are tailored to the work function's individual value proposition with free combinations of input and actuation technologies, supporting advanced control logic and customer-specific programming.

Classified as Business

Low-cost voice coil technology is one element of Danfoss' future valve actuation platform. It significantly improves machine productivity, energy efficiency, and control precision, positioning it as a superior alternative to conventional solenoid-driven valves. It offers:

- **High Dynamics & Precision**: Performance comparable to servo valves, enabling faster and more accurate response to sensor-driven commands.
- **Software-Based Flexibility**: Facilitates the transition from mechanical-hydraulic features to software-enabled controls, improving customization and productivity.

Danfoss conducted a design study with the goal of demonstrating a cost-effective PVG valve actuator using voice coil technology**.** Functional tests carried out in the lab showed encouraging results. These confirm excellent controllability and dynamic performance of voice coil technology.

9. REFERENCES

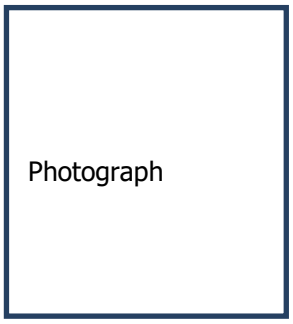
[1] Moog Fluid Power Journal April 2003, Proportional and Servo Valve Technology

[2] D. Wroblewski, 'New Electronic Proportional Valve System with Dual Spool Architecture', Kolloquium Mobilhydraulik, Braunschweig, 2014

[3] MikeRun, 'lorentz-force-horseshoe-field.svg', License: CC-BY-SA-4.0, Source: commons.wikimedia.org

Classified as Business

Biographies



Dirk Wroblewski received the master's degree (Dipl.-Ing.) in mechanical engineering from University of Rostock, Germany, in 2002, and the philosophy of doctorate degree in Engineering from University of Rostock in 2010. He currently serves as a Senior Manager overseeing technology development in electro-hydraulic valve control at Danfoss Power Solutions.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Classified as Business

Push-Pull Electromagnet in Hydraulic Systems

Filip VLACH

Jaroslav DVOŘÁČEK

ARGO-HYTOS s.r.o. R&D, Dělnická 1306, Vrchlabí, Czech Republic
f.vlach@argo-hytos.com
j.dvoracek@argo-hytos.com

Abstract

Electromagnets are one of the key components in modern hydraulic systems, enabling precise and efficient control of hydraulic valves. The use of electromagnets in hydraulic valves offers several advantages, including fast response times, high forces with low power consumption and reliable linear motion. A disadvantage is that they only move in one direction, hence two electromagnets are required for full fluid control of 4/3 valves, which means two coils and, among other things, double control wiring. This disadvantage can be overcome by using a push-pull electromagnet, which can move in either direction depending on the polarity of the supply voltage.
This paper presents a study of push-pull electromagnets and the usage in various hydraulic valves to replace the standard two-coil solution.

The push-pull electromagnet consists of a setup with permanent magnets together with discs made of magnetically soft material. The armature is housed in plastic sleeves that allow it to move smoothly.

Keywords: Push-pull electromagnet, Hydraulic valve, Electromagnetic actuator, Bidirectional motion, Coil reduction, Mobile and industry hydraulics

1. INTRODUCTION

In the field of modern hydraulics, key requirements are not only reliability and precise control but also energy efficiency and component compactness. Traditional designs of electromagnetic actuators for 3-position hydraulic valves often employ two separate coils, one for each directional function. While this solution is tried and tested, it brings certain disadvantages, larger structural volume and more complex wiring.
An alternative to this classical approach is the use of a so-called push-pull electromagnet, which enables control of both directions of armature movement using a single coil with polarity reversal. This concept offers potential for design simplification, reduction of energy demands and optimization of system response. The paper will describe the functional principle of the push-pull electromagnet, its advantages compared to the traditional solution and practical aspects of its implementation in hydraulic applications.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. PRINCIPLE OF THE PUSH-PULL ELECTROMAGNET

The push-pull electromagnet utilizes a reversible magnetic field to achieve movement of the armature in both directions. By changing the polarity of the supply voltage, both the “push” and “pull” effects can be realized. This makes it possible to replace two coils with a single one, reducing the number of components and simplifying the valve design.

The armature of the electromagnet consists of strong permanent magnets combined with discs made of magnetically soft material. The armature is mounted on plastic sleeves that allow smooth movement.

Although the push-pull principle is well known, it is rarely used in the field of oil hydraulics. Most electrohydraulic valves still rely on conventional single-direction electromagnets or a pair of coils for bidirectional movement. In contrast, the push-pull configuration is commonly applied in other areas, such as pneumatics, precision mechanics, automation, and drives for non-oil hydraulic systems (e.g., water hydraulics). These applications take advantage of the ability to generate bidirectional motion with a single coil to simplify design, reduce weight, and increase dynamics. As a result, the push-pull principle has proven effective where fast response or compact design are essential.

3. DESIGN SOLUTION AND PATENTED INNOVATION

One of the primary technical challenges in using a push-pull electromagnet for hydraulic applications is hysteresis – caused by the difference in magnetic behaviour during current switching on and off – which can adversely impact the accuracy and repeatability of the valve control. During the development of a new type of actuator, an innovative solution was proposed and patented, enabling active influence on hysteresis through a refined armature design.

The armature of the push-pull electromagnet consists of a series of magnetic and permanent elements arranged in a specific configuration: magnetic disc – permanent magnet – disc – magnet – disc. A key feature of this design is the variability of the diameters of the magnetic discs at the front and rear ends of the armature (Figure 3.1). By adjusting these diameters, the magnetic field distribution can be purposely influenced, thereby controlling the magnitude of hysteresis within the system.

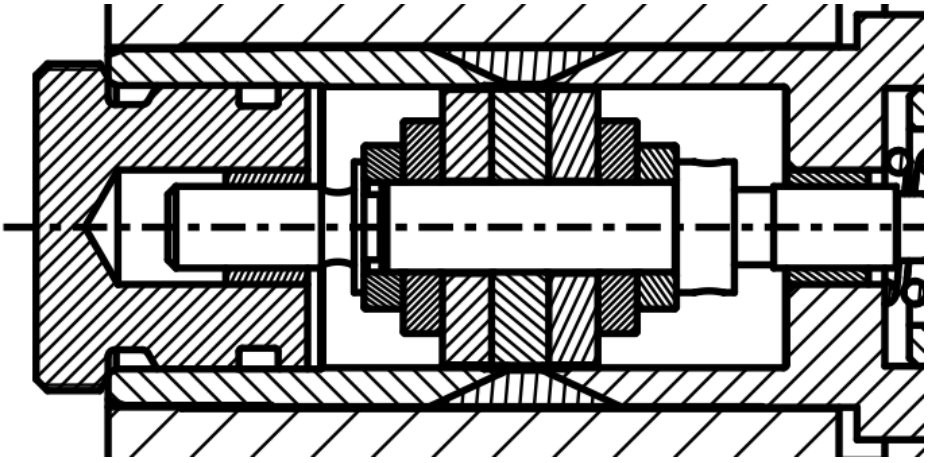


Figure 3.1. – Asymmetric armature design

Figures 3.2 (uncompensated hysteresis) and 3.3 (compensated hysteresis) illustrate an exemplary effect of hysteresis compensation. As seen in the uncompensated case, the valve opens abruptly without any sign of proportionality and remains fully open until the polarity of excitation changes; then it jumps to the opposite side and, with a slight indication of proportionality, stays open until the excitation ends, after which it closes. With hysteresis compensation, these abrupt changes in opening are eliminated, and the valve behaves proportionally throughout the entire operating range and for all pressure levels.

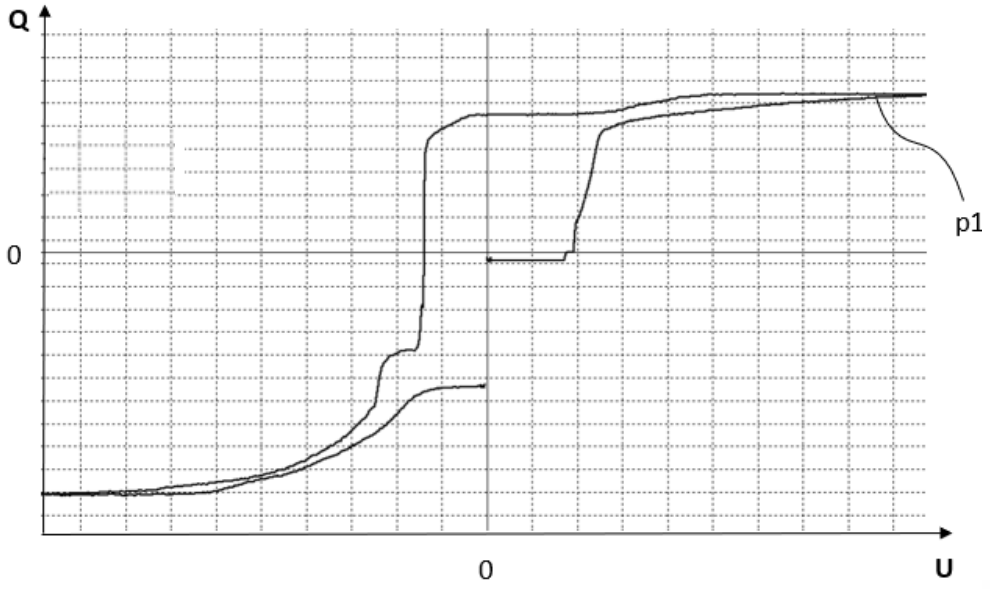


Figure 3.2. – Hysteresis of an indirectly controlled valve without hysteresis compensation (pressure p1)

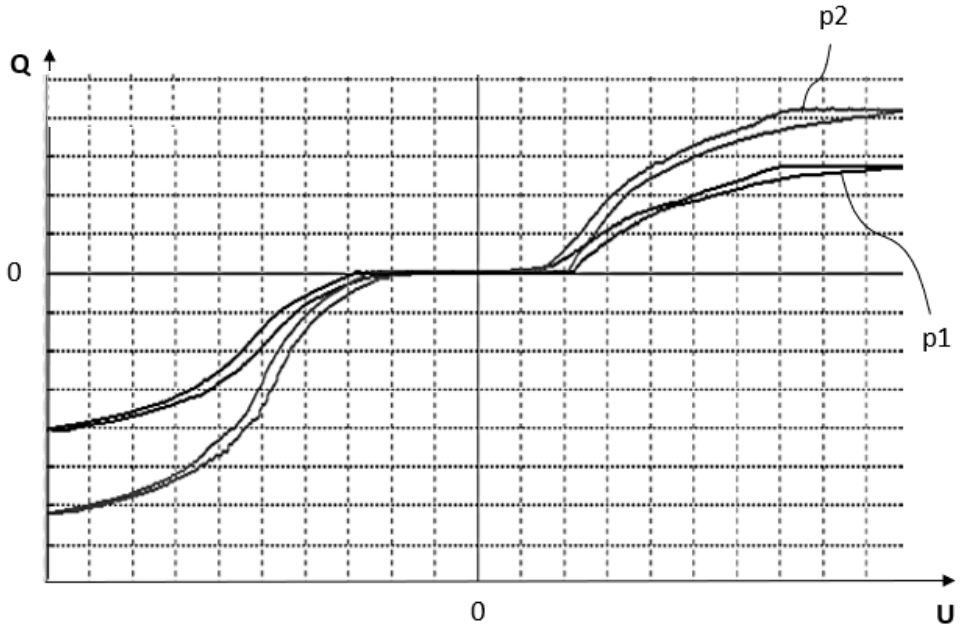


Figure 3.3. – Hysteresis of an indirectly controlled valve with hysteresis compensation (at pressures p1 and p2)

3.1. Advantages Compared to Two-Coil Systems

Overall dimensions of the valve are reduced, allowing it to be used in applications where the two-coil version cannot be installed due to space limitations.

The usage in CETOP valves opens up other possibilities for this type of valve, such as the possibility of mounting a position sensor on the spool and using its sensor output, either as a feedback function or as a safety function.

There are numerous additional to this technology:

- **Reduced material consumption** – Eliminating the second coil means less copper, plastic, and insulating materials. Copper mining and plastic production are energy-intensive processes with high CO_2 emissions, so using fewer of these materials directly improves the ecological footprint of manufacturing and supports circular economy principles.

- **Lower energy demands** – Push-pull systems operate with a single coil and optimized magnetic flux control, which reduces current load and idle-state losses.

- **Smaller carbon footprint** – Lower material and energy consumption during both production and operation significantly reduce emissions throughout the entire life cycle. Optimized hydraulic systems can achieve emission reductions over their lifetime, especially when combined with intelligent electronics and regenerative control strategies.

- **Longer component life and less waste** – Fewer parts mean fewer potential failures, reducing maintenance frequency and the need for spare parts. This not only saves costs but also reduces industrial waste, contributing to sustainability goals.

- **Economic competitiveness** – In addition to environmental benefits, simplified design reduces assembly time, wiring complexity, and storage costs. Combined with lower energy bills, this represents a significant advantage in total cost of ownership for OEMs and end users.

3.2. Challenges of Integration into Control Systems

One of the main obstacles preventing wider adoption of push-pull electromagnets is the need to adapt control systems for bipolar control. Although most modern control systems already support polarity reversal or can easily implement it through a simple firmware update, the use of push-pull electromagnets in the hydraulic sector is not common.

The key factor, therefore, is awareness and cooperation with system integrators [1].

4. CASE STUDIES AND APPLICATIONS

Push-pull electromagnets can be applied both in stationary equipment and in mobile hydraulics, for example in agricultural machinery, construction equipment, or mining systems. Thanks to their compact dimensions and high reliability, they are ideal for environments with high vibrations and limited space. In some applications, they have also proven effective in systems requiring fast response and low energy consumption. The proportional valves were controlled by the EL7-IA3 electronics [2], which is specifically designed for use with valves equipped with push-pull electromagnets.

4.1. Tested Applications

4.1.1. Electromagnetic Direct Operated Screw-In Valve (SD2E-A4/H3)

Successful implementation of a bidirectional push-pull electromagnet has been achieved on the originally two-coil, electromagnetically operated screw-in valve SD2E-A4/H3 (Figure 4.2.). The valve is designed for installation in a standard A4 cavity, 3/4-16 UNF thread. An example of the valve used in an application where it controls the direction of fluid flow and thus the direction of actuator movement - or stops it - can be seen in Figure 4.1..

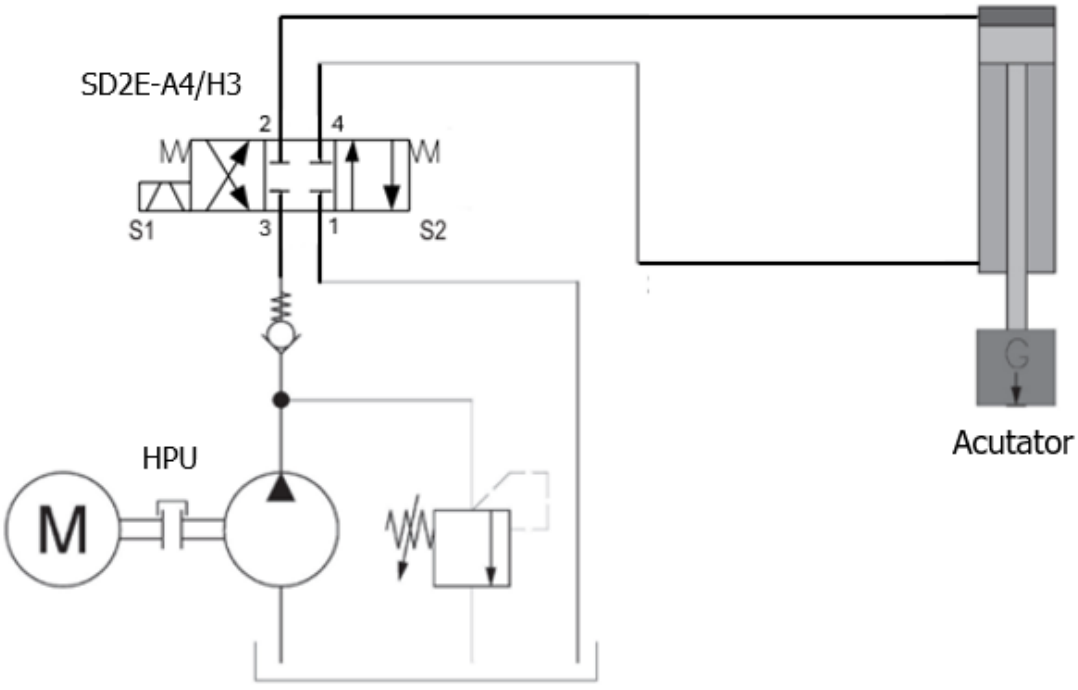


Figure 4.1. Example of using the SD2E-A4/H3 valve in a hydraulic circuit

Comparison of the original design and the new solution with a push-pull electromagnet, Figures 4.2 [2] and 4.3. The hydraulic part of the valve, including the spring and shims, is identical for both types; the push-pull version has a special chamber in the control system for integrating components from the standard two-coil version.

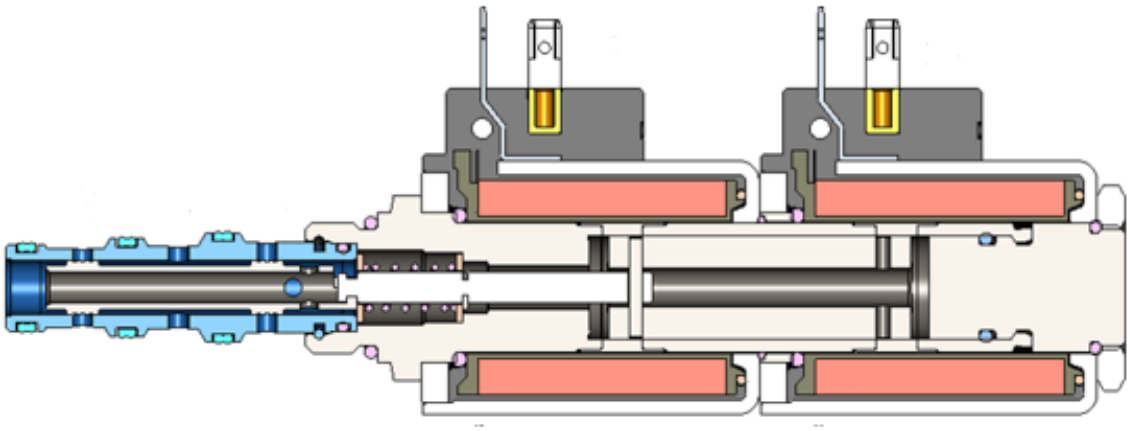


Figure 4.2. – SD2E-A4/H3

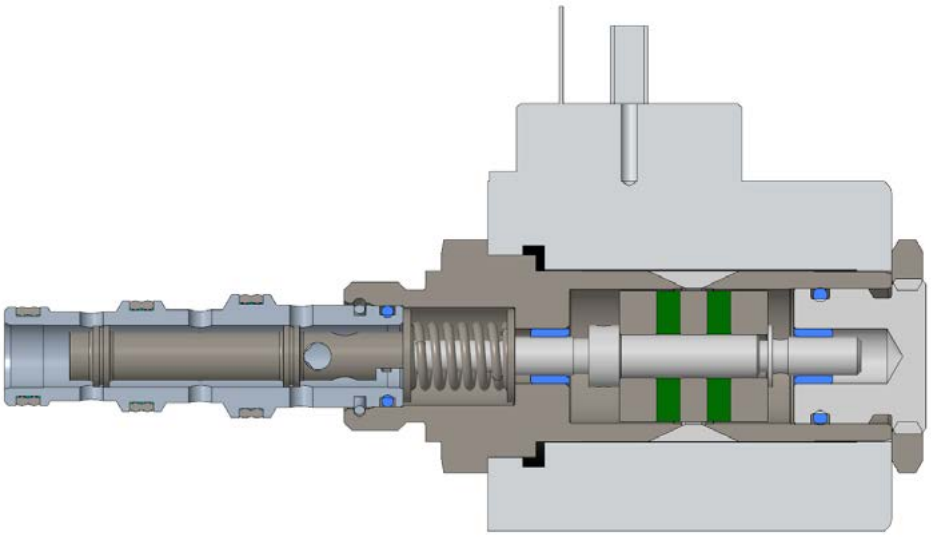


Figure 4.3. – SD2E-A4/H3 with push-pull electromagnet

The valve was tested with different spool types (thus with the effect of different flow forces). A full-fledged replacement – that means same power limits – for the standard version with two coils could successfully be proven.

4.1.2. Electromagnetic Direct Operated Proportional Valve (PRM2-063Z11/15)

A directly controlled proportional CETOP valve (Figure 4.4), combined with a push-pull electromagnet, exhibited significant hysteresis (Figure 4.5), which would need to be reduced, for example, by using electronics and internal feedback from the spool position.

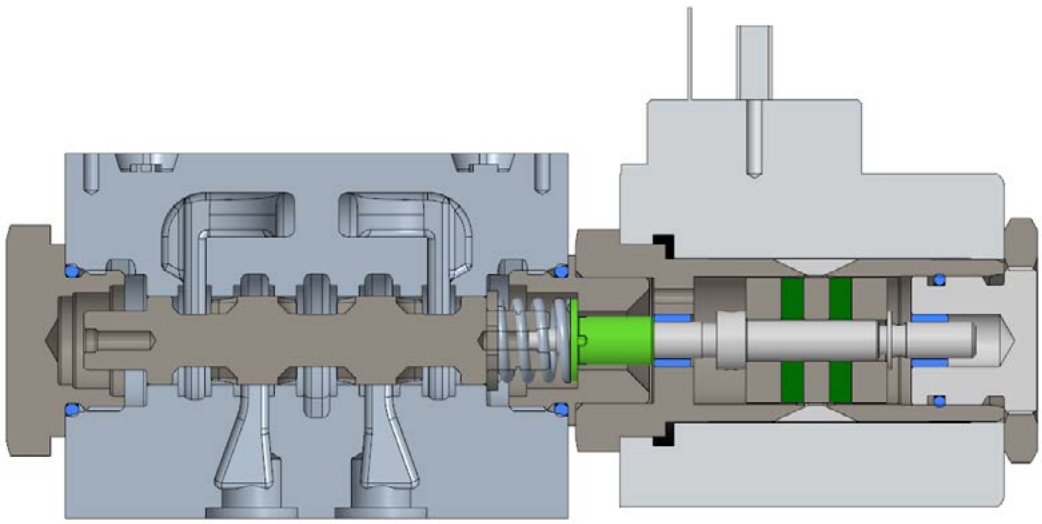


Figure 4.4. – PRM2-063Z11/15 with push-pull electromagnet

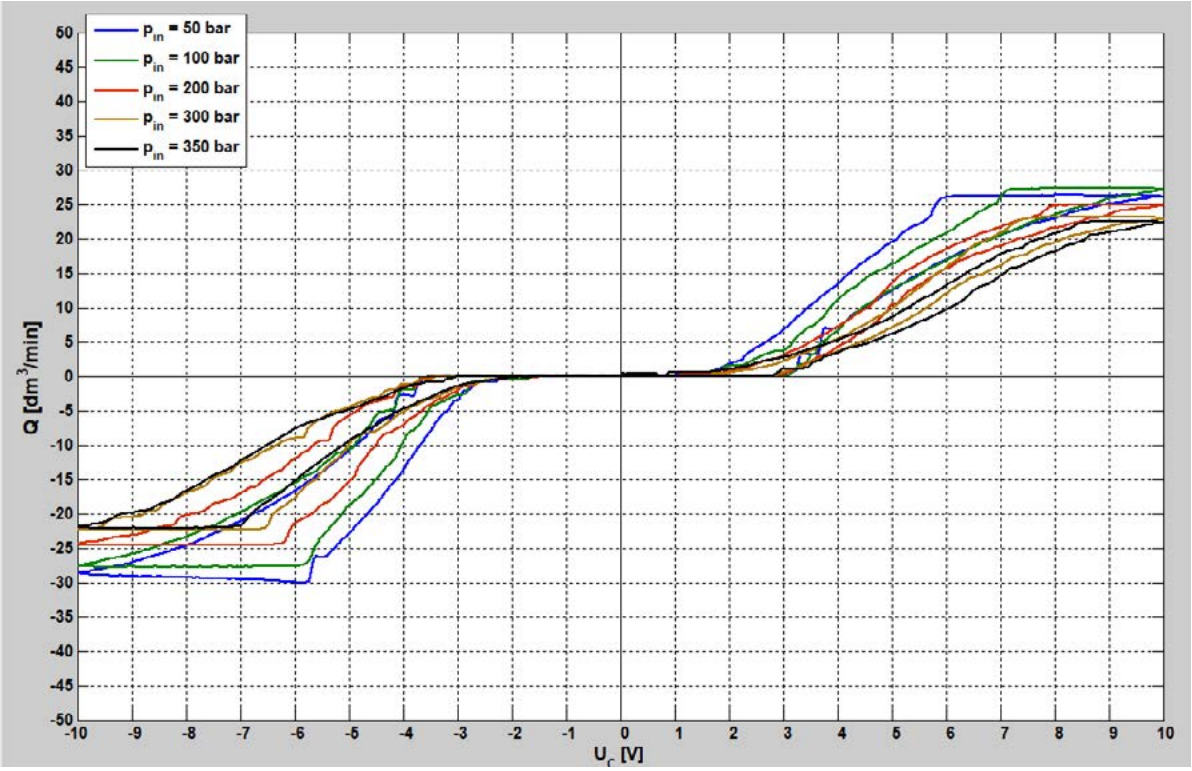


Figure 4.5. Flow curves of the PRM2-06 valve with a push-pull electromagnet

4.1.3. Electromagnetic Pilot Operated Proportional Valve (PRM8-063Z11/15)

Valves that are particularly suitable for the use of push-pull electromagnets are pilot operated valves (Figure 4.6.), because the principle of the pilot and follow-up spool [3] compensates for the lower force of the push-pull electromagnet compared to the two-coil design. Thanks to the indirectly controlled spool, equivalent performance to the two-coil variant can be achieved, even with the lower available force of the electromagnet. At the same time, this type of valve allows hysteresis reduction using an armature with discs of different diameters (Chapter 3). Hysteresis and performance of the valve modified in this way are then identical to the two-coil version (Figure 4.7.).

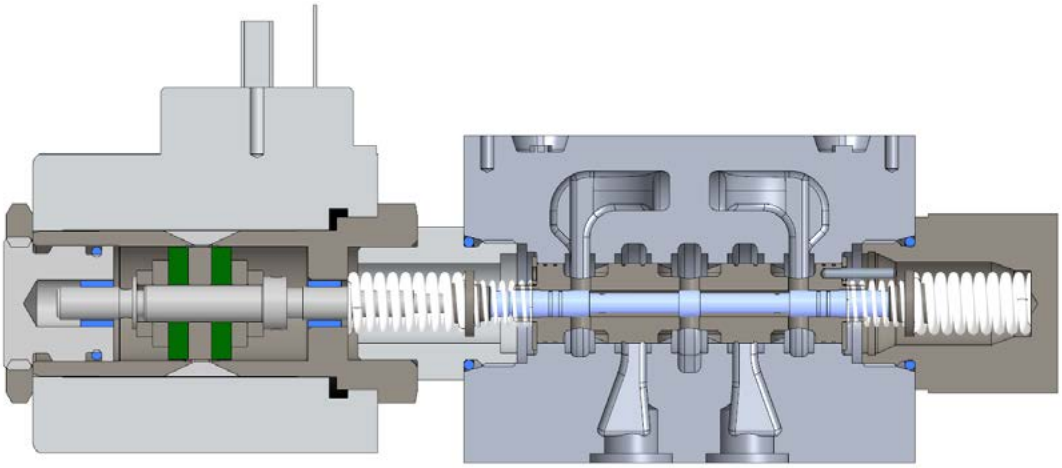


Figure 4.6. – PRM8-063Z11/15 with push-pull electromagnet

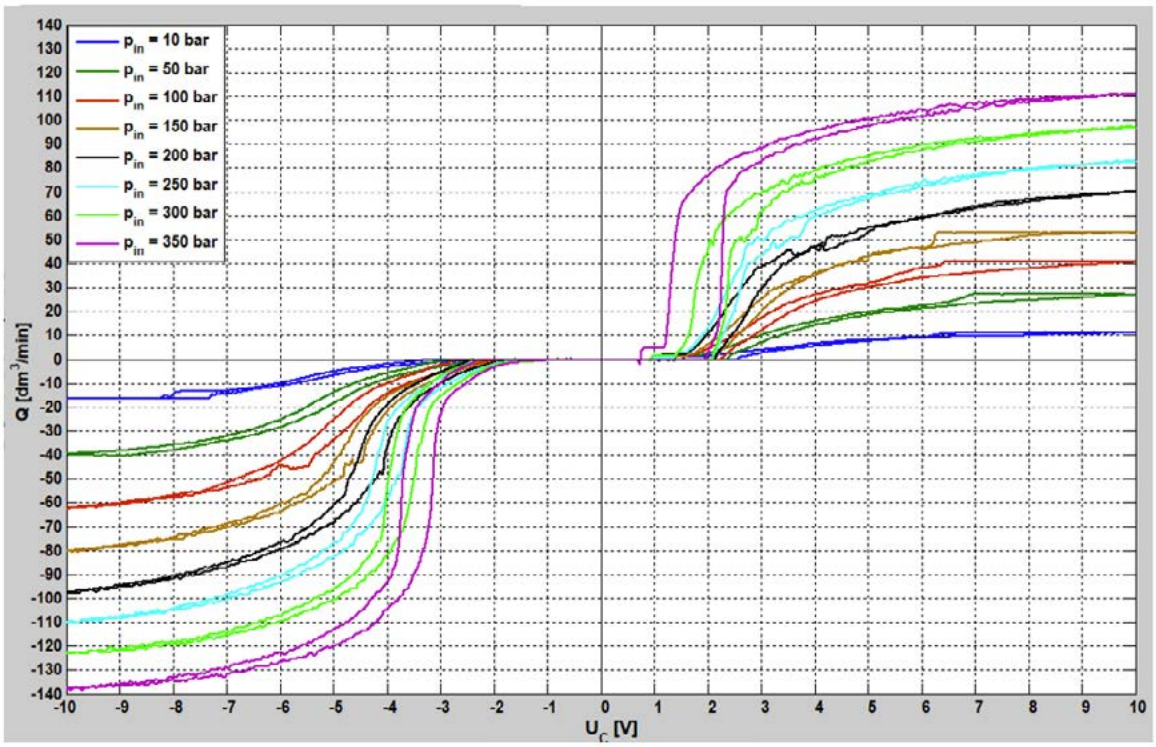


Figure 4.7. Flow curves of the PRM8-06 valve with a push-pull electromagnet

5. CONCLUSION

Push-pull electromagnets represent an interesting, modern, efficient, and environmentally friendly solution for controlling hydraulic valves. Thanks to successful tests, case studies and the patented design solution, their full-scale deployment in industrial practice can be confirmed. The remaining challenge is awareness and cooperation with integrators, but the potential of this technology is significant. Their ability to provide fast response, high performance, and reliable linear motion is essential for automation and safety applications. Their broader adoption depends on overcoming communication barriers between component manufacturers and system integrators.

6. REFERENCES

[1] Danfoss (2024). *Directional Control Valves – Push Pull Valves*.
[2] Website Argo-Hytos *https://www.argo-hytos.com*
[3] Mejsnar P., Englberth E., Schuster G., *Design of an Internally Pilot Operated Proportional Valve by Use of the Floating Spool Principle, in: Proceedings of the 8th International Fluid Power Conference (8th IFK), 2012, Dresden*

Biographies



Filip Vlach works as an R&D Designer specializing in hydraulic systems. His expertise includes the development of electromagnetic actuators, particularly push-pull designs, with a focus on energy efficiency, hysteresis reduction and ecological aspects. He is also engaged in designing electronics for proportional valves, position sensors, and hydraulic components, covering calculations, simulations, and testing. Filip has experience in creating innovative solutions for both mobile and industrial applications and is a co-author of patented technologies. Currently, his work focuses on integrating electromagnetic and electronic components into hydraulic systems to enhance reliability and optimize performance.

Recuperation on a transmission test bench

Dr.-Ing. Siegfried Rotthäuser, Dr. Nico van de Sandt

Dr.-Ing. Siegfried Rotthäuser, Ingenieurgemeinschaft IgH, rt@igh.de

Dr.-Ing. Nico van de Sandt, ZF Industrieantriebe Witten GmbH, nico.vandesandt@zf.com

Abstract.

This article presents a recuperation concept for an industrial gearbox test bench. The concept of hydraulic testbench was realized in the past already, the fully automated system is a new technology. Here the concept is presented witch combines to the hydraulic circuitry and the control technology implementation.

Keywords. Recuperation, Gear Box, Multiple variable control, efficiency optimization, Open Source control software

1. INTRODUCTION

The test bench presented here is a combination of classic industrial hydraulics and modern control technology. The test bench has been developed as a research and development test bench. It offers a very high degree of flexibility both in the design of test sequences and in the selection of test specimens for industrial and mobile gearboxes as well as drivetrains or hydraulic motors. The basic concept can be seen in Figure 1.

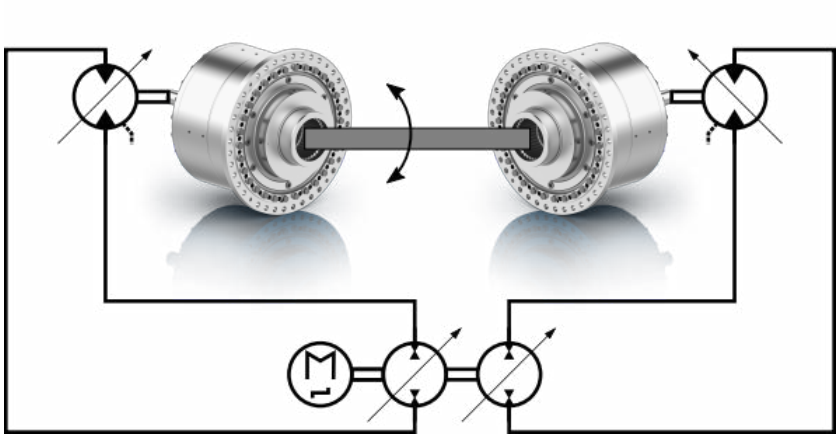


Figure 1: Figure 1: Concept of the Test Bench

An asynchronous motor operated by a frequency converter drives two identical A4V pumps that are installed in tandem. Type A6V hydraulic motors are installed on the test gearboxes. The two test gearboxes are braced against each other, with one gearbox operating as a speed-controlled drive and the second gearbox as a torque-controlled brake.

The test bench allows tests to be carried out on a variety of different purposes. Its main tasks are:

- Investigation of thermal behavior
- Identifying mechanical efficiency
- Investigation of wear and fatigue behavior
- Measuring vibrations and noise
- Testing new materials, manufacturing processes, and lubricants
- Testing gearboxes, hydrailic motors or complete drive trains

2. TEST BENCH DESIGN

When designing the test bench, the focus was on energy efficiency and flexibility in addition to suitability for the planned test specimens. For this reason, automation does not rely on ready-made solutions from component manufacturers, but rather on an open source solution. The EtherLab [1] project offers numerous functions that are used here.

The selection of hydraulic components is based on the performance data of the gearboxes. Typical input speeds of 1500 to 2500 RPM are required, which are then also applied to the output of the test setup due to the back-to-back arrangement. The torques can exceed 3000Nm and the transmitted power can range up to 600kW. With a gearbox efficiency of around 97,5% each, this results in a maximum waste heat output of about 60 kW.

These key data were used for the design of the pumps, motors, and gearbox cooling system [2]. Figure 2 shows the entire test bench; the soundproof enclosure and the hydraulic container are not shown.

Since this is a research and development test bench, the hydraulic connection options are designed to be very flexible. Figures 3 to 5 show three examples of frequently used arrangements. For large test specimens, the two built-in main pumps can be operated in parallel, see figure 3. However, two load tests can also be run simultaneously on smaller gearboxes, figure 4. Free-running tests with no load can be performed on up to four test specimens simultaneously, figure 5. Mixed operation between the operating modes is also possible.

All auxiliary functions, such as control oil pressure supply, filtration, temperature control, housing flushing, and feed and flushing functions, are implemented via separate systems. These systems are working independently of the main pumps.

The feeding of the closed hydraulic circuits is temperature-controlled in order to minimize the demand for feed oil. For this purpose, oil is taken from the low-pressure section [3] via a proportional throttle, and the pressure-controlled feed oil supply pushes fresh oil through accordingly.

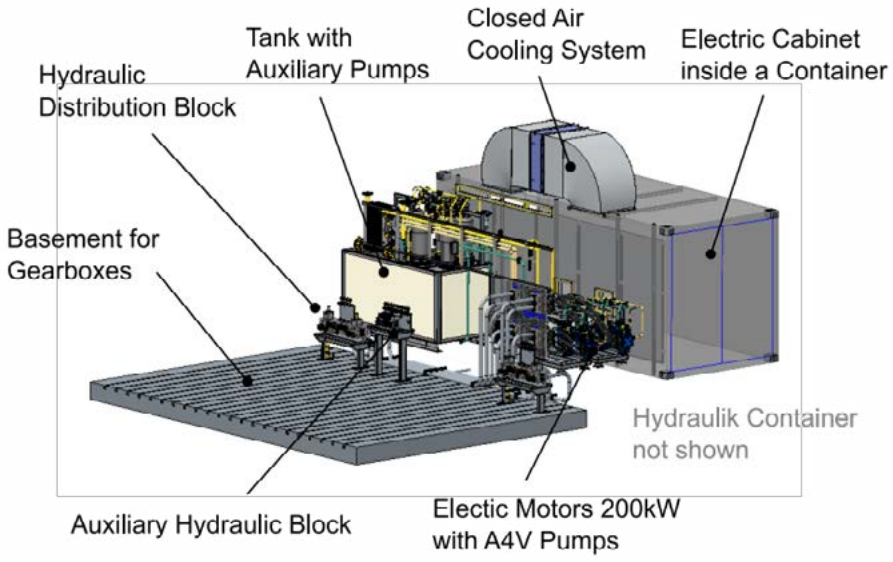


Figure 2: Test Bench Overview without Specimen

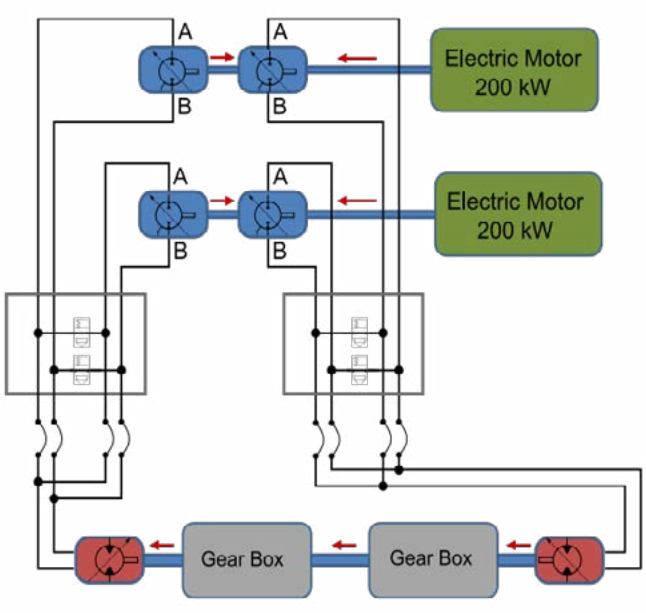


Figure 3: Parallel Pumps operating for big Specimen

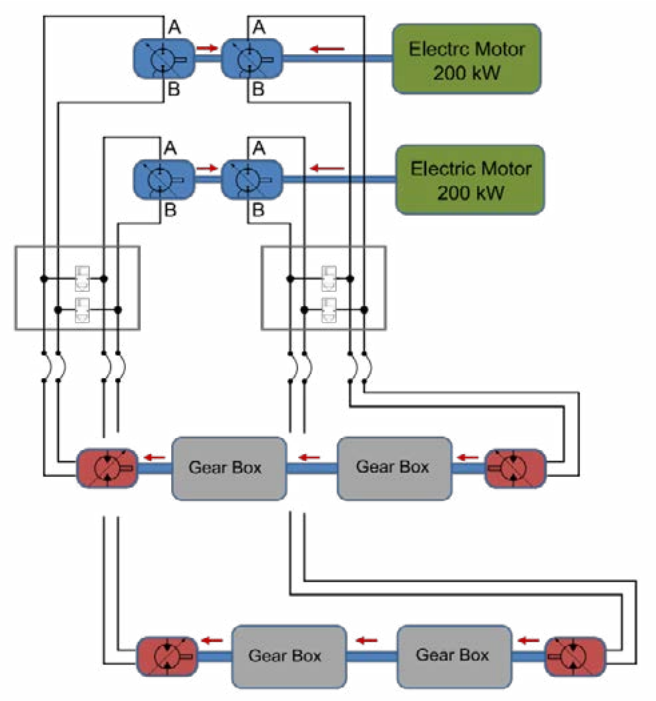


Figure 4: Configuration for two Tests at the same Time

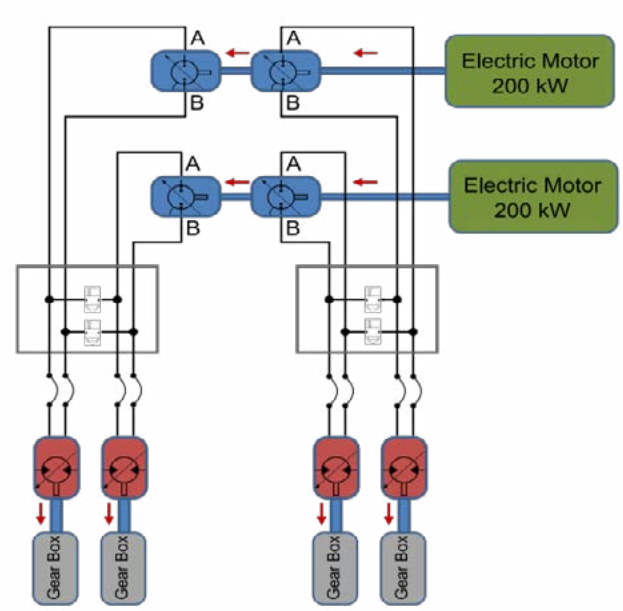


Figure 5: Multiple Specimen in free running Mode

3. EFFICIENCY OF A TEST BENCH

What is the efficiency of a test bench in which all of the energy used is ultimately dissipated as heat? For the test bench implemented here, the definition shown in the image was chosen in order to implement an optimization strategy.

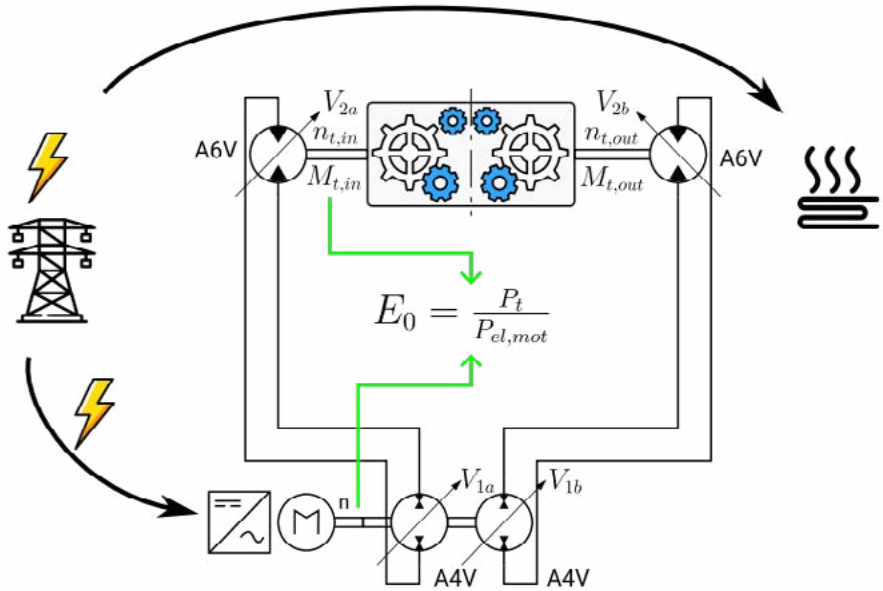


Figure 6: Definition of Efficiency at the Test Bench

Efficiency is defined as the ratio between benefit and effort and is defined here as the ratio between the mechanical input power in the test specimen and the active power used in the electric pump drive.

$$E_0 = \frac{\eta_{A4Va}\eta_{A6Va}}{1-\eta_{geara}\eta_{gearb}\eta_{A6Vb}\eta_{A4Vb}\eta_{A4Va}\eta_{A6Va}} \quad (1)$$

Since many hydraulic adjustment units operate in series in this test bench, each individual efficiency level has a very strong influence on the overall efficiency. The overall efficiency can be determined using the equation 1 and is in the order of magnitude of 1,5. This means that the power flow through the gearboxes can be about one and a half times greater than the electrical power input. If the components are not operated at their optimum operating point, the overall efficiency drops significantly. A deterioration of around 2% in a single component reduces the overall efficiency by around 10%.

4. CONTROL SYSTEM

The entire automation of the test bench, including all auxiliary systems, is implemented using the open source tool kit EtherLab®. EtherCAT® is used as the fieldbus system, which easily allows sampling rates better than 1 kHz, even for large systems. As shown in the figure, the control system for the main functions is cascaded.

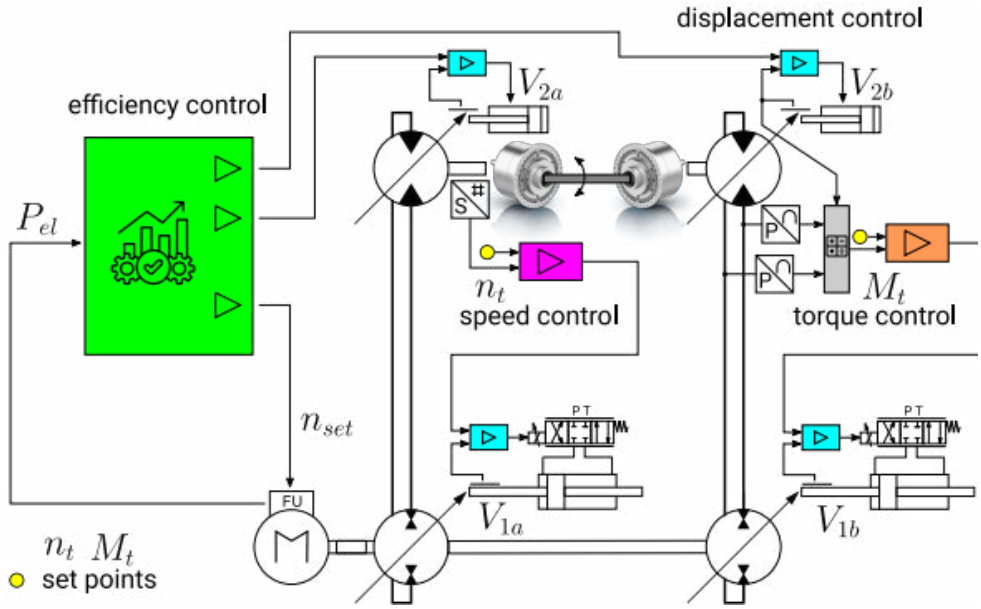


Figure 7: Structure of the Control System

All units with variable displacement have a subordinate swivel angle control loop [5]. The setpoints for this control loops are generated in the higher-level control loops for the input speed n_t and for the torque M_t.

The efficiency of displacement units is highly dependent on their operating points. The operating points are largely determined by the pressure, speed, and swivel angle. The general relationship [3] is shown in the figure 8. Looking at the basic efficiency curves helps to set a good starting configuration. On the drive side, the desired test speed at the gearbox input determines the ratio of the displacement volumes at a given drive speed of the electric motor.

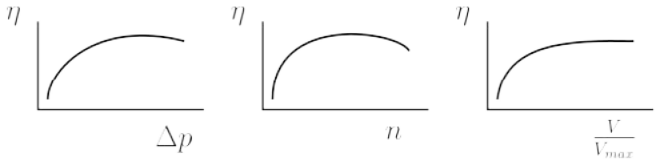


Figure 8: Efficiency and Operating Point

A mathematical description of these relationships, which are mostly determined experimentally, is only possible to a limited extent and is not useful due to the large number of different configurations. Instead, an evolutionary method is used here for the efficiency controller in order to achieve the most efficient overall operation.

Since the test runs last several hundred hours, there is enough time for optimization to implement a robust, i.e., slow algorithm. This algorithm uses small increments when changing the available control variables V2a, V2b, and n. Taking the operating limits into account, the speed of the drive is changed as long as the efficiency increases. The displacement volumes of motors V2a and V2b are then manipulated. The algorithm for increasing efficiency is activated by the operator during a test run and then runs automatically. The algorithm stops when no further improvement can be achieved by changing the control variables.

5. CONCLUSION

The test bench presented here combines the capabilities of classic hydrostatic drives with modern automation technology. The control concept with higher-level efficiency optimization and detailed improvements in the area of auxiliary functions has resulted in a system that is unique in this form.

It offers a high degree of flexibility in test design as well as a high degree of robustness through the use of high-quality components.

This investment in the Witten development site will help to secure expertise in industrial gear manufacturing in the long term and develop new technologies.

6. REFERENCES

[1] www.etherlab.org
[2] Findeisen Dietmar, Ölhydraulik, Springer-Verlag Berlin Heidelberg 2005
[3] Schmitz K., Murrenhoff H., Grudlagen der Hydraulik Teil 1 Hydraulik, RWTH Aachen University 2018
[4] Murrenhoff Hubertus, Servohydraulik 4.Auflage, Shaker Verlag 2012
[5] Haas, H.-J.: Sekundär geregelte hydrostatische Antriebe im Drehzahl- und Drehwinkelregelkreis. Diss. RWTH Aachen University 1989
[6] Niemann G., Winter H., Maschinenelemente Band 2: Getriebe allgemein, Zahnradgetriebe-Grundlagen, Stirnradgetriebe, 2. Auflage, 2003

Biographies



Siegfried Rotthäuser

was born in Essen in early 1961 under the sign of Aquarius. After attending school and completing his military service, he studied general mechanical engineering at RWTH Aachen University from 1981 onwards and then worked as a research assistant at the IHP under Prof. Wolfgang Backe. After obtaining his doctorate, he founded the engineering firm IgH in 1993 together with Dr. Torsten Finke († 2025), where he continues to serve as managing director and board member within AXXERON Technologies.



Nico van de Sandt

born in Bocholt on December 8, 1971, studied mechanical engineering at RWTH Aachen University between 1993 and 1998 and then worked as a research assistant at the Institute for Machine Elements and Machine Design. He received his doctorate in 2004, specializing in rolling bearings and lubrication. From 2003, he held senior development positions in the gearbox industry. Since December 2015, he is head of R&D for gearboxes for mobile and industrial applications at ZF Industrieantriebe Witten GmbH.

Next Generation Fixed Displacement Pumps for Electrified Mobile Applications

Patrick Fischer[1], Adrien Gambet [1], Shanmukh Sarode[2], Howard Zhang[2]

[1]Parker Hannifin Corporation, Motion Systems Group (MSG), Hydraulic Pump and Motor Europe Division (HPME), patrick.fischer@parker.com

[2]Parker Hannifin Corporation, Motion Systems Group (MSG), Motion Technology Center (MTC)

Abstract.

Electrification of mobile machinery imposes new demands on hydraulic components, notably higher operating speeds, greater power density, and stricter noise requirements. This paper presents developments in the next generation of Parker fixed displacement pumps – gear, bent axis, and vane – designed specifically for electrified applications. We introduce a pragmatic low complexity gear pump approach retaining straight teeth while leveraging advanced manufacturing methods for higher shaft speeds at acceptable costs. For bent axis pumps, we demonstrate an integrated ripple chamber that reduces flow ripple yielding multi-decibel reductions in system noise while valve plate and suction channel refinements enable significant higher self-priming speeds, up to 17% depending on the displacement. The next generation vane pump incorporates tribological coatings for higher pressures and exceptional self-priming speed capabilities. Virtual and physical validation qualify the proposed concepts. Finally, the paper outlines a novel approach to integrate these pumps into ePump architectures while keeping modularity.

Keywords. Hydraulic pumps, electrification, fluid power electrification, fixed displacement pumps, noise, efficiency, integration, semi-integration, ePumps.

1. Introduction

Electrification is reshaping mobile fluid-power systems by removing the masking effect of Internal Combustion Engines (ICE) on one hand and by driving demand for higher power density due to batteries reducing available space and target weight for the hydraulics. As a result, hydraulic pumps – the primary flow sources in these systems – must deliver higher shaft speeds while meeting increasingly stringent Noise, Vibration, Harshness (NVH) requirements, higher efficiency while becoming smaller and lighter. Traditional pump designs, optimized for ICE-driven platforms, are therefore being re-evaluated for the new constraints imposed by electrified architectures. Although variable piston pumps dominate the hydraulic pump markets, fixed displacement units remain foundational in many mobile

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

applications. However, the shift to electric prime movers alters the trade space for pump selection and optimization. Recent literature reports a broad spectrum of hydraulic pump concepts under investigation for designing electrified flow sources (ePumps), including vane, internal and external gear, bent axis piston, axial piston and radial piston designs to address the abovementioned requirements. Concurrently, to achieve the required power density of electrified mobile applications, system level integration considerations become more important as well, e.g. Integrated ePumps (IEPs).

This paper focuses on next generation fixed displacement pumps tailored for electrified mobile machinery, more specifically for open circuit work functions. Section 2 reviews the state of the art, summarizing the electrification-driven requirements and surveying relevant pump technologies. Different ePump integration levels are explained. Section 3 presents design concepts and validation results for the proposed next generation gear, bent axis, and vane pumps. Section 4 outlines a novel approach of integration for ePumps by combining modularity of non-integrated and the advantages of highly integrated ePumps. Section 5 summarizes the present contribution to provide an outlook on further research needs.

2. State of the Art

2.1. Mobile application perspective: Electrification leaves less space for hydraulics

Fischer et al. [1] provide an extensive overview on the state of the art, challenges and chances of electrification for hydraulic pumps and their operating limits. They summarize key literature dealing with different layouts of electrified mobile applications (centralized vs. de-centralized hydraulics as dealt in Casoli et al. [2]) and their maturity in terms of whether it is a phase/generation 1 electrification (pure conversion of existing ICE application as explained by Myers [3]) or a fully de-centralized layout in phase/generation 3 systems with one dedicated Electrohydraulic Pump (EHP) or ePump per function (as outlined by Nagel & Zhang [4]). Overall, it appears that the more electrified and de-centralized the hydraulics in the application are, the lesser space is available for the hydraulics due to increased electrical content.

2.2. System perspective: EHPs with higher power density

Both variable and fixed pump technologies have their use-cases in today's applications. This applies also to tomorrow's electrified mobile applications as well. While the Electric Motor (EM) has the flexibility of variable speeds, a Hydraulic Machine (HM) – this more generic term will be used in the following to consider multi-quadrant operation and not only pumping operation – using fixed displacement would require many and high accelerations and decelerations for the pump to implement an electric Load-Sensing (eLS) functionality. Using variable displacement technology adds mechanical complexity, though it provides an additional degree of freedom, as the flow can not only be adjusted by varying the EM's speed, but also through adjusting the displacement within the pump, e.g. through Electronic Displacement Control (EDC). Those fewer accelerations and deceleration of the pump's, its drivetrain and the EM's inertia enable efficiency optimized operating strategies as outlined by Myers [3]. Since the focus of this paper is fixed displacement technology, variable displacement pumps will not be further detailed. Ultimately, the choice for the technology depends on many factors to be weighted, the work function and the desired duty cycle to be performed.

In both cases however and as described by Fischer et al. [1] electrification demands more integration to compensate for decreasing available space in the mobile applications as well as tougher requirements to the target weight of hydraulics, as batteries occupy the bulk of the vehicle's size and increase vehicle weight accordingly. One first step for moderate integration is the direct coupling of EM to the HM as proposed by Cailteux [5] and Pelz et al. [6]. Having this direct connection instead of a coupling allowing compensation of misalignments and/or damping oscillations in circumferential direction requires however specific lubrication solutions for the splines to avoid fretting corrosion and ensure appropriate durability over lifetime. [1]

In the context of Parker's SMART Electrification portfolio (refer to [7]), pumps directly connected to Parker electric motors are referred to as Configured ePumps. Despite having the advantage of direct coupling with appropriate spline lubrication, Configured ePumps are highly modular, as they feature a broad spectrum of different pump technologies, displacements and power ranges as well as motor types, featuring not only Parker's GVM but also its NX motor both as low (LV) and high voltage (HV) versions. They can further be combined with Parker's Global Vehicle Inverter (GVI, refer to Parker Hannifin [8]), depending on the voltage level. Another direction is a more consequent type of integration, which can be seen on Figure 2.1, called Integrated ePump (IEP).

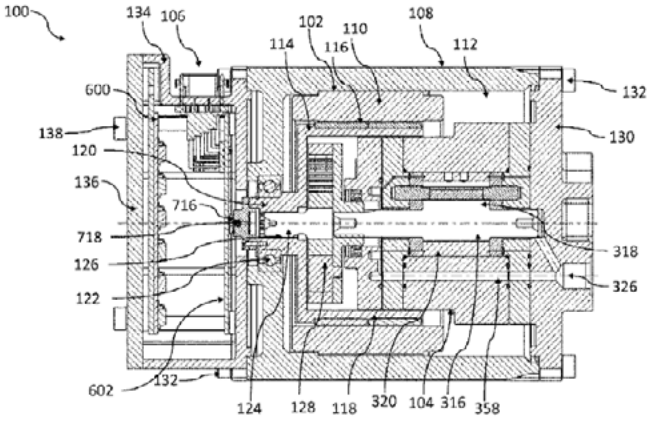


Figure 2.1: Integrated ePump (IEP) with Planetary Gearset (PGS) acc. to Zhang et al. [9]

There are multiple approaches on how to integrate HM with the EM, as a recent publication by Heeger, West & Ericson [10] outlines. The authors describe systematically whether HM and EM are axially or radially stacked and then whether the HM is direct-driven or gearbox-coupled. Combinations are also presented, e.g. radially integrated gearbox and HM. In related work, Lindh et al. [11] demonstrated a gearbox-coupled configuration connecting a bent axis pump with a permanent-magnet motor and examined the feasibility and performance of hydraulic-oil based cooling strategies for the motor. Ponomarev et. al [12] integrated a permanent magnet synchronous machine with a bent axis piston pump. The thesis also explores use of hydraulic fluid to cool the electric motor. Pietryzk et al. [13] investigated the potential of implementing high speeds at the EM with a Planetary Gearset (PGS) between EM and HM for a high-power density Electrohydraulic Actuator (EHA). Gadiyar et. al [14] designed a single casing integrated axial flux electric motor with a radial

piston pump unit, allowing a reduction in the overall package volume and removing shafts, bearings, and bearing seals. Along the same lines, Sarode [15] have designed and experimentally validated an integrated fixed displacement axial piston machine with a permanent magnet electric motor. These single-casing motor–pump integrations enable highly compact packaging well suited to electrified mobile machinery. In the concept of Figure 2.1, the HM and the gearbox no. 128 are radially integrated into the EM's rotor no. 114. With that, a concept like the gearbox type of integrated ePumps from Heeger, West & Ericson [10] arises and has in addition a motor controller no. 106, which is also integrated axially with joint cooling system. This approach of combining hydraulic pump, electric motor, gearbox, and inverter represents one of the most highly integrated ePump architectures considered in this work, delivering a compact package and an in-situ thermal management solution.

2.3. Component perspective: Higher speed and low noise pumps

While the previously mentioned integration reduces size of the ePump thus weight, it appears that if each of those sub-components is as well optimized itself towards higher power density, a further optimized application size and weight is enabled. Hence, it becomes crucial to increase Self-Priming Speed (SPS) of HM to reach higher speeds for higher flows, compensating for a smaller and lighter pump with lower displacement. The displacement is a key factor defining the torque needed to drive a HM in pump mode at a certain pressure, as depicted in the following basic formula (refer to Ivantysyn & Ivantysynova [16]):

$$T_P = \frac{\Delta p \cdot V_{D,P}}{2 \cdot \pi \cdot \eta_{hm}} \qquad (2.1)$$

At a given pressure differential Δp, reducing the displacement $V_{D,P}$ of a pump compared to a reference with lower SPS, it appears, that the needed torque T_P is significantly lower, assuming the hydromechanical efficiency η_{hm} staying constant. This drives secondary downsizing potential for the EM connected to the HM, since – as outlined by Schweigert et al. [17] – the required torque has the main influence on the size and hence the costs of an EM, which is – amongst others – one of the drivers of high-speed concepts for EHPs.

A wide variety of hydraulic pump architectures are employed in ePump designs: Sarode et al. [18] systematically compared several ePump configurations that combine commercially available hydraulic pumps to meet the demands of a representative mobile-hydraulic application. The study evaluated drive-cycle efficiency across four utility drive cycles and benchmarked mass, volume, cost, and power density for each architecture. By providing this generalized, utilization-based comparison, the work offers a practical guideline for selecting appropriate ePump architectures in electrified mobile hydraulic systems. These conclusions inform trade-off decisions when matching pump architecture to application requirements e.g., favoring bent axis designs where power density and efficiency are paramount, gear solutions where cost and simplicity dominate and vane where noise needs to be reduced.

Indeed, while external gear pump technology is cost-efficient and robust, it is usually rather loud at higher speeds and pressures, when having straight teeth. Using helical teeth for a more continuous meshing of the gears brings advantages in terms of NVH behavior. Transitioning to alternative profile geometries (e.g. trochoidal) instead of conventional involute profile allows to reduce dead-volume hence reduces flow ripple as outlined for example by Morselli [19] or Zhao & Vacca [20]. The latter ones discuss a Continuous-

Contact Helical Gear Pump (CCHGP) design combining both helical teeth and alternative profile geometries yielding improved NVH behavior. A more complex overall pump design compared to straight teeth gear pumps is usually resulting, as then either complexity increases in the design by adding balancing features or in the manufacturing by adding a double-helical gear layout (e.g. Morselli [21]) similarly to herringbone gears. Despite external gear technology, internal gear technology can also be used in medium pressure areas, typically being less noisy than external straight gear technology.

Vane pump technology operating in medium pressure often features a symmetric layout in circumferential direction, hence it is radially balanced, reducing nominal loads on the shaft compared to other technologies with a similar displacement. In addition to that, vane technology is rather quiet and offers great NVH behavior, due to a different operating principle bringing however disadvantages in terms of robustness and overall efficiency. [1]

Ultimately, bent axis pump technology typically offers best-in-class efficiency, however at the price of higher noise emissions inherent to its operating principle. Further details on advantages and disadvantages of each pump concept and type can be looked upon at Vacca & Franzoni [22], Sarode et al. [18] or Fischer et al. [1].

3. Next Generation Fixed Displacement Pumps for EHPs

3.1. Next Generation Gear Pumps for Electrified Applications

3.1.1. Concept and/or design

As mentioned in section 2.3, novel gear pump designs using helical teeth to enable lower noise may become more complex as they need special features to ensure equalization of the axial forces generated by the helical angle. The proposed design here features minor changes on today's gear pumps having straight gears, i.e. major process and quality improvements of the critical tribological interfaces, that cannot be detailed further due to confidentiality reasons. Those modifications enable achieving best-in-class efficiencies for straight gear pumps as well as improved durability at higher speeds, hence higher SPS.

3.1.2. Virtual and/or physical validation and limitations of this study

An exemplary SPS assessment performed for the 14cc sample can be seen in Figure 3.1 (left). The SPS gain potential for different displacements is outlined in Figure 3.1 (right).

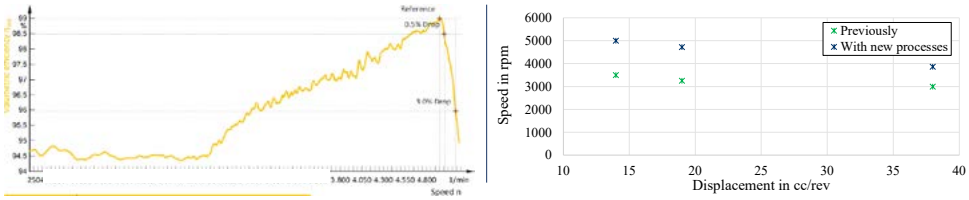


Figure 3.1: Exemplary SPS measurement at 200 bar and 0.8 bara inlet pressure (left) and SPS comparison before and after using the new manufacturing processes (right)

However, it must be acknowledged that those SPS gains are currently purely based on the formal definition of SPS. Additionally, the product maturity is at an A-Sample stage, hence a full endurance validation is to be performed. This will give more insights into the

practically reachable SPS, whether this value can be driven continuously, intermittently or is just to be considered as a peak value.

3.1.3. Exemplary use-case related to mobile electrification

Since this new process is in A-Sample stage, there are currently no real use-cases demonstrating its advantages in mobile applications being in the field. However, any cost-sensitive electrified application where the noise performance with straight gears is acceptable, but an increased SPS is required while benefitting from the robustness and simplicity of a gear pump could be applicable.

3.2. Next Generation Bent Axis Pumps for Electrified Applications

3.2.1. Concept and/or design

The next generation bent axis pumps features an integrated Ripple Chamber (RC), which is novel for this type of pump. It has been applied to both the F1 truck pump, making it the F1e and the heavy-duty F12 pump, making it the F12e (refer to Parker Hannifin [23]). While the RC is integrated in the end cap of the F1e, the RC on the F12e is integrated into the barrel housing. The RC lowers significantly the flow ripple, hence also the dB(A) noise emissions in the lab but yields even more noise reduction potential in real applications. Additional design improvements on valve plate and suction channels increase the SPS by up to ≈ 17%, depending on the developed displacement of the F12e.

3.2.2. Virtual and/or physical validation and limitations of this study

The optimizations have been firstly virtually conceptualized and validated by means of Computational Fluid Dynamics (CFD) and then physically validated through extensive, physical testing, covering all typical qualification demands. It must be highlighted, that as opposed to the gear pump improvements described in section 3.1, for those new bent axis pump series F1e and F12e, full qualification took place. This includes a very demanding Steady-State Tests (SST), where a factor of the maximum pressure is applied – increasing the pressure beyond maximum pressure – and the pump is operated at the same time at the increased SPS for the whole testing duration covering several weeks. The SPS values of the F12 vs. the F12e having passed this demanding duty-cycle are on Figure 3.2 (right).

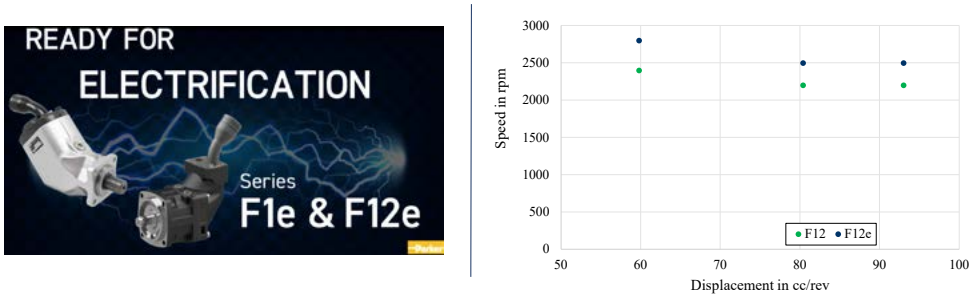


Figure 3.2: The new F1e and F12e series (left) and exemplary validated SPS values of the F12e vs. F12 pump versions at 1 bara inlet pressure (right)

There are no known limitations to this study at this stage, as the product has passed in C-/D-Sample maturity both all internal lab validations as well as the field testing successfully.

3.2.3. Exemplary use-case related to mobile electrification

The F1e has been tested on a truck application (tipper) to be compared to the standard F1. To avoid disruption of running operation where the measurement took place, it was not possible to compare on the same layout F1 vs. F1e during tipping operation. This can be seen in Figure 3.3 (left) where using the F1e, it appears that parasitic noises dominate further away from the pump compared to the same analysis with the F1. Hence, this made the F1e measurement look worse, than it actually was, though still achieving an improvement compared to the F1. The F12e however could be measured in exactly the same refuse truck side loader application to enable a side-by-side comparison. Representative measurement results can be seen on Figure 3.3 (right), showing excellent noise emission reductions.

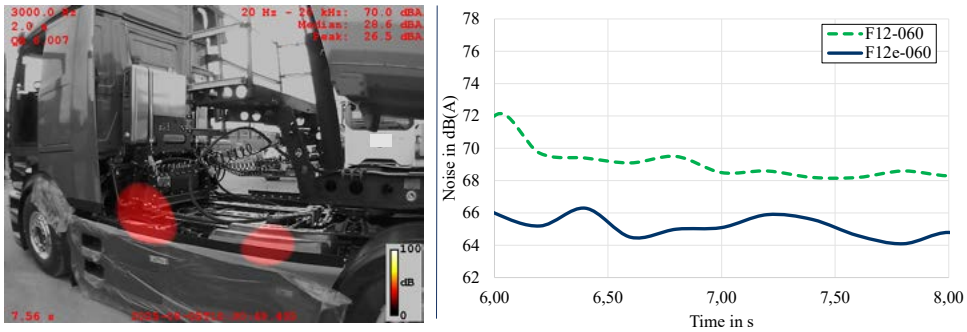


Figure 3.3: Exemplary F1e (left) measurements with acoustic camera during tipping operation and F12 vs. F12e noise measurement (right) results with acoustic camera performed on a refuse truck (ejection beam in operation at driver side)

In both cases, the external truck operators mentioned that they perceived significant noise improvements when switching from the F1 to the F1e and from the F12 to the F12e, which shows a real gain from psychoacoustic perspective in addition to the measurements results.

3.3. Next Generation Vane Pumps for Electrified Applications

3.3.1. Concept and/or design

Vane pumps are typically medium pressure pumps, but they still can reach higher pressures than today. Parker's new T8mini (refer to Parker Hannifin [24]) features 280 bar continuous pressure, while allowing 300 bar intermittent. This is made possible – amongst other changes – through novel coatings on the critical tribological interfaces between pressure and rear plates that interface with the rotor side faces. Additionally, a novel design of the port plates and housing in conjunction with additional radial holes in the cam profile enable excellent filling behaviour, yielding exceptional SPS performance. Together with a simpler mechanical design – for instance 10 instead of 12 vanes – the hydromechanical efficiency of this pump as well achieves highest levels. This enables a new breakthrough of overall efficiency for vane technology compared to the already rather high efficiency of the T7

series. The T8mini reaches on all his cam profiles enabling $V_{D,P} = 6 \ldots 12\ cc/rev$ an SPS of 4,800 rpm at 0.8 bara. [24]

3.3.2. Virtual and/or physical validation and limitations of this study

The T8mini has been conceptualized by means of virtual validation (e.g. 1D-Simulation) and detailed using modern CFD simulations, e.g. to improve filling behaviour as described in Jeganathan et al. [25] and high volumetric efficiencies. The standard validation program in lab environment has been performed, e.g. from performance measurements (efficiency), noise emissions in reverberant rooms as well as high-pressure and endurance tests. Additional tests have been performed to meet special requirements demanded in new applications, e.g. BEV trucks both in on- and off-highway applications. Here, tests with high-viscosity oils (e.g. ISO 680 oils) were performed to reach kinematic viscosities of $\nu \approx 5{,}000\ cSt$ to simulate the effect of cold temperatures < 0° C. It must be mentioned that such tests do not replace specific in-vehicle tests at such low temperatures, nevertheless, they provide good-enough insights on cold-start capabilities of the pump. The test layout for this specific lab test can be seen on Figure 3.4 (left). Tests have been performed both in standard configuration (Figure 3.4, right) as well as submerged in the oil tank (not shown below).

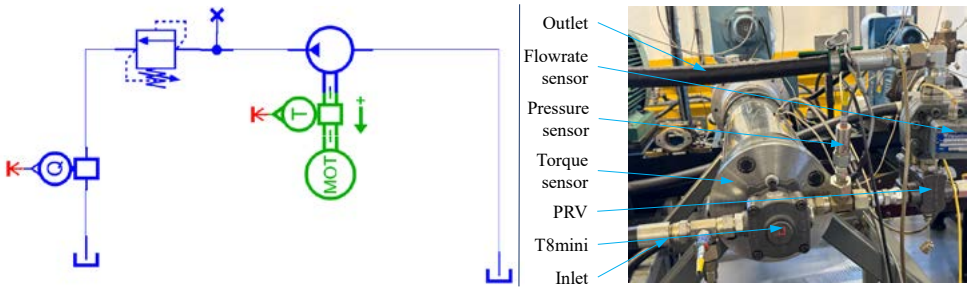


Figure 3.4: Hydraulic scheme & standard testing of T8mini at high viscosity

The submerged variant considered the intended layout in the final application by mimicking the application's oil tank volume and showed particularly good heat-up behaviour of the oil as well as an even better subjective noise assessment due to the surrounding oil, damping noise emissions. Ultimately, submerged pumps typically yield improved suction performance. As those investigations have been performed for a specific displacement of the T8mini and at a B-Sample maturity, further investigations are to be done for C-Samples at multiple displacements and also yielding higher production volumes.

3.3.3. Exemplary use-case related to mobile electrification

Since the T8mini has been optimized towards high-speeds and higher pressures, the detailed characterization at combinations of low speeds (≤ 1,000 rpm) at low pressures (≤ 50 bar) as well as high-speeds (> 4,000 rpm) at low pressures (≤ 50 bar) was not in the initial scope of the development. Nevertheless, special tests have been performed to check compliance with requirements demanded for an auxiliary low pressure 24 V ePump for On- and Off-Highway BEV Truck applications. Here, the T8mini's high-speed and cold-start capabilities as well as its resistance against high pressures spikes were key to achieve the high-power density, low noise, low envelope size and weight requirements as well as reliability for such a heavy-

duty application. While the B-Sample featured a bell housing between the Brushless Direct Current (BLDC) motor with integrated motor controller rated for 1.85 kW (refer to Figure 3.5), the C-Samples have been designed featuring a direct connection. The potential of axial space saving resulting by that can be seen in Figure 3.5. Further reduction of envelope size however requires more extensive integration, that will be outlined in the next section.

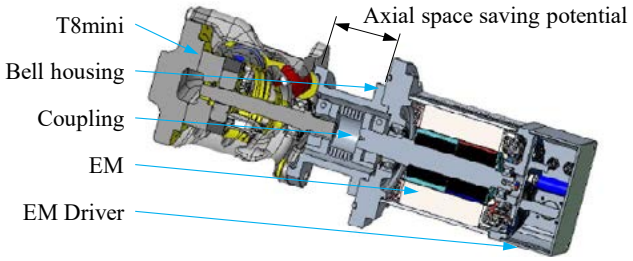


Figure 3.5: T8mini HM as 24 V auxiliary EHP for HV BEV Trucks, here in B-Sample maturity with moderate integration (still using a bell housing with coupling)

4. Discussion: A Novel Approach to EHP Integration

4.1. Limitations of increasing solely the pump's power density

The previous section demonstrated that with nowadays development tools like Computer Aided Engineering (CAE) – more specifically CFD – further advancements can be done on existing pump technologies, to increase their power density or to develop new pump generations. Increasing the SPS while improving the NVH behaviour at the same time enables reducing the pump size. In other words, assuming a certain flow Q_P is required from the application's perspective, if the pump's SPS n_P can be increased, its displacement $V_{D,P}$ can be subsequently reduced, as it can be seen in the following basic formula:

$$Q_P = n_P \cdot V_{D,P} \cdot \eta_{vol} \qquad (4.1)$$

If an intermittent flow of around 160 l/min is required, with the standard F12-060 for example, the requirement could not be reached without boosting the inlet, and it would be needed to take the next bigger size, i.e. the F12-080. With the F12e-060, this flow could be met through the increased SPS (refer to path a) in Figure 4.1) assuming the same volumetric efficiency η_{vol} for F12 and F12e at those different speeds: A significant reduction in displacement, hence pump size would be enabled. Another option would be, to maintain the F12-080 and operate it at lower speed to achieve similar flow Q_{des} at reduced noise (refer to path b) in Figure 4.1). However, having this example in mind, the HM downsizing remains marginal, compared to the size of the EM that typically represents the bulk of the EHP's size, weight and costs. Additionally, the speed limiting factor often being the HM's SPS, it remains difficult to operate the EM frequently in its efficiency optimum areas that typically are at higher speeds.

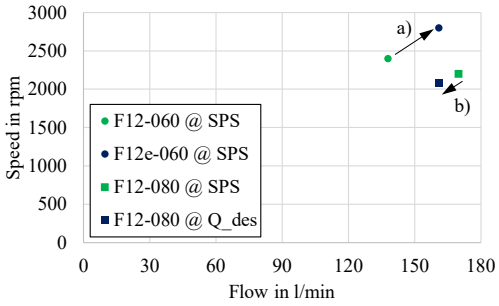


Figure 4.1: Exemplary calculation study to assess benefits and limitations of SPS increases

4.2. Semi-integration: High power density EHPs that can remain modular

One approach to tackle the beforementioned challenge to match HM and EM speeds, is to add a gearbox between EM and HM. The added complexity, losses and costs of the gearbox can only be justified, if the EM's downsizing is such, that from system perspective the overall efficiency remains constant or even improves through operating the EM in more efficient areas of its maps. Additionally, the system costs should ideally be lower than before. While Heeger, West & Ericson [10] showcase most possibilities to combine HM, EM and gearbox in Figure 4.2 (concepts (a) to (e), there may be another option (f) (Figure 4.2, below) that may appear counterintuitive at first sight.

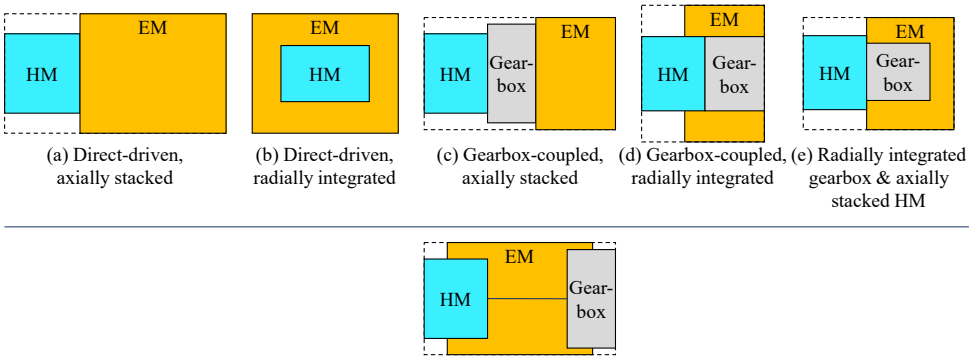


Figure 4.2: Concepts for combining Hydraulic Machines (HM) and Electric Motors (EM) into electro-hydraulic energy converters according to Heeger, West & Ericson [10] (left, with concepts (a) to (e)) and novel approach (f) for semi-integration (patent pending)

This option is to deliberately separate the gearbox and HM from the EM, so that both are arranged on opposed sides of the EM. This requires a through-shaft going through the EM which subsequently needs to have a hollow-shaft rotor and resolver design to have the gearbox connected to the HM. What appears to be more complex and costly at first sight, brings a lot of advantages as such a layout still enables a certain degree of integration, while offering more flexibility for modular design of the overall system and its components: Hence it can be referred to as "semi-integrated" EHP concept. The advantages of such an unconventional approach to integration are the following:

- Easier scalability of the outer diameter of the EM, while keeping the possibility to scale the EM in axial length. This gives additional flexibility for the design of a modular system comprising certain EM, HM and gearbox variants.
- The EM's rotor diameter can be much smaller, as it is not necessary to integrate gearbox and HM inside of the rotor, hence the only limiting factor of the EM's inner diameter is the very moderate diameter of the through-shaft. This is essential for high-speed EM concepts that require small diameters.
- The radial dimensions of HM and gearbox are less limited in radial direction than in other integrated concepts, offering additional possibilities to improve the scalability of those components to increase modularity of the EHP system.
- An oil cooling of the EM's rotor can easily be implemented via the through-shaft and/or the hollow rotor shaft, if a common oil sump between gearbox and HM is possible.
- If a common oil sump between gearbox and HM is chosen, the through-shaft layout extends the size of the oil sump, enabling using the oil sump as integrated tank.
- The through-shaft further enables integration of features to compensate radial and axial misalignments between gearbox and HM.
- In case a Planetary Gearset (PGS) is used as a gearbox – which has the benefit of highest power density and radially equalized forces – the overall EHP is fully co-axially arranged (concentric). This simplifies design of housing parts and provides an axially extending layout, that is beneficial if the radial extension of the available space is limited. Also, such a co-axial design supports a common part strategy.
- Gearbox and HM can be assembled nearly at the same time in parallel on each side of the EM, which reduces assembly time.
- Gearbox and HM can be dis-assembled and/or serviced independently from each other, which also eases maintenance, if required.

Such a concept is shown schematically in Figure 4.3 (i) using a 2-stage PGS on the left side with high PGS ratio, that is partly fitted under the windings of the EM. The HM is to be thought of a cartridge type of pump, that can be either of fixed or variable displacement type and can have 1- to 4-quadrant-capability, or any in-between step, depending on the requirements. The suction side of one or both pumps can be fed through the oil sump of the system or be led to an external tank. The outlets can be separated or together, depending on the application.

Another variant with only one minimal change can be seen in Figure 4.3 (ii). Here the only difference is that the hollow gear of the PGS stage I is connected to the carrier of the PGS stage II. This enables a lower ratio with only a minor modification of the PGS, supporting a common parts strategy.

Using only stage I of this same PGS – consistently driving common parts strategy – enables a further reduced ratio, that may be sufficient for intermediate displacements, see Figure 4.3 (iii).

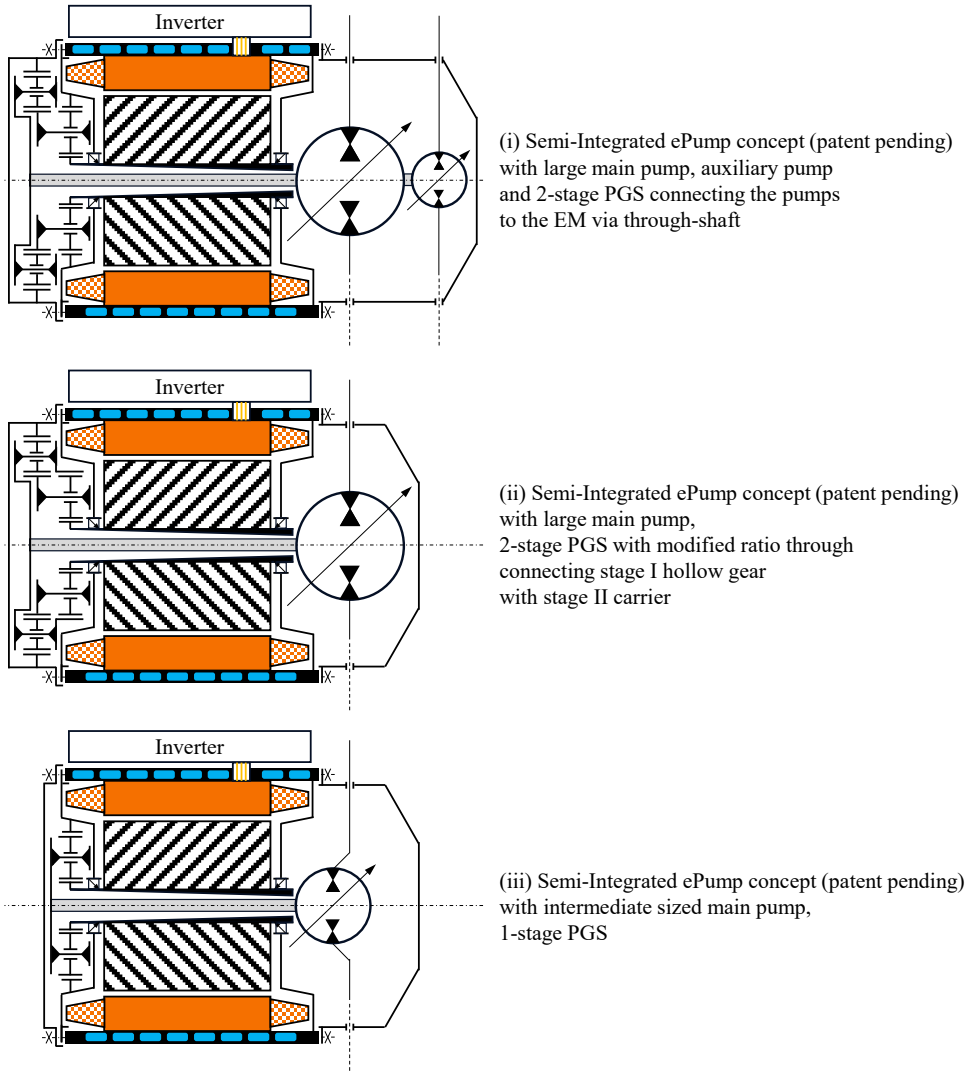


Figure 4.3: Semi-Integrated ePump concept and exemplary variants (patent pending[1])

Further variants can be thought of, e.g. increasing torque capacity of the PGS through increasing either the number of planets, their width, or both together. Changes in the teeth numbers with or without addendum modifications enable slightly modified ratios. The schematically indicated position of the inverter is deliberate, as being on top allows proximity to the EM, shortening connection distances on one hand. On the other hand, having it not axially integrated ensures that gearbox and HM can be radially scaled. Ultimately, the use of a gearbox does not only limit itself to be used to allow higher speeds hence higher efficiencies at the EM (EM "high-speeding") but may also be used to lower or

[1] A patent application covering aspects of this work has been filed; details will be available upon publication of the application.

at least keep the pump speed constant for purpose of à posteriori noise reduction in mobile applications particularly sensitive to noise emissions (HM “low-speeding”).

4.3. Potential limitations to tackle when dealing with semi-integration

While the abovementioned approach for semi-integration is rather new and shows considerable advantages on many aspects, it brings also new challenges with it that need to be further investigated. Certainly, achieving the right balance between integration and modularity represents one of the biggest challenges. Adding a gearbox will increase inertia, which may be a limiting factor for highly dynamic applications. Depending on the lubrication concept, using a mono-fluid for the gearbox and the hydraulics might require compromises on either one of those sub-components’ performance. Choosing separate, dedicated fluids on the other hand would complexify the sealing and eventually the assembly concept. All those aspects need to be investigated both theoretically and experimentally before implementation of semi-integration.

5. Summary

The present contribution shows advancements in the development of the next generation fixed hydraulic pumps for electrified mobile applications, more specifically for electrified open circuit work functions. In a first step, the current state of the art summarizes chances and challenges of electrification for hydraulic pumps. A main takeaway is, that due to increased need of space for batteries on HEV and BEV, the hydraulics need to be more power dense than ever to compensate a bigger portion of the vehicle’s weight being accounted by the battery. Additionally, noise emissions of hydraulic pumps need to be reduced, being not masked by ICE noise anymore. In a second step, a moderate approach outlines, how to achieve those development targets from pump perspective presenting three examples of Parker’s eReady pumps from lowest to highest degree of modifications: A new combination of manufacturing processes enabling higher SPS on gear pumps; an integrated ripple chamber in the new bent axis pumps F1e and F12e to reduce flow ripple, hence noise and further optimizations to increase SPS to up to 17% and ultimately a completely new generation of vane pumps designed for electrification. System level considerations have then been performed to reflect on the shown pump’s power density improvements and to enable even higher density of EHPs. A novel approach was presented – that tries to reconcile high integration and high modularity – described as “semi-integration”. Further research is needed in terms of endurance validation of the outlined SPS improvements for gear pumps as well as on the design and implementation of the outlined modular EHP system leveraging semi-integration.

6. References

[1] P. D. Fischer, M. Keil, R. Pelz, D. Frankiewicz, M. Wetjen and H. Zhang, “Electrified Pumps for Mobile Hydraulics - State of the Art, Challenges and Chances,” in *Karlsruher Schriftenreihe Fahrzeugsystemtechnik, Band 119, 13. Kolloquium Mobilhydraulik*, Karlruhe, 8th-9th October 2024, DOI: 10.5445/KSP/1000173591.

[2] P. Casoli, F. Scolari, T. Minav and M. Rundo, “Comparative Energy Analysis of a Load Sensing System and a Zonal Hydraulics for a 9-Tonne Excavator,” *actuators,* vol. 39, no. 9, 2020, DOI: 10.3390/act9020039.

[3] K. Myers, *The role of the ePump when transitioning from an ICE to an Electric System in Mobile Equipment,* Cleveland/Marysville: Parker Hannifin Corporation, Hydraulic Pump And Power Systems Division (HPS), 2024.

[4] R. Nagel and H. Zhang, “Electrification Challenges & Solutions in Mobile System Design,” in *ConExpo-CON/AGG and IPFE*, Las Vegas, NV, 2023.

[5] A. Cailteux, “Selecting a Shaft Connection Method for a Long ePump Life,” Parker Hannifin Corporation; Hydraulic Pump and Power Systems Division, 15th November 2023. [Online]. Available: https://blog.parker.com/site/usa/hydraulics-home/selecting-a-shaft-connection-method-for-a-long-epump-life-us. [Accessed 3rd January 2024].

[6] R. Pelz, M. Keil, K. J. Myers, D. E. N. Mack, F. Magnusson and O. K. Kayani, “System comprising an electric machine and a hydraulic unit, as well as electric machine and hydraulic unit”. United States of America Patent US 12,180,957 B2, 31st December 2024.

[7] Parker Hannifin Corporation, “SMART Electrification,” Motion Systems Group, [Online]. Available: https://discover.parker.com/smart-electrification-2. [Accessed 24th October 2025].

[8] Parker Hannifin Corporation, Motion Systems Group (MSG), Electric Motion and Pneumatic Division (EMPD), “GVI - Global Vehicle Inverter - Mobile Inverters for Traction, Electro-Hydraulic Pumps (EHP) and Auxiliary Systems 24 to 650 VDC,” May 2024. [Online]. Available: https://www.parker.com/content/dam/Parker-com/Literature/Electromechanical-Europe/Literature/192_300118_GVI_catalogue.pdf. [Accessed 25th October 2025].

[9] H. Zhang, S. K. R. Kalindindi, S. Huard, Y.-S. Chu, S. Natarajan, B. Burgess, F. Iannizzaro, P. Toops and R. Fisher, “Electric Motor with Integrated Hydraulic Pump and Motor Controller”. United States of America Patent US 11,788,529 B2, 17th October 2023.

[10] T. Heeger, M. West and L. Ericson, “Comparative Analysis of Direct-Drive and Gearbox-Coupled Electro-Hydraulic Energy Converters,” *SSRN,* 30th September 2025, DOI: 10.2139/ssrn.5548911.

[11] P. Lindh, J. Tiainen, A. Grönman, T. Turunen-Saaresti, C. Di, L. Laurila, E. Scherman, H. Handroos and J. Pyrhönen, “Two Cooling Approaches of an Electrohydraulic Energy Converter for Non-Road Mobile Machinery,” in *IEEE Transactions on Industry Applications*, 2023, DOI: 10.1109/TIA.2022.3207983.

[12] P. Ponomarev, T. Minav, R. Aman and L. Luostarinem, “Integrated Electro-Hydraulic Machine with Self-Cooling Possibilities for Non Road Modile Machinery,” *Strojniski vestnik - Jounral of Mechanical Engineering,* vol. 61, no. 3, pp. 207-213, 2015, DOI: 10.5545/sv-jme.2014.2017.

[13] T. Pietrzyk, D. J. Roth, K. Schmitz and G. Jacobs, “Design study of a high speed power unit for electro hydraulic actuators (EHA) in mobile applications,” *Fluid power networks : proceedings : 19th - 21th March 2018 : 11th International Fluid Power Conference,* vol. 2, pp. 232-245, 2018, DOI: 10.18154/RWTH-2018-224632.

[14] N. Gadiyar, G. Bohach, M. M. Nahin, J. Van de Ven and L. Severson, “Design of an Axial Flux Machine With an Integrated Hydraulic Pump for Off-Highway Vehicle Electrification,” in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*, Detroit, MI, USA, 2020, DOI: 10.1109/ECCE44975.2020.9235731.

[15] S. Sarode, Design and Analysis of An Integrated Electrohydraulic Axial Piston Machine (Version 1), West Lafayette, Indiana: Faculty of Purdue Universtiy, School of Mechanical Engineering, 2023, DOI: 10.25394/PGS.23421326.v1.

[16] J. Ivantysyn and M. Ivantysynova, Hydrostatische Pumpen und Motoren: Konstruktion und Berechnung, Vogel Communications Group GmbH & Co. KG, 1993.

[17] D. Schweigert, M. E. Gerlach, A. Hoffmann, B. Morhard, A. Tripps, T. Lohner, M. Otto, B. Ponick and K. Stahl, “On the Impact of Maximum Speed on the Power Density of Electromechanical Powertrains,” *Vehicles,* vol. 2, no. 2, pp. 365-397, 2020, DOI: 10.3390/vehicles2020020.

[18] S. Sarode, H. Assaf, P. Tawarawala, A. Vacca and L. S. S. Shang, “A Survey Study on ePump, Design Architectures for Mobile Hydraulics,” in *Proceedings of the ASME/BATH 2023 Symposium on Fluid Power and Motion Control*, Sarasota, FL, USA, 16th-18th October 2023, DOI: 10.115/FPMC2023-113009.

[19] M. A. Morselli, “Rotary positive-displacement pump with meshing gear wheels without encapsulation, and gear wheel for such a positive-displacement pump”. USA Patent US 2003/0231972 A1, 3rd June 2003.

[20] X. Zhao and A. Vacca, “Analysis of Continuous-Contact Helical Gear Pumps through Nunmerical Modeling and Experimental Validation,” *Mechanical Systems and Signal Processing,* no. 109, pp. 352-378, 2018, DOI: 10.1016/j.ymssp.2018.02.043.

[21] M. A. Morselli, “Double-Helical gear rotary positive-displacement pump”. United States Patent Patent US 9,464,632 B2, 19th July 2012.

[22] A. Vacca and G. Franzoni, Hydraulic Fluid Power: Fundamentals, Applications, and Circuit Design, Hoboken, NJ: Wiley, 2021.

[23] Parker Hannifin Corporation, Motion Systems Group (MSG), Hydraulic Pump and Motor Division Europe (HPME), “eReady Pumps - Series F1e/F12e - Fixed

Displacement," September 2025. [Online]. Available: https://www.parker.com/content/dam/Parker-com/Literature/PMDE/Catalogs/eready/MSG30-8253-UK.pdf. [Accessed 25th October 2025].

[24] Parker Hannifin Corporation, Motion Systems Group (MSG), Hydraulic Pump and Motor Division Europe (HPME), "Hydraulic Vane Pump T8mini - Parker Denison Vane Technology," [Online]. Available: https://www.parker.com/content/dam/Parker-com/Literature/PMDE/Catalogs/Vane_Pumps/t8mini/MSG30-0113_UK_T8MINI.pdf. [Accessed 25th October 2025].

[25] V. Jeganathan, R. Ahmed, Y. Chen, S. S. Pasunurthi, A. Pandey, E. Gnesi, R. Kalidindi, B. Kashid and H. Zhang, "Enhancing System Performance in Subsea and Off-Highway Applications: A High-Fidelity, Transient Dynamic CFD Analysis Approach," *Conference Proceedings of the 2024 International Maha Fluid Power Conference,* vol. 27, no. 1, 3-5 September 2024, DOI: 10.13052/rp-9788770047456.

[26] Parker Hannifin Corporation, Motion Systems Group (MSG), Hydraulic Pump and Motor Division Europe (HPME), "Gear Pumps/Motors - Series PGP/PGM - Fixed Displacement Pumps, Cast-Iron and Aluminium Designs," November 2017. [Online]. Available: https://www.parker.com/content/dam/Parker-com/Literature/PMDE/Catalogs/Gear_Units/PGP_PGM/HY30-3300-UK.pdf. [Accessed 25th October 2025].

[27] Parker Hannifin Corporation, Motion Systems Group (MSG), Electric Motion and Pneumatic Division (EMPD), "Electrified Vehicle Motor Series - Permanent Magnet (PMAC) Motor for Mobile Systems," 2025. [Online]. Available: https://www.parker.com/content/dam/Parker-com/Literature/Electromechanical-Europe/Literature/192_300140_Vehicle_Motor_Series.pdf. [Accessed 25th October 2025].

Biographies



Patrick Fischer graduated as Diplom-Ingenieur in Mechanical Engineering from Technical University of Munich (TUM). He worked several years as Research Associate at the Gear Research Centre (FZG, TUM) before taking on R&D roles of increasing responsibility in the automotive industry in Europe and PR China. He joined the Parker Hannifin Corporation in 2023 as Division Engineering Manager of HPME Division, part of the Motion Systems Group (MSG). His research areas feature motion control technologies and electrification.



Adrien Gambet graduated with a general engineering degree (equivalent to a M. Sc.) from École Centrale de Lille in 2025. After working on 1D simulation models and on a process of creation of digital twins for Parker pumps, he permanently joined Parker Hannifin's HPME Division in 2025 as product development engineer, focusing primarily on the design and industrialization of the next generation of Parker's vane pumps.



Shanmukh Sarode, Ph.D. graduated with a Ph.D. in Mechanical Engineering from Purdue University, USA, and did his graduate research at the Maha Fluid Power Research Center. He has over five years of experience in the design and manufacturing of hydraulic and electric pumps. He currently serves as an R&D engineer at the Parker Hannifin Motion Technology Center, focusing on new product development.



Howard Zhang, Ph.D. is a chief engineer and Parker engineering fellow for Parker Motion Systems Group, with over 20 years of motion control industry experience. He has also led and coordinated various advanced mobile electro-hydraulic components/systems development and commercialization activities. Currently, his focus is advancing Parker's engineering innovations around intelligent mobile hydraulic systems controls with power-dense and efficient ePumps.

Experimental study of the influence of the channel geometry of regenerative pumps on their internal flow

VUCEMILOVIC Alexandre[1], RAVELET Florent[2], KOUIDRI Smaïne[3], ESPANET Christophe[4]

1. *MMT (Besançon, France) and LIFSE (ENSAM Paris, France), alexandre.vucemilovic@movingmagnet.com*
2. *LIFSE (ENSAM Paris, France), florent.ravelet@ensam.eu*
3. *LIFSE (ENSAM Paris, France), smaine.kouidri@ensam.eu*
4. *MMT (Besançon, France), Christophe.espanet@movingmagnet.com*

Abstract.

Regenerative pumps are capable of generating high pressures at relatively low flowrates, making them suitable for applications such as electric vehicle cooling systems and cleaning systems for autonomous vehicle sensors. However, their nominal efficiency remains relatively low. The objective of this study is to improve the understanding of how the channel geometry influences the internal flow and performance of regenerative pumps, with particular emphasis on the empirical coefficient of the analytical model. A regenerative pump prototype with variable channel geometry was designed and manufactured. The channel height and width were modified while maintaining a constant side-channel area to enable direct comparison among configurations. The hydrodynamic characteristics, including pressure-flowrate and efficiency-flowrate curves, were experimentally measured. Based on both CFD simulations and experimental data, the empirical coefficients of the analytical model developed by Badami and Mura [1, 2] were estimated, and the corresponding performance predictions were analysed.

Keywords. Regenerative pump, side-channel pump, internal flow, regenerative pump geometry.

1. INTRODUCTION

Pumps can be classified into three main categories: centrifugal pumps, axial pumps and volumetric pumps. Centrifugal pumps use the centrifugal force generated by their rotor to project the fluid outward into the volute and toward the outlet. These pumps can produce high flow rates but have small pressure differential between the inlet and outlet (Figure

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1.1(a)). Axial pumps have similar characteristics but can generate higher flowrate (Figure 1.1(b)). In contrast, volumetric pumps rely on a variation in volume to draw liquid through the inlet and discharge it through the outlet. They are more efficient at high pressure and large flow rates (Figure 1.1.(c)).

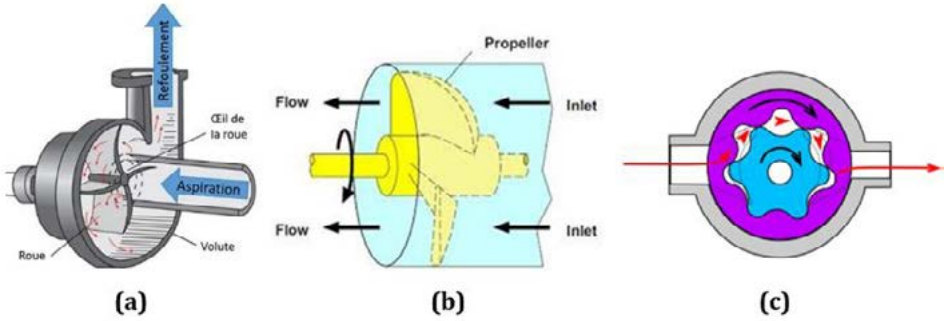


Figure 1.1. Centrifugal pump (a), Axial pump (b) and volumetric pump (gerotor pump) (c)

The pumps described above do not operate efficiently under conditions of low flowrate and high pressure. Regenerative pumps, which combine characteristics of both centrifugal and positive displacement machines, are well adapted to such operating regimes. Applications such as cleaning sensors on autonomous vehicles require compact pumps capable of generating high pressure at low flowrates with good efficiency, conditions for which regenerative pumps are particularly suitable. Introduced in the 1920s by Siemen and Hinsch [3], these pumps are currently used in fluid transfer (oil and gas, boiler feed, cooling and refrigerant systems) and can also be used in mixing and misting applications.

One advantage of this type of pump is its simple design. Its rotor consists of a central bore used for the mounting on a rotating shaft and a series of radially distributed blades around the peripheral surface (Figure 1.2). The symmetry and regular distribution of the blades ensure that the rotor is balanced which minimize vibration and wear. The housing of the pump has a side channel integrated along its inner circumference, aligned with the blades of the rotor (Figure 1.3). It plays a crucial role in the pump operation, that will be explained later. Another key feature of the housing is the stripper, which is a component that separates the inlet and the outlet sections of the pump. This part ensures that the leakages from outlet to inlet are minimal thus improving pump performance.

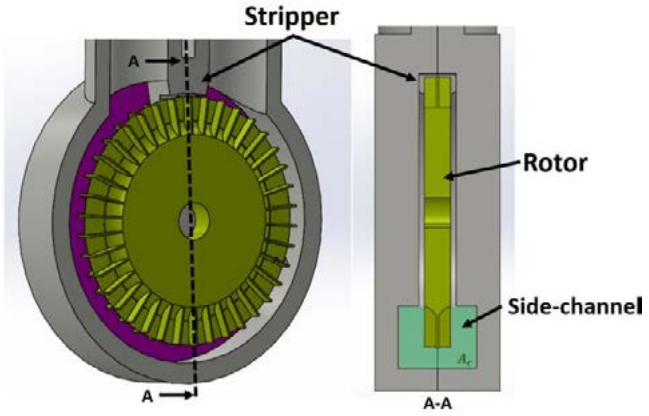


Figure 1.2. Regenerative pump basic design

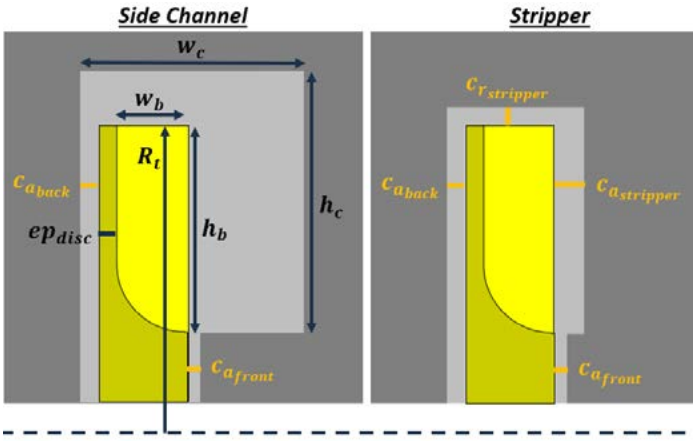


Figure 1.3. Regenerative half-pump geometric parameters

The flow within such a pump is complex and three-dimensional (Figure 1.4). Initially, the fluid enters the pump and is transported by the rotor to the outlet of the housing. Observing from a radial cross-section aligned with a blade, one can also identify a flow referred to as "regenerative" (Figure 1.5). During this flow, the fluid enters at the inner radius of the blade, and the rotation of the wheel increases the fluid's energy. Subsequently, the fluid moves into a channel where various physical phenomena convert a portion of the gained energy into pressure. These phenomena will be explained later. The fluid then returns to the inner radius of the wheel, and the cycle begins anew.

This flow behaviour has been validated by numerous experimental studies and numerical simulations [4-9]. A notable advantage associated with this flow behaviour is the pump's pulsation-free operation. Another advantage is that regenerative pumps are less prone to cavitation compared to centrifugal or volumetric pumps, the gradual increase of the fluid pressure along the pump circumference minimizes turbulent zones that could lead to low-pressure areas. However, regenerative pumps are not immune to cavitation, as shown by

Senoo [4]. This flow behaviour also allows the pumping of liquid/gas mixtures, as gas bubbles are thoroughly mixed with the liquid, breaking down larger gas pockets and evenly distributing the gas within the liquid.

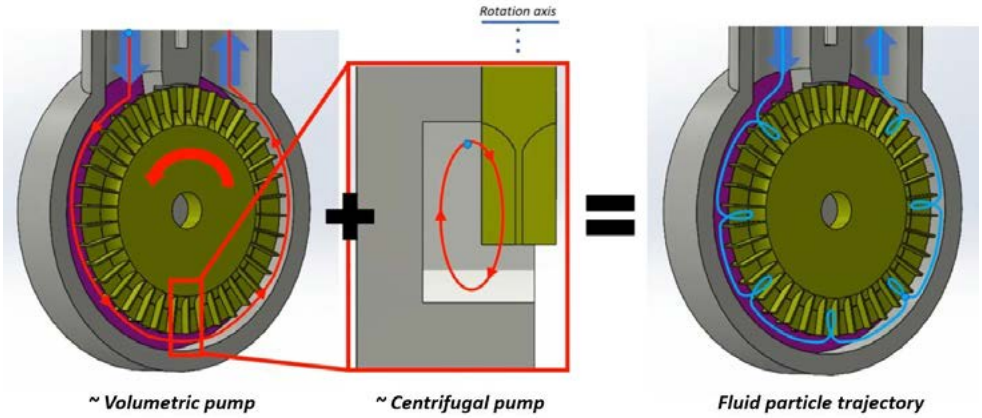


Figure 1.4. Internal flow inside of a regenerative pump

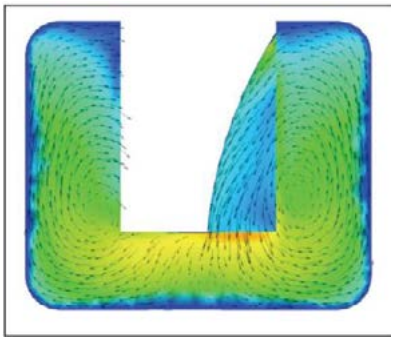


Figure 1.5. Flow inside of the side-channel of a regenerative pump [9]

To predict the characteristic of a regenerative pump, the development of models is essential. Multiple theories have been developed that try to predict the performances of regenerative pumps.

One of these theories is the Momentum Exchange Theory (MET) which was introduced in the 50s by Hopkins and Wilson [5, 6]. In this theory, the mechanism responsible for the pressure generation is the transfer of momentum between the impeller blades and the fluid. The momentum energy imparted to the fluid by the impeller blades propels the fluid to the tip of the blades. Upon entering the pump channel, the fluid's momentum energy is partly converted into pressure through interaction with the turbulence generated by the fixed wall of the outer diameter. Nevertheless, MET is not without limitations. Turbulence mechanism responsible for the pressure rise are simplified and their modelling rely on the use of empirical coefficient. The model then requires experimental data to give accurate results.

The model also relies on several idealized assumptions about flow conditions which can introduce deviations from the actual pump performance.

Since this initial work, several research groups have extended and refined the model [1, 2, 10]. Notably, it has been established that the flowrate in regenerative pumps is strongly influenced by the cross-sectional area of the side channel. Consequently, following definitions for the dimensionless flowrate φ and pressure coefficients ψ of regenerative pumps have been proposed:

$$\varphi = \frac{Q}{UA_c} \quad ; \quad \psi = \frac{2\Delta P}{\rho U^2} \tag{1.1}$$

Here, Q and ΔP are respectively the flowrate and pressure differential between the inlet and the outlet of the pump, A_c the cross-sectional area of the side channel, U the blade tip tangential velocity and ρ the fluid density.

φ and ψ are dimensionless performance coefficients that enable consistent comparison across pumps of varying sizes and rotational speeds. In this study, the definitions given in eq. (1.1) are employed to evaluate and contrast pump configurations featuring different side-channel geometries. By expressing the flowrate and the pressure generated by the pump in nondimensional form, these coefficients isolate the intrinsic dynamic behaviour of each design, allowing performance trends to be interpreted independently of scale or operating conditions.

The paper first introduces the analytical modelling of regenerative pumps. The experimental setup and the tested pump configurations are then described, followed by a presentation of the experimental results. Using both experimental data and CFD simulations, the empirical coefficients of the analytical model are determined and applied to predict pump performance. Finally, a comparison between the predicted and measured efficiency curves is conducted to evaluate the accuracy of the analytical model.

2. REGENERATIVE PUMP MODELLING

The model presented in this section was originally developed by Badami and Mura in 2010 [1] based on Wilson et al. work [6].

To determine the pressure coefficient of the pump, the momentum equation is applied to the fluid in the side-channel:

$$\frac{dK}{dt} = T_p + T_w \tag{2.1}$$

Where K is the angular momentum of the fluid, T_p is the torque associated to the increase in pressure along the pump circumference and T_w is the torque linked to the effect of the shear forces on the side channel walls.

By assuming that the fluid is not compressible and that the torque T_w is proportional to the square of the mean velocity in the channel, equation (2.1) can be rewritten as:

$$Q_m \rho \left(R_t C_{u_2} - r_1 C_{u_1}\right) = A_c r_c \Delta P + \frac{1}{2} K_p A_c r_c \rho u_c^2 \tag{2.2}$$

With R_t the radius of the tip of the impeller, r_c the mean radius of the side channel, r_1 the mean radius of the inlet meridian flow stream at the impeller as defined in Figure 2.1, A_c the cross-section of the side channel, Q_m the recirculation flowrate, C_{u_1} and C_{u_2} the tangential components of the absolute velocity at the entry and exit point of the blades and K_p a coefficient used to model the losses linked to the friction forces on the side channel wall.

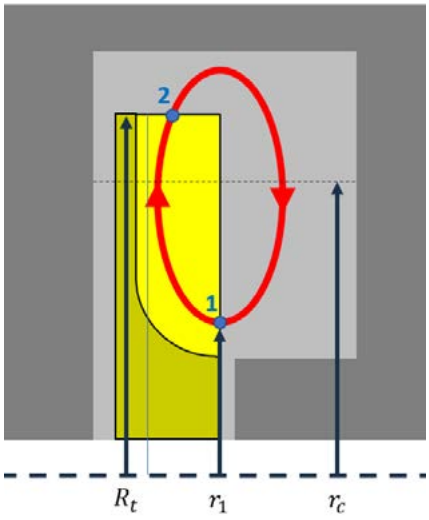


Figure 2.1. Mean flow streamline

the non-dimensional pressure coefficient ψ of the pump can then be determined from equation (2.2):

$$\psi = 2\left(\frac{Q_m}{A_c U}\right)\left(\frac{r_1}{r_c}\right)\left(\frac{R_t}{r_1}\frac{C_{u_2}}{U} - \frac{C_{u_1}}{U}\right) - K_p\varphi^2 \tag{2.3}$$

With φ and ψ the non-dimensional flowrate and pressure coefficients as defined in equation (1.1) and U the tangential velocity of the tip of the blades.

The mean radius of the inlet flow of the impeller can be estimated using the following equation:

$$r_1 = R_t - kh_b \tag{2.4}$$

Where h_b is the height of the blade (Figure 1.3) and k is an unknown coefficient. Badami & Mura use a value of $k \approx \frac{2}{3}$.

Assuming that the angular velocity of the fluid in the side channel is constant along the pump circumference:

$$\frac{C_{u_1}}{U} = \frac{U_{c_1}}{U} = \frac{r_1}{r_c}\varphi \quad ; \quad \frac{C_{u_2}}{U} = \frac{U_2}{U} - \frac{\Delta U_2}{U} = \left(1 - \frac{\Delta U_2}{U_2}\right) \tag{2.5}$$

Here, $\frac{\Delta U_2}{U_2} = \sigma$ is a slip coefficient that considers the velocity difference between the fluid exiting the blades and the fluid in the side channel. Badami & Mura gives the following approximation to estimate its value:

$$\sigma = \frac{2.6}{Z\left(1 - \left(\frac{r_1}{R_t}\right)^2\right) + 2.6} \tag{2.6}$$

Substituting equation (2.5) in equation (2.3):

$$\psi = 2\left(\frac{Q_m}{A_c U}\right)\left(\frac{R_t}{r_c}(1-\sigma) - \left(\frac{r_1}{r_c}\right)^2 \varphi\right) - K_p \varphi^2 \tag{2.7}$$

In order to estimate the value of the recirculation flowrate Q_m, the energy equation must be applied to the fluid in the vane grooves and in the side channel, with reference to a coordinate system moving with the impeller and the fluid respectively. By summing the obtained equations:

$$K_m \left(\frac{A_c}{A_2}\right)^2 \left(\frac{Q_m}{A_c U}\right)^2 + \left(\frac{r_1^2 - R_t^2}{R_t^2} + \left(\frac{U_{c_2}}{U}\right)^2 - \left(\frac{U_{c_1}}{U}\right)^2\right) = 0 \tag{2.8}$$

With the assumption that the fluid has a constant angular velocity in the side channel:

$$\frac{Q_m}{A_c U} = \frac{1}{\sqrt{K_m}} \frac{A_2}{A_c} \sqrt{\frac{R_t^2 - r_1^2}{R_t^2} - \frac{R_t^2 - r_1^2}{r_c^2}\varphi^2} \tag{2.9}$$

Where K_m is a loss coefficient that considers the energy losses due to restrictions and enlargements of the flow in the meridian plane and the losses due to the turbulent shear forces in the vanes, and A_2 is the outlet area of the blades.

This model supposes that there are no internal leakages in the pump. Badami & Mura [1] propose that the real flowrate coefficient φ_R is equal to the difference between the theoretical flowrate coefficient φ and the leakage coefficient φ_f:

$$\varphi_R = \varphi - \varphi_f \tag{2.10}$$

In their pump geometry, there are two main leakages: the leakage through the axial mechanical clearance between the impeller and the housing (φ_{f_1}) and the leakage through the stripper clearances (φ_{f_2}).

Using Bernoulli's equation and assuming the leakage induced by the drag of the impeller through the stripper can be correlated by a triangular drag velocity distribution in the clearance, the leakage flowrate coefficients can be determined:

$$\varphi_{f_1} = \frac{A_{f_1}}{A_c} \frac{1}{\sqrt{K_{f_1}}} \sqrt{\frac{\psi}{2}} \quad ; \quad \varphi_{f_2} = \frac{A_{f_2}}{A_c}\left(\sqrt{\frac{\psi}{Z_s K_{f_2}}} + \frac{1}{2}\frac{r_c}{R_t}\right) \tag{2.11}$$

Where A_{f_1} and A_{f_2} are the leakage areas and Z_s is the number of blades facing the stripper. In those equations, two new loss coefficients appear: K_{f_1} and K_{f_2}. Those coefficients consider the pressure drops in the different leakage areas.

Finally, the efficiency η of the regenerative pump is defined as the ratio between the hydraulic power P_h generated by the pump and the input mechanical power P_m. The

hydraulic and mechanical powers can be estimated using the flowrate and pressure coefficient of the pump:

$$P_h = \frac{1}{2}\rho A_c U^3 \varphi_R \psi \tag{2.12}$$

$$P_m = \frac{1}{2} A_c \rho U^3 \frac{r_c}{R_t}\left(\psi + K_p \varphi^2\right) \tag{2.13}$$

$$\eta = \frac{P_h}{P_m} = \frac{R_t}{r_c}\frac{\varphi_R \psi}{\psi + K_p \varphi^2} \tag{2.14}$$

To predict the hydraulic performances of a regenerative pump using this model, the coefficients K_m, K_p, k, K_{f_1} and K_{f_2} must be determined. In [1] those coefficients are considered as constants for a given geometry.

In this study, the effect of the leakages will be neglected ($K_{f_1} = K_{f_2} = 0$) and K_p and k will be determined using CFD simulations. The value of K_m will then be estimated from experimental data.

3. EXPERIMENTAL SETUP AND RESULTS

The previously described model will be applied to the following pump prototype (Figure 3.1).

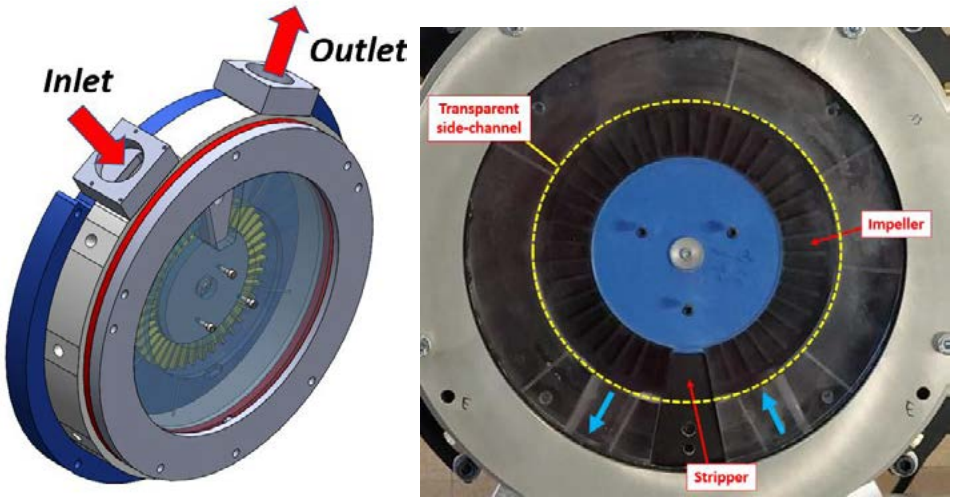


Figure 3.1. Regenerative pump prototype

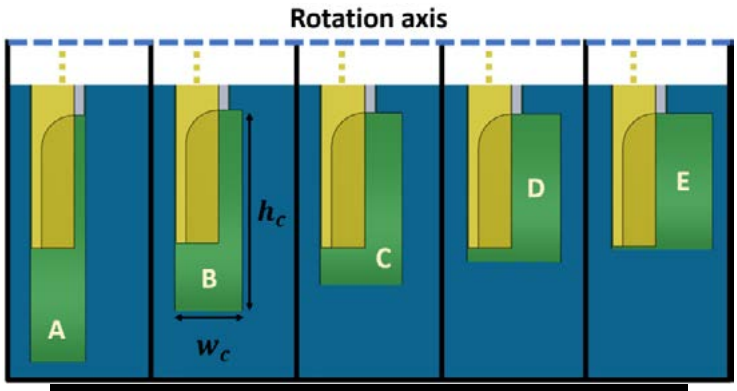
This prototype is a half regenerative pump (Figure 1.3). Unlike a full regenerative pump, in which the regenerative flow is on both side of the impeller, the half regenerative pump only features one regenerative flow in the side channel. This design simplifies the manufacturing and allows for focused investigation of regenerative flow behaviour while retaining the core operating principles of regenerative pumps. Theoretically, a half regenerative pump can

achieve the same pressure head as a full regenerative pump, but its flowrate is reduced by half due to the single-sided channel. However, this design introduces an axial clearance c_{a1} between the impeller and the housing (Figure 1.3), in which the fluid can flow. This creates internal leakages and viscous frictions that do not exist in the full pump configuration, impacting the performances of the pump.

Table 3.1. Constant geometrical parameters

$\boldsymbol{R_t}$	$\boldsymbol{82.5\ mm}$	$\boldsymbol{c_{a_{back}}}$	$\boldsymbol{0.7\ mm}$
$\boldsymbol{w_b}$	$8.5\ mm$	$c_{a_{front}}$	$0.25\ mm$
$\boldsymbol{h_b}$	$26.25\ mm$	$c_{a_{stripper}}$	$0.5\ mm$
$\boldsymbol{ep_{disc}}$	$2\ mm$	$\boldsymbol{c_{r_{stripper}}}$	$0.5\ mm$

Despite the constraints of the design, the channel geometry can be easily modified, and multiple pump configurations have been manufactured (Table 3.1 and Figure 3.2).



Configuration	w_c	h_c
A	$11\ mm$	$48.75\ mm$
B	$13.5\ mm$	$39.7\ mm$
C	$16\ mm$	$33.5\ mm$
D	$18.5\ mm$	$29\ mm$
E	$20\ mm$	$26.8\ mm$

Figure 3.2. Geometric configurations of the channel

To evaluate the performance of this half regenerative pump prototype, a dedicated hydraulic test bench was developed. This setup allows for the measurement of key hydraulic and mechanical parameters.

The test bench consists of a closed-loop circuit equipped with a variable speed electric motor driving the pump. A torquemeter equipped with an incremental encoder allows for the measurement of the input torque T_{exp} and rotational speed ω_{exp} of the impeller leading to the computation of the input mechanical power. The flowrate Q_{exp} is measured using an ultrasonic flowmeter installed downstream of the pump, while pressure sensors mounted upstream and downstream of the pump provide measurements of the pressure differential of the pump $\Delta P_{exp} = P_{outlet} - P_{inlet}$. A data acquisition system records all sensor outputs for post-processing and performance evaluation of the different pump configurations. The circuit also includes a fluid tank to minimize air entrainment as well as a regulation valve used to modify the flow' resistance and simulate different operating conditions (Figure 3.3). A temperature sensor and a barometer mounted on the fluid tank allows for the control of the fluid temperature and of the hydrostatic pressure inside of the hydraulic circuit. Pressure sensors mounted on the circumference of the pump also allows for the measurement of the pressure gradient inside of the pump but this experimental data will not be investigated in this paper.

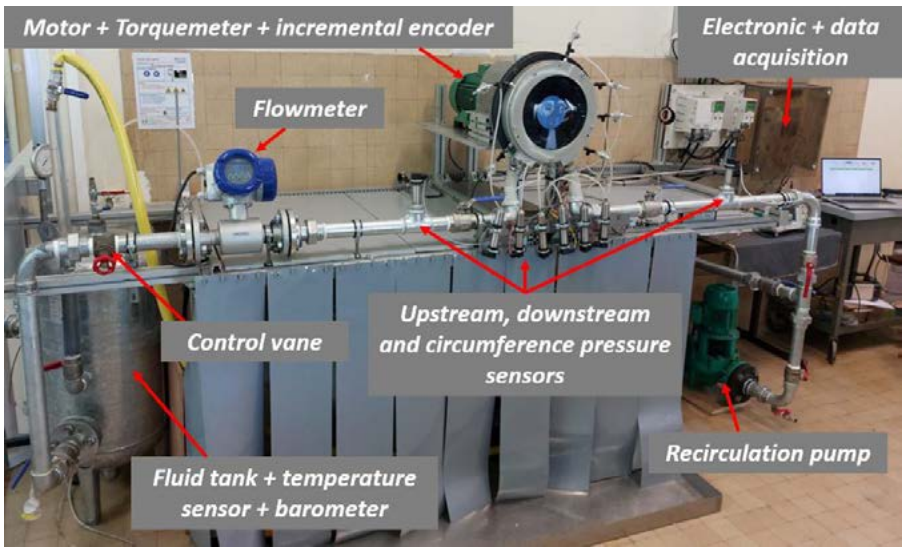


Figure 3.3. Hydraulic test bench

For each configuration, the pump rotational speed was maintained at a constant value, and the pressure drop of the hydraulic circuit was progressively increased by adjusting the control valve, enabling measurement of the pump flow/pressure and flow/efficiency characteristics.

Experimental results at 900 RPM (configurations A to D) are presented in Figure 3.4. Due to manufacturing issues encountered with configuration E, its performance could not be measured at 900 RPM; therefore, the results obtained at 600 RPM are reported instead.

φ_{exp} and ψ_{exp} have been calculated from experimental flowrate, pressure differential and rotational speed data using the formulas presented in eq.(1.1). The experimental efficiency was calculated from flowrate, pressure, torque and rotational speed measurements:

$$\eta_{exp} = \frac{Q_{exp}\Delta P_{exp}}{T_{exp}\omega_{exp}} \tag{3.1}$$

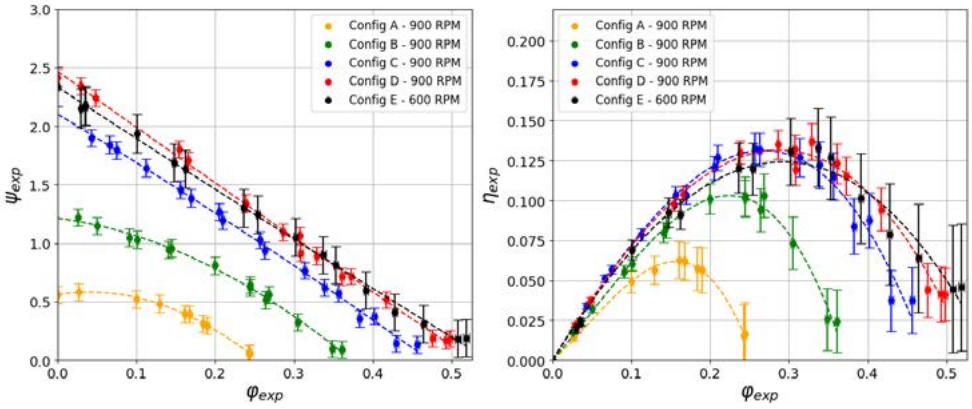


Figure 3.4. Experimental results

The study of the coefficient K_m will be conducted on configuration C at 900 RPM (blue curves on Figure 3.4).

4. ESTIMATION OF K_m

As mentioned in Chapter 2, leakage effects are neglected in this study. Using equations (2.3) and (2.9), an expression for the recirculation loss coefficient K_m can be derived:

$$K_m = \frac{4A_2^2(R_t^2 r_c^2 - R_t^2\varphi^2(R_t^2 - r_1^2) - r_1^2 r_c^2)(R_t r_c(\sigma - 1) - r_1^2\varphi)^2}{A_c^2 R_t^2 r_c^6\left(\psi + K_p\varphi^2\right)^2} \tag{4.1}$$

In this formulation, K_m is expressed in terms of known geometric parameters (R_t, r_c, A_2), dimensionless flowrate and pressure coefficient (φ, ψ) and unknown parameters σ, r_1 and K_p. The slip coefficient σ can be evaluated from eq. (2.6). The remaining two coefficients, K_p and k are determined via CFD simulations.

K_p represents the pressure drops in the pump's inlet, outlet and side channel. A simplified CFD simulation of these regions, excluding the impeller, enables the evaluation of the pressure drop generated by the pump. Experimental measurements were also performed using an impeller without blades and the recirculation pump installed on the test bench to produce the required flowrates (Figure 4.1).

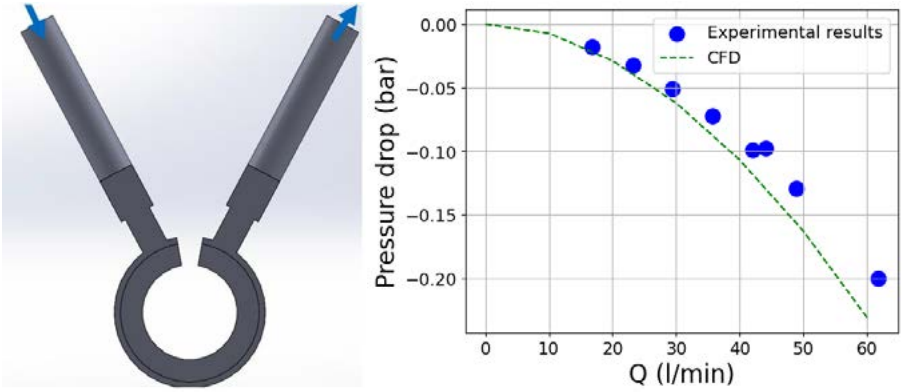


Figure 4.1. CFD geometry and comparison of experimental and CFD results

The experimental data and CFD predictions are in close agreement, with a minor deviation of approximately 25 mbar at high flowrates, attributable to leakage paths not included in the simulation. This agreement enables an accurate estimation of the pressure drop in complex channel geometries.

From the pressure drop curve estimated by CFD simulation, the value of K_p can then be determined:

$$K_p = \frac{2\Delta P_l}{\rho v^2} = \frac{2\Delta P_l}{\rho (Q/A_c)^2} \tag{4.2}$$

Where ΔP_l is the pressure drop at the flowrate Q and $v = \frac{Q}{A_c}$ is the fluid velocity in the side-channel. For configuration C, $K_p = 4.5$. Badami & Mura [2] reported a value of $K_p = 3.5$ for a regenerative air compressor with similar geometry, deriving the coefficient through conventional pressure-drop analysis.

The coefficient k is then evaluated from the velocity field in the side channel within a meridional plane. The radius r_1 is assumed to correspond to the center of the impeller inlet height. For configuration C, $k \approx 0.66$ (Figure 4.2) corresponding to the value used by Badami & Mura [2].

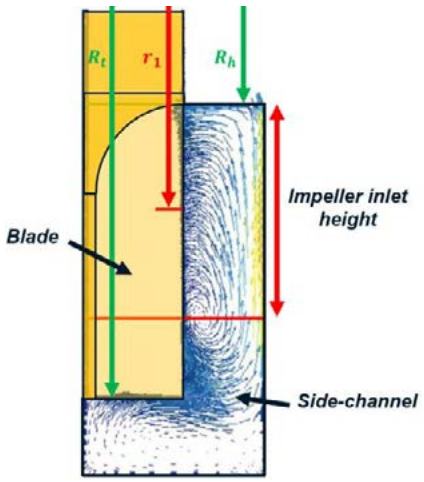


Figure 4.2. Flow in the meridional plane in configuration C at 900 RPM

With the values of K_p and k determined, K_m can be calculated from eq.(4.1) using the experimental φ and ψ data (Figure 4.3). In Badami's model [1], K_m is a constant loss coefficient. The mean value of the obtained K_m curve is then used. For configuration C, $K_m \approx 24.9$.

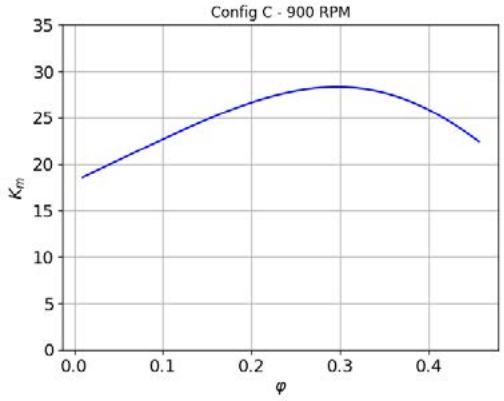


Figure 4.3. Calculated K_m for configuration C at 900 RPM

The obtained values of K_p, k and K_m enable prediction of pump performance through eqs. (2.3) and (2.14) (Figure 4.4).

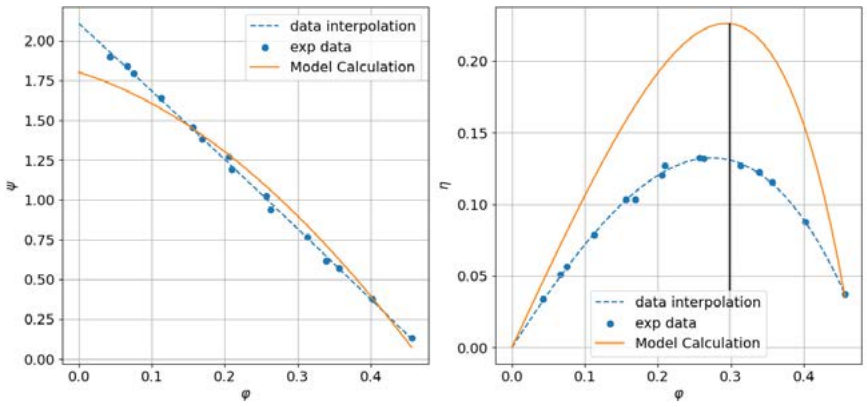


Figure 4.4. Computed pump performance and comparison to experimental data

Overall, the analytically derived $\psi = f(\varphi)$ curve (Figure 4.4) fit closely with experimental data. The predicted efficiency of the pump is overestimated because of neglected losses. However, the optimal flowrate of the pump is accurately captured.

Applying the same methodology to the remaining pump configurations demonstrates how channel geometry influences overall pump performance.

5. CHANNEL GEOMETRY INFLUENCE

In equation (2.3), the influence of the channel geometry is reflected in r_c, the mean channel radius of the pump, and in the loss coefficient K_p. Both parameters exhibit similar trends, decreasing from configuration A to configuration E (Figure 5.1).

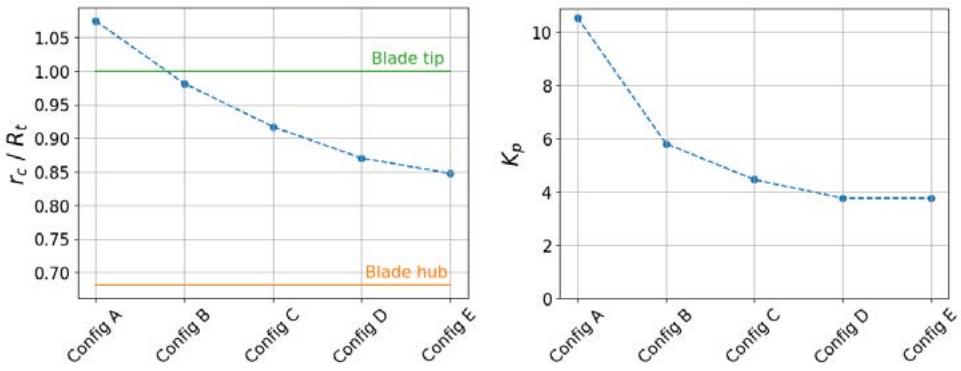


Figure 5.1. r_c and K_p values for each configuration

Variations in the channel geometry also affect the pump's efficiency curve. Figure 5.2 illustrates that configurations with smaller r_c exhibit higher maximum efficiency and a larger optimum flowrate, with an asymptotic trend observed around configuration D. The tendencies are consistent with both experimental and analytical results. A slight decrease in η_{max} is observed for configuration E, which can be attributed to the difference in pressure

resulting from the lower rotational speed, leading to a larger measurement error in the pressure sensors. This asymptotic behaviour around configuration D can also be seen in the variations of K_m (Figure 5.3).

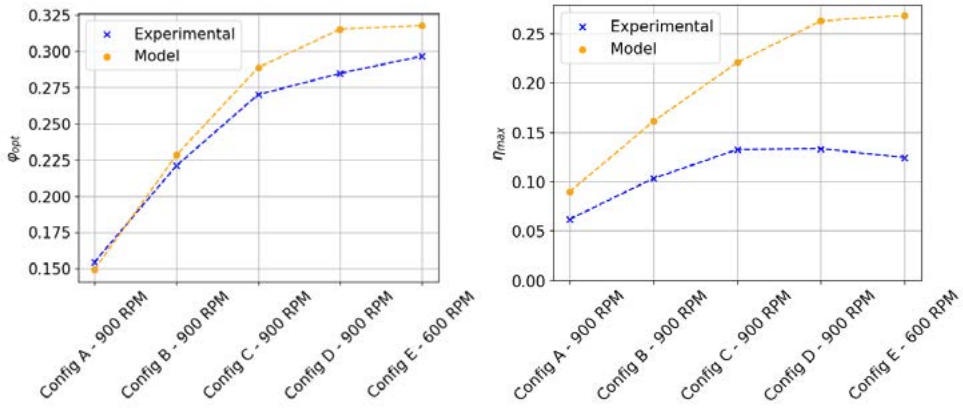


Figure 5.2. η_{max} and associated flowrate coefficient for each configuration

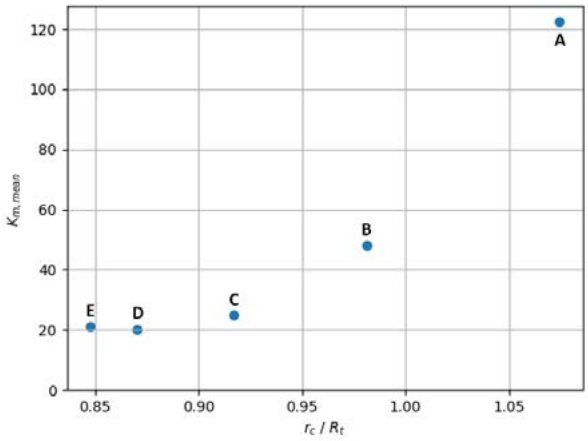


Figure 5.3. Mean K_m calculated for each configuration

6. CONCLUSION

This study provided further insight into the influence of side-channel geometry on the performance of regenerative pumps, with particular focus on the empirical coefficients of the model developed by Badami and Mura [1], based on the earlier work of Wilson et al. [6]. The loss coefficient K_p and the geometric coefficient k were estimated using CFD simulations, enabling the subsequent computation of the recirculation loss coefficient K_m. The model's efficiency predictions did not accurately reproduce the absolute efficiency

values due to neglected losses, but they provided a good estimation of the optimum flowrate coefficient when compared with experimental data.

In the present study, leakage effects were neglected and the coefficient k was assumed constant. Future work will address the role of mechanical clearances in the pump, aiming to refine the characterisation of their impact on internal flow dynamics and overall performance, while also examining the variability of k through CFD simulations.

7. REFERENCES

[1] M. Badami and M. Mura, "Theoretical model with experimental validation of a regenerative blower for hydrogen recirculation in a PEM fuel cell system," *Energy Conversion and Management,* vol. 51, pp. 553-560, 2010.

[2] M. Badami and M. Mura, "Setup and validation of a regenerative compressor model applied to different devices," *Energy Conversion and Management,* vol. 52, pp. 2157-2164, 2011.

[3] O. Siemen and J. Hinsch, "Teilweise beafschlagte umlaufpumpe mit dichtender, kreisender hilfsflüssigkeit". Patent DE413435C, 1920.

[4] Y. Senoo, "Theoretical research on friction pump," *Reports of the Research Institute for Fluid Engineering, Kyushu University,* vol. 5, pp. 23-38, 1948.

[5] T. J. Hopkins and L. R. Lazo, "Theoretical and experimental analysis of a regenerative turbine pump," *Bachelor of Science thesis, Massachusetts Institute of Technology,* 1953.

[6] W. A. Wilson, M. A. Santalo and J. A. Oelrich, "A Theory of the Fluid-Dynamic Mechanism of Regenerative Pumps," *Journal of fluid engineering,* vol. 77, no. 8, pp. 1303-1311, 1955.

[7] D. Dewitt, "Rational design and development of the regenerative pump," *PhD thesis, Massachusetts Institute of Technology,* 1957.

[8] F. Quail, T. Scanlon and M. Stickland, "Study of a regenerative pump using numerical and experimental techniques," *8th European Turbomachinery Conference,* 2009.

[9] J. Nejadrajabali, A. Riasi and S. A. Nourbakhsh, "Flow Pattern Analysis and Performance Improvement of Regenerative Flow Pump Using Blade Geometry Modification," *International Journal of Rotating Machinery,* vol. 2016, pp. 1-16, 2016.

[10] F. Quail, F. Scanlon and A. Baumgartner, "Design study of a regenerative pump using one-dimensional and three-dimensional numerical techniques," *European Journal of Mechanics - B/Fluids,* vol. 31, pp. 181-187, 2012.

Biographies



Alexandre Vucemilovic received the engineering diploma in Mechatronics and Robotics from ENSMM (Besançon, France) and the master's degree in Microsystems, Instrumentation and Robotics from the University of Bourgogne-Franche-Comté in 2022. He is currently working as a R&D engineer at moving Magnet Technologies (Besançon, France) and pursuing a Ph.D. at LIFSE, ENSAM (Paris, France), focusing on improving the understanding and modelling of regenerative pumps.



Florent Ravelet received the master's degree in marine engineering from ENSTA in 2002, and the philosophy of doctorate degree in Fluid Mechanics from Polytechnique in 2005, respectively. He is currently working as a Full Professor at the Laboratory LIFSE, Arts et Metiers Science and Technology. His research areas include turbulent flows, cavitation and turbomachinery.



Smaïne Kouidri is a Professor at Arts et Métiers Institute of Technology (ENSAM). He graduated in 1990 from Arts & Métiers ParisTech where he received his Ph.D. in 1993. He then became an associate professor at ENSAM until 2007, when he joined the Computer Sciences Laboratory for Mechanics and Engineering Sciences (LIMSI), a CNRS laboratory UPR 3251. He later became the director of the ENSAM Paris Campus (2021-2023) before joining the Laboratory of Fluid Engineering and Energy Systems (LIFSE). His research activities fall within the general area of internal flows in turbomachinery, aeroacoustics and thermoacoustics and his main research focus is the development and the implementation of models for reducing noise pollution and improving the efficiency of conversion energy systems such as Pumps, compressors and fans.



Christophe Espanet received his Ph.D. degree in Electrical Engineering from the University of Franche-Comté, France, in 1999. His doctoral research focused on the optimal design of permanent magnet brushless DC in-wheel motors. From 1999 to 2015, he served as Assistant Professor and later Full Professor at the University of Franche-Comté. He is currently the Scientific Director of the Sonceboz Group, where he leads research teams at Moving Magnet Technologies, developing innovative mechatronic solutions.

Navigating Electrification Challenges: Low Speed Behavior of External Gear Pumps

Nandakumar P S[1], Vishnudas Ramaswamy[1], Stefan Kock[2], Martin Laube[3]

[1]Bosch Rexroth India, Advanced Engineering External Gear Units, Bengaluru, India

[2]Bosch Rexroth AG, Advanced Engineering External Gear Units, Nürnberg, Germany

[3]Bosch Rexroth AG, Head of Product Area External Gear Units, Lohr a. Main, Germany

E-Mail: npa4ban@bosch.com, Ramaswamy.Vishnudas@boschrexroth.co.in, stefan.kock@boschrexroth.de, martin.laube@boschrexroth.de,

Abstract.

Electrification in mobile hydraulics, particularly for material handling applications like electric forklifts, demands external gear pumps capable of reliable operation at very low speeds, high pressures, and elevated temperatures. The low-speed requirement arises from the low-flow needs of the hydraulic system. However, the operational speeds required by material handling applications often fall below the current minimum speed limit for external gear pumps (typically 500 rpm). This can lead to insufficient lubrication of tribological contacts, excessive wear of pump components (e.g.: bearings and gears), and significant volumetric losses. Bosch Rexroth has addressed these low-speed challenges in external gear pumps through the development of specialized low-speed characterization parameters and ongoing research into various pump optimization methodologies. This paper is a comprehensive compilation of low-speed operation requirements and challenges (e.g.: volumetric losses and wear), low-speed pump performance characterization parameters available with Bosch Rexroth, and ongoing research to address low-speed challenges (e.g.: surface coating technology). The paper introduces and explains two parameters to characterize the low-speed behaviour of external gear pumps: “zero delivery speed” and “hydrodynamic speed limit”. Zero delivery speed defines the minimum operating speed of the pump at which the theoretical flow equals volumetric losses. Hydrodynamic speed limit describes the operating speed of the pump corresponding to lubrication regime change from mixed film lubrication regime to hydrodynamic lubrication regime.

The research topics explored in this paper offer practical solutions to optimize external gear pumps for low-speed operation in mobile machinery. These studies directly address the wear on bearing and gear faces, and volumetric loss associated with low-speed operation of external gear pumps.

Keywords. Electrification, External gear pump, Low-speed, Mobile machinery

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Bosch Rexroth is one of the world's leading suppliers of drive and control technologies, significantly impacting the development in the industrial hydraulics sector. The company's core focus areas span mobile and industrial applications, factory automation, and digitization and software development, providing comprehensive solutions that integrate electric drives and valves, linear motion technologies, and hydraulics. Bosch Rexroth's extensive product portfolio encompasses hydraulic pumps, motors, valves, cylinders, and power units, catering to diverse industrial and mobile machinery requirements.

Hydraulic systems are integral part of mobile machinery deployed in agriculture, material handling, construction, and mining, to convert energy into fluid power and then into mechanical energy, enabling various actuators to produce force or motion. A typical hydraulic system comprises of a power source (e.g.: batteries), drive (e.g.: servo motors), pump, control elements (e.g.: valves), actuators (e.g.: cylinders, hydro motors), auxiliary components (e.g.: filters, accumulators), and a working medium (e.g.: hydraulic oil). A hydraulic system can either have a mechanical-hydraulic drive system where an internal combustion engine is used to drive the pump, or an electric-hydraulic drive system where an electric motor is used to drive the pump. Hydraulic systems are currently undergoing transformational change towards electrification where an electric-hydraulic drive system replaces the conventional mechanical-hydraulic drive system, which ensures improved energy utilization and reduced air and noise pollution. This is primarily driven by the electrification trends in non-road mobile machineries [1]. The material handling sector is in the forefront of electrification and has undergone much electrification in last decade, especially smaller forklifts, and the trend is continuing to bigger forklifts and reach trucks.

Applications such as lifting at very low speeds or idle machine operation require hydraulic systems capable of generating low flow rates while keeping the working pressure on the desired level. Electric-hydraulic drive systems, which utilize electric motors, can operate effectively at low speeds [2] compared to internal combustion engines. To manage low-flow criteria, traditional valve control systems are often implemented. These systems deploy additional valves to regulate flow, but this approach increases system complexity and cost. Furthermore, the energy losses associated with these valves significantly reduce the overall hydraulic system efficiency [3].

An alternative is a pump-controlled system that employs a pump capable of operating efficiently at low speeds, thereby eliminating the need for additional valves. External gear pumps (EGPs) are widely used in hydraulic systems due to their low cost, and simple and compact design. However, continuous operation of EGPs for electric-hydraulic drive systems under low-speed conditions implies significant challenges, including operating points with reduced volumetric efficiency and wear of components due to inadequate lubrication.

The objective of this paper is to:

1. Present the challenges and requirements from low-speed mobile machinery applications emerging due to electrification of hydraulic systems.
2. Evaluate the current capabilities of Rexroth EGPs under low-speed conditions to highlight the challenges.

3. Elaborate potential design optimizations to meet future market requirements in the electrified mobile machinery sector.

This paper compiles a set of operational requirements for EGPs operating at low speeds, specifically derived from mobile material handling applications. The work then investigates the low-speed performance of these pumps. To evaluate the minimum operational speed limit and to identify the lubrication regime corresponding to operating conditions, two parameters will be described– "zero delivery speed" and "hydrodynamic speed limit". Variations of these results with operating conditions (e.g.: speed, pressure, and temperature) are presented to illustrate the application of these parameters in low-speed performance analysis. Subsequently, the key challenges inherent in low-speed EGPs operation and their potential causes are identified. Finally, this paper consolidates relevant research addressing these challenges.

The paper concludes by synthesizing the findings and identifying directions for future research. By providing a comprehensive analysis of low-speed behaviour and presenting concrete optimization strategies, this paper aims to promote the broader adoption of EGPs in a wider range of electrification applications, thus paving the way for more efficient and reliable solutions.

2. EXTERNAL GEAR PUMPS AND LOW-SPEED OPERATION: STATUS QUO

The previous section introduced the integration of EGPs into electric-hydraulic drive systems. This section will delve deeper into EGPs, exploring the applications, the challenges, and the specific requirements for their operation under low-speed conditions.

2.1. External gear pumps

EGPs are positive rotary-displacement pumps. Meshing of drive and driven gears creates a low-pressure region at the suction side to which working fluid enters and is carried along with gear teeth and squeezed out at the outlet side. **Figure 1(b)** is the exploded view of an EGP with the major components highlighted. Two pairs of bearing and sleeve, housing, drive and driven gear, and covers constitute an EGP. EGPs stand out for their:

- Simple, robust and compact design.
- High volumetric efficiency and power density.
- Broad operational speed capabilities (500-6000 rotations per minute (rpm)) depending on pump size and pressure (0-280 bar).
- Low cost and maintenance.

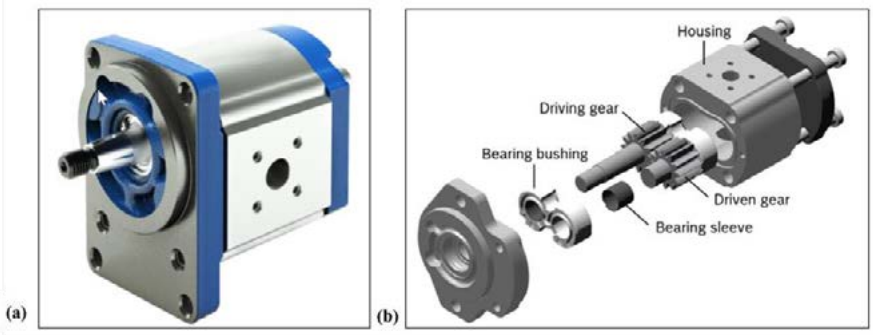


Figure 1: (a) Picture of a Rexroth external gear pump, (b) external gear pump components

Rexroth EGPs are widely used across diverse applications, particularly due to the low-noise operation achieved through their Silence and Silence Plus technologies, even under high-pressure conditions. While these pumps typically operate across a broad speed range, from 500-6000 rpm (for 1 cubic centimetre (ccm) pumps), many mobile and industrial applications specifically utilize pumps with displacements ranging from 7-28 ccm, where the operational speed range is often constrained to 500-4000 rpm. Anything below 500 rpm is treated as a low-speed requirement. The next section evaluates the general low-speed requirements from electrified mobile machinery applications.

2.2. Low-speed operation of external gear pumps: Application, Requirements, and Challenges

The demand for low-rotational-speed pumps is prominent in electrified material handling applications, which includes reach trucks, forklifts, scissor lifts, mobile cranes etc. **Figure 2** depicts a general speed-load diagram (or duty cycle) for EGPs in material handling application, where a significant portion of EGPs operation occurs at 100-300 rpm for short durations, even with intermittent pressures up to 280 bar. All these requirements exist within a demanding product lifecycle of 20000 operating hours for the pump. Apart from the mentioned conditions, it is well documented that the start-stop operations of the system pose significant threat to mechanical components like bearings [4]. EGPs have the capability to meet these requirements for short durations or limited load cycles, but continuous operation under low-speed conditions is to be investigated. The combination of low speed (below 500 rpm), high pressure (above 250 bar), and long and continuous operation (20000 operating hours) defines emerging low-speed requirements in the electric forklift market. These low speeds are crucial for tasks like exact positioning of loads, stacking, and automated movements, highlighting the need for highly precise movement and fine control of the forks. But most EGPs are not designed for continuous operation under such load collectives in the electrified forklift market; this is where active research is required.

Beyond material handling, the necessity for low rotational speed extends to applications like agriculture and construction demanding precise movements and control of actuators. Agricultural machinery requires precise planting and harvesting, where forward speed must be matched to implement function and ground conditions. Furthermore, road construction

equipment exemplifies this requirement, as paving and compacting demand consistent, low forward speeds for optimal material laydown and compaction density, effectively requiring machines to "crawl."

Figure 2 highlights the need for both low rotational speed operation and high rotational speed operation, under varying pressures, which are prevalent in material handling and agriculture applications that typically utilize fixed displacement pumps. Other pumps most used in mobile machinery applications include internal gear pumps and axial piston pumps. While internal gear pumps demonstrate comparable performance at lower speeds [5], EGPs surpass them in high-speed operation. Axial piston pumps, though capable, are generally costly and noisy to serve as a viable alternative to EGPs. Furthermore, axial piston pumps are more susceptible to oil contamination, posing an additional challenge compared to the more robust EGPs. Among the available options, EGPs emerge as the most suitable choice for the applications that require both low-speed and high-speed capabilities, specifically due to their cost effectiveness and simple construction.

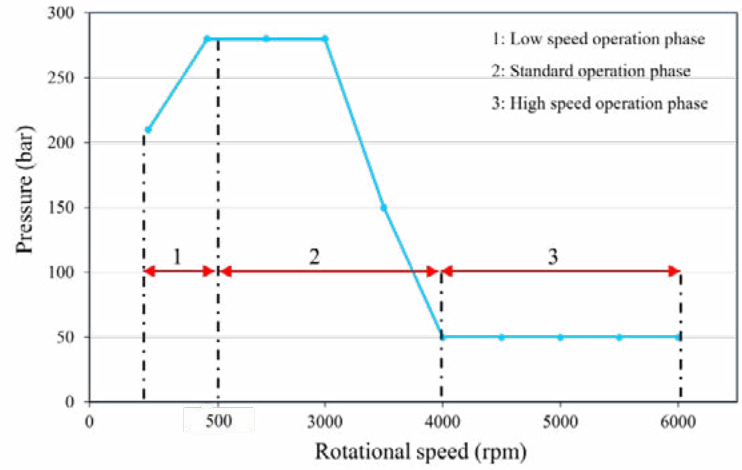


Figure 2: Example of a specific speed-load diagram for external gear pumps in material handling application

In short, EGPs have considerable potential to meet new requirements from mobile machinery application. Electrification of mobile machinery require pumps with both low-speed and high-speed capability. Pumps are expected to operate continuously under low-speed and high-pressure conditions. This brings new challenges and opportunities for external gear pump. This opens the possibility and requirement to study and optimize the low-speed behaviour of EGPs.

3. LOW-SPEED OPERATION: CHALLENGES

The preceding section focused on EGPs, and requirements for low-speed operation. The subsequent paragraphs will detail the specific challenges that arise during low-speed operation. Two major challenges are identified through literature review, field experience, and pump tests: volumetric losses inside the pump and wear of the pump components.

3.1. Volumetric losses

Volumetric losses describe the flow loss from the pump through internal leakage paths. Major internal leakage paths include the gap between housing and gear teeth tip, through gear flank contact region, between bearing face and gear face and between bearing contact faces. Internal leakage paths reduce the pump's net flow rate, thereby decreasing volumetric efficiency. Higher volumetric loss limits the minimum operating speed of the pump as the theoretical flow rate of the pump will equal the volumetric loss of the pump.

3.2. Wear of components

It is a fundamental principle that the flow rate from an EGPs is directly proportional to its operating speed. Consequently, at very low speeds, the resulting flow rate is significantly reduced. Low flow rate affects the lubrication and cooling of pump components. Apart from low flow rate, the hydrodynamic force generated by the rotation of the shaft is not sufficient to ensure a hydrodynamic lubrication regime between tribological contacts (explained in next section). During low-speed operation, the pump will be operating in mixed film lubrication regime which accelerates the wear of components. Along with lubrication, the pump working medium (oil) carries the chips generated due to the surface contacts inside the pump and prevents further wear or crack formation in pump components. Consequently, low flow rate leads to component wear, with bearings and gears being particularly susceptible. Abrasive and adhesive wear are a well-documented tribological phenomenon [6], which significantly diminishes the endurance life of these components and of the pump. Especially, continuous operation under low-speed conditions has been observed to cause partial removal of the polymer layer from polymer sleeves.

Apart from volumetric losses and wear, low flow rate leads to an elevated local pump temperature (i.e., the temperature of the working fluid inside the pump) as the cooling rate is directly proportional to flow rate. This elevated temperature further reduces the working fluid's viscosity or can cause permanent damage, such as oil degradation [7] thereby intensifying the lubrication deficiency. This creates a destructive cycle that continues until noticeable damage to the pump occurs**.**

In conclusion, most challenges observed during the low-speed operation of EGPs are due to volumetric losses, and wear of the components.

4. VOLUMETRIC LOSS: EVALUATION AND CHARACTERIZATION METHODOLOGY

The preceding section underscores challenges during low-speed operation and the need to investigate volumetric loss of the pump. Addressing these requirements requires testing capabilities and matured evaluation methodologies. The following section explains:

- Testing capability to measure and calculate flow rate and volumetric losses (Volumetric performance investigation test)
- Evaluation of low-speed performance (Volumetric loss characterization, zero delivery speed (hereafter ZDS))

Volumetric efficiency (η_{Vol}) describes the capability of the pump to deliver a flow rate close to its theoretical displacement, indicating the quantity of internal fluid leakage (Q_l in litre

per minute (lpm)). Volumetric efficiency is calculated from the theoretical flow rate (Q_{th}, in lpm) and real or measured flow rate (Q_{real}, in lpm). The measured flow rate is the flow rate generated by the EGPs during the operation. Theoretical flow rate is calculated from rotational speed (n, in rpm) and theoretical flow (V_{th}, in litre per revolution) as:

$$Q_{th} = n \times V_{th} \tag{4.1.1}$$

Volumetric efficiency (η_{Vol}) is calculated as:

$$\eta_{Vol} = \frac{Q_{real}}{Q_{th}} \times 100\ \% = \left(1 - \frac{Q_l}{Q_{th}}\right) \times\ 100\ \% \tag{4.1.2}$$

Volumetric losses reduce the real flow rate of an EGPs, leading to elevated temperatures and reducing the flow available for operating the working hydraulic components like hydraulic cylinders. These losses are approximately independent of rotational speed but highly dependent on operating pressure [8]. At low-speed operation, the real flow rate from the pump is inherently small and volumetric losses further diminish the performance of the pump. Therefore, accurately monitoring and characterizing a pump's volumetric losses is crucial for understanding its low-speed behaviour, necessitating mature testing and characterization methodologies.

4.1. Volumetric loss: Testing facility

Test benches capable for the measurement and recording of operating speed, operating pressure, flow rate, torque loss, and suction and outlet temperature are used to calculate results including volumetric efficiency, and volumetric losses. The same test benches are used to derive ZDS and hydrodynamic speed limit (hereafter HSL). The test bench includes a servo motor (e.g.: 45 kW for 16 ccm pump), a coupler to attach pump with motor, flow meter, rotational speed sensor, suction and outlet temperature sensor, pressure relief valve and an external leakage sensor. The test benches can test pumps from wide range of sizes (e.g.: 1 ccm to 45 ccm), wide range of pressure (e.g.: 0 to 300 bar) and temperature (e.g.: 100 °C). Testing is conducted on different batches and different number of pumps to ensure statistical confidence.

4.2. Volumetric loss characterization: Variation with pressure and temperature

Volumetric losses, flow rate, and volumetric efficiency are derived from the test bench data. **Figure 3(a-c)** presents the volumetric losses, flow rate and volumetric efficiency characteristics for different temperatures and pressures (P), keeping the operational speed constant. From **Figure 3(a)** it is observed that the maximum volumetric loss occurs in the medium pressure range. It is expected that volumetric loss of an EGPs will increase with increase in pressure, because of the higher-pressure gradient between suction and pressure side. However, the volumetric losses reduce from medium pressure to high pressure. This is primarily due to the process of "run-in", which every pump undergoes during initial operation. Run-in is the process of deepening the leakage path between gear teeth tip and housing wall during the initial operation of the pump in high pressure condition. Run-in occurs due to the bending and misalignment of gears (towards suction side) at high pressure operation. This causes excessive material removal from the suction side housing wall, and closing the leakage path for selected "run-in" pressure (usually 10% above the maximum operating pressure). For operation under lower pressures of the pump, volumetric efficiency will be reduced due to the enlarged leakage path. However, as the gear starts to bend or

misalign towards suction side due to higher pressure, the leakage path created by run in process is being closed and this phenomenon enhances the volumetric efficiency. The closing of above-mentioned leakage gap happens only for "Run-in" pressure. As a drawback, the flow and volumetric efficiency will be maximum only at run-in pressure compared to other pressure conditions. This is the reason for smaller volumetric efficiency and higher volumetric losses at mid-pressure range. The closure of run-in track due to high pressure is called passive radial gap compensation.

Figure 3(a) also shows the impact of temperature on volumetric losses. Comparing the results for 50 °C and 80 °C, a significant increase in volumetric loss is observed with increase in temperature. The maximum volumetric loss at 80 °C is more than twice that at 50 °C. As temperature increases, viscosity of oil decreases and reduction in viscosity leads to higher volumetric losses.

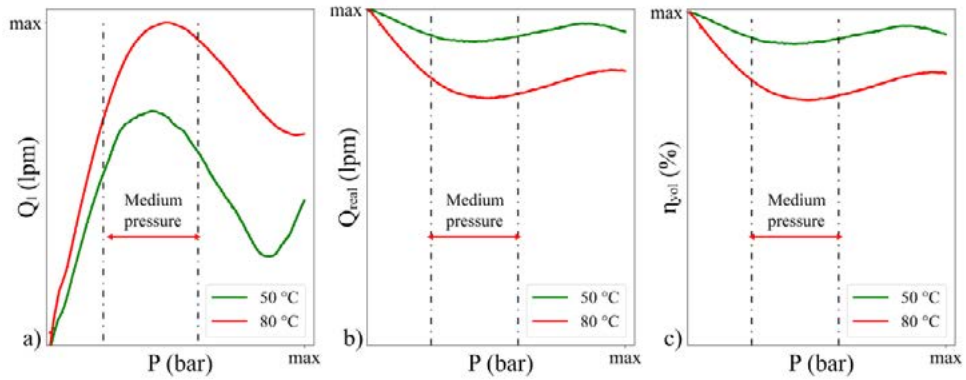


Figure 3: Variation of (a) volumetric flow loss, (b) flow rate, and (c) volumetric efficiency with pressure and temperature

Figure 3(b) and Figure 3(c) discusses the flow rate variation and volumetric efficiency variation with temperature and pressure. Flow rate and volumetric efficiency attains a minimum value at medium pressure range and maximum is observed during low- and high-pressure operation. With respect to temperature, a considerable drop observed at elevated temperature cases. The findings are aligned with volumetric loss findings as volumetric losses are maximum at medium pressure cases and minimum at low- and high-pressure cases, and volumetric loss increases with temperature. The observed results clearly indicate that medium pressure and high temperatures are the critical operating conditions for EGPs. These behaviours are valid for all modern external gear units using passive radial gap compensation over "run-in track".

4.3. Definition and calculation of zero delivery speed (ZDS)

Volumetric loss is important as it restricts the minimum operational speed of the pump. Understanding the minimum possible speed by the EGPs is required to define the operating boundary conditions. In this context, ZDS is the rotational speed at which the pump is operating without any measurable flow for a defined pressure and temperature. ZDS helps to identify the lower speed boundary for an EGP. Below ZDS, the volumetric losses from the pump equals flow from the pump and hence the pump is unable to produce any measurable flow at the outlet.

The calculation of ZDS for a given temperature and pressure, is derived from the flow rate - speed curve, for a fixed pressure and temperature condition as depicted in **Figure 4**. The speed (x-axis) corresponding to zero flow (y-axis) is calculated from the experimental data. The intercept point represents the state of no flow though the rotational speed of the pump is not zero.

ZDS is proportional to the flow capabilities of the pump. A pump with higher outlet flow (or with low volumetric losses) can provide flow at lower speeds. Flow rate and thereby ZDS is influenced by both operating temperature and pressure.

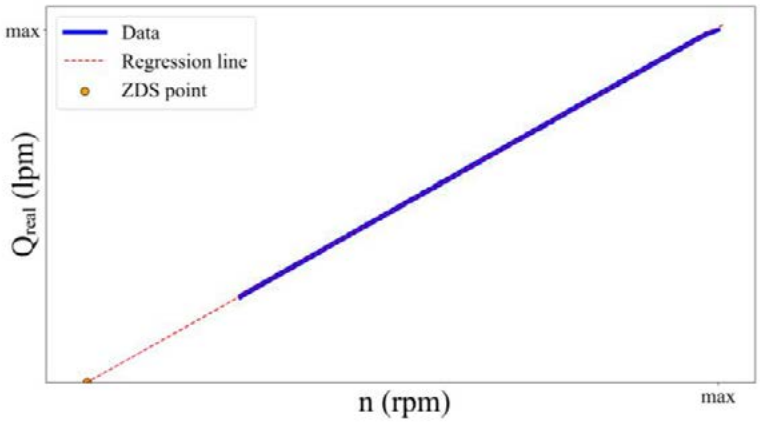


Figure 4: Calculation zero delivery speed from the flow versus speed curve

Figure 5 presents the ZDS variation with pressure for two temperature cases. As depicted in **Figure 5**, the variation of ZDS with pressure exhibits a parabolic profile, with its peak observed in the "medium pressure" or "centre pressure" range. This behaviour is consistent with the pump volumetric efficiency, flow rate, and volumetric loss characteristics, as shown in **Figure 3(a-c)**. The influence of temperature on ZDS is also presented in **Figure 5**. An increase in operating temperature leads to a reduction in flow rate, consequently elevating the ZDS. **Figure 5** conveys that at medium pressure and high temperature cases, pump should be run at higher speeds to have flow from the pump.

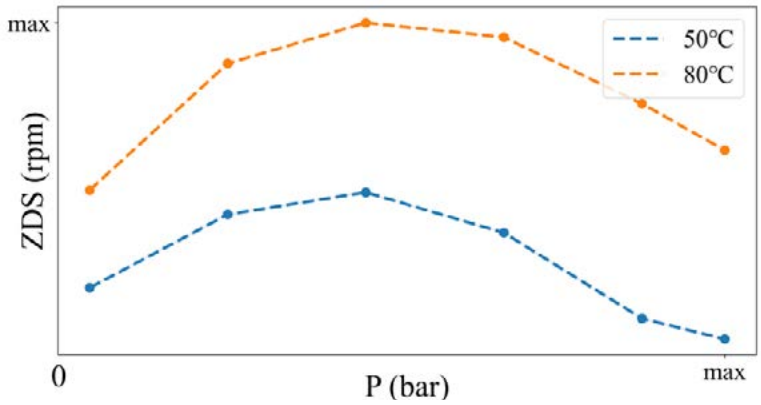


Figure 5: Variation of zero delivery speed with pressure and temperature

5. WEAR: EVALUATION AND CHARACTERIZATION METHODOLOGY

Apart from volumetric losses, wear is identified as a challenge during low-speed operation. The following section details about the characterization and evaluation methodology for wear.

Major tribological contacts of a pump include that between housing and gears, gears and bearings and between shaft and sleeve. During operation, the pump components and these tribological contacts are immersed in working oil which acts as the lubrication medium. Depending on the operating parameters like rotational speed, pressure, oil viscosity and temperature, a pump could be operating under boundary lubrication regime, mixed lubrication regime, or hydrodynamic lubrication regime. Hydrodynamic lubrication regime is characterized by a thick lubrication oil film between tribological contacts and ensures minimum contact wear during operation, while mixed film lubrication regime is characterized by thin film lubrication oil film, and metal to polymer contact between tribological contacts. Components operating under mixed film lubrication could encounter high wear and lubrication issues. The transition from mixed film lubrication to hydrodynamic lubrication is identified as the transition speed point, and adapted Stribeck curve is used to identify the transition point and lubrication regime [9]. Stribeck curve is usually deployed for bearing lubrication regime identification, but the concept is, under simplification and assumptions, applicable to pumps even though a pump consists of multiple tribological contacts including gear and bearings. The Stribeck curve for a pump discusses about the overall torque loss from the pump, which includes torque loss through all the tribological contact and viscos loss due to working oil. Transition speed from the Stribeck curve describes the minimum torque loss point for the pump, it doesn't mean all the components of the pump will be working in the hydrodynamic lubrication regime if the operating speed is above transition speed. At low-speed operating conditions, the pump speed could be below the transition speed point, and this will induce wear of the components and eventual failure of the pump. The load cycle capability of the pump for continuous operation under low-speed is restricted by the mixed film lubrication regime and the wear associated with the lubrication regime.

Above sections detailed about the importance of transition speed point and wear associated with mixed film lubrication regime. Extracting transition speed point from the Stribeck curve for the pump presents practical challenges. To address this, the paper is proposing a model that simplified the Stribeck curve into a function of torque loss and rotational speed by fixing operating pressure and temperature. The transition speed between mixed film lubrication regime and hydrodynamic lubrication regime is identified from this curve. This speed is defined as HSL. **Figure 6** represent a simplified Stribeck curve with lubrication regime and HSL illustrated.

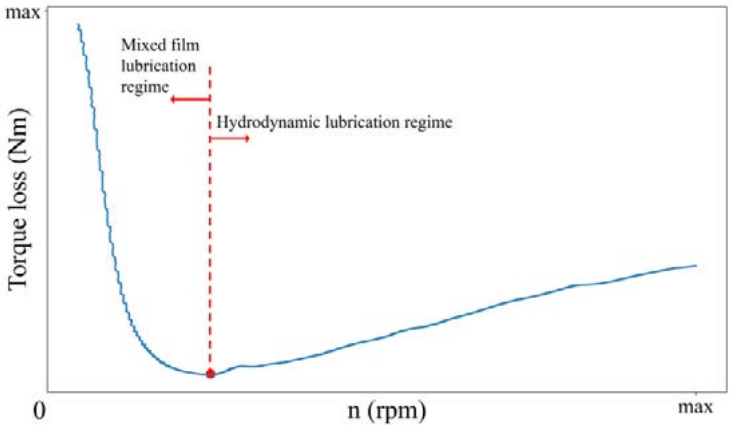


Figure 6: A simplified version of Stribeck curve with lubrication regime distinguished

5.1. Definition and calculation of hydrodynamic speed limit (HSL)

HSL is a parameter to determine the speed corresponding to minimum torque loss. Minimum torque loss directly correlates to minimum coefficient of friction (friction due to surface contact and viscosity of oil) and thereby to the speed at which the wear is minimum. Hypothetically pumps operating above the HSL could be under hydrodynamic lubrication regime and have better durability life in the long run. However further research is required to evaluate this hypothesis. It is possible that the pump components will be operating under different lubrication regimes even though the overall torque loss from the pump is corresponding to hydrodynamic lubrication. **Figure 7** shows a simplified Stribeck curve for EGPs. **Figure 7** illustrates HSL value for different pressure cases. **Figure 7** show a steep increase in torque loss value as operating speed falls below HSL. Continuous operation of pump under mixed lubrication regime will significantly affect the pump durability life, and for continuous operation, HSL is the minimum speed condition for the pump.

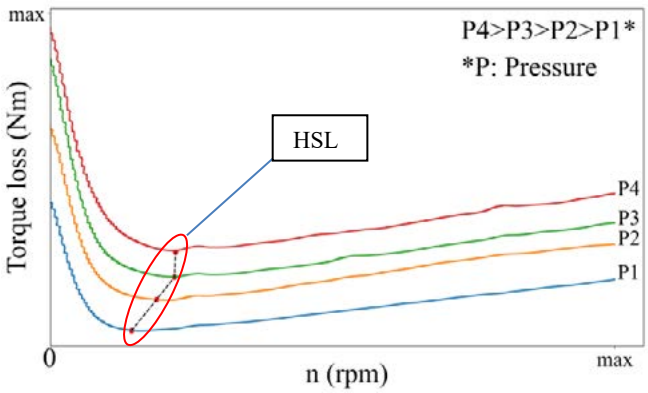


Figure 7: Torque loss versus speed curve for different pressures

HSL is a function of the operating pressure, temperature and viscosity of working medium. **Figure 8** presents the variation of HSL with pressure for a different temperature case. As the pressure increases, the hydrodynamic force required to generate an oil film between the

components also increases. The shaft rotation speed and hydrodynamic force generated are proportional. At higher pressure cases, higher hydrodynamic force is required to ensure hydrodynamic lubrication and operating speed is required to generate this force. Hence the HSL increases with pressure.

Figure 8 illustrates impact of temperature on HSL. An increase in operating temperature leads to a reduction in fluid viscosity, consequently reducing the generated hydrodynamic force and thus elevating the HSL required for hydrodynamic lubrication.

The last two sections described the various low-speed characterization parameters available including testing capabilities. It is observed that the ZDS and HSL ranges from 400 to 750 rpm (for 14 ccm external gear pump with passive radial gap compensation). However, the low-speed applications demand operational speed to be below 400 rpm. The decision if the pump is suitable for low-speed operation condition need to be evaluated carefully by considering the requirements and results from parameters like ZDS and HSL.

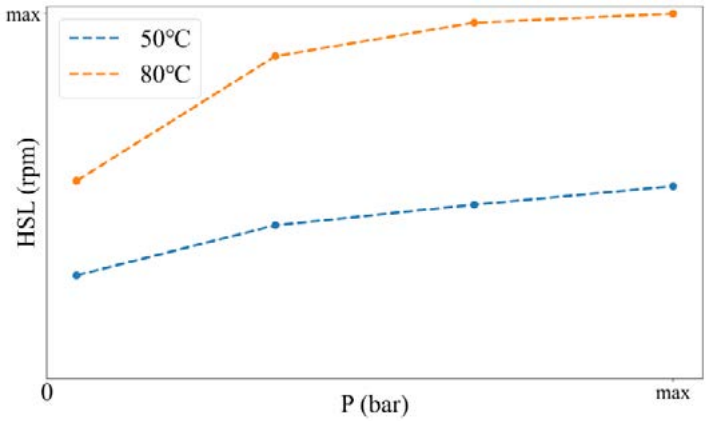


Figure 8: Variation of hydrodynamic speed limit with pressure and temperature

To study the wear performance and durability life of pump and components under low-speed operation, specific tests with focus on continuous operation of the pump and its components are developed. These tests are described in the pump durability testing section.

5.2. Wear measurement: Pump durability test

Pump durability tests are used to study the performance of the pump during and after continuous operation. These tests help to evaluate the wear on components, and variation in performance of the pump with continuous operation. Depending on the study requirements, a loading cycle is formulated for every test. This loading cycle will be executed for long periods of time (e.g.: 1000 hour) after which the pump will undergo through evaluation including performance, and tribological assessments. Tests are completed using a volumetric performance investigation test bench. Pressure, flow, external leakage and torque are continuously monitored and recorded for further analysis.

All the pumps that undergo durability test need to be investigated for their tribological parameters, including measurement of wear and microstructural analysis.

6. LOW-SPEED OPERATION: OPTIMIZATION MEASURES

Advanced research into the low-speed performance of EGPs is ongoing at scientific communities. These research addresses two critical issues: optimizing the pump to reduce volumetric losses, and enhance the wear robustness and reduce friction between components. The subsequent paragraphs detail some of the key research initiatives.

6.1. Pump optimization measures to reduce volumetric losses

High volumetric losses directly contribute to high ZDS for a pump. This limits the minimum operating speed for the pump and forces the pump to operate at higher rotational speeds. Following section compiles the research on this.

6.1.1. Precise bearing technology

Polytetrafluoroethylene (PTFE) based metal-polymer sleeves currently represent the state-of-the-art for radial bearings in EGPs. However, these metal-polymer sleeves exhibit large inner diameter tolerance fluctuation, making it challenging to achieve tight dimensional control. This difficulty is partly due to the inability to machine polymer sleeves post press-fitting into the bearing housing. Furthermore, recent research has highlighted the adverse environmental impacts associated with PTFE products.

Alternative solutions, such as bi-metal sleeves and Polyether Ether Ketone (PEEK) sleeves are actively evaluated. These offer the distinct advantage of allowing post-installation machining, enabling much tighter control over the inner diameter tolerances of the bearing. Achieving optimized clearance reduces the misalignment of gears [10];thereby mitigating stress on both the shaft and the sleeve, and also preventing creation of larger leakage paths between gear teeth tip and housing. Precise bearing technology can contribute to lowering the volumetric losses and thereby lowering the ZDS of the pump.

6.1.2. Radial gap compensation

Volumetric loss poses a significant challenge during low-speed pump operation, largely due to leakage across the radial gap between the housing and gear teeth tips. To mitigate this, novel designs actively reduce radial gap leakage. One promising approach involves incorporating a radial gap compensation element within the pump. This concept, state-of-the-art in internal gear pumps, dynamically adjusts the clearance between the gear teeth and the housing wall in response to operating pressure, thereby minimizing the leakage path [11]. **Figure 9** illustrates a proposed design featuring a radial gap compensation lid integrated into the pressure-side housing wall [12]. This lid is engineered to automatically adjust its position based on varying operating pressures. During high-pressure operation, the lid is pushed towards the gears, effectively reducing the clearance and thus dynamically controlling and minimizing volumetric loss through this pathway. Numerous iterations of this concept explore the optimal shape, position, and impact of the lid on internal leakage.

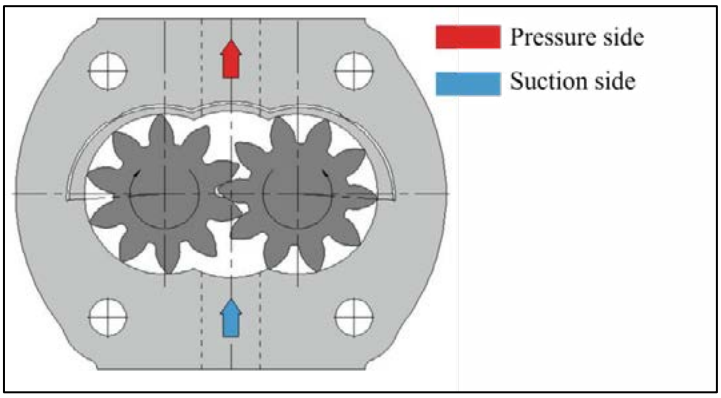


Figure 9: An example for radial gap compensation technology [12]

6.2. Enhancing the wear robustness and reduce friction between components

Enhanced wear robustness and reduced friction are crucial for improving both mechanical performance (e.g.: lower torque loss) and overall pump durability (e.g.: extended pump durability life). These improvements can be achieved through various methods, ranging from material property enhancements to advanced coating technologies. The subsequent section focuses on surface coatings that have demonstrated superior tribological properties – such as reduced wear, lower friction coefficients, and enhanced hardness – as documented in the literature.

6.2.1. Surface coatings for enhanced wear robustness of hydraulic components

Surface coatings are critical in tribology for reducing friction, mitigating wear, and enhancing the overall tribological performance of mechanical components [13]. For low-speed operation, wear on critical components like bearings and gears remains a significant challenge. Various coating materials, including Diamond-Like Carbon (DLC), Molybdenum Disulfide (MoS2), and Nickel-Molybdenum-Phosphide (Ni-Mo-P), are known to improve the tribological properties of coated parts [14] [15].

Specifically, applying DLC coatings to gears is documented to reduce friction at interfaces such as the gear-to-bearing contact and between the gear shaft and bearing sleeve. This friction reduction consequently minimizes torque loss and mitigates wear on these critical components. Moreover, reduced friction is expected to reduce the wear, enabling more prolonged and reliable operation under challenging low-speed conditions.

7. CONCLUSION

The electrification of mobile hydraulic applications, particularly in material handling, requires efficient electric-hydraulic systems capable of low-flow operation without traditional valving inefficiencies. This demands that EGPs operate effectively at low speeds, directly enhancing system efficiency and reducing costs. However, continuous operation (20000 operating hours) of these pumps at low speeds (< 500 rpm), high pressures (up to 280 bar), and elevated temperatures (80 °C) presents significant challenges, including volumetric losses and wear on bearings and gears.

This paper systematically addresses these requirements and challenges. The investigation utilized Bosch Rexroth's specific low-speed parameters: "zero delivery speed (ZDS)" and "hydrodynamic speed limit (HSL)". ZDS, a critical determinant of minimum pump operation, is significantly affected by temperature and pressure, increasing with temperature and following a non-linear (parabolic) curve with pressure. The HSL is crucial for minimizing torque loss and reducing component wear; operating below this limit drastically increases torque loss. Similar to ZDS, the HSL also rises with both temperature and pressure.

To overcome these low-speed challenges, the paper reviews and compiles practical, research-backed solutions. Specialized surface coatings are shown to directly mitigate wear, while advanced bearing technology and radial gap compensation offer promising approaches to minimize volumetric losses.

In conclusion, this work synthesizes the essential requirements and operational challenges of low-speed EGPs. It further highlights Bosch Rexroth's capabilities for their thorough investigation and identifies practical solutions for optimizing performance. Integrating these insights provides a clear path for EGPs to reliably and efficiently meet the stringent low-flow demands of electrified mobile hydraulic applications. This advancement will enable more sustainable and cost-effective electrified material handling systems.

8. REFERENCES

[1] A. Lajunen, P. Sainio, L. Laurila, J. Pippuri-Mäkeläinen and K. Tammi, "Overview of powertrain electrification and future scenarios for non-road mobile machinery," *Energies,* vol. 11, no. 5, p. 1184, 2018.

[2] M. d. F. Ramos, D. A. d. L. Brandao, D. P. V. Galo, B. d. J. Cardoso Filho, I. A. Pires and T. A. C. Maia, "A Study on the Performance of the Electrification of Hydraulic Implements in a Compact Non-Road Mobile Machine: A Case Applied to a Backhoe Loader," *World Electric Vehicle Journal,* vol. 15, no. 4, p. 127, 2024.

[3] L. J. Love, E. Lanke and P. Alles, "Estimating the impact (energy, emissions and economics) of the US fluid power industry," Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2012.

[4] N. D. S. A. Santos, V. R. Roso and M. T. C. Faria, "Review of engine journal bearing tribology in start-stop applications," *Engineering Failure Analysis,* vol. 108, p. 104344, 2020.

[5] M. Rundo, "Models for flow rate simulation in gear pumps: A review," *Energies,* vol. 10, pp. 1996-1073, 2017.

[6] F. Du, D. Li, X. Sa, C. Li, Y. Yu, C. Li, J. Wang and W. Wang, "Overview of Friction and Wear Performance of Sliding Bearings," *Coatings,* vol. 12, no. 9, p. 1303, 2022.

[7] B. Novaković, L. Radovanović, N. Zuber, D. Radosav, L. Djordjević and M. Kavalić, "Analysis of the influence of hydraulic fluid quality on external gear pump performance," *Eksploatacja i Niezawodnosc - Maintenance and Reliability,* vol. 24, no. 2, pp. 260-268, 15 03 2022.

[8] P. Michael, H. Khalid and T. Wanke, "An Investigation of External Gear Pump Efficiency and Stribeck Values," *SAE Technical Papers,* vol. 8, September 2012.

[9] M. Kalin, I. Velkavrh and J. Vižintin, "The Stribeck curve and lubrication design for non-fully wetted surfaces," *Wear,* vol. 267, no. 5-8, pp. 1232-1240, 2009.

[10] P. Szwemin and W. Fiebig, "The Influence of Radial and Axial Gaps on Volumetric Efficiency of External Gear Pumps," *Energies,* vol. 14, no. 15, p. 4468, 2021.

[11] P. Osiński and U. Warzyńska, "FEM Strength Analysis of Circumferential Compensation with Integrated Lips in Gear Pumps," *Energies,* vol. 15, no. 7, p. 2691, April 2022.

[12] P. Osiński, "Strength Calculation Methodology for Circumferential Backlash Compensation with Integrated Lips," in *Advances in Hydraulic and Pneumatic Drives and Control 2020*, Cham, 2021.

[13] I. P. Okokpujie, L. K. Tartibu, H. O. Musa-Basheer and A. O. M. Adeoye, "Effect of Coatings on Mechanical, Corrosion and Tribological Properties of Industrial Materials: A Comprehensive Review," *Journal of Bio- and Tribo-Corrosion,* vol. 10, no. 1, p. 2, 2023.

[14] A. K. Sharma, N. Kumar, A. K. Das and M. Kumar, "Ni-Mo-P coatings on thrust plate of a gear pump to enhance its mechanical, tribological, and corrosion resistance properties," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* vol. 238, no. 8, pp. 1196-1211, July 2023.

[15] D. K. Rajak, A. Kumar, A. Behera and P. L. Menezes, "Diamond-Like Carbon (DLC) Coatings: Classification, Properties, and Applications," *Applied Sciences,* vol. 11, no. 10, p. 4445, May 2021.

Biographies

Photograph

Nandakumar P S received the bachelor's degree in Mechanical engineering from Vellore Institute of Technology in 2020, and the master's degree in Energy engineering and management from National Institute of Technology-Calicut in 2024. He is currently working as an Engineer at the Advanced engineering department of Bosch Rexroth. His research areas include hydraulic systems with a focus on external gear pumps and thermo-fluid interaction.



Vishnudas Ramaswamy received the bachelor's degree in Mechanical engineering from Amrita School of Engineering in 2010, the master's degree in Fluids Engineering from the National Institute of Technology Allahabad in 2012 , and the philosophy of doctorate degree in Fluid Mechanics from the Indian Institute of Technology Madras in 201, respectively. He is currently working at the Advanced Engineering Department of Bosch Rexroth India Pvt. Ltd. His research areas include fluid power systems, fluid mechanics, computational fluid dynamics, and multiphase flow.



Stefan Kock received his Mechanical Engineering degree from the Technical University of Dresden, Germany, and his Ph.D. from RWTH Aachen University, where his research focused on wind turbine drive trains and torque measurements. He is currently Team Leader for Advanced Engineering for External Gear Units at Bosch Rexroth. His research areas include drive train technology, mechatronic systems, fatigue assessment, finite element method (FEM), and fluid power drives.



Martin Laube received his Diploma degree as Dipl.-Ing. (univ.) in mechanical engineering from the Technical University of Munich (TUM) in 1996. He specialized during his studies in the field of Design and Development with a focus on machine elements, vehicle technology and fluid technology. He started his industrial career in 1996 as a development engineer for railway systems at Mannesmann Rexroth, Lohr a. Main. After holding various positions in sales and engineering for industrial applications at Mannesmann Rexroth and Bosch Rexroth (since 2001), he assumed the position of Engineering Director for Hydraulic Power Units, Manifolds and Systems at Bosch Rexroth AG in 2013. Since July 2024, he has been Vice President and Head of Product Area External Gear Units in the Business Unit Mobile Components at Bosch Rexroth AG.

Geometry-Aware Physics-Informed Neural Network: A Framework for Solving the Reynolds Equation with Starvation for Lubricated Contacts

Faras Brumand-Poor, Artem Konstantinov, Darren Lam Ming Hui, Eric Guntermann, Jakob Wirtz, Katharina Schmitz

RWTH Aachen University, Institute for Fluid Power Drives and Systems (ifas), faras.brumand@ifas.rwth-aachen.de

Abstract.

Machine learning has gained increasing attention in fluid power applications, particularly in condition monitoring and control. A promising development in this field is physics-informed machine learning (PIML), and specifically physics-informed neural networks (PINNs), which integrate governing physical laws directly into the training process. Unlike traditional networks, PINNs do not require data collection and can significantly accelerate computationally demanding simulations. This capability is valuable for studying lubricated contacts, such as seals in pneumatic valves, which operate pneumatic systems by controlling flow rate and pressure, and where accurate yet efficient modeling is essential. Seals are critical for maintaining pressure and preventing leakage; their malfunction can lead to costly failures, downtime, and environmental risks. However, seal behavior remains difficult to characterize due to the coupled hydrodynamic (HD) and elastodynamic (ED) effects within the sealing interface. Conventional elastohydrodynamic lubrication (EHL) simulations based on fluid-structure interaction models provide accurate insights. However, they are computationally expensive, limiting their usability for real-time or large-scale investigations. To address these challenges, this work extends a validated PINN framework for modelling lubrication behavior inside a pneumatic seal by introducing an unsupervised learning autoencoder (AE) to compress complex gap geometries into low-dimensional latent representations. The framework, consisting of a PINN, a loss balancing, and a hyperparameter tuning algorithm, has already been validated for modelling hydrodynamic lubrication simulations with transient phenomena and cavitation. The addition of the AE to the framework yields a geometry-aware PINN (GAPINN) capable of fast, accurate, and geometry-adaptive analysis of sealing systems, thereby enhancing the applicability of PIML in tribological research.

Keywords. Physics-Informed Machine Learning, Hydrodynamic Lubrication, Geometry-Awareness, Autoencoder, Unsupervised Learning, Physics-Informed Neural Networks.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Machine Learning has emerged as a promising class of numerous algorithms that can enhance different applications in fluid power, e.g., condition monitoring and control [1, 2].The field of physics-informed machine learning, with its novel class of data-free solvers, the so-called physics-informed neural networks, has shown great potential for accelerating distributed-parameter simulations with the integration of physical laws in the network and without the time-consuming collection of data [3–6]. Accelerating simulation, thus decreasing computational time, is especially relevant for investigating lubricated tribological contacts. One prominent example of such systems is a seal inside a housing and a counter surface, such as a pneumatic valve spool and a lubricant between the seal and the counter surface. Seals are usually less costly than other fluid power components, especially in pneumatic valves, but are crucial for the system. A non-functional seal might result in operational disruptions or a system failure, leading to expensive repairs, delayed production time, and possibly environmental hazards. Seals in fluid power systems are mainly responsible for maintaining operational pressure and preventing leakage. However, the operation of seals generates friction, causing a loss of overall efficiency, especially in pneumatic valves, which exhibit a substantial adverse effect due to increased friction on their dynamic behavior. Despite their significant role in fluid power components and systems, seal behavior is not completely understood due to the complex interactions of hydrodynamic and elastodynamic phenomena [7].

Elastohydrodynamic lubrication simulations or experiments are required to understand these tribological interfaces better. Experiments are rather time-consuming and costly due to the high measurement accuracy needed to obtain valuable information [8]. On the other hand, EHL simulations represent a different approach to obtaining a deeper understanding of these tribological systems by solving the Reynolds equation to compute the hydrodynamic pressure distribution and the deformation within the sealing interface. These EHL simulations primarily consist of a fluid-structure-interaction (FSI), which combines the hydrodynamic pressure buildup of the lubricant with a Finite Element Analysis (FEA) for the deformation. These FSI models provide accurate solutions but exhibit high computational costs [9, 10], which prevents real-time application or a detailed investigation of a significant number of operation points of a sealing system.

These limitations prevent optimal exploitation of sealings for fluid power systems and demand novel approaches for accurate and fast evaluation of different sealing system scenarios. PINNs represent a promising alternative to traditional EHL simulations since they incorporate the same underlying physical equations in the network training process and further solve the investigated scenario in a fraction of the time required by the EHL simulation [11]. Integrating physical laws also dismisses the necessity of data, which is not easily obtained for lubricated contacts, required for regular neural networks.

The first published work applying PINNs to the Reynolds equation was presented by Almqvist in 2021, demonstrating a fundamental implementation of PINNs for this equation [12]. More advanced approaches were later introduced by Zhao et al. and Li et al., who addressed the two-dimensional Reynolds equation for linear sliders and gas bearings, respectively [13, 14]. Rom was the first to develop a PINN capable of solving the stationary Reynolds equation in combination with the Jakobsson–Floberg–Olsson (JFO) cavitation model [15]. Building on this progress, Cheng et al. implemented a PINN framework that

solves the Reynolds equation using either the JFO or the Swift–Stieber (SS) cavitation model, incorporating three different multi-task learning strategies to balance the respective loss terms [16]. Recent work focused on the use of hard constraints to deal with the problem of cavitation [17, 18]. Further research was conducted on the application of hybrid PINNs, combining data-driven approaches and physical laws for the solution of the Reynolds equation [22, 23]. The prior work focused on stationary scenarios; Ramos investigated PINNs for solving the Reynolds equation with time-dependent height changes in journal bearings [25]. Rimon implemented a PINN to solve an elastohydrodynamic lubrication simulation using a simplified stationary one-dimensional Reynolds equation, without cavitation, and incorporating linear pressure-dependent sealing deformation based on the Lamé equation [27]. The mentioned studies focus either on stationary cavitation problems or dynamic problems without cavitation. Furthermore, most developed solutions do not focus on a PINN's ability to understand the investigated geometry. In the work of Rom, the eccentricity of the journal bearing was provided as an input to the PINN, thus adding some geometry-awareness to the PINN. To sum up, the current state of the art in PINNs for modelling hydrodynamic lubrication solves variants of the Reynolds equation; however, there has been little focus on implementing geometry-awareness in PINNs for such applications. The application of geometry-aware PIML has already been successfully used in deformation by implementing geometry-aware PINNs [32, 33] , or directly developing physics-informed graph neural networks (PIGNNs) [34-36].
In this work, a previously validated PINN framework (*ifas HD-Framework*) for seals [3-6] is used. The framework has been applied to solve the Reynolds equation, including transient cavitation modelling, surface roughness, and non-Newtonian lubricant, for interpolation and extrapolation tasks. Interpolation and extrapolation differ in whether the evaluated scenario was included during the training process. The framework has also been applied to solve deformation tasks [34]. The aim of the framework is to eventually model a complete EHL simulation with a PINN-based FSI structure. The hydrodynamic PINN in the framework requires geometric awareness to communicate with the deformation PINN; thus, the framework is extended with an unsupervised learning autoencoder that compresses the high-dimensional gap geometry of a pneumatic seal into a low-dimensional latent vector. This vector is passed to the PINN to create a geometry-aware PINN, which distinguishes itself from the prior framework by enabling the application to different, more complex, and more realistic sealing geometries.

2. HYDRODYNAMIC LUBRICATION – THE REYNOLDS EQUATION WITH CAVITATION MODELING

Elastohydrodynamic lubrication simulations are essential for analyzing wear, friction, and leakage in lubricated mechanical interfaces. These simulations investigate the dynamic interactions between lubricants and contacting surfaces, focusing on the computational modeling of surface deformations and the resulting hydrodynamic pressure within the contact area. They have proved to be pivotal for the design and optimization of tribological contacts across various industrial applications. The DDS (Dynamic Sealing Simulation) model is an advanced simulation framework that has been validated against experimental data under different operating conditions [6]. This model enlightens the complex interactions between a seal and its mating surface, emphasizing the lubricating film that separates them and significantly influences seal behavior. It employs the finite-element software ABAQUS

to simulate seal deformation under various operating conditions accurately, integrating the Reynolds equation through custom user subroutines to calculate hydrodynamic phenomena. This research encompasses the solution of the Reynolds equation as well as the simplification of the model by excluding surface deformation, contact mechanics, and friction. The objective is to study lubrication aspects without the added complexity of deformations. Given its rigorous consistency with experimental results, an accordingly simplified variant of the DDS model will serve as a benchmark for validating the solution of the Reynolds equation in this work.

The Reynolds equation (2.1), as integrated into the DDS model, is based on Osborne Reynolds' original formulation from 1886 and extended by the Jakobsson-Floberg-Olsson cavitation model. It is described as follows [5]:

$$\underbrace{\frac{v}{2}\frac{\partial}{\partial x}((1-\theta)(R_q\Phi^{\tau}+h))}_{Couette\ flow} - \underbrace{\frac{1}{12\eta}\frac{\partial}{\partial x}(\Phi^{p}h^{3}\frac{\partial p}{\partial x})}_{Poiseuille\ flow} + \underbrace{\frac{\partial}{\partial t}((1-\theta)h)}_{Transient\ term} = 0 \tag{2.1}$$

This extension introduces the cavitation fraction concept, represented by the variable θ [8]. The value of θ traditionally denotes the local volume fraction of the gas phase, ranging from zero (no cavitation) to one (complete cavitation). In this work, since a pneumatic system is examined, the cavity fraction is utilized to model the starved film height of the lubricant in the investigated gap, where 0 represents a filled gap and 1 represents an empty gap. The Reynolds equation describes the hydrodynamic pressure p and the cavity fraction (starved lubricant) θ in thin liquid films, accounting for the dynamic viscosity η , the relative velocity v between the contact surfaces, and the gap height h. The variables of time t and position x are also essential to its formulation. Additionally, the model incorporates the shear and pressure flow factors Φ^{τ} and Φ^{p} according to the work of Patir and Cheng [19, 20] allowing for the consideration of surface topography effects on lubrication. Said flow factors are finally complemented by the root mean square roughness R_q, culminating in a partial differential equation that assumes a crucial role in the analysis of lubricated contacts with incompressible fluids.

The Fischer-Burmeister equation (2.2), as employed by Woloszynski et al. [21], establishes the relationship between the p and θ:

$$p + \theta - \sqrt{p^2 + \theta^2} = 0 \tag{2.2}$$

The value of θ provides insight into the occurrence of a starved lubricant within the tribological contact, with the pressure dropping to zero (cf. (2.2)). The JFO cavitation model and its implementation make minimal assumptions about the physical mechanisms leading to a starved lubricant (cavitation), resulting in a non-zero cavity fraction. This allows the model to track lubricant distribution in tribological contacts with limited lubricant supply, such as grease-lubricated sealing contacts in pneumatic spool valves. In these cases, a non-zero cavity fraction indicates partial filling of the sealing gap. Both the Reynolds and the Fischer-Burmeister Equations are structured to equate to zero, establishing the groundwork for the computation of physics-based losses for neural networks, as will be demonstrated during this paper.

3. MATERIALS AND METHODS

3.1. *Hydrodynamic Lubrication and the Potential of Physics-Informed Neural Networks*

As introduced in Chapter 2, one essential mathematical tool for simulating the pressure distribution in lubricated contacts is the Reynolds equation. Numerical techniques like finite volumes, finite elements, or finite differences are usually used to solve tribological issues in situations like the Reynolds equation, where an analytical solution is not available. However, using numerical approaches frequently necessitates a significant amount of computing power and has proven to be very time-intensive. The solution may lie in the realms of PIML, which has recently shown considerable promise in tribology, facilitating notable advancements [10, 24].

To improve the depiction of phenomena like wear, lubrication, and friction, PIML, more specifically PINNs, incorporates models with physical principles to direct the learning process, in contrast to typical machine learning techniques that only use data-driven methodologies [10, 26]. PINNs employ a neural network for predictive purposes but are trained with applicable laws of physics by including residual terms derived from PDEs in the loss function [28]. Therefore, these models frequently produce results that are more accurate and reliable than those obtained using only data-driven approaches and are used in a variety of tribological applications, ranging from the anticipation of wear or damage to the assessment of lubrication conditions on hydrodynamic surfaces.

Like conventional neural networks, PINNs process inputs like position x, time t, and further case-dependent parameters. The network's output is produced after these input values are processed by the network's hidden layers that comprise a predefined number of neurons. Each neuron carries out a series of mathematical operations that culminate in a complex function, which is then characterized as the network's output.

Once the network has produced an output, the training commences, involving the computation of various supervised and unsupervised losses that are then utilized in the enhancement of neuron parameters, namely the weight and bias.

Figure 3.1 illustrates an exemplary PINN. The network depicted is a hybrid PINN consisting of three physics-informed and one data-driven components. The available data is employed to accelerate the convergence of the governing equations, thereby facilitating the attainment of a more accurate solution.

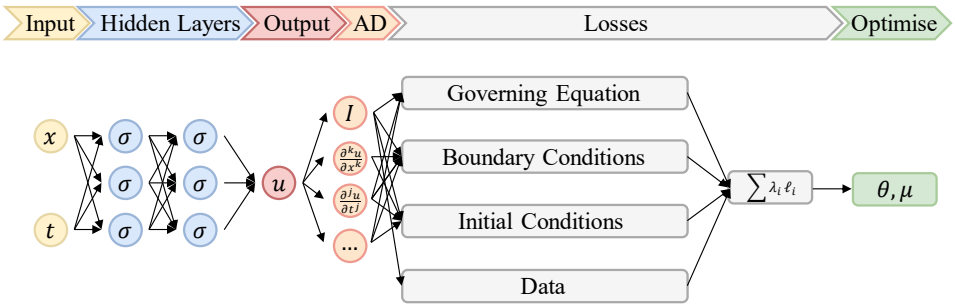


Figure 3.1: Schematic illustration of a PINN [3]

The residual loss (or Governing Equations-Loss, as depicted in Figure 3.1) coincides with the residuals of the governing physical equations, making this an unsupervised loss since there is no direct comparison to a reference value [29]. The evaluation of the loss is conducted at designated spatial and temporal locations, referred to as collocation points, by inserting the PINN's outputs at said points into the governing physical equations and determining the residuals. At this point, it becomes apparent why equations (2.1) and (2.2) must equate to zero (as mentioned in Chapter 2), because the residual between the PINN's output and the governing equations at specific collocation points can only be computed if the right side equals zero.

In addition to the residual losses, PINNs also rely on losses associated with boundary and initial conditions. The objective of said losses is to guarantee that the prescribed boundary and initial conditions for the partial differential equation are satisfied. The presence of a direct value comparison in this loss computation makes these two losses subject to supervisory control. The data loss term corresponds to the classical data-driven loss. Following the determination of the individual losses, the loss terms are then each multiplied by a weight factor and added together. Subsequently, the network parameters $\vec{\theta}$ are optimized so that the weighted loss sum is minimized.

3.2. The ifas HD-Framework

Having established the fundamental principles of hydrodynamic lubrication and the use of PINNs in this realm, the following section will present the structure of the ifas HD-Framework. For a more detailed description also regarding the parameter of the framework and the whole training process the reader is advised to refer to this work [3]. The ifas HD-Framework comprises three main pillars that are all implemented in the Python programming language: the HD-PINN, the Bayesian Optimizer from the Keras Tuner Library [30] and the Adaptive Moment Estimation Optimizer from the Keras Library [31].

The training in this framework begins by employing the Bayesian Optimizer to find the optimal hyperparameters (HP) for the HD-PINN. These include all the parameters that are necessary for the initial construction of the PINN. After that, the PINN computes the initial pressure and cavity fraction distribution for the given input and initializes the loss functions. The iterative body of the training process involves adapting the network parameters $\vec{\theta}$ by the Adam Optimizer based on the loss functions and a consequential recalculation of the pressure and cavitation distribution by the PINN. Once the desired number of epochs has been achieved, the training has completed, and the PINN's losses are transferred back to the Bayesian Optimizer to complete the final optimization of the hyperparameters for that training.

Contrary to previous studies, the HD-PINN in this work consists of a PINN that is preceded by an autoencoder. Figure 3.2 illustrates the novel hydrodynamic framework with the AE, which processes the sealing and counter surface's high-dimensional height profile, provided by the finite-element software ABAQUS, through a Python interface with the encoder to a low-dimensional latent vector consisting of n_{latent} height parameters. To validate the compression, the decoder reconstructs the latent space into a high-dimensional height profile, which is then compared to the original profile. Subsequently, the latent vector

$$\vec{h}_{latent} := \left[h_1, h_2, \dots, h_{n_{latent}}\right] \tag{3.1}$$

is passed to the hydrodynamic PINN in an input vector $\vec{\mathcal{P}}$

$$\vec{\mathcal{P}} := \left[\vec{x}, t, p_l, p_r, \theta_l, \theta_r, \rho, \eta, R_q, \Phi^\tau, \Phi^p, v, h_{lubricant}, \vec{h}_{latent}\right] \text{ with} \tag{3.2}$$

$$\vec{x} := [x_1, x_2, \dots, x_{200}] \tag{3.3}$$

which is then presented to each neuron of the first hidden layer.

Within each of the $n_{neurons}$ neurons in the n_{layers} hidden layers of the PINN, the respective inputs of a single neuron are multiplied by a weight w, summed up, and finally offset with the bias b of the current neuron.

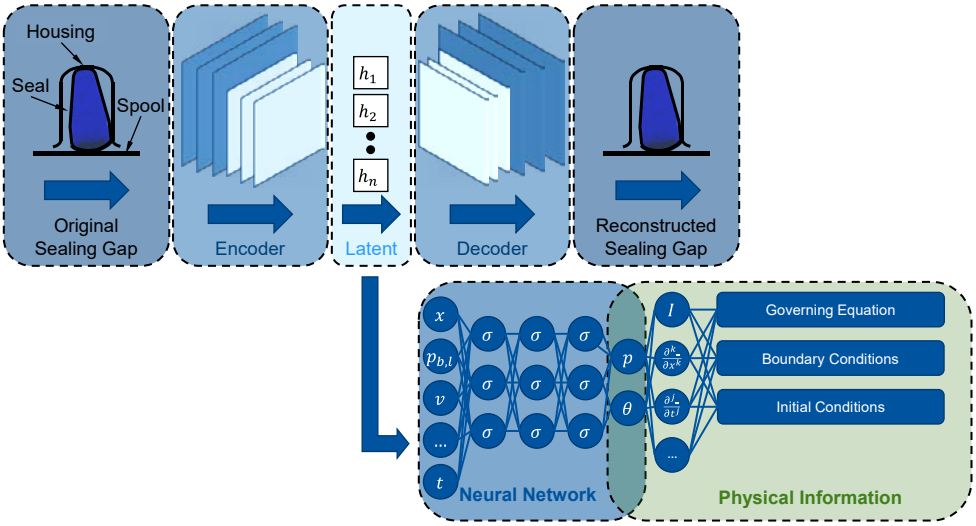


Figure 3.2: The ifas HD-GAPINN

The result will then serve as the input of the neuron's activation function f whose evaluation is then characterized as the neuron's output σ. These individual outputs are required for the solution of the governing equations, namely the Reynolds and Fischer-Burmeister equations, as well as the boundary and initial conditions. In this work, the layer size $n_{neurons}$ will stay constant across all n_{layers}. Additionally, all neurons in one layer will host the same activation function f, thus a different activation function can only be achieved in a different layer. Eventually, the PINN computes the pressure p and the cavity fraction θ, depending on the network parameters $\vec{\Theta}$:

Once the PINN has produced its output, various losses are calculated to ensure its accuracy. These losses include a residual loss for both the Reynolds- and Fisher-Burmeister equation (domain Ω), a boundary condition loss (domain Γ) and an initial condition loss (domain Υ). Partial derivations that may occur in the complex differential equations during loss computation ($\frac{\partial^k u}{\partial x^k}, \frac{\partial^j u}{\partial t^j}, \dots$), are usually determined by employing automatic differentiation (AD), which efficiently computes gradients of any order with machine accuracy by applying differential rules such as the chain and product rule. However, in all its efficiency, the computation of partial derivatives with AD relies on the prerequisite of a continuous function. It thus introduces a limitation into loss calculations, as it does not adequately

account for inputs of a discrete nature. When aiming for a high degree of realism in the underlying simulation, the processing of discrete inputs becomes indispensable because realistic geometries are captured through discrete measurements into a height profile. Thus, to address this limitation, this work will use the finite-difference method (FDM), more specifically, the central finite-difference scheme, for the computation of partial derivatives.

4. RESULTS

Having outlined the methodology used to achieve the research objectives of this work in the previous section, this section presents the obtained results. The results of this study encompass a variation of two geometries ("*Slight Seal Contact*" and "*Seal Geometry with Kinks*") across two different scenarios, namely both with and without the consideration of cavitation modelling. In each test case ("*Slight Seal Contact*", "*Seal Geometry with Kinks*"), the autoencoder was trained with latent-space sizes of 8, 16, and 32, leading to height latent vectors of the exact dimensions that were consequently provided to the PINN.

4.1. Slight Seal Contact without Cavitation Modelling

This first presented test case is characterized by a slightly stressed seal with a force exerted on the housing and prolonged onto the seal. Since cavitation is not taken into consideration, the loss calculation for this test case solely revolves around the residual and pressure boundary condition loss $\mathcal{L}_{\Omega_{Residual}}$ and $\mathcal{L}_{\Gamma_p}$. This specific test case is relevant for assessing the accuracy of the PINN's geometry-aware loss computation, which relies on finite difference schemes to approximate the derivatives of the height profile. Figure 4.1 shows the results of the AE with different latent size (8, 16, 32), the results of the PINN and the relative error of the PINN ($p_{PINN} - p_{DDS}/p_{DDS}$). For each latent size, the reconstruction of the height and the pressure computation of the PINN yield good results. In Figure 4.1 (left column), the PINN configured with an 8-dimensional latent space shows a maximum absolute percentage pressure error of 1.4%. By expanding the latent representation to 16 dimensions, the peak absolute percentage discrepancy decreases to 1.2%, indicating a slight improvement in accuracy. Further increase of the latent size to 32 yields a maximum absolute percentage error of around 1.5%, thus worse than the smaller latent sizes. An increase in the latent size does not necessarily result in a better performance of the PINN. The latent size primarily affects the performance of the compression and reconstruction of the AE. Since all three variants show a good match between the original and reconstructed height profiles, the use of a smaller latent size allows for the possibility of using a smaller network size, as the layer sizes should be proportional to the number of inputs. Thus, the size of the latent space should be selected rather smaller and can be checked by the reconstruction accuracy of the AE. The increase in latent size does not necessarily translate into better physical fidelity and may, in fact, introduce unnecessary complexity that hinders generalization. Overparameterization can lead to overfitting or suboptimal convergence, especially when the problem geometry does not warrant such a high-capacity representation.

In summary, this test case demonstrates that increasing the latent dimension from 8 to 16 improves the quality of the predicted pressure field by enabling the PINN to better resolve small variations in geometry. However, further increase to 32 dimensions offers no clear advantage and slightly worsens performance, emphasizing the importance of balancing representational power with the complexity of the underlying physical domain.

Nevertheless, each investigated latent size shows good results in combination with the PINN.

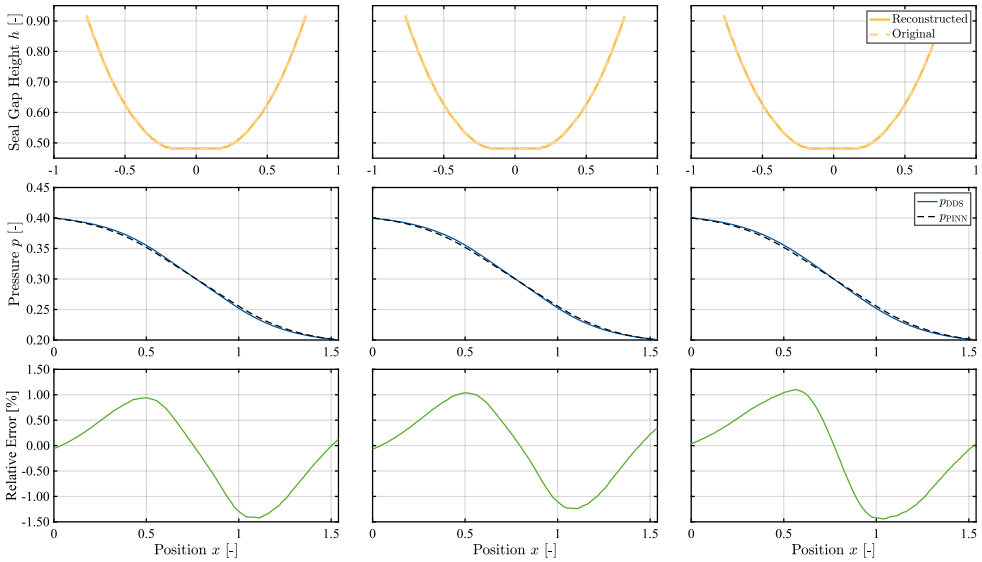


Figure 4.1: Pressure Distribution for Geometry "*Slight Seal Contact*"

4.2. Seal Geometry with Kinks without Cavitation Modelling

This second investigated test case is characterized by a seal that is pressed against a diagonal slider and thus contains kinks and steep gradients. Analogously to the test case "*Slight Seal Contact*" without the consideration of starved film height, the loss calculation for this test case revolves around the residual and pressure boundary condition loss $\mathcal{L}_{\Omega_{Residual}}$ and $\mathcal{L}_{\Gamma_p}$.

This test case was designed to evaluate the PINN's capabilities in handling high height gradients, which is particularly relevant for this geometry-aware PINN because its loss computation relies on a finite-difference scheme for the derivatives of the height profile. Finite-difference schemes, being numerical, introduce an error into the computation that can severely alter the result if said computation proves to be vulnerable to error propagation. This is especially the case when error-prone intermediate results experience a potentiation. According to Equation (2.1), such a potentiation of the height profile in combination with its x-derivative is introduced in the Reynolds equation and thus also in the computation of the Reynolds loss term (scaled with the factor three, cf. x-derivation of h^3 in Reynolds equation (2.1)).

The results shown in Figure 4.2 demonstrate a slightly worse approximation of the pressure distribution for lower latent space sizes than the previous test cases; the maximum absolute percentage error reaches up to 2% for latent space sizes of 8 and 16. For the 32-LS case, however, the PINN manages to obtain excellent results, leaving the maximum absolute percentage deviation from the ifas DDS-reference below 0.7%.

Clearly, the PINN profited from a higher resolution of the challenging height profile. This may be attributed to a better performance of the finite-difference Method for higher latent space sizes, resulting in a more accurate loss computation and thus a more effective training.

Nevertheless, the performance in all three test cases is sufficient for a precise modelling of the hydrodynamic pressure.

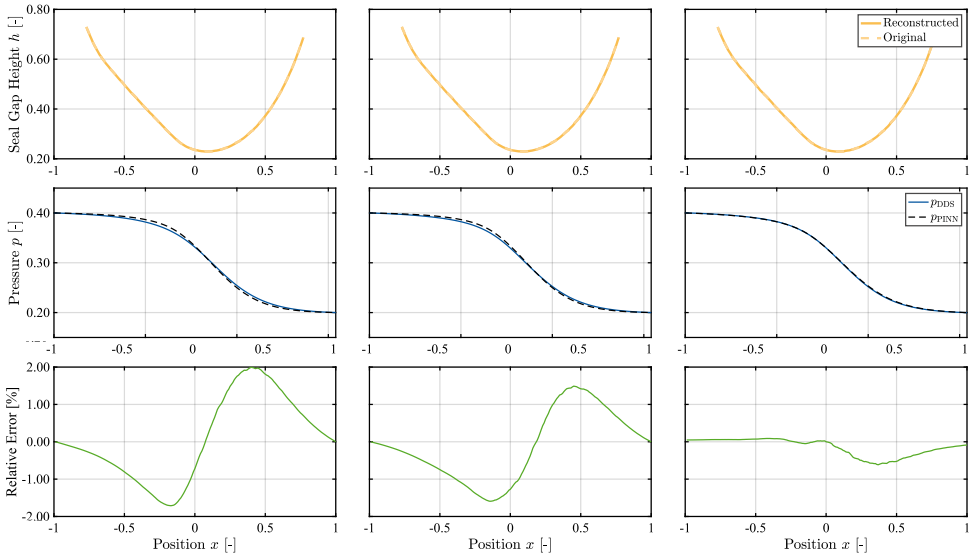


Figure 4.2: Pressure Distribution for Geometry "*Seal Geometry with Kinks*"

4.3. Slight Seal Contact with Cavitation Modeling

The following section extends the first test case by including the starved film height, which increases the complexity of the PINN's training. In contrast to the previous test cases, the PINN's training now incorporates all losses described in Section 3.2. In addition, a Filtered Absolute Error metric is used to compare the results between the PINN and the DDS for the cavity fraction and pressure. This accounts for differences in numerical precision between the two methods. For instance, in regions where cavitation should ideally be zero, the DDS may yield values on the order of 10^{-20}, while the PINN may only approximate this with values around 10^{-5}. The usage of a relative error in these areas would not adequately illustrate the accuracy of the PINN. By using the "Rolling Mean" filter, the mentioned outliers from the absolute error $|(p,\theta)_{PINN} - (p,\theta)_{DDS}|$ are eliminated.

In this test case, shown in Figure 4.3, the PINN does not perfectly reproduce the pressure and cavitation peaks. However, for the 16-LS and 32-LS cases, the pressure peaks are well captured, with absolute errors of less than 0.05. The sharp cavitation peak remains a challenge across all latent space sizes, with maximum absolute errors ranging between 0.01 to 0.15. It becomes apparent that the inclusion of cavitation significantly increases the complexity of the problem, making it more challenging for the PINN to correctly identify the regions of maximum pressure and cavitation. Nonetheless, the performance of the PINN could potentially be improved through extended Bayesian Optimization training, allowing for the discovery of more optimal hyperparameters and a better global minimum.

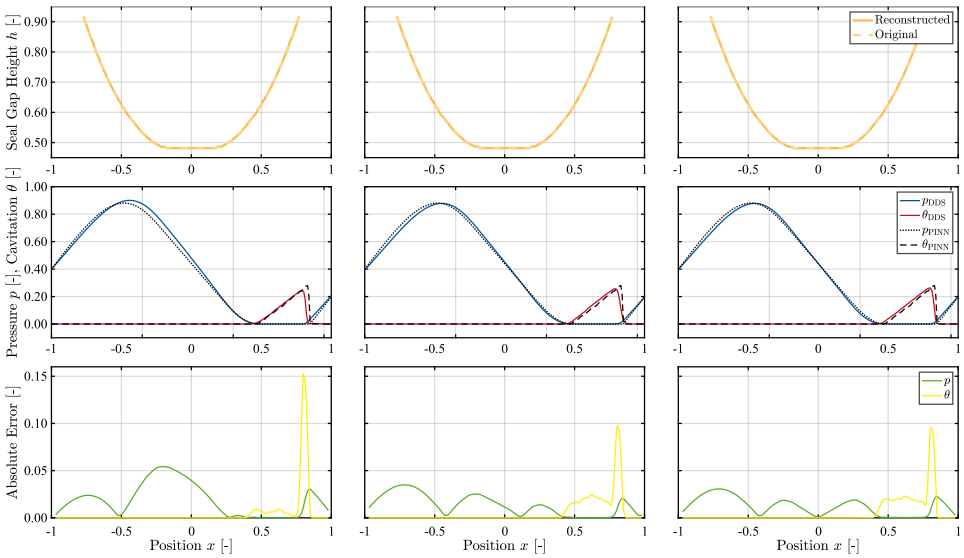


Figure 4.3: Pressure and Cavitation Distribution for Geometry "*Slight Seal Contact*"

It is worth noting that the application of soft-constraints is not expected to improve the overall results. This is because soft-constraints primarily assist in the transition regions between pressure and cavitation, rather than directly enhancing the PINN's ability to identify the locations of maximum pressure and cavitation. Therefore, they do not offer a complete solution for addressing the PINN's limitations in capturing these critical features. Nevertheless, all three setups display good agreement between the PINN and the numerical solver, especially when considering the speedup from four milliseconds to one to two seconds (Intel I9 13th-generation processor, 128 GB RAM DDR5, and Nvidia RXT 4090) in terms of computation time for PINN and DDS. Table 4.1 summarizes the results of the different test case by displaying the mean absolute error (MAE) and the root mean squared error (RMSE) and underline the high accuracy for each test case and each latent size.

Case	**Latent**	**MAE p**	**MAE θ**	**RMSE p**	**RMSE θ**
1	32	0.0050	-	0.0060	-
1	16	0.0024	-	0.0029	-
1	8	0.0027	-	0.0033	-
2	32	0.0020	-	0.0018	-
2	16	0.0027	-	0.0020	-
2	8	0.0028	-	0.0021	-
3	32	0.0118	0.0056	0.0151	0.0205
3	16	0.0134	0.0057	0.0172	0.0193
3	8	0.0185	0.0057	0.0254	0.0279

Table 4.1: Mean absolute error and root mean squared error for each test case

To sum up, the results show that the extension of the AE enables the PINN to obtain geometry-awareness and solve each test case with high accuracy for all three latent sizes. Since all latent sizes displays sufficient accuracy, the usage of a smaller latent size is promising since it would less increase the input size of the PINN and thus keep the whole framework as lightweight as possible.

5. CONCLUSION

This work focused on the conception, implementation, and testing of a geometry-aware physics-informed neural network, comprising the HD-PINN from the ifas HD-Framework and an extension that enables it to process more complex and realistic geometries. The extension included the finite-element software ABAQUS for the seal model, a Python interface that extracted height profiles from ABAQUS, and an autoencoder that enabled the compression of those height profiles to make the input processable for the PINN. The PINN's capabilities in the pressure and cavitation prediction for lubricated sealing gaps in pneumatic valves were then assessed in three test scenarios. All pressure and cavitation distributions produced by the PINN were then systematically compared to those generated by the ifas DDS for the identical test scenarios.

This work represents the next step in implementing a PINN-based fluid-structure interaction for elastohydrodynamic lubrication simulations. The novel framework extends the current state-of-the-art of PINNs in hydrodynamic lubrication by adding geometry-awareness and enables the coupling with the deformation part of the elastohydrodynamic simulation. Furthermore, the findings show that the PINN can process a latent representation of a high-dimensional variable, which is a promising result, since the PINN for the deformation also needs to process the high-dimensional load, described by the hydrodynamic pressure vector, coupled in the fluid-structure-interaction. To sum up, the findings underline the potential of PINNs to model elastohydrodynamic behavior in lubricated contacts in pneumatic seals, and the next steps are to implement an AE for the deformation PINN and, eventually, couple all components in a fluid-structure-interaction.

6. ACKNOWLEDGEMENTS

The authors thank the Research Association for Fluid Power of the German Engineering Federation VDMA for its financial support (grant: FKM No. 7058400). Special gratitude is expressed to the participating companies and their representatives in the accompanying industrial committee for their advisory and technical support.

7. REFERENCES

[1] Brumand-Poor, F., Makansi, F. et al. 'Implementation of Variational Autoencoder for Dimension Reduction of a Hydraulic System', 2022 Global Fluid Power Society Ph.D. Symposium, 2024. DOI: 10.13052/rp-9788770047975.

[2] Brumand-Poor, F., Kauderer, L. et al. 'Application of Deep Reinforcement Learning Control of an Inverted Hydraulic Pendulum', International Journal of Fluid Power, 2023. DOI: 10.13052/ijfp1439-9776.2429.

[3] Brumand-Poor, F., Bauer, N. et al. 'FAST COMPUTATION OF LUBRICATED CONTACTS: A PHYSICS- INFORMED DEEP LEARNING APPROACH', 14th International Fluid Power Conference, pp. 254–266, 2024. DOI: 10.13052/rp-9788770042222C19.

[4] Brumand-Poor, F., Barlog, F. et al. 'Advancing Lubrication Calculation: A Physics-Informed Neural Network Framework for Transient Effects and Cavitation Phenomena in Reciprocating Seals', 22nd International Sealing Conference 2024, 2024. DOI: 10.61319/J7I2HNKR.

[5] Brumand-Poor, F., Rom, M. et al. 'Physics-Informed Deep Learning for Lubricated Contacts with Surface Roughness as Parameter', Tribologie und Schmierungstechnik, Vol. 71, Nr. 4, 2024. DOI: 10.24053/TuS-2024-0021.

[6] Brumand-Poor, F., Azanledji, F. K. et al. 'Extrapolation of cavitation and hydrodynamic pressure in lubricated contacts: a physics-informed neural network approach', Advanced Modeling and Simulation in Engineering Sciences, Vol. 12, Nr. 1, 2025. DOI: 10.1186/s40323-025-00283-9.

[7] Bauer, N., Baumann, M. et al. 'Elastohydrodynamic Simulation of Pneumatic Sealing Friction Considering 3D Surface Topography', Chemical Engineering & Technology, Vol. 46, Nr. 1, pp. 167–174, 2023. DOI: 10.1002/ceat.202200471.

[8] Wohlers, A., Heipl, O. et al. 'Numerical and Experimental Investigation on O-Ring-Seals in Dynamic Applications', International Journal of Fluid Power, Vol. 10, Nr. 3, pp. 51–59, 2009. DOI: 10.1080/14399776.2009.10780988.

[9] Dakov, N. 'Elastohydrodynamische Simulation von Wellendichtungen am Beispiel der PTFE-Manschettendichtung mit Rückförderstrukturen', Universität Stuttgart, 2020. DOI: 10.18419/opus-10997.

[10] Marian, M., Tremmel, S. 'Current Trends and Applications of Machine Learning in Tribology—A Review', Lubricants, Vol. 9, Nr. 9, pp. 86, 2021. DOI: 10.3390/lubricants9090086.

[11] Marian, M., Tremmel, S. 'Physics-Informed Machine Learning--An Emerging Trend in Tribology', Lubricants, Vol. 11, Nr. 11, pp. 463, 2023. DOI: 10.3390/lubricants11110463.

[12] Almqvist, A. 'Fundamentals of Physics-Informed Neural Networks Applied to Solve the Reynolds Boundary Value Problem', Lubricants, Vol. 9, Nr. 8, pp. 82, 2021. DOI: 10.3390/lubricants9080082.

[13] Li, L., Li, Y. et al. 'ReF-nets: Physics-informed neural network for Reynolds equation of gas bearing', Computer Methods in Applied Mechanics and Engineering, Vol. 391, pp. 114524, 2022. DOI: 10.1016/j.cma.2021.114524.

[14] Zhao, Y., Guo, L., Wong, P. P. L. 'Application of physics-informed neural network in the analysis of hydrodynamic lubrication', Friction, Vol. 11, Nr. 7, pp. 1253–1264, 2023. DOI: 10.1007/s40544-022-0658-x.

[15] Rom, M. 'Physics-informed neural networks for the Reynolds equation with cavitation modeling', Tribology International, Vol. 179, pp. 108141, 2023. DOI: 10.1016/j.triboint.2022.108141.

[16] Cheng, Y., He, Q. et al. 'HL-nets: Physics-informed neural networks for hydrodynamic lubrication with cavitation', Tribology International, Vol. 188, pp. 108871, 2023. DOI: 10.1016/j.triboint.2023.108871.

[17] Xi, Y., Sun, R. 'Inverse problems in hydrodynamics lubrication: Parameter identification in the Reynold equation by using physics-informed neural networks', Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, Vol. 239, Nr. 4, pp. 423–435, 2025. DOI: 10.1177/13506501241291403.

[18] Xi, Y., Deng, J., Li, Y. 'A new method to solve the Reynolds equation including mass-conserving cavitation by physics informed neural networks (PINNs) with both soft and hard constraints', Friction, 2024. DOI: 10.1007/s40544-023-0791-1.

[19] Patir, N., Cheng, H. S. 'An Average Flow Model for Determining Effects of Three-Dimensional Roughness on Partial Hydrodynamic Lubrication', Journal of Lubrication Technology, Vol. 100, Nr. 1, pp. 12–17, 1978. DOI: 10.1115/1.3453103.

[20] Patir, N., Cheng, H. S. 'Application of Average Flow Model to Lubrication Between Rough Sliding Surfaces', Journal of Lubrication Technology, Vol. 101, Nr. 2, pp. 220–229, 1979. DOI: 10.1115/1.3453329.

[21] Woloszynski, T., Podsiadlo, P., Stachowiak, G. W. 'Efficient Solution to the Cavitation Problem in Hydrodynamic Lubrication', Tribology Letters, Vol. 58, Nr. 1, 2015. DOI: 10.1007/s11249-015-0487-4.

[22] Zhao, Y., Wong, P. P. L. 'A hybrid data-driven approach for the analysis of hydrodynamic lubrication', Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, Vol. 238, Nr. 3, pp. 320–331, 2024. DOI: 10.1177/13506501231214584.

[23] Shutin, D., Kazakov, Y. et al. 'Data-driven and physics-informed approaches for improving the performance of dynamic models of fluid film bearings', Tribology International, Vol. 191, pp. 109136, 2024. DOI: 10.1016/j.triboint.2023.109136.

[24] Paturi, U. M. R., Palakurthy, S. T., Reddy, N. S. 'The Role of Machine Learning in Tribology: A Systematic Review', Archives of Computational Methods in Engineering, Vol. 30, Nr. 2, pp. 1345–1397, 2023. DOI: 10.1007/s11831-022-09841-5.

[25] Ramos, D. J., Cunha, B. Z., Daniel, G. B. 'Evaluation of physics-informed neural networks (PINN) in the solution of the Reynolds equation', Journal of the Brazilian Society of Mechanical Sciences and Engineering, Vol. 45, Nr. 11, 2023. DOI: 10.1007/s40430-023-04418-0.

[26] Raissi, M., Perdikaris, P., Karniadakis, G. E. 'Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving

nonlinear partial differential equations', Journal of Computational Physics, Vol. 378, pp. 686–707, 2019. DOI: 10.1016/j.jcp.2018.10.045.

[27] Rimon, M. T. I., Hassan, M. F. et al. 'A Design Study of an Elasto-Hydrodynamic Seal for sCO2 Power Cycle by Using Physics Informed Neural Network', ASME Power Applied R&D 2023, 2023. DOI: 10.1115/POWER2023-108802.

[28] Nabian, M. A., Meidani, H. 'Physics-Driven Regularization of Deep Neural Networks for Enhanced Engineering Design and Analysis', Journal of Computing and Information Science in Engineering, Vol. 20, Nr. 1, 2020. DOI: 10.1115/1.4044507.

[29] Cai, S., Mao, Z. et al. 'Physics-informed neural networks (PINNs) for fluid mechanics: a review', Acta Mechanica Sinica, Vol. 37, Nr. 12, pp. 1727–1738, 2021. DOI: 10.1007/s10409-021-01148-1.

[30] O'Malley, Tom and Bursztein, et. al. 'KerasTuner', https://github.com/keras-team/keras-tuner, 2019.

[31] Chollet, F. e. a. 'Keras', https://github.com/fchollet/keras, 2015.

[32] Oldenburg, J., Borowski, F. et al. 'Geometry aware physics informed neural network surrogate for solving Navier–Stokes equation (GAPINN)', Advanced Modeling and Simulation in Engineering Sciences, Vol. 9, Nr. 1, 2022. DOI: 10.1186/s40323-022-00221-z.

[33] Nguyen, T. N. K., Dairay, T. et al. 'Geometry-aware framework for deep energy method: an application to structural mechanics with hyperelastic materials', http://arxiv.org/pdf/2405.03427v1, 2024.

[34] Brumand-Poor, F., Trautmann, M., Schmitz, K. 'Advancing deformation calculation: a physics-informed deep graph learning framework for hyperelastic materials', Advanced Modeling and Simulation in Engineering Sciences, Vol. 12, Nr. 1, 2025. DOI: 10.1186/s40323-025-00304-7.

[35] Scarselli, F., Gori, M. et al. 'The graph neural network model', IEEE Transactions on Neural Networks, Vol. 20, Nr. 1, pp. 61–80, 2009. DOI: 10.1109/TNN.2008.2005605.

[36] Dalton, D., Husmeier, D., Gao, H. 'Physics-informed graph neural network emulation of soft-tissue mechanics', Computer Methods in Applied Mechanics and Engineering, Vol. 417, pp. 116351, 2023. DOI: 10.1016/j.cma.2023.116351.

Biographies



Faras Brumand-Poor received a bachelor's degree in electrical engineering from RWTH Aachen University in 2017, a master's degree in electrical engineering from RWTH Aachen University in 2019, and a master's degree in automation engineering from RWTH Aachen University in 2020, respectively. He is the Group Leader of the research groups Fluids and Smart Systems and the Deputy Chief Engineer at the Institute for Fluid Power Drives and Systems at RWTH Aachen University. His research areas include machine learning, particularly deep learning, physics-based learning, fluid transmission lines, and virtual sensory systems.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz’s awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

AI Assisted Engineering: Information Extraction from Various Formats for Engineering Decisions

Vaishnavi Kulkarni, Anwesha Sen, Anik Kumar Samanta*, Shrinivas Kulkarni

Danfoss Power Solutions, Danfoss India Innovation Centre, Pune, India.,

vaishnavi.kulkarni@danfoss.com,

anwesha.sen@danfoss.com,

anik.samanta@danfoss.com,

shrinivas.kulkarni@danfoss.com

Abstract

Engineering drawings and 3D (3-Dimensional) models store vital information that is essential for any industrial product life cycle – right from designing and manufacturing to order placement and inventory management. Accurate retrieval of the information requires intensive manual efforts and is prone to human bias and error. In this paper, we present two novel methods to extract information from 3D models and 2D engineering drawings. This information is contained in the models and drawings with different file formats such as portable document format (PDF) and computer aided design (CAD) models. This makes it difficult to integrate the drawing data into the company's data management system. Hence, we have used optical character recognition (OCR) to gather dimensions and specifications from the drawings and models of different file formats. The information is also present in different fonts and locations across the drawing or models. Although these documents contain a lot of relevant data, they also include unnecessary information. The fine tuning of character recognition models is done in the context of mechanical drawings, which improves the quality of data reading. Additional quality checking parameters such as nearest character matching and template matching of required text areas are also integrated to avoid a trade-off between speed and accuracy. The accuracy of the correctly read tokens using a combination of above-mentioned OCR strategies gives an accuracy of approximately 93%, while the detections with low confidence are automatically filtered out. This leads to the error proof output of the data in a format that is expected by the end user. An OCR based strategy for information extraction in this manner significantly reduces manual efforts and the same methodology is widely scalable across different types of drawings and models.

Keywords: 2D-drawings, 3D-models, Assembly Drawings, Drawing information retrieval, EasyOCR, machine learning, Tesseract OCR, Optical character recognition, text recognition.

1. INTRODUCTION

Hydraulic drawings contain critical information that is used from the design and development stage to manufacturing and ordering. In this era of digitalization, it is an

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

organizational prerogative to arrange, clean, label, and store data for inferences and insights. However, this requires tremendous human effort, and the repetitive manual entry is prone to errors. The drawings and documents have accumulated over the years through internal innovative developments and external mergers. The format of the data is varied, and the numbers are huge. The drawings contain information about different child parts, dimensions, specifications, and material details. Drawings / any relevant fluid power circuitry are stored as three-dimensional (3D) models or two-dimensional (2D) images. Moreover, some of the drawings are scanned and stored in document formats like PDF as discussed in Jameison et. al. [1]. The scanned documents are noisy and extraction of useful information from such documents is challenging. Despite many advances in OCR-based document analysis, significant challenges remain in industrial engineering contexts. A large portion of legacy mechanical drawings exists as old, noisy PDF files, many of which are scanned images containing faded print, and handwritten annotations. Such documents have low contrast, background noise, and inconsistent text orientation, making text extraction difficult for standard OCR systems. The current methods reviewed in the literature cannot be directly applied to accurately extract structured engineering information from these documents, highlighting the need for a robust, domain-specific OCR framework.

The interpretation of 3D model textual data on a large scale without metadata is a hugely time-consuming task. OCR has not been used in the popularly available literature to interpret area, mass and other data for 3D models with high accuracy, this is something that we have attempted for the first time in our use case. The paper presents two OCR-based frameworks to extract engineering information from existing 2D drawings and 3D CAD models. For 2D drawings, we convert pages to images, split them into quadrants, run OCR, and match the text against internal databases to identify materials and surface finishes. For 3D models, we have automated Creo View Lite to open files, capture the “Summary” window, apply OCR, and parse values relevant to the model’s design. OCR in our use case has been finetuned to handle text, numeric and specials symbols data as high accuracy is a requirement for our case. The goal is to reduce manual work and errors in feeding data into company systems.

Information extraction from 2D Drawings: Mechanical drawings often contain critical information that must be accurately identified for further analysis in manufacturing industries. They have critical specifications, dimensions, and material details that directly influence design and production quality. Traditionally, these drawings have been manually interpreted by engineers. However, a vast number of legacy drawings exists only as PDFs or scanned images, making automated extraction of textual information tricky.

Based on literature, very recent research introduces eDOCr2 [2], a hybrid framework that combines traditional OCR, image processing, and Vision-Language (VL) models to extract structured and semantic information from mechanical engineering drawings (EDs). To improve semantic understanding and reasoning, eDOCr2 integrates Vision Language models (Qwen2-VL-7B and GPT-4o) after segmentation to verify, filter, or retrieve information [3]. Some literature proposes an OCR-based system using Tesseract OCR [4] to extract text from general digital images through detection, segmentation, and binarization. It separates the text from the background and then detects the text. Although these techniques extract text by applying different methods like text localization, binarization, etc., the method operates on entire drawings and does not explicitly consider sub-region OCR to boost recognition performance. Unlike standard documents, mechanical drawings contain multi-scale annotations, symbols, non-standard fonts. Optical character recognition (OCR)

systems trained on printed or handwritten text often underperform on these noisy, skewed, and symbol-heavy images. Consequently, there is a need for robust and domain-specific extraction pipelines that can bridge the gap between raw images and structured databases.

In this work, we propose a quadrant-based OCR technique for mechanical PDF drawings. The approach converts each drawing page into an image, divides it into four quadrants, applies OCR independently, and then maps extracted text to a structured excel database. Dividing image into more than four quadrants can result into system processing overhead, which will take long time for processing. The resulting information is stored in Excel along with the drawing names, creating a structured dataset. This workflow is lightweight and adaptable, provides improved accuracy by localizing OCR to smaller image regions.

Information extraction from 3D models: 3D models across Danfoss and other multinational corporations dealing in mechanical models and assemblies require the area, mass, volume and other data of the parts to be saved across different product lifecycle management software such as SAP, Siemens Teamcenter, etc. The manual entry of a large number and variety of components consumes longer man hours and is a non-value-added activity. Additionally, the handling of different types of files such as step (STP), part (PRT) and assembly (ASM) may employ different software tools and strategies to ensure the correct file is used and its subcomponents are all loaded properly. Repeating this monotonous process leads to human error, especially since area, volume and other values are required to be in single point precision and a single wrong entry may affect the system data in the long run. The quality control of such processes additionally needs to be performed manually, which again may contain human bias and discrepancies.

Hence, there is the requirement of an automation system that can process many files with high accuracy, extract relevant information from the same and return the digitized version of the results. All of this should be done in a fraction of the time required to repeat the same manually. Different software packages were that were popularly available were compared for analysis [5]. Most software either did not contain metadata for viewing or displayed incorrect values. To address this lacuna, a non-parametric method was needed. Thus, CREO View and Python based OCR dependent logic was developed to access and digitize area, volume, mass, and envelope information for STP, PRT and ASM type mechanical model files. This can return accurate information without any human intervention and segregate healthy and unhealthy files on its own, thus producing additional insights.

The main contributions made in this paper can be listed as following: (1) Automated information extraction from old, noisy, and sometimes scanned mechanical drawings in PDF format, (2) proposed a quadrant-based OCR technology to improve the accuracy of text extraction of existing methods, (3) applying OCR to 3D model snapshots, overcoming challenges of graphical overlays and extracting engineering-specific dimensions and annotations, (4) proposed a model viewer based software for 3D model data extraction, (5) combining two OCR engines to achieve optimally best results for small fonts and numeric data of 3D models.

2. INFORMATION EXTRACTION FROM OLD, NOISY DOCUMENTS AND IMAGES

Any leading manufacturing industry player always strives to make their products environment friendly and sustainable. So, the materials that are used to make mechanical products play a pivotal role. Mechanical CAD drawings contain the data of dimensions and tolerances along with critical information regarding which materials are to be used, and surface finishes required for each component in product manufacturing. Accurate identification and evaluation of constituent materials and finishing ensure Danfoss products are aligned with current sustainability goals. We developed a robust OCR approach to identify material specifications and surface finishes. We have used Tesseract OCR as it is open source and all the computation happens locally, which helps to maintain the confidentiality of data.

The proposed solution is implemented in four major steps:

i. Image segmentation and optical character recognition
ii. PDF to Image conversion
iii. Text consolidation
iv. Pattern matching and classification

2.1. Image Segmentation and Optical Character Recognition

The first step involves opening a PDF document with the help of a default PDF viewer of the operating system. The system attempts to identify and activate the system's default PDF viewer window. Now the system's screen dimensions are calculated to identify the best capture regions. The screenshot capture process implements a quadrant-wise approach for capturing the screenshots of PDF content. Now based on the document's width and height, which is extracted from PDF metadata, it calculates the optimal zoom level. This is done to maintain the resolution of the text present in the PDF. The viewport manipulation works in the clockwise direction starting from the top-left quadrant. Now the system applies to the calculated zoom level and captures the screenshot. Following this, it captures the screenshots of all the quadrants ending at the top-right corner. The captured screenshots are then saved with the detailed name of the file followed by the quadrant name for further analysis as seen in Fig. 1.

As an initial step, OCR is applied separately to each quadrant image. Each image maintains its resolution loaded using image processing libraries that provide image format support and image manipulation techniques. Based on the image quality, different methods are used like denoising and contrast enhancements to improve text definition. The images are converted into high contrast black and white images which are optimal for OCR. The Tesseract OCR engine is configured with required parameters for technical document processing, along with including language models appropriate for engineering terminology and character recognition settings. In this process, accuracy is preferred over speed. The extracted text from each screenshot is stored in a temporary text file, which is further used to post-process the data to extract the relevant information.

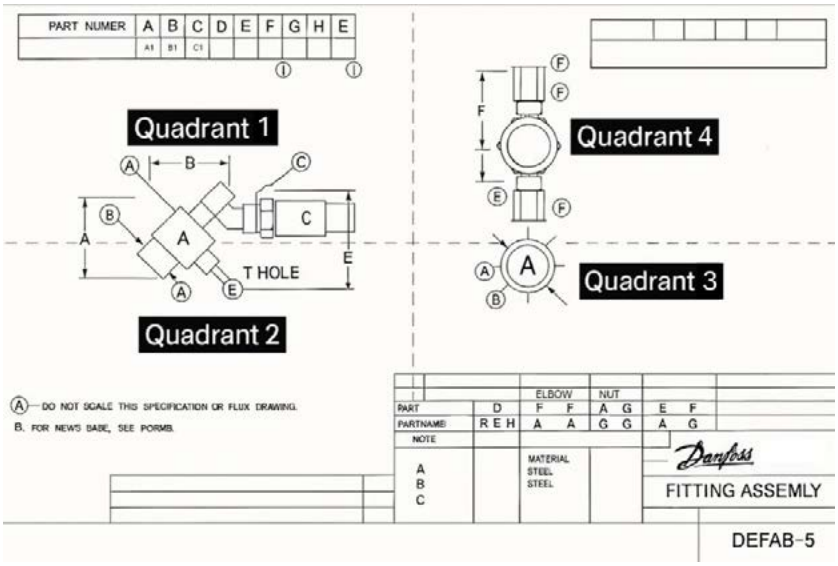


Fig. 1. Representative Dummy Drawing with quadrant references

2.2. PDF to Image Conversion

This step includes direct conversion of PDFs into images. This process makes use of specialized PDF-to-image conversion libraries that interpret the PDF format and generate bitmap representations of the document. The conversion process converts all pages within each PDF document, creating joint photographic experts' group (JPEG) images at a resolution suitable for OCR analysis. The output images are organized in such a way that it maintains the directory hierarchy which has clear association between image name and original source PDF name. Now all JPEG images are converted to portable network graphic (PNG) image format. PNG is a lossless format that preserves image quality throughout the process [6].

Now the image cropping phase is implemented, which applies strategic segmentation to full-page images, focusing on relevant content that is present in the PDF. By leveraging domain specific knowledge, we targeted the selection of regions where most critical information lies. The regions include a title block, material specifications, and annotations block. The cropping method begins by analyzing the dimensions of the full-page image to set proportional dimensions for the interested regions. The primary extraction focuses on the upper half of the document, which commonly contains detailed specification table and material information. After that, it extracts focused regions like the bottom-right area of the document, which typically has title blocks and primary material specifications in mechanical drawings.

Now each cropped image is given a descriptive name as "original filename + region-identifier" and saved as seen in Fig. 2 in the regions highlighted by quadrants 3 and 4. This

segmentation approach allows for region specific OCR, which not only reduces computational overhead but also improves the detection accuracy.

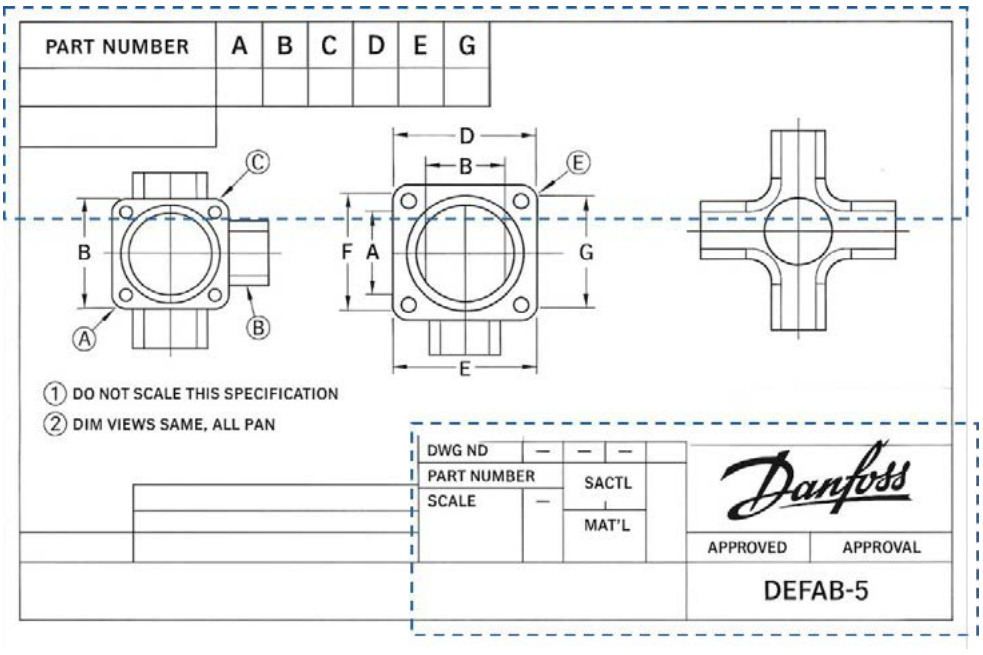


Fig. 2. Representative Dummy Image of Drawing with Identified Region

2.3. Text Consolidation

Text consolidation phase includes concatenation of different text extraction outputs in analyzable collection of text data. This process begins with collecting text data from three different methods:

i. Native PDF text extraction using PyPDF2 library of Python.
ii. From quadrant-based screenshot OCR results from the first step.
iii. Direct image conversion OCR results from the second step.

Each text data source provides unique characteristics. Native PDF text extraction usually provides high accuracy for digitally created documents, preserving exact characters. The screenshot-based OCR results are important, as it extracts text from specific information in dense regions of the drawings. The integration process concatenates these different text sources while maintaining source identification markers that make analysis of data easier.

The consolidated text undergoes initial cleaning process that removes whitespace patterns, normalizing line break characters. Additionally, the code generates an uppercase variant of the text to reduce case insensitive pattern matching operations. This duplicate representation ensures that material specifications written in various case combinations can be identified.

After consolidating and removing inconsistencies from the text, now comes character level error correction phase. This step is specifically performed to handle common misidentifications. This process applies domain-specific correction rules, that account for the visual similarities between certain characters, in the context of industrial specifications.

The most fundamental confusion is between letters and numbers that are visually the same. The letter 'O' is replaced by number '0' in context where numeric value is expected. Lowercase 'l' characters are converted into numeric '1' when there are alphanumeric specification codes. Also, some technical specification needs corrections for misidentification such as 'AIST' replaced by 'AISI' to correct common misreading of 'American Iron and Steel Institute' abbreviation. Apart from this character level correction, this solution also takes care of spacing irregularities, hyphens, dashes, and punctuation marks that commonly occur in technical document processing. Although rule-based standardization resolves many character-level ambiguities, it does not fully capture contextual intent. Future mitigation strategies should integrate contextual AI models and pattern-based validation to disambiguate visually similar characters using semantic cues and domain knowledge, thereby improving accuracy in technical document interpretation.

2.4. Pattern Matching and Classification

2.4.1 Material Name Recognition

The material recognition process implements a database-driven approach to correctly identify the materials. This process begins with accessing the material column from the predefined database of excel format. Material names are normalized to lowercase format to avoid case-insensitive matching. The matching algorithm works on exact string matching within collected text from different phases. For each material name in the database, the code searches for its presence in the collection of data.

Some domain specific logic rules are implemented to handle ambiguous material terminology that may appear to be the same. For example, when the material name 'anneal' is found, the algorithm applies additional check to ensure that 'annealing' is not present in the same document, as these terms represent different material specifications and should not be conflicted. Valid matches are stored into result set along with its document name. The identified materials are stored in the form of comma-separated values.

2.4.2 Material Specification

Material specification is one of the most complex aspects of pattern matching. It requires a valid pattern recognition capability to identify and validate the material specification codes commonly used in industrial applications. This process identifies various material specification codes like AISI steel standards, W-codes, and generic specifications.

The American Iron and Steel Institute (AISI) specification identification starts with pattern matching, which will first identify the characteristics format of 'AISI' prefix followed by numeric codes ranging from three to four digits. Each AISI code undergoes the verification of material specification codes against the database, ensuring that only valid AISI designations are stored in the results.

W-codes are processed based on the pattern matching which begins with letter ‘W’ followed of alphanumeric characters. The W-code processing algorithm implements a comprehensive character correction system that addresses common misreads. The character correction process replaces frequently confused characters with the most likely intended value in the context of material specification codes. The letters 'o' and 'O' are standardized to '0', because the material specification does not contain letter ‘o’ or ‘O’. The W-code validation process includes length verification to ensure that processed codes are of five-character format.

2.4.3 Surface Finish

Surface finish recognition focuses on identifying surface treatment and finishing specifications. The surface finish defines the final appearance, texture, and protective characteristics of materials. This process begins with text normalization performed to address the common OCR misinterpretation of finish specifications.

These include the replacement of 'ES _' with 'ES', and the systematic replacement of punctuation marks that are frequently misread during OCR processing. Exclamation marks are converted to '1' when they appear in specification contexts, addressing a common OCR confusion pattern. For each finish specification in the database, the system applies string normalization and searches within the processed OCR text using both exact matching and the character-substituted variants.

Special case handling addresses industry-specific terminology. For instance, when the term 'protective' is identified within the document text, the system automatically adds 'protective oil' to the finish specifications, recognizing this as a standard industry term. This attention to terminological precision reflects the critical importance of accurate finish specification in manufacturing and compliance contexts.

All extracted information related to material names, material specifications, and surface finish details is stored in a structured Excel-based database. The process begins by uploading a folder containing the required PDF mechanical drawings, after which the automated extraction workflow is initiated. Each PDF drawing is processed sequentially, and the extracted results are systematically stored in the Excel database. The structured database maintains key attributes for every processed drawing, including the PDF drawing name, material name, material specifications, and associated surface finish information. This organized representation of extracted data enables efficient retrieval and analysis. As a result, the proposed system significantly supports product engineers in making faster and more informed design and material-related decisions, while reducing the need for manual interpretation of engineering drawings.

3. INFORMATION EXTRACTION FROM 3D MODELS

The 3D model parts of any hydraulic or relevant industry are primarily of 3 types – STEP, PRT, and ASM. These files may be generated by different software and contain different metadata and compositions. The STP files are normally generated through the Autodesk platform software and are widely usable across different cross platform 3D model handlers. However, being a proprietary data format, use with other software requires export and conversion into another suitable format. The PRT and ASM file types are more common

and can be generated using a wider variety of software, such as PTC Creo Parametric and Siemens NX.

In terms of complexity, the STP files are easiest to handle as they contain suppressed parts and hence can be directly visible and accessed. In most cases, these types of files contain metadata as well, as they are built using a standardized method and hence require some information related to area and material to be already present in the model. The PRT files are the second easiest type of file to handle as they contain all required sub parts arranged under the parent part. However, the issue occurs when the subparts have failures during building or are damaged as part of the migrations from one platform to another. This can occur either if the model was not downloaded or imported incorrectly, or if the model itself has errors while building. The ASM files are the most difficult to handle as directly opening the ASM alone provides no information other than envelope; it requires the individual PRTs inside it to be handled as separate files. Also, in case of nested ASMs it should be ensured that the incorrect PRT is not chosen as seen in Fig. 3. Additionally, any damaged or broken ASMs should be automatically filtered out and not allowed to proceed with data extraction.

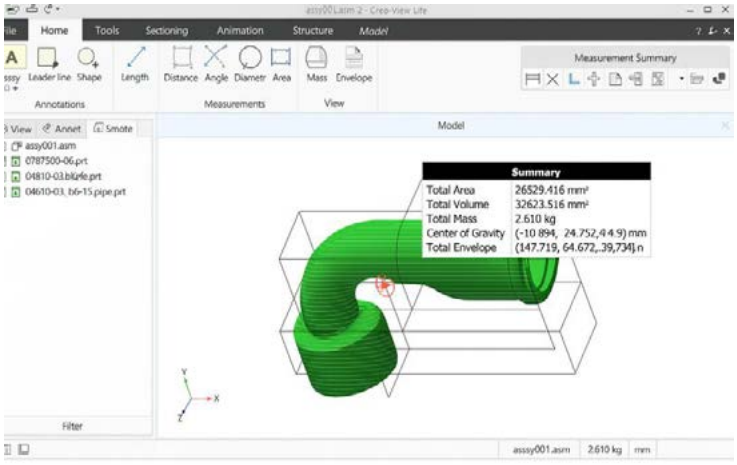


Fig. 3. Representative dummy structure of assemblies and Summary window

3.1. Software to view 3D model data

To get the area, volume, and other descriptive information about the 3D models, a software is required where one can access and open every type of file and additionally, give indicators of damaged cases without any false positives. Generic options such as Creo Parametric 3.5.0, SolidWorks and Autodesk Inventor were explored and referred to through literature. Creo Parametric allows easy access to PRT and ASM files; however, it requires manually intensive conversions to open STP files. Additionally, it offers viewing of metadata parameters, provided that the built file has metadata. For SolidWorks, the file formats are not always handled properly and give a lot of launching errors. For the Autodesk Inventor,

it is good for viewing and accessing STP files, but for PRT and ASM files it does not contain a lot of the information needed for data entry, like volume, mass, and more [7].

The reason that these software do not work with optimal efficiency is because they are predominantly parametric software, i.e., they work best with models which already have predefined settings and parameters as a part of their design. However, since most models made in a corporate environment are a result of the interaction of multiple software and their versions, the parametric measures often are hidden or removed with user interaction. Hence, there was a requirement for software that could not damage the entities of the model itself but just give the information in a collected way.

For this purpose, the most optimal software to be found was Creo View Lite 8.1. This software comes as a part of the PTC suite of software and is a viewer that can open STP, PRT and ASM type files. Additionally, it can indicate which subparts of these models are damaged using warning marks and it can indicate "broken" assemblies in red as well as observed in Fig. 4. The Creo View window has an interesting feature known as the Summary feature under the markups section, that shows information about the model regarding the Area, Volume, Mass, Centre of Gravity and Envelope. This information is highly accurate and useful to us as it contains data about all the parameters that can be needed as part of the SAP data entry. However, this tool does not save the results of the summary anywhere directly. Hence, an approach was required to save the summary data in some format for use later.

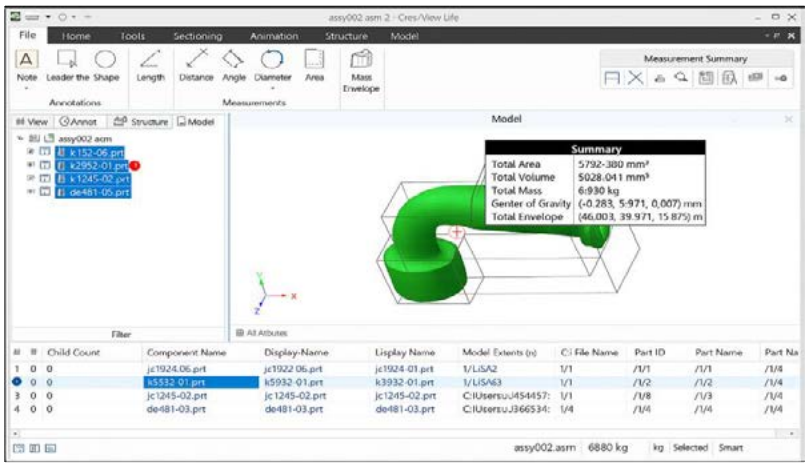


Fig. 4. Representative locations of Summary window with broken parts marked in red

3.2. Using Metadata Extraction

An initial attempt to extract the data was done using Creo Parametric's metadata extraction. This involved linking Creo Parametric with a Creoson API server that could intermittently ping the software in the background to access metadata on its own. However, this process observed multiple limitations. The metadata is only present for around 10% of total files (30,000 STP and 20,000 ASM files that we processed), which amounts to a very less

percentage. Also, the API can only handle up to 15 fields at once instance, requiring the instances to be stopped and restarted after every 15 files.

The metadata was affected by certain software defined settings which might not be correct for every model. For example, the automatic material assigned for a lot of the models was stainless steel; however, this was the wrong material and hence manipulated the mass value from the actual value.

Owing to all these issues, the OCR-based approach was chosen as the best method to extract the Summary information from the Creo View lite software's summary window.

3.3. The Process - Extracting files

The first step involves choosing the correct file for information extraction. The arrangement of the files is in a nested folder–inside-folder arrangement. The correct file is the one that is inside the folder starting with the keyword "MD" and contains the latest revision. At times, the file may be of a different name than the one inside the folder, hence a smart adaptive logic is needed to find the naming inside which is closest to the given part name. For this purpose, match-by-similar name logic was used, and this was verified with the domain experts to ensure the correct file was being loaded.

The file loads up and opens in Creo View. The first thing to do is to tick the checkbox next to the filename, which makes the file visible.
This is where the OCR process starts.

3.4. Optical Character Recognition Techniques for Different Texts

Optical character recognition finds different ways of working in different use cases. For the clicking of the checkbox at the first box, an approximate coordinate region is given to the system within which it looks for the text entry. In Creo View, the checkbox is at the upper left quadrant region of the window, so the OCR is made to act in that region initially only. The text is of the font Calibri, with a font size of approximately 12-14 pixels. As per Park and Shin [8], for such conditions, Tesseract OCR or EasyOCR provide the best results. For recognizing small Calibri text in paragraph form, Tesseract OCR generally offers better accuracy than EasyOCR due to its optimized handling of clean, printed fonts and structured layouts. Multiple methods of OCR were also compared, such as TrOCR (Transformer OCR, used by Microsoft Tools) [9], and Mistral OCR [10]. A comparison was done between all the different possibilities using a variety of parameters.

Feature	**Tesseract (v4/5)**	**EasyOCR**	**TrOCR (Base/Large)**	**Mistral OCR (LLM)**
Architecture	LSTM + Statistical	CRNN (Deep Learning)	Vision Transformer	Large Language Model
Best For	Scanned Docs, High Res	Noisy/Scene Text	Handwritten/Curved	Semantic Understanding

Inference Speed	Extremely Fast (<0.2s)	Fast (~0.5s - 1s)	Slow (~3s - 8s)	Slowest (API/GPU heavy)
Model Size	~20 MB	~80 MB	~1.5 GB+	~4 GB+ (or API call)

While EasyOCR is faster and easier to set up, it tends to struggle with small font sizes and produces more errors in such cases where there are multiple words close to each other. Tesseract, though slower, benefits from extensive customization options and consistently delivers more precise results for small, well-formatted text like Calibri. This is due to the differences in the way they target individual frames, as seen in Fig. 5.

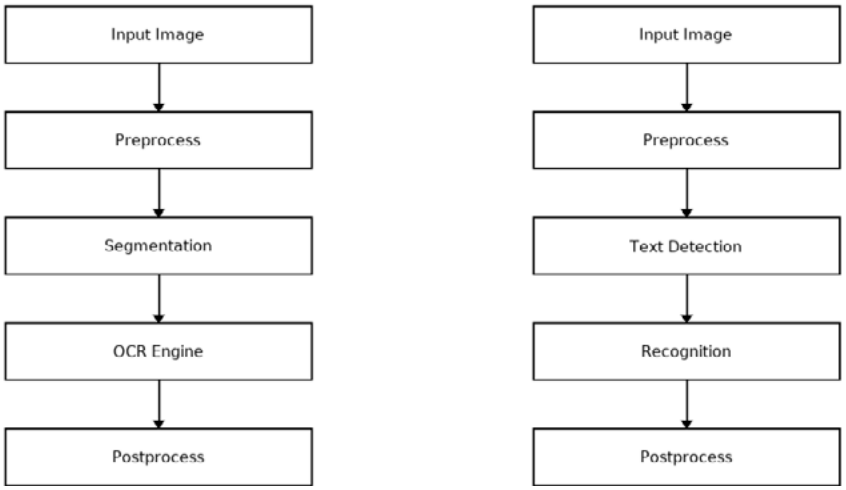


Fig. 5. Comparison of Tesseract OCR and EasyOCR in terms of working

Thus, a combination of EasyOCR and Tesseract OCR is used. First, Tesseract OCR detects smaller text like names of file and subparts, and EasyOCR detects the bigger text like markup subtab and summary option.

Tesseract OCR includes a set of configurable modes known as page segmentation modes (PSMs), which are essential for tailoring how the engine interprets the layout of text in an image. These modes are particularly useful when dealing with different types of documents or image structures, such as full pages, single lines, or isolated words. For example, PSM 3 is the default mode and works well for general documents with multiple blocks of text, while PSM 11 is ideal for images containing a single, uniform block of texts such as a cropped paragraph in Calibri font. These modes help Tesseract decide whether to look for multiple columns, lines, or just a single word, which directly impacts the accuracy of OCR results. Selecting the right PSM is crucial, especially when working with structured fonts like Calibri, as it allows the engine to better understand the spatial arrangement of characters and lines, reducing recognition errors and improving overall performance [11].

For ASM files, the OCR can find files which have broken PRTs as present in it as well. Often, the name of the PRT may contain a drop-down triangle next to it, indicating the presence of 1 or more nested ASMs. In such nests, the name of the PRT and the ASM may be the same, hence a 50 to 100-pixel margin to the bottom of the last entry needs to be reread for confidence. The broken ASM is found by applying a red mask over the images of the summary. If a red region is identified in the region of the ASM, it means it is a broken file.

3.5. OCR for Data Extraction from Snapshots

To extract the data, pattern matching is performed to get entries of area, volume, and other information from the summary snap. Generally, the data follows a fixed structure of ~ symbol followed by the value followed by 3 decimal points. However, it may also contain exponential marks like +/- e^0, which need to be interpreted as scientific symbols. This is done using Tesseract OCR, which finds good application in this case as it works for mathematical text.

Additionally, the text might be present in multiple regions like top right, middle, bottom right, etc. Hence, a fallback logic is added starting in anticlockwise order across the image to read if not found in any region. For the envelope, it may contain data such as “Enve”, “ENVELOP” and other formats, hence all possible combinations of case and letter mappings are considered. Once the data is extracted for each file it is automatically added to an excel. This principle can also be extended to cases involving fluid power drawings specifically – involving complex wiring and/or component symbols. The relevant character / symbol matching in this case simply changes but follows the similar generic principles of searching by character in multiple locations of a document as fall back, or mapping the mis detected characters to their nearest matches based on historical interpretation of that character based on testing with a sample dataset.

4. RESULTS AND DISCUSSION

The entire pipeline of the OCR process for both 2D PDF drawings and snapshots of 3D models involves the need of a dedicated system with stable to high configuration settings. A PC with a high configuration of not less than 32GB RAM and a GPU was used to perform the OCR for the drawings and to capture the snapshots for the 3D model information extraction. The pipeline follows the mass download of all required files onto the dedicated system, followed by the identification of the relevant files to be processed using a naming convention as specified by a product engineering team.

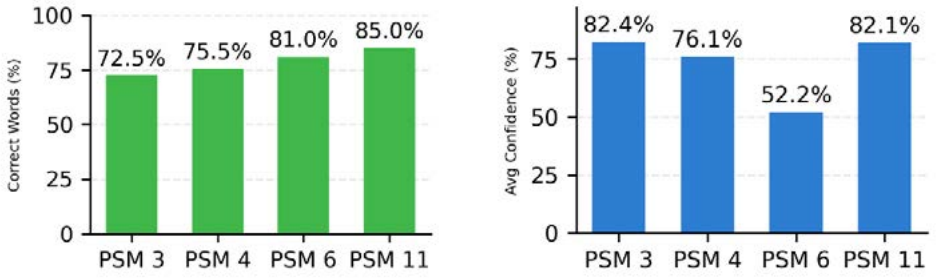
In case of the PDF files, they are converted into images (as discussed in Section 2) and processed. Overall, each file takes around 20-30 seconds to be processed from end to end. The proposed system for text extraction from PDF drawings was tested on a collection of PDFs mechanical drawings. This solution was tested on a sample of 30 engineering drawings in PDF format. Material information was successfully extracted from 90% of the drawings with 79% accuracy. For material finish information, data was extracted from 83% of the drawings with 72% accuracy. By implementing the proposed solution, manual effort can be reduced by approximately 76%, resulting in significant time savings and improved efficiency in processing engineering drawings. Unsuccessful extraction of information from PDF was due to low-quality and unreadable drawings. The database matching step helped

filter out irrelevant or misrecognized text, ensuring high-quality results. The final structured output in excel format enabled engineers to take decisions related to the material specifications.

As per the observations and discussed Section 3.4, for the 3D models text extraction, EasyOCR performs better in individual character cases, but Tesseract OCR performs better in cases of repeated words. For example, for occurrences like “Volume: 12345.67”, EasyOCR is better at interpreting the relevant data; however, when it comes to compound words such as “Total Volume”, Tesseract OCR performs better. Hence, EasyOCR is first used to quickly extract text from the snapshot for an initial overview and to give a general idea about locations of words. Then, Tesseract OCR with different PSM settings is applied to the same file for more detailed results.

To accurately detect word combinations like ‘Area’, ‘Total Area’, and ‘Volume’, analysis of different PSM modes of Tesseract was done. Fig. 6(a) shows that higher PSM values generally result in more correctly detected words for a given file. However, as seen in Fig. 6(a) when comparing based on confidence of the individual PSM modes independently, it was observed that PSM 3 and PSM 11 gave good results in terms of detection confidence. The only case that seemed to act poorly in this case was the PSM 6, with a confidence of around 52%. This did not correlate with the accuracy-based findings in Fig. 6(a). This meant that no PSM configuration alone would give accurate and confident results.

Thus, the best approach was using the average result of all the different modes. When tested on a sample batch of 100 files, it was observed that this method gave balanced results. The logic was finetuned and scaled by mentioning the exact possible combinations of words that could be seen in the screen capture and estimating fallback regions across the image to determine which region contained the numeric information. For example, the image could contain the information in either the middle region or the upper right region of snapshot, and hence the code was arranged to look at those regions in order of possibility and determine the final reading from only one.



(a) Accuracy of detecting correct words for different modes

(b) Corresponding confidence score for each mode in detecting the words

Fig. 6. Performance of Tesseract OCR page segmentation for accuracy of detecting correct words and corresponding confidence score for different modes

Once the information was extracted, it was seen that the extracted text and numeric values for the different classes like area, volume etc., have some amount of overlap due to their occurrences appearing very close to each other and the OCR not being able to understand which unit at the end of the numeric value belongs to which class. For example, as seen in Fig. 7, some of the values of Area are interpreted as Volume and some values of Volume are often interpreted as Envelope. The cases with more than 2 overlaps are highlighted in green. However, they are filtered out on the basis of the error handling strategies which involves extracting the values at least twice in one iteration to check if the value extracted belongs to that class actually or not. Coupled with code to declare some preset formats of the text (E.g., volume values start with ~, integral value of the number will always be followed by 3 decimal points, etc), it highly improves the accuracy of text extraction from the snapshots irrespective of the location of the text.



Fig. 7. Confusion Matrix indicating the accuracy of detection of words as per category

The logic was finalized on a batch of 10,000 ASM files, of which only 4700 were healthy unbroken files whose information we could enter into the system. Out of a total of 23,500 words/phrases (4,700 healthy file entries * 5 columns), 21,900 were detected correctly. This was verified using an automation script checking generic conditions (incomplete words, white spaces, etc) and manual quality checking. The same method when scaled to all ~50,00 files led to an overall accuracy of 93.1% for the text interpretation logic, which was in-line with the accuracy seen on the batch of 10,000 files. While the paper shows samples of dummy parts, the logic has been tested on actual industrial partner parts across multiple divisions and types.

5. CONCLUSIONS

This system demonstrates a practical approach for extracting textual information from mechanical drawings using OCR-based techniques. The methodology of converting PDFs into images, segmenting them into four quadrants, and applying OCR provides significant improvements in recognition accuracy. By matching extracted results with a database, the system ensures that only relevant and valid information is retained. This helps in reducing errors commonly associated with OCR in technical drawings. The data extraction from 3D models offers an opportunity to expedite a process that would otherwise require extensive manual intervention with high chances of error. Additionally, a large database of readily available dimension and weight possibilities might also help in creating models for interpolation or estimation in cases of those models wherein a stable build is not present. In the field of fluid power and its related disciplines, the adoption of methods that minimize repetitive manual tasks and automate data collection offers significant advantages. By streamlining these processes, organizations can not only improve the efficiency of data management but also enhance their ability to estimate and identify optimal design solutions for future projects. Automation reduces human error, accelerates workflows, and ensures that valuable engineering data is captured consistently and accurately.

Importantly, the scope of this approach is not confined solely to technical drawings within the fluid domain. The same strategy can be extended to a wide range of fluid power documentation, including power circuitry diagrams, hydraulic schematics, pneumatic layouts, and design review charts. Because the methodology does not rely on project-specific retraining or exhibit bias toward a particular type of drawing, it remains flexible and broadly applicable across different design contexts.

This universality means that engineers and designers can apply automated data handling techniques to diverse aspects of fluid power systems without the need for specialized customization.

6. FUTURE SCOPE

The strategy discussed is highly scalable and can be replicated with minor fine tuning for different types of drawings and models across industries. OCR-powered vision systems observe and automate conversion systems for electronics components as well, providing detailed information of components at each process step and saving thousands of engineering hours. The collected information can be implemented in generative AI to generate drawings and models of a given configuration on its own. To integrate additional fluid power circuitry and documentation, a modification would be needed to understand symbolic text instead of regular numeric text. With a combination of OCR powered text matching and location-based template matching, the symbols and nearby text can be found in a similar manner. Additionally, conversion of the relevant drawing circuitry into a suitable image or PDF format will also be viable to simplify the complexity of working with different file types.

7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support from the different Danfoss divisions, whose 3D models and drawings were instrumental in validating this solution. The domain expertise

and technical guidance provided by the different divisions were invaluable to the successful validation of this work. The authors thank Satish T., Shilpi S., Ketan M., Bernie B., Hemant S. and Prajakta K. We would also appreciate support from Jeff H., Rahul V., Prashant J. and leadership of Danfoss India Innovation Centre (DIIC). The authors highly appreciate Dr.-Ing. Robert Rahmfeld for his unwavering help and support.

8. REFERENCES

[1] Jamieson, L., Francisco Moreno-García, C. & Elyan, E. A review of deep learning methods for digitisation of complex documents and engineering diagrams. Artif Intell Rev 57, 136 (2024). https://doi.org/10.1007/s10462-024-10779-2

[2] A. Kae, G. Huang, C. Doersch and E. Learned-Miller, "Improving state-of-the-art OCR through high-precision document-specific modelling," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 1935-1942, doi: 10.1109/CVPR.2010.5539867

[3] L. Li, Y. Wei, Z. Xie, X. Yang, Y. Song, P. Wang, C. An, T. Liu, S. Li, B. Y. Lin, L. Kong, and Q. Liu, "VL-RewardBench: A challenging benchmark for vision-language generative reward models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025, pp. 24657–24668.

[4] R. Smith, "An Overview of the Tesseract OCR Engine," Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Curitiba, Brazil, 2007, pp. 629-633, doi: 10.1109/ICDAR.2007.4376991.

[5] E. Tyflopoulos and M. Steinert, "A comparative study of the application of different commercial software for topology optimization," Applied Sciences, vol. 12, no. 2, Art. no. 611, 2022, doi: 10.3390/app12020611.

[6] E. Öztürk and A. Mesut, "Performance Evaluation of JPEG Standards, WebP and PNG in Terms of Compression Ratio and Time for Lossless Encoding," 2021 6th International Conference on Computer Science and Engineering (UBMK), Ankara, Turkey, 2021, pp. 15-20, doi: 10.1109/UBMK52708.2021.9558922.

[7] J. Villena Toro and M. Tarkian, "Optimizing text recognition in mechanical drawings: A comprehensive approach," Machines, vol. 13, no. 3, Art. no. 254, 2025, doi: 10.3390/machines13030254.

[8] Y. Park and Y. Shin, "Gradual OCR: An effective OCR approach based on gradual detection of texts," Mathematics, vol. 11, no. 22, Art. no. 4585, 2023, doi: 10.3390/math11224585.

[9] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, "TrOCR: Transformer-based optical character recognition with pre-trained models,"

arXiv:2109.10282 [cs.CL], Sep. 2021, last revised Sep. 2022. [Online], https://doi.org/10.48550/arXiv.2109.10282

[10] A. Scius-Bertrand, A. Fakhari, L. Vögtlin, D. R. Cabral, and A. Fischer, “Are layout analysis and OCR still useful for document information extraction using foundation models?,” in Proc. Int. Conf. on Document Analysis and Recognition (ICDAR)*, 2024, Lecture Notes in Computer Science, vol. 14807. Cham, Switzerland: Springer, pp. 171–186, doi: 10.1007/978-3-031-70546-5_11.

[11] Majumder, Mujibur Rahman, et al. "Offline optical character recognition (OCR) method: An effective method for scanned documents." 2019 22nd International Conference on Computer and Information Technology (ICCIT). IEEE, 2019.

Biographies

Vaishnavi Kulkarni is an Associate Engineer at Danfoss India Innovation Centre at Pune. She had received her B.Tech. degree in Mechanical Engineering from Vishwakarma Institute of Technology, Pune. Her research interests include natural language processing and image processing and more.

Anwesha Sen is an Associate Engineer at Danfoss India Innovation Centre at Pune. She had received her B.Tech. degree in Instrumentation and Controls Engineering from Cummins College of Engineering for Women in Pune. Her research interests include computer vision and image processing.

Anik Kumar Samanta is the Lead Data Science and Machine Learning Engineer at Danfoss India Innovation Centre at Pune. He received his Ph.D. (2021) and Masters (2016) degrees from the Indian Institute of Technology Kharagpur. He serves on the editorial board as associate editor of the IEEE Transactions on Instrumentation and Measurement and is an Executive Committee Member of the IEEE Signal Processing Society, Pune Section. His research interests include high-resolution spectral estimation, signal-based fault diagnosis of induction motors and hydraulic pumps, detection theory, and embedded signal processing.

Shrinivas Kulkarni is leading ‘Data Science Technologies’ team as Senior Manager at Danfoss India Innovation Centre at Pune. He received his Ph.D. in Machine Learning from the Indian Institute of Technology Bombay, India. He completed his master’s (M.Sc.) degree in ‘Mechatronics’ from University of Duisburg-Essen Germany. He has 17+ years of experience in reputed organizations like, German Aerospace Centre (DLR)- the Institute of Robotics & Mechatronics, Moog GmbH (Germany), Moog India Innovation Centre, Eaton India Innovation Centre and Danfoss. He has expertise in the field of Mechatronics and machine learning techniques. He has patents to his credit and authored many technical papers in international conferences and journals. His research focus includes mathematical modelling of mechatronic components, intelligent mechatronic systems, Autonomous Systems, Control Systems, blending AI & Data Science Technologies with Industrial requirements.

Enabling Vibroacoustic End-of-Line Testing of APUs Across Varying Fluid Temperatures through ML-based Conditional Anomaly Detection

Maximilian Romeser[1,2,*] and Reinhold von Schwerin[1]

[1]Faculty for Informatics, Ulm Technical University of Applied Sciences, Ulm, Germany
[2]Development of Axial Piston Units, Bosch Rexroth AG, Elchingen, Germany
E-mail: maximilian.romeser@boschrexroth.de, reinhold.vonschwerin@thu.de
**Corresponding Author*

Abstract.

Axial piston units (APUs) are key components of hydrostatic drivetrains, where End-of-Line (EoL) testing ensures product quality by detecting manufacturing or assembly defects before delivery. Vibroacoustic (VA) measurements offer potential for fault detection in roller bearings, but their robustness is challenged by process variations, particularly fluctuating fluid temperatures that influence hydraulic pulsations and lubrication behavior. This study investigates these effects in a replicated EoL scenario, comparing two conditional anomaly detection (AD) approaches from the field of machine learning (ML): a residual based method and a feature augmentation strategy that incorporates fluid temperature into the feature space of a One Class Support Vector Machine (OCSVM), against a baseline using only VA features. The results demonstrate that, while temperature-induced effects are less pronounced than natural serial variability, their impact on VA behavior remains detectable. Importantly, residual-based AD yields improved fault detection performance compared to both the baseline and feature-augmented approaches, provided that sufficient data is available. These findings underline the relevance of accounting for temperature effects in data-driven EoL testing and highlight the potential of ML-based methods to enhance the robustness of VA-guided quality assurance in APUs.

1. Introduction

Axial piston units (APUs) are the central elements of high–power-density hydrostatic drivetrains and hydraulic work drives [1]. Their wide application spectrum is driven by key advantages, including robustness, high pressure capability, precise controllability, and compatibility with electrified drivetrain topologies such as hybrid and fully electric vehicles [1, 2]. These features establish APUs as indispensable components in both established and emerging mobile-hydraulic systems. As the operational heart of such systems, APU condition is decisive for sustained performance and service life; deviations typically show up as elevated noise, reduced efficiency, lowered power delivery, or accelerated wear [3]. Ensuring consistently outstanding acoustic and performance characteristics therefore requires

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

rigorous quality assurance throughout production. In parallel, enhancing the reliability of fluid power systems remains a central research priority [4], addressing both optimized materials and components as well as advanced data analysis methods.

1.1. Motivation for End-of-Line Quality Assurance

While advances in materials and design remain essential, verifying every manufactured unit is crucial to ensure consistently high quality. This final check, known as End of Line (EoL) testing [5], functions as a 100 % inspection gate that confirms acoustic and performance compliance, assembly conformity, and contractual tolerances before delivery. Unlike development tests that explore robustness and failure mechanisms under defined conditions, EoL testing verifies that each product instance meets the validated design and enables targeted rework or sorting when deviations occur [5]. Such measures are particularly relevant during ramp up of new product families, after specification updates, or following supplier changes for critical parts like roller bearings, where process variability is elevated and customer confidence must be secured. Comparable situations often arise in mobile working machines, where new variants are introduced under tight schedules and reliability and acoustic performance expectations are especially high [6]. In addition, EoL results can be linked with Statistical Process Control (SPC) [7] to track hydraulic or performance indicators over time and detect systematic process trends.

Beyond these conventional checks, vibroacoustic (VA) measurement has become a key element of EoL testing, complementing functional and hydraulic evaluations with information on dynamic behavior. VA metrics support quality assurance on two levels: directly, by quantifying noise emissions or tonal content relevant for perceived quality and compliance, and indirectly, by providing sensitivity to internal rotating components [8, 9]. In this way, VA data can reveal anomalies in gears, roller bearings, or tribological interfaces such as the slipper swashplate or valve plate–cylinder block contact [10]. Roller bearings are particularly critical, as they operate under substantial radial and axial loads and respond sensitively to even small deviations in geometry, lubrication, or assembly. Such deviations leave characteristic VA signatures that influence smoothness, efficiency, and wear. By capturing these effects, tailored VA-based verification at EoL not only secures acoustic requirements but also highlights subtle deviations in component behavior. Consequently, fault detection in roller bearings using VA and advanced signal processing has become the subject of several research studies [11, 12].

1.2. Problem Statement and Contribution

The implementation of a robust VA EoL test, however, is challenging, especially in the context of manufacturing hydraulic components, which often involves a complex and diverse product portfolio. Robustness is compromised by two main factors: first, frequent batch- and test-stand-related variations in VA behavior causing abrupt data shifts and drifts [5, 13], and second, the high sensitivity of VA measurements to environmental influences, which can be significant in an industrial setting. Direct disturbances like structure-borne noise from adjacent machinery affect VA measurements [14]. However, a more natural source of variation within the hydraulic system is the fluctuating fluid temperature, which this study focuses on. In analogy to the fluid type [14], it can have a major influence on the VA behavior of APUs since it directly affects the fluid viscosity, which in turn has a known effect on hydraulic pulsations [15]. Furthermore, the fluid also lubricates the roller bearings, and it is known

that their VA behavior is influenced by the lubricant's properties and the fluid temperature, consequently [16]. While this effect may be negligible for healthy bearings compared to the dominant hydraulic pulsations, it is no longer trivial in the case of faulty bearings. For the detection of defective roller bearings at the EoL test stand, it is therefore important to investigate the sensitivity of VA measurements to bearing faults relative to the natural process variation, to which the fluctuating fluid temperature is a major contributor.

However, characterizing such complex system behavior and detecting faults in the rotary group of an APU with conventional SPC methods is not trivial, as these methods are typically designed for monitoring single, univariate process parameters (e.g., via control charts). The high-dimensional and non-stationary nature of raw VA signals makes them poorly suited for such an approach. This challenge motivates the use of machine learning (ML) and deep learning, which has become a growing research area for handling complex data patterns, aligning with both Industry 4.0 initiatives and advanced trends in condition monitoring and fault detection of rotating machinery [17, 18]. This data-driven approach, as supported by the trends in [4], enables new possibilities for ensuring quality and reliability at the EoL stage, where the application of ML techniques offers substantial potential [19].

Based on these aspects, our previous work [20] focused on the implementation of an ML-based pipeline for detecting faulty roller bearings using VA sensors. Different sensor applications, load points, and signal processing techniques were investigated, thereby addressing research gaps related to EoL testing of highly varied products [5], and particularly APUs. To the best of our knowledge, only Gaugel et al. [6] have previously addressed the EoL testing of APUs, though their work focused on optimizing test cycle objectives rather than explicitly modeling environmental influences and considering VA signals. While they mention varying fluid temperatures as a relevant factor, it was not systematically investigated. In contrast, this work asks whether and how ML-based anomaly detection (AD) can be made robust against fluctuating fluid temperatures in a realistic EoL scenario. Thus, the objective is to evaluate different strategies for conditional AD under these varying conditions. This forms the foundation of our study. Accordingly, the contribution of this paper comprises the following aspects:

- Analysis of the influence of fluctuating fluid temperatures on the VA behavior of APUs based on a new dataset from a laboratory EoL testing scenario with manipulated roller bearings
- Evaluation of the robustness of ML-based AD for faulty roller bearings against these temperature variations, using the framework from our previous work [20]
- Comparison of approaches for conditional AD to compensate fluid temperature effects using a One-Class Support Vector Machine (OCSVM).

Compared with [20], the present study fixes load conditions, employs a single sensor, and evaluates an AD approach based on the same overall framework.

2. METHODOLOGY

It is is necessary to provide a more detailed explanation of the individual techniques and methods within the applied framework. For the subsequent analyses and experiments, it is further important to describe the dataset used, even though it originates from the same experimental test campaign as the dataset employed in the investigations presented in [20].

Hence, an in-depth description of the test campaign is omitted, and only the most relevant information as well as the differences between the data used here and those of the previous study are provided. To assess the robustness of fault detection, we additionally apply a consistent evaluation scheme for AD, which is likewise explained in the following.

2.1. Experimental Test Campaign Dataset

An experimental test campaign on a laboratory test rig was conducted to reproduce a realistic EoL scenario. This was necessary because obtaining meaningful data from real EoL tests is difficult, requiring large numbers of faulty units and varied test conditions. The essential setup of the test rig and the sensor arrangement is shown in Figure 1. However, for the present study only the triaxial accelerometer mounted on the test rig is considered. According to [20], measurements in the y-direction provide reproducible results and represent a solid basis for robust VA fault detection. Accordingly, only data from this measurement axis are analysed. This sensor remained installed throughout the entire campaign, whereas the test specimen, i.e., the APU under investigation, was replaced for each test in order to reproduce a realistic EoL scenario. Bent axis type units of the same specification were used, of which seven were available. Nevertheless, by exchanging the rotary groups to imitate serial dispersion due to assembly tolerances, 30 different rotary group-housing combinations could be realized, all of which underwent the same automated test procedure. In a second phase of the campaign, 14 faulty tapered roller bearings were mounted on new rotary groups and installed into the APUs. These were primarily the large tapered roller bearings in the O arrangement that support the rotary group, which were abrasively manipulated at the outer ring as well as at the rolling elements. This procedure resulted in a broad spectrum of fault intensities and localizations, leading to correspondingly different anomalous VA behaviors.

The standardized test procedure consists of traversing a full factorial set of operating points defined by rotational speed, swash plate angle with respect to displacement, and pressure, with the APU operating in pump mode. Preceding this procedure, however, is a 15-minute run-in phase at a stationary load point specified in Table 1, which, on the one hand, enables comparable conditions within the operating map. On the other hand, this phase more realistically reflects an actual EoL scenario, since the test rig is restarted for each experiment. In particular, the fluid temperatures at the beginning of each test vary, leading to significant differences depending on whether the test rig is started for the first time at the beginning of a day or whether tests are conducted in quick succession with preassembled rotating group-housing combinations. An overview of the distribution of temperatures, measured at the leakage oil drain (see Figure 1 in the hydraulic circuit), is provided in Figure 2. During the 15-minute run-in phase, one interval per minute was recorded at 10 Hz over a duration of 5 seconds, from which the mean value was calculated. Although temperature control within the operating map ensures reasonably stable fluid temperatures, this is not possible during the run-in phase due to self-heating caused by the load valve, resulting in the fluid warm up behavior shown. A notable effect of increased warm up due to the faults cannot be observed.

Table 1: Operating point during the run-in phase

Rotational speed	Displacement	High-pressure	Low-pressure
1000 rpm	Maximum	70 bar	30 bar

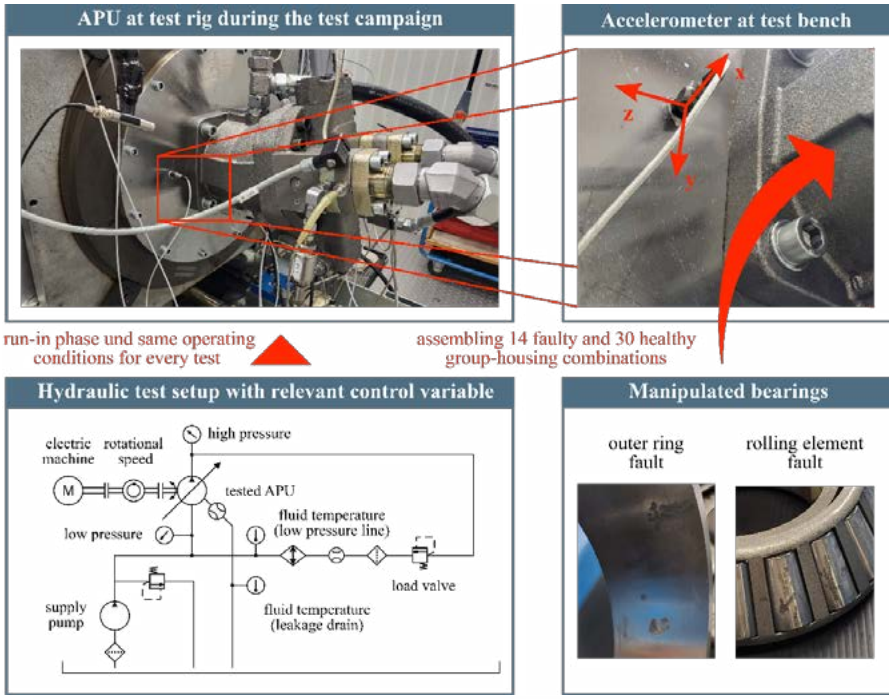


Figure 2: Setup for the experimental test campaign imitating a real EoL sce-

In this study, we explicitly use the VA measurements that were recorded simultaneously with the case drain temperature during the run-in phase, here with a sampling rate of 20 kHz. Since interval measurements during the run-in phase were not available for every test, the dataset is limited to 39 measurements from 25 healthy and 14 faulty units. In total, this results in 39 × 15 time series of length 100000 samples each, together with the corresponding fluid temperatures, which were already averaged and are included as scalar values.

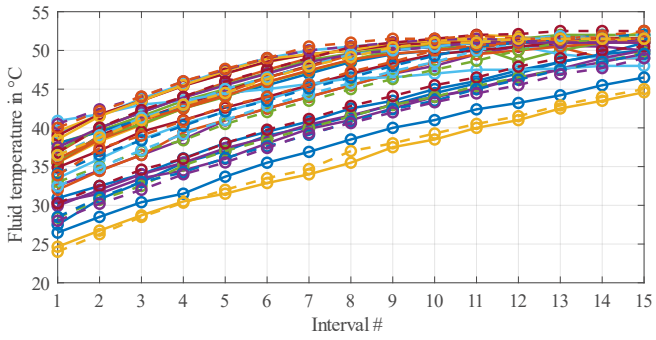


Figure 1: Trends of the fluid temperatures at the leakage drain during the run-in phase, with each color-line-style combination representing

2.2. Vibroacoustic Feature Extraction

For robust fault detection, we must construct informative features that effectively characterize VA behavior while being independent of external disturbances. This reflects

conventional approaches in ML, whereas current research also considers deep learning based methods that do not rely on predefined features. Instead, these features are learned directly by the model, which, however, requires large amounts of data. As such data cannot be assumed in this scenario, we follow previous work on (VA) fault detection and quality monitoring by relying on predefined features, which also enhances interpretability and robustness. In this regard, it is worth noting studies such as [21], which show that the performance of such conventional ML approaches can be competitive with more recent deep learning methods when it comes to detecting roller bearing faults based on VA behavior.

Signal processing plays a decisive role in this context, as it can enhance the fault-related information contained in the data. As demonstrated in our previous work [20], a fundamental step is to transform the time-domain signal into the frequency domain, for instance using a Fast Fourier Transform (FFT). This makes the periodic impulse sequences caused by defect overrollings more visible in the data. Moreover, envelope analysis of the analytic signal—combined with an empirically chosen band-pass filter—yielded almost perfect results in [20], thereby confirming the recommendation of Randall et al. [22] that the character becomes even more evident in the envelope spectrum. Nevertheless, the choice of an appropriate frequency band remains a general challenge and requires substantial domain knowledge, which is not always available in practical applications [11]. For this reason, the present study focuses exclusively on the raw spectrum-based approach. This choice is deliberate to isolate temperature effects on the fundamental signal characteristics and to create a baseline method that does not depend on domain-specific knowledge**.**

For the subsequent feature extraction, only the raw spectrum is utilized. Following the band definitions in Table 2, the spectrum is divided into several frequency regions. For each band, the mean value and the kurtosis are calculated. The mean reflects the overall energy distribution within a frequency band and is sensitive to shifts in load or operating conditions, while the kurtosis quantifies the impulsiveness of the signal content and is well established as an indicator for localized bearing faults. To clearly distinguish both types of features, the mean values corresponding to the frequency bands are referenced as F1, F3, F5, F7, and F9, whereas the kurtosis values are denoted as F2, F4, F6, F8, and F10. This feature set provides complementary information: the mean captures gradual changes in spectral energy, while the kurtosis highlights the presence of impulsive fault signatures. In this way, temperature-induced effects on the spectral characteristics can be systematically captured without relying on application-specific band selection as required in envelope analysis.

Table 2: Frequency bands used for the calculation of the features

Frequency Band	**1**	**2**	**3**	**4**	**5**
Range [Hz]	0 - 120	120 - 360	360 - 840	840 - 1800	1800 - 3200

2.3. Conditional Anomaly Detection using One-Class Support Vector Machine

In the context of AD, we employ a ML model that interprets healthy units as a single class; hence the AD task can be regarded as one class classification. Because only healthy units are used for training and no mapping to multiple labels is performed, while faulty units are reserved solely for evaluation, the approach may also be described as semi-supervised AD. As the base model in this work, we use an OCSVM, which adapts the standard SVM for binary classification to the one class setting and, with a radial basis function kernel,

effectively generates a closed decision region around the input data. The basic operation is as follows: the OCSVM learns a decision function that accepts regions of feature space populated by the training data (healthy units) and rejects regions that deviate from this distribution. A small fraction of training samples may be allowed outside the decision region through a hyperparameter, which controls the trade-off between sensitivity and robustness. For further details regarding the method and its implementation in MATLAB R2022b [23], refer to our previous study [20] and to [24]. Specifically, a radial basis function kernel was used. The kernel scale was selected using a subsampling-based heuristic procedure [23].

Train-test-partitioning is applied accordingly to Figure 3, with the test portion used for evaluation as described below. To prevent data leakage from the test set, the standardization parameters (mean and standard deviation) were computed exclusively from the training data. These parameters were then applied to standardize both the training and the corresponding test data. In addition to the VA features, the model is extended in a second approach by including the mean fluid temperature for each interval. This variant is referred to as temperature-augmented AD (TAD), in contrast to the standard AD approach using only the VA features (SAD). The rationale is that including temperature should allow the model to represent shifts in the data driven by temperature and to learn possible correlations between temperature and the VA features. In this sense the approach can be seen as conditional AD: temperature does not define the nominal behaviour itself but provides context under which a more robust fault detection is to be achieved. Note that alternatives exist: Wißbrock et al. [25], for example, apply conditional isolation forest variants to incorporate specifications and test run conditions, while other studies propose splitting the dataset by test rigs [26].

An alternative or complementary approach to conditional AD involves the use of linear mixed-effects (LME) models. Each VA feature y_{ij} related to an i-th unit or test and its j-th interval can be expressed as

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 n_{ij} + u_i + \epsilon_{ij} \qquad (2.1)$$

where β_0 is the global intercept, β_1 and β_2 are fixed effects representing the systematic influence of fluid temperature T_{ij} and test interval $n_{ij} \in [0; 15]_{\mathbb{N}}$, u_i is the random effect capturing serial dispersion, e. g. unit-to-unit variability, and ϵ_{ij} is the error representing unexplained variation. The residuals r_{ij}, i.e., the deviation of each observation from its expected value given temperature, interval, and unit, can then be computed as

$$r_{ij} = y_{ij} - \left(\beta_0 + \beta_1 T_{ij} + \beta_2 n_{ij}\right) = u_i + \epsilon_{ij} \qquad (2.2)$$

based on the estimated coefficients β_0, β_1 and β_2. These coefficients were estimated for each feature separately using the fitlme function, which implements linear LME models based on maximum likelihood estimation [27]. The residuals r_{ij} serve as compensated features for subsequent AD (see Figure 3). By training an OCSVM on these residuals (i.e. residual-based AD or RAD), the model can learn to detect anomalies relative to the expected behavior under the given operating conditions, thereby reducing false predictions due to temperature shifts.

2.4. *Evaluation Approach*

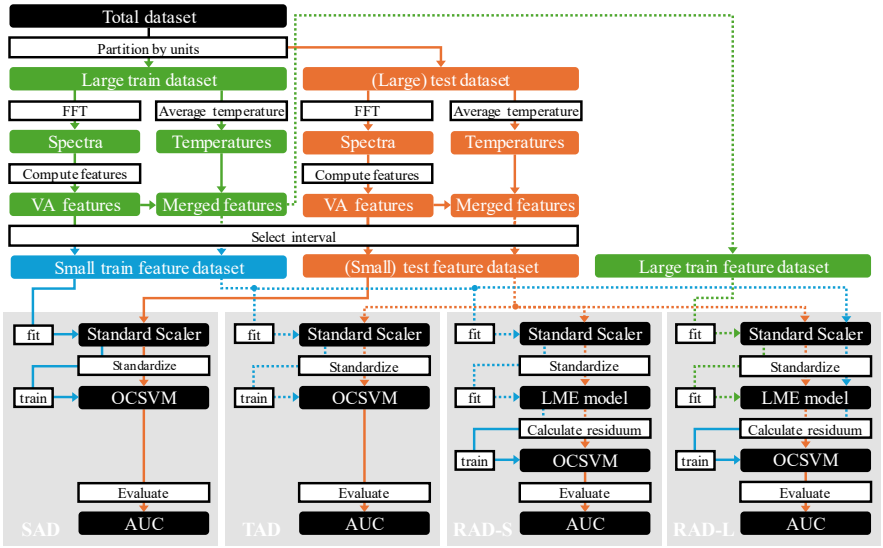


Figure 3: Pipeline for training and evaluating different anomaly detection models. The process includes feature extraction from raw data, data partitioning, and model-specific training and evaluation steps.

The evaluation of the models and of the overall is based solely on the test data, which include a subset of the healthy instances (disjoint from training data instances) and all the faulty instances. Given the limited dataset, five different train-test-splits are generated with respect to the healthy data, in order to obtain statistically more robust statements regarding model performance. The exact partitioning of the datasets depends on the experiments and will be described in detail later. In general, however, training and test data are kept disjoint not only at the instance level but also at the unit level. If multiple intervals from the same unit are considered, they must not be distributed across training and test data within the same split. Furthermore, it should be emphasized that train-test-partitioning applies to the entire pipeline, and consequently, the LME model, i.e. the coefficients β_0, β_1 and β_2, must be estimated exclusively based on training data.

For evaluation, an appropriate metric is employed, which is computed for each of the four splits and then averaged arithmetically. Specifically, the Area Under the Receiver Operating Characteristic (ROC) Curve is used (AUC). The ROC curve characterizes the relationship between the true positive rate (TPR)—i.e., the proportion of correctly identified faulty units—and the false positive rate (FPR)—i.e., the proportion of healthy units incorrectly flagged as faulty. Thus, the AUC is defined as

$$AUC = \int_0^1 TPR(FPR)\partial FPR, \tag{2.3}$$

with

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \tag{2.4}$$

where TP and FP denote the number of true positive or false positive instances, respectively, and FN and TN the number of false negative and true negative instances. Since this relationship inherently depends on the threshold applied to the anomaly scores, integrating over the ROC curve to compute the AUC provides a single threshold-independent value that serves as a robust performance indicator for each model.

3. EXPERIMENTS

Based on the described method for extracting VA features, the following section first presents an analysis of the data from the underlying experimental test campaign with respect to fluid temperature effects, while also accounting for potential time-dependent influences such as run-in, in accordance with the LME model framework. Specifically, an analysis of variance (ANOVA) is performed on the model to quantify the significance and relative strength of both temperature and interval effects across the features and principal components. Subsequently, the results of three distinct experiments are presented and discussed in terms of the robustness of fault detection using the conditional AD approach.

3.1. Analysis of Fluid Temperature Effects in Vibroacoustic Data

For the evaluation of temperature effects, we initially consider only the 25 healthy units in order to exclude fault-related VA anomalies and isolate the inherent variability of the process, without the need to perform train-test-partitioning. To gain a general understanding of the variation within the experimental test campaign, it is useful to apply dimensionality reduction techniques to interpret the ten VA features, derived from the raw spectrum, with respect to temperature effects. For this purpose, principal component analysis (PCA) is employed, which identifies orthogonal principal components (PCs) in descending order of explained variance, spanning the low-dimensional PC space, with the proportion of variance illustrated in Figure 4, alongside the factor loadings represented by arrows in the biplot, which indicate the contribution of each feature to the linear combination of the respective PCs. In particular, the second PC exhibits a correlation with rising temperature whereas isolating other time-dependent effects based on this representation is not feasible. It is, however, evident that the second PC is largely determined by features F1 and F2. Regarding the Pearson correlation shown in Table 3 for both the first two PCs and the individual features, high correlations with fluid temperature can be seen in general for F3 and the second PC, whereas the first principal components appear to reflect other, more random sources of variability, which account for a larger portion of the overall variance in this context.

Table 3: Pearson correlation coefficients of each feature with respect to the fluid temperature

PC1	PC2	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
-0.02	0.39	0.29	0.04	0.47	0.27	0.09	0.03	-0.25	0.12	0.19	0.06

In order to quantitatively assess the extent to which the temperature effect contributes relative to random serial dispersion, while simultaneously separating this effect from other potential time-dependent influences, the aforementioned ANOVA is employed to the mixed-effects model according to (2.3) from the features as well to the first two PCs. According to the p-values summarized in Table 4, a statistically significant effect of temperature is evident for most features as well as for the first two PCs. By contrast, the residual time-related effect

does not reach significance in any case and can therefore be considered irrelevant in the context of this study.

Table 4: P-values from ANOVA indicating the significance of effects (values below 0.05)

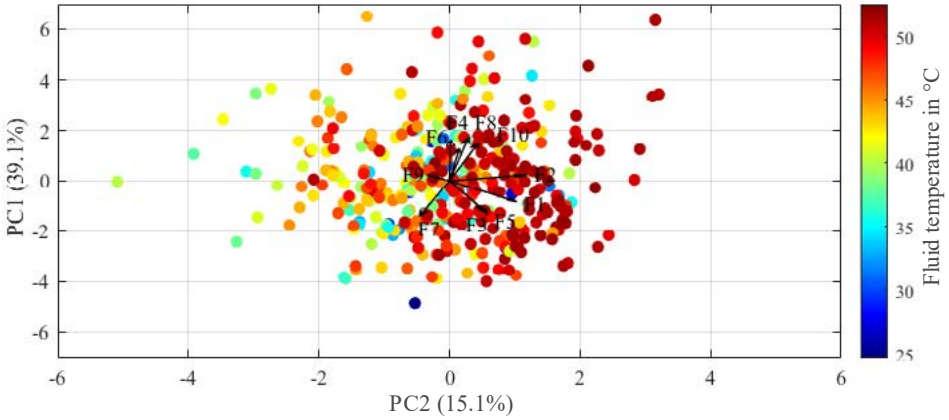


Figure 4: Low-dimensional PC spaces resulting from the PCA of the healthy instances across all intervals, including factor loadings and explained variance

Effect	PC1	PC2	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Temp.	0.00	0.00	0.00	0.49	0.00	0.00	0.09	0.57	0.00	0.02	0.00	0.22
Time	0.33	0.65	0.92	0.64	0.75	0.66	0.73	0.55	0.48	0.62	0.75	0.73

3.2. Fault Detection Results

In this study, two different experiments are conducted to assess the effect of varying fluid temperatures on fault detection and to evaluate the resulting implications for the comparison of the employed conditional AD methods. In the first experiment, training and testing are performed for each interval in isolation. In the second experiment, scenarios are compared with either similar or strongly differing temperature distributions across the training and test sets. In both cases, all intervals are available for selection. The five data splits, as described in Chapter 2.4, are maintained consistently regarding the units in both training and test sets.

In the context of RAD, two approaches for constructing the LME model are considered within the experiments (see Figure 3). In the first approach, i.e. RAD-S, only the data used for training the OCSVM according to the respective split are included (20 instances per split, e.g. one instance per unit for training). In the second approach, i.e. RAD-L, all intervals of the units within the train portion are used, providing a larger basis for characterizing the temperature effect (20 × 15 instances per split, e.g. all instances of one unit for training).

3.2.1. Experiment 1 – Interval-wise Comparison

As shown in Figure 2, the spread of fluid temperatures decreases over the course of the run-in period, suggesting that the VA behavior in later intervals would similarly exhibit reduced variability based on prior observations. However, when considering the AUC values in Figure 5, no clear trend is apparent, implying that here the influence of temperature and other run-in effects is not pronounced. While the second interval tends to show slightly lower

AUC values, overall, the AUC remains relatively stable across intervals when averaged over the splits. In most cases, the fifth split is particularly critical and occasionally yields AUC values below 0.8 when the residual-based AD is applied using the LME model with the larger dataset. For the other splits, however, and on average, the AUC values are considerably higher than those obtained with the smaller dataset or with the standard OCSVM (with or without temperature as an additional feature). Moreover, the larger-dataset LME model exhibits less overall variability, consistently achieving mean AUC values between 0.958 and 0.989, and does not show the negative outlier observed in the second interval.

In analogy to the trends in Figure 5, the exemplary ROC curves averaged over the splits in

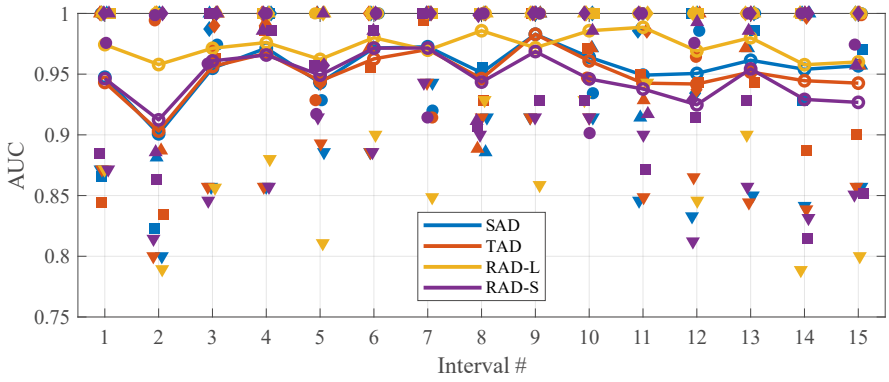


Figure 5: Mean AUC values for all intervals averaged over the splits, with individual AUCs of the splits additionally shown as scatter points with different markers

Figure 6 do not differ strongly from interval to interval. Nevertheless, it can be observed that RAD-L achieves higher true positive rates at the same false positive rate (FPR). For practical EoL testing, performance at a low FPR is crucial, as this minimizes pseudo faults that would lead to unnecessary rework. Notably, at an FPR of 5%, the RAD-L method maintains a true positive rate (TPR) over 90%, a clear advantage over the other methods. Notably, for the last interval, the ROC curve of RAD-S is lower, resulting in a reduced AUC (see also Figure 5). This can be attributed to the narrower temperature spread in this interval, as shown in Figure 2, which limits the LME model for the residuals to a small temperature range.

3.2.2. Experiment 2 – Similar vs Mixed Temperature Level

The previous experiment eliminated non-temperature-related time effects by using a constant interval. However, this approach also limited the variance in the considered temperature levels. To more clearly assess the influence of fluid temperature on fault detection performance, we now compare scenarios with very constant temperature levels to those with highly varying temperatures. In the first case, a temperature level is chosen according to Figure 2 that is traversed by every unit during the run-in; consequently, for each unit, the interval closest to 42 °C is selected. In the second case, all instances—that is, data from all units, runs, and intervals—are pooled, and for each unit a random interval is selected, ensuring that the full spectrum of fluid temperatures observed during the run-in phase is represented in the dataset. To ensure statistical reliability, ten repetitions with different randomly selected intervals per unit are performed for this latter scenario.

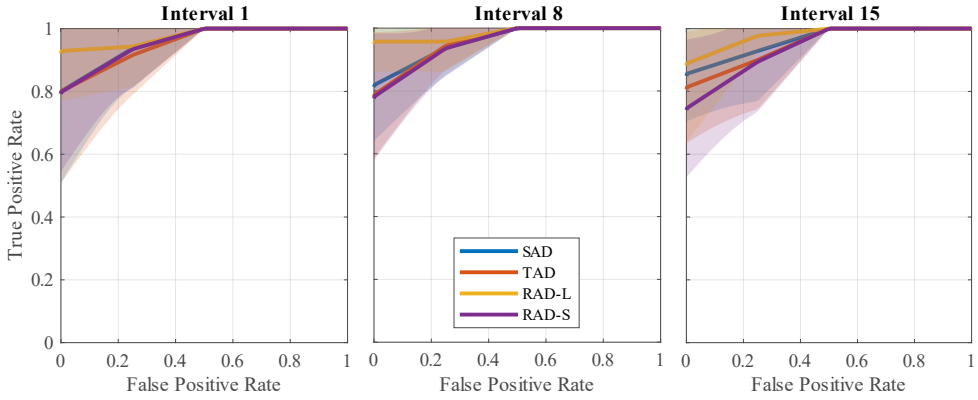


Figure 6: ROC curves averaged over the splits, with the shaded area representing the standard deviation and exemplary curves for three of the 15 intervals shown

The results of this experiment regarding the fault detection performance, measured in terms of AUC, are summarized in Table 5. In line with the expectations about general process variability due to temperature effects, it becomes evident that fault detection works better under comparable temperature levels than when the wide range of temperatures occurring during the run-in phase is taken into account. Importantly, further time-dependent effects cannot be concluded from this experiment, since, as shown by the temperature distributions in Figure 2, different intervals may correspond to the same fluid temperature level of approximately 42 °C across different units. Thus, for the case of comparable temperatures, nearly perfect results can be achieved: with TAD, every split yields an AUC of 1.0, while SAD and RAD-L also reach perfect performance in all but split 5. In contrast, for the mixed-temperature scenario, RAD-L achieves the highest AUC in each split, outperforming the other methods. Consistent with the findings of Experiment 1, split 5 is again the most problematic, likely due to the specific units contained in this subset. Across both scenarios, however, RAD-S consistently underperforms, reaching an average AUC of 0.946 in the mixed-temperature case and 0.892 in the constant-temperature case. Indeed, as shown in Figure 7, this is mainly related to a lower TPR at a FPR of 0, indicating that some faulty units do not exhibit anomalous patterns in the features when using RAD-S, leading to an overlap in the anomaly scores. At this point, the RAD-L–based method, in contrast, achieves a relatively high TPR for mixed temperature levels.

3.3. Discussion and Limitations

The experimental results indicate that fluid temperature during the run-in phase has a measurable effect on the VA behavior of APUs. Feature F9, in the highest frequency band, is particularly relevant for distinguishing healthy from faulty units, as suggested by the choice of the bandpass filter in the envelope-based analysis in [20]. However, the impact of temperature is limited: Table 4 shows significant effects, but correlations in Table 3 and the first principal component (see Figure 4) suggest only minor contributions compared to inherent production scatter. Similarly, experiment 2 or Table 5 shows that fault detection is only slightly affected by mixed fluid temperature levels, and temporal effects across intervals in Experiment 1 (see Figure 5) are minimal.

Table 5: AUC for similar and mixed fluid temperature levels with respect to different splits as well as the average. Since in the case of mixed levels repetitions regarding the mixing of the intervals are performed, the standard deviation is additionally report.

Levels	Method	Average	Split 1	Split 2	Split 3	Split 4	Split 5
Similar	SAD	0.974 ± 0.052	**1.000**	**1.000**	**1.000**	**1.000**	0.871
	TAD	**1.000** ± 0.000	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	RAD-L	0.966 ± 0.071	**1.000**	**1.000**	**1.000**	**1.000**	0.831
	RAD-S	0.946 ± 0.075	0.971	0.957	**1.000**	**1.000**	0.800
Mixed	SAD	0.950 ± 0.057	0.893 ± 0.008	0.987 ± 0.008	**1.000** ± 0.000	0.997 ± 0.006	0.871 ± 0.000
	TAD	0.944 ± 0.048	0.913 ± 0.005	0.950 ± 0.014	**1.000** ± 0.000	0.986 ± 0.000	0.871 ± 0.000
	RAD-L	**0.976** ± 0.050	**1.000** ± 0.000	**1.000** ± 0.000	**1.000** ± 0.000	**1.000** ± 0.000	**0.879** ± 0.023
	RAD-S	0.892 ± 0.069	0.883 ± 0.006	0.886 ± 0.000	0.986 ± 0.000	0.929 ± 0.000	0.779 ± 0.015

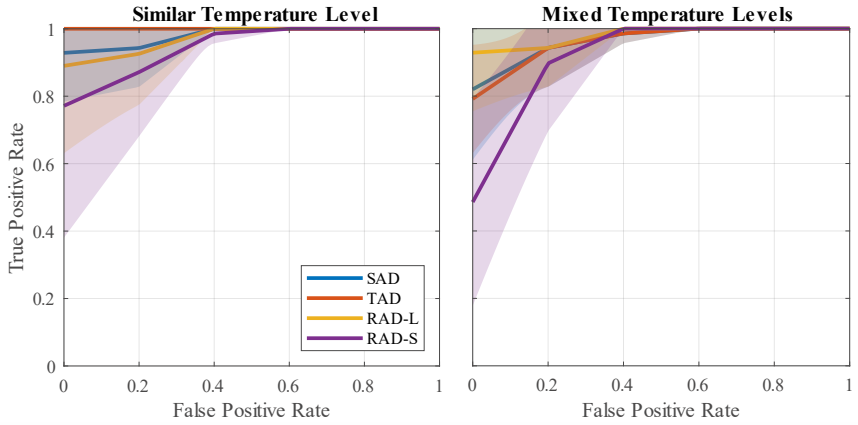


Figure 7: ROC curves averaged over the splits (and repetitions), with the shaded area representing the standard deviation and curves for similar and mixed temperature levels

Adding temperature as an explicit feature in the OCSVM provides little additional benefit. Performance does not deteriorate, and in the constant-temperature scenarios in experiment 2, results occasionally improve, likely coincidentally, since the feature is effectively constant. Limited variance and the implicit encoding of monotonic temperature effects in the VA features reduce the value of including temperature alone. In contrast, residual-based preprocessing or RAD clearly improves fault detection. By modeling monotonic temperature effects using a linear residual model, the method reduces systematic shifts before

OCSVM evaluation. Robustness requires a sufficiently large dataset: The limited data from a single interval makes the RAD-S approach prone to unstable coefficients in (2.2). The resulting model risks underfitting, as it cannot reliably separate systematic temperature trends from the dominant production scatter. In contrast, RAD-L, with all intervals included, provides a simple and interpretable data-driven correction. While this reduces the risk of underfitting, the general challenge with small industrial datasets remains a potential for overfitting the AD model itself, a risk that was mitigated here by using a simple OCSVM.

Nevertheless, several limitations must be considered for industrial EoL testing. Variability in APU specifications, tolerances, and supplier components may cause distributional shifts in features that neither OCSVM nor residual models can fully accommodate. Operational conditions such as pressure, rotational speed, swash angle, and initial oil temperature are more variable in production than in the laboratory. While temperature was explored as a conditioning variable, real EoL scenarios involve multiple interacting factors, limiting the predictive power of a single conditioning feature. Finally, the limited number of faulty units constrains statistical generalization, particularly for residual-based models requiring sufficient coverage to estimate systematic effects accurately.

4. Conclusion and Outlook

This study investigated the detection of APUs with faulty roller bearings in a laboratory-based EoL testing scenario, focusing on VA measurements and the influence of fluid temperature variations. Across two experimental setups, semi-supervised ML methods were employed for AD, comparing a conventional OCSVM approach with temperature-augmented and residual-based strategies for conditional AD. While temperature-induced variations were relatively small compared to the inherent production scatter, residual-based preprocessing proved effective in compensating for monotonic temperature effects, leading to improved fault detection performance and more stable AUC values compared to standard OCSVM or temperature-augmented OCSVM. Importantly, this approach not only enhances robustness against systematic disturbances but also increases interpretability, as the linear residual model clearly quantifies the temperature contribution to the VA features. These results highlight that even in the presence of moderate environmental variations, a semi-supervised residual-based pipeline can reliably identify faulty units within the scope of this study, demonstrating its potential relevance for real industrial EoL testing, where multiple sources of variability must be accounted for.

Future research should address the study's limitations by investigating additional sources of variability, such as background noise or the diversity of product variants, which may induce substantial data shifts. In the latter case, alternative conditional AD approaches could be explored, for instance methods capable of compensating for variations in categorical variables such as component suppliers or APU specifications. Regarding the VA behavior itself, non-linear modeling techniques may also be promising, provided that sufficiently large datasets are available. More generally, systematic, in-line data collection under realistic production conditions—including comprehensive sensing of potential disturbance factors—would enable more robust, data-driven fault detection and provide a foundation for extending these methods to industrial EoL testing.

5. REFERENCES

[1] S. Haug, Ganzheitliche Optimierung einer Axialkolbenpumpe durch bedarfsangepasste Entlastung tribologischer Kontakte, Bd. 95, Karlsruhe, Germany: KIT Scientific Publishing, 2022. DOI: 10.5445/KSP/1000140024.

[2] P. C. Palavicino, B. Sarlioglu, D. Bobba, W. Lee and T. Minav, „Electrification of hydraulic systems using high efficiency permanent magnet motors," in *12th IFPC*, Dresden, Germany, 2020. DOI: 10.25368/2020.53.

[3] C. Lu, S. Wang and V. Makis, „Fault severity recognition of aviation piston pump based on feature extraction of EEMD paving and optimized support vector regression model," *Aerospace Science and Technology,* Bd. 67, pp. 105-117, 2017. DOI: 10.1016/j.ast.2017.03.039.

[4] „NFPA Technology Roadmap," NFPA, 2023. [Online]. Available: https://www.nfpa.com/technologyreports. [Accessed September 4, 2025].

[5] V. Hirsch, P. Reimann and B. Mitschang, „Data-Driven Fault Diagnosis in End-of-Line Testing of Complex Products," in *6th IEEE DSAA*, Washington D.C., USA, 2019. DOI: 10.1109/DSAA.2019.00064.

[6] S. Gaugel and M. Reichert, „Data-driven multi-objective optimization of hydraulic pump test cycles via wrapper feature selection," *CIRP Journal of Manufacturing Science and Technology,* Bd. 50, pp. 14-25, 2025. DOI: 10.1016/j.cirpj.2024.01.004.

[7] W. A. Shewhart, Economic Control of Quality of Manufactured Product, Mansfield Centre: Martino Fine Books, 1931. DOI: 10.1002/j.1538-7305.1930.tb00373.x.

[8] T. Bischof, „Root Cause Analysis of Acoustically Suspicious Transmissions During EOL Testing," in *7th ISNVH*, Graz, Austria, 2012. DOI: 10.4271/2012-01-1528.

[9] T. Bertolini and T. Fuchs, „Schwingungen Und Geräusche," in *Handbuch Elektrische Kleinantriebe*, Bd. 1, Berlin, Germany, De Gruyter Oldenbourg, 2020, pp. 358-375. DOI: 10.1515/9783110565324.

[10] P. Casoli, M. Pastori, F. Scolari and M. Rundo, „A Vibration Signal-Based Method for Fault Identification and Classification in Hydraulic Axial Piston Pumps," *Energies,* Bd. 12, Nr. 5, 2019. DOI: 10.3390/en12050953.

[11] C. Xiao, H. Tang, Y. Ren, J. Xiang and A. Kumar, „A fault frequency bands location method based on improved fast spectral correlation to extract fault features in axial piston pump bearings," *Measurement,* Bd. 171, 2021. DOI: 10.1016/j.measurement.2020.108734.

[12] J. Du, S. Wang and H. Zhang, „Layered clustering multi-fault diagnosis for hydraulic piston pump," *Mechanical Systems and Signal Processing,* Bd. 36, pp. 487-504, 2013. DOI: 10.1016/j.ymssp.2012.10.020.

[13] J. A. E. Bonart, P. Huemmer, J.-B. Maurice, F. Trautmann and W.-G. Drossel, „Transfer of NVH limits for highly-integrated electrical drives between end-of-line test benches," in *13th IEEE SDEMPED*, Dallas, USA, 2021. DOI: 10.1109/SDEMPED51010.2021.9605505.

[14] B. Wieczorek, L. Wargula, O. Zharkevich, L. A. Gierz, L. Wojciechowski and K. Perz, „The study of the impact of oil type and throttling pressure on the vibrations of a gear pump," *Advances in Science and Technology Research Journal,* Bd. 19, Nr. 5, 2025. DOI: 10.12913/22998624/201216.

[15] X. Yimin, L. Zhenghui, T. Shunhui, L. Wei and X. Shiqi, „Influence of Oil Characteristics on Flow Pulsation of Axial Piston Pump of Large Displacement Under Solid-Liquid-Temperature Coupling," *Journal of South China University of Technology,* Bd. 51, Nr. 9, 2023. DOI: 10.12141/j.issn.1000-565X.220651.

[16] I. Chatzisavvas, P. Koutsovasillis and B. Schweizer, „Influence of the oil temperature of thrust bearings on the vibratory behavior of small turbochargers," in *11th ICoEV*, Ljubljana, Slovenia, 2015.

[17] M. Hakim, A. A. B. Omran, A. N. Ahmed, M. Al-Waily and A. Abdellatif, „A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations," *Ain Shams Engineering Journal,* Bd. 14, 2023. DOI: 10.1016/j.asej.2022.101945.

[18] J. Yu and Y. Zhang, „Challenges and opportunities of deep learning-based process fault detection and diagnosis: a review," *Neural Computing and Applications,* Bd. 35, pp. 211-252, 2023. DOI: 10.1007/s00521-022-08017-3.

[19] C. Nunes, E. J. Solteiro Pires and A. Reis, „Machine Learning and Deep Learning applied to End-of-Line Systems: A review," *WSEAS Transactions on Systems,* Bd. 21, 2022. DOI: 10.37394/23202.2022.21.16.

[20] M. Romeser and R. von Schwerin, „Data-Driven Serial Testing of Axial Piston Units: Integrating Signal Processing and Machine Learning for Roller Bearing Fault Detection," in *19th SICFP*, Linköping, Sweden, 2025. DOI: 10.13052/rp-9788743808251A30.

[21] C. Bienefeld, F. M. Becker-Dombrowsky, E. Shatri and E. Kirchner, „Investigation of Feature Engineering Methods for Domain-Knowledge-Assisted Bearing Fault Diagnosis," *Entropy,* Bd. 25, Nr. 9, 2023. DOI: 10.3390/e25091278.

[22] R. B. Randall and J. Antoni, „Rolling element bearing diagnostics - A tutorial," *Mechanical Systems and Signal Processing,* Bd. 25, pp. 485-520, 2011. DOI: 10.1016/j.ymssp.2010.07.017.

[23] „fitcsvm," MathWorks, [Online]. Available: https://de.mathworks.com/help/stats/fitcsvm.html. [Accessed December 8, 2025].

[24] B. Schoelkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson, „Estimating the Support of a High-Dimensional Distribution," *Neuro Computing,* Bd. 13, Nr. 7, 2001. DOI: 10.1162/089976601750264965.

[25] P. Wißbrock, L. Koschek, Z. Ren and W. Nejdl, „Enhancing quality inspection of highly variant geared motors," *Applied Acoustics,* Bd. 235, 2025. DOI: 10.1016/j.apacoust.2025.110687.

[26] D. Muhr, S. Tripathi and H. Jodlbauer, „Divide and conquer anomaly detection: A case study predicting defective engines," *Procedia Manufacturing,* Bd. 42, pp. 57-61, 2020. DOI: 10.1016/j.promfg.2020.02.090.

[27] „fitlme," MathWorks, [Online]. Available: https://de.mathworks.com/help/stats/fitlme.html. [Accessed December 30, 2025].

BIOGRAPHIES



Maximilian Romeser received his B.Eng. in Automotive Engineering from Ulm Technical University of Applied Sciences in 2021 and his M.Sc. in Automotive Engineering from the University of Stuttgart in 2023. Since 2023, he has been pursuing a Ph.D. at Ulm Technical University of Applied Sciences in cooperation with Bosch Rexroth AG. His research focuses on data-driven methods for industrial serial testing of axial piston units.



Reinhold von Schwerin received his Ph.D. in Computational Science from Heidelberg University and is now a professor for Data Science, Machine Learning and Artificial Intelligence at Ulm Technical University of Applied Sciences. Furthermore, he serves on the board of Directors of DASU - Transferzentrum für Digitalisierung, Analytics & Data Science Ulm which strives to strengthen the ties between academia and industry in the fields of Data Science, ML and AI.

Case Study on Independent Metering Hydraulics for Energy-Efficient Battery-Electric Telescopic Loaders

Fabian Kreutmayr, Lucas Cayé, Markus Imlauer

HAWE Hydraulik SE, {f.kreutmayr, l.caye, m.imlauer}@hawe.de

Abstract.

This paper discusses the topic of energy-efficient hydraulics on battery-electric telescopic loaders. For this, two hydraulic systems are compared. The first one is a standard setup that can be found in classical machines that are driven by a combustion engine. The second one is a proposal using a decentralized independent metering approach. Through a comprehensive case study, the energy efficiency of both systems is evaluated. Thereby, it is shown that modern hydraulic setups can generate a great benefit for electrically driven machines by extending battery load cycles or reducing the required energy capacity to be installed, respectively.

Keywords. Mobile machines, energy efficiency, independent metering, battery-electric telescopic loaders, electrification.

1. INTRODUCTION

The transition from diesel-hydraulic to fully battery-electric mobile working machines is a process that is fundamentally reshaping the system-level energy architecture of such equipment. In this new paradigm, it is imperative that every kilowatt-hour drawn by the hydraulic circuit be meticulously balanced against a series of stringent constraints on battery capacity, including weight, installation space, upfront cost, and charging intervals. Consequently, the hydraulic subsystem, which was previously regarded as a passive energy consumer, has evolved into a pivotal mechanism for attaining an uninterrupted, eight-hour operating window within a designated energy budget [1].

Conventional hydraulic architectures – including open-center and load-sensing systems – are predicated on centralized valve blocks and extensive pressure lines. These designs introduce inherent inefficiencies, including high stand-by losses, throttling across metering orifices, and dynamic mismatches between pump output and actuator demand [2]. These inefficiencies are becoming increasingly unacceptable in battery-electric platforms, where energy is a scarce and costly resource. This comes from the fact that additional battery packs

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

are adding both weight and cost to machines, which is undesirable from an economic point of view.

In order to address these limitations, this case study explores a decentralized independent-metering (IM) architecture, in which compact, seated valve cartridges are mounted directly onto each hydraulic actuator [3]. This physical reconfiguration of components eliminates the necessity for external load-holding and counterbalance valves. Additionally, it offers more direct control, thereby reducing pressure losses. Notably, this system facilitates native support for gravity-assisted lowering, a feature that is crucial for the system's functionality. When paired with a variable-speed electric motor-pump unit (MPU), the IM concept provides access to advanced supervisory functionalities, including electronic flow sharing, adaptive pressure control, and load-aware energy scheduling.

In this work, a full-scale experimental evaluation is conducted using battery-electric telescopic loader. The machine is tested in two configurations: (i) a baseline layout in an open-center system setting, and (ii) the reengineered IM-based hydraulic circuit. Both systems are assessed on the exact same machine using repeatable, standardized loading cycles that are representative of typical material handling tasks in logistics, construction, or agricultural environments.

In summary, the study establishes a foundation for the advancement of intelligent energy-efficient hydraulics by demonstrating its potential as a component-level enhancement and a fundamental enabler for the cost-efficient, space-conscious, and operationally viable electrification of next-generation mobile machinery. This work aims to demonstrate the transformative potential of rethinking hydraulic systems from a first-principles energy-efficiency perspective, tightly integrated with modern electric powertrain technologies.

1.1. Literature Review

Recent shifts toward higher energy efficiency in the mobile machinery market are driven by regulatory requirements and ongoing electrification. However, research on more energy-efficient hydraulic circuits has been active for many years. A brief literature review highlights some significant work in this area.

Opgenoorth et al. [4] present an advancement of the STEAM hydraulic system originally developed at IFAS, RWTH Aachen. Their work introduces a multi-pressure hydraulic architecture applied to an excavator platform, featuring three pressure lines that are dynamically regulated via electric motors. The system employs two separate electric drives – one powering a conventional pump and the other driving a pump-motor unit. These pressure sources, in combination with the tank pressure, are selectively routed to the actuators through a configurable valve network. A central focus of the study is the design of a supervisory controller responsible for optimal state selection, determining both the active pressure lines and their respective pressure levels based on actuator demand. In a simulated 90° dig-and-dump cycle, the proposed system demonstrates a 29% increase in efficiency of hydraulic System compared to a conventional reference model, underscoring the potential of multi-pressure architectures and reduction of throttle losses for energy-optimized mobile hydraulics.

A novel independent metering valve configuration for excavator boom control, integrating three electrohydraulic 2/2-way proportional poppet valves with a 4/3-way directional control valve is proposed by Nguyen et al. [5]. The control strategy combines pump speed modulation with valve actuation, wherein the inlet valve is fully opened to minimize throttling losses, while the outlet valve is regulated based on a predefined flow rate and position feedback. The architecture supports multiple operating modes, including regenerative functions such as low-side regeneration retraction, in which the boom is retracted solely by gravitational force without active pump contribution. This functionality is enabled by an additional short-circuit valve integrated into the system. Experimental evaluations demonstrate energy savings of up to 22% in power retraction mode and 6.5% in high-side regeneration mode compared to a conventional reference system. Moreover, the configuration achieves improved position accuracy while reducing overall system complexity.

Hansen et al. [6] present a system design aimed at reducing energy consumption in mobile hydraulic applications. Their approach is implemented on a JCB telescopic handler and utilizes a separate-meter-in/separate-meter-out valve configuration embedded within a load-sensing (LS) hydraulic system. The conductive subsystem is realized using two 3/2-way proportional valves, each governed by distinct PI controllers. The study focuses on the development of a control strategy that enables precise modulation of flow paths and pressure levels to match actuator demands. Comparative analysis against a conventional pressure-compensated proportional valve setup reveals significant energy-saving potential. For the evaluated duty cycle, the proposed configuration achieves a reduction in energy consumption exceeding 30%, demonstrating the effectiveness of the control-oriented valve architecture in enhancing hydraulic efficiency.

1.2. Main Contribution and Outline

Although there is a wealth of literature on the advantages and disadvantages of modern IM systems, this work focuses on a standardized comparison with a conventional hydraulic circuit on a real-world telescopic loader. The main efforts are:

- The Definition of a standardized test procedure for both the conventional system and the IM system.
- The equipment of a battery electric telescopic loader with a conventional hydraulic circuit and a modern decentralized IM system.
- A comprehensive testing phase followed by a thorough analysis of the results.

The structure of the work is as follows: First, the telescopic loader and the hydraulic circuits are introduced. Next, the experimental setup and the type and number of experiments are presented. Next, the results are presented and interpreted. The work concludes with a summary and a recommendation for future research.



Figure 1: Image of the telescopic loader used in this work.

2. SYSTEM

In this section, the hydraulic system used in the telescopic loader is introduced. An image of the overall machine can be seen in Figure 1. We start with an explanation of the machine's conventional configuration, which serves as the basis of our study. Afterwards, an introduction of the adapted IM system is given.

Maximum load	3200 kg
Maximum lifting height	7 m
Battery capacity	63 kWh
Peak power overall	120 kW
Peak hydraulic power	60 kW
Volume flow	20 – 100 l/min
Maximum pressure	240 bar
Pump rotational speed	500 – 2200 rpm
Motor rotational speed	0 – 6600 rpm

Table 1: Technical data of the telescopic loader used in this work.

2.1. Technical Data

The telescopic loader that is used for this case study has a maximum load capacity of 3200 kg and a maximum lifting height of 7 m. Moreover, it is equipped with a fixed-displacement axial piston pump that delivers flows from 20 to 100 l/min with a rotational speed from 500 to 2200 rpm. The rotational speed of the motor that is driving the pump ranges from 0 to 6600 rpm, i.e. there is a gear with a ratio of 3 in between the two components. The maximum pressure is 240 bar.

The drivetrain of the machine is fully electrified, i.e. it contains no combustion engine. For this, eight battery packs with a total capacity of 63 kWh are installed. It is assumed that the machine lasts an average working day of 8 hours with only one loading cycle. The system can provide a peak power of 120 kW, from which the hydraulic system accounts for 60 kW. Table 1 provides an overview of those key points.

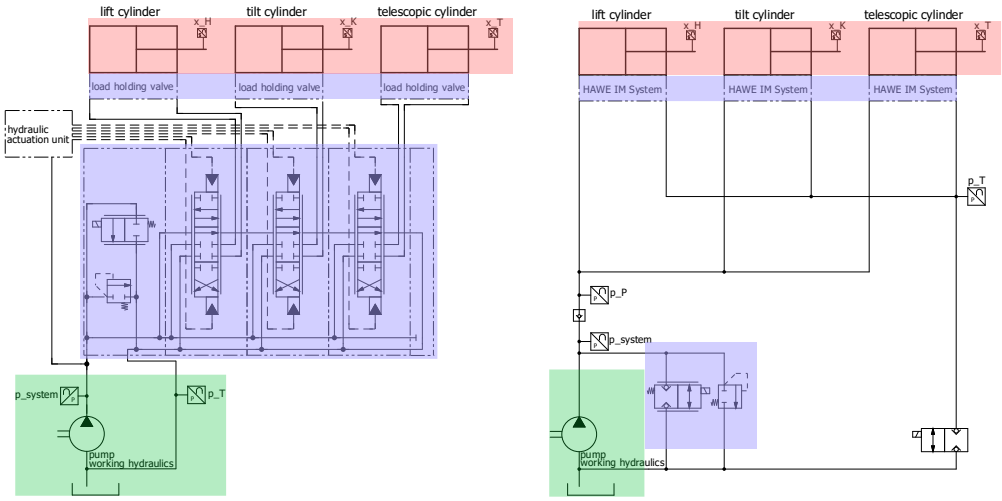


Figure 2: Excerpt of the hydraulic circuits used in this case study. On the left-hand side, the conventional configuration is shown. On the right-hand side, the IM system is shown. The colors indicate the different parts of the system, i.e. actuator, control valves, and pump.

2.2. Conventional Configuration

In conventional mobile machinery – particularly within the lower-cost segment – open-center valves in constant-flow hydraulic systems are commonly employed. This comes from the easy-to-use and cost-effective nature of such system. The reference machine used in this study is likewise equipped with such a baseline hydraulic architecture. This constant-flow system, characterized by a fixed volume flow, serves as the benchmark for evaluating alternative hydraulic system topologies.

On the energy generation side, the baseline system utilizes an axial piston pump with fixed displacement. This can be seen in the green marked sections of the hydraulic circuits depicted in Figure 2. In the electrified machine, which is used in this work, this pump is driven by an electric motor, which also powers additional consumers mounted on the same

shaft. The electric motor can be operated at discrete rotational speeds, which can be set by the driver and are regulated by the machine's control unit.

On the actuator side, the telescopic loader features three individually controlled consumers: a lift cylinder, a telescopic cylinder, and a tilt cylinder as depicted in Figure 2. These are responsible for manipulating the kinematic chain.

So, this investigation focuses on quantifying the impact on the hydraulic systems' impact on the overall energy balance of the machine, using working cycles that include the three cylinders. On the left-hand side of Figure 2 the conventional hydraulic configuration for the cylinders, which are marked in red, is shown. Hereby, the conductive section, which is marked by blue color, is realized as an open-center valve system. The valves of each cylinder are implemented as 6/3-way proportional spool valves with hydraulic actuation. An additional component of the conductive path is formed by load-holding valves, which are positioned between the spool valves and the actuators. Those are necessary to prevent free falling loads in case of a hydraulic hose failure.

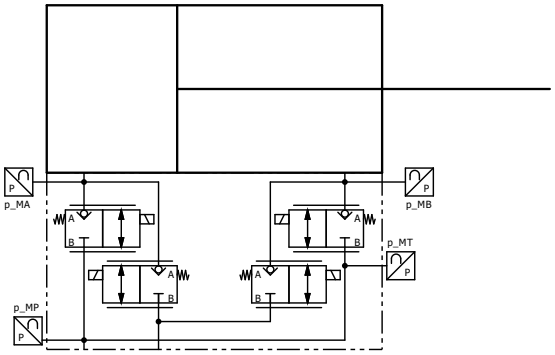


Figure 3: Detailed view on the used IM system. The valve and sensor configuration is the same for all the three equipped cylinders.

2.3. Independent Metering Configuration

As mentioned previously, the second configuration relies on a decentralized independent metering approach. In the selected system design, actuator control is achieved via four 2/2-way proportional seated valves, as shown in Figure 3. This configuration is applied to all three cylinders shown on the right-hand side of Figure 2.

Each valve of the IM system governs a single flow direction, necessitating the simultaneous and interdependent operation of multiple valves. To achieve this, all the seated valves are electronically actuated by a controller applying a PWM signal on the valves' coils. Direction and openness of the valves is controlled by the machine's operator via a joystick. To extend a cylinder, the inflow valve on the A-side and the outflow valve on the B-side are opened, while the retract direction requires to actuate the inflow valve on the B-side and the outflow valve on the A-side, respectively. While the design and tuning of the low-level control strategy is out of scope for this work, interested readers are advised to review e.g. [7, 8, 9].

To emulate a constant-flow system in conjunction with the IM valve architecture, an additional flow-controlled 2/2-way proportional seated valve is employed parallel to the

pump. This configuration ensures that the desired flow is delivered from the pump to the actuated function at any time.

It is to note that there are no load-holding valves used in the IM configuration. This comes from the fact that the valve blocks are directly mounted to the cylinders. This decentralized approach enables the seated-valves to additionally fulfill the load-holding functionality. Moreover, it is to note that additional shock-valves are used to protect the cylinders, and no specific pressure compensator is used. To ensure readability, those are not drawn in the right circuit of Figure 2.

3. EXPERIMENTAL SETTING

To enable a consistent and comparable evaluation of the hydraulic systems implemented on the telescopic loader, a standardized test cycle is defined. This cycle is designed to facilitate a comprehensive assessment of both the machine's hydraulic subsystem and its overall energy consumption. The evaluation is conducted using the machine depicted in Figure 1 with the hydraulic circuits from Figure 2 installed. In the following section, the load cycle and evaluation metrics are introduced.

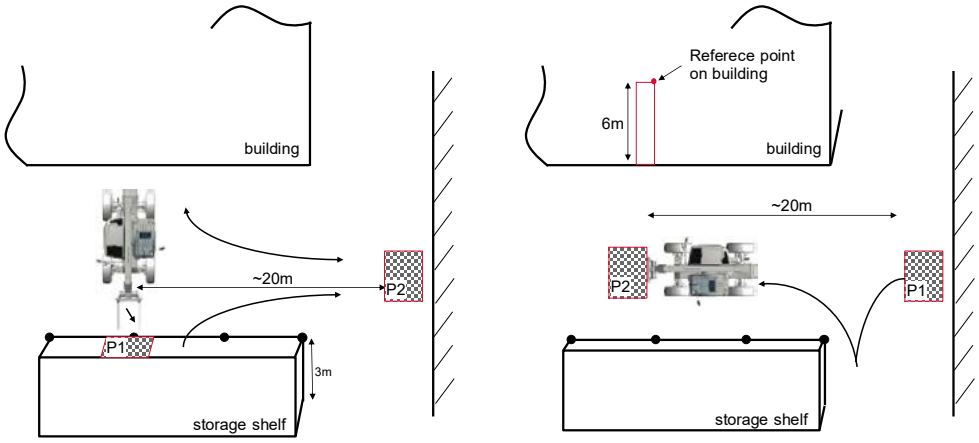


Figure 4: Illustration of the reference cycles used for the benchmark tests. The left image shows the "Pick and Go" task. The right image shows the lifting task.

3.1. Load cycles

Telescopic handlers exhibit highly variable load profiles during operation, depending on their specific application domain in e.g. construction or agriculture. Related machines, such as wheel loaders, are typically evaluated using so-called "Pick and Go" tasks, which are represented by a standardized Y-cycle [10, 11].

Since the aim of this study is to test both hydraulic systems using a representative working scenario, two adapted Y-cycles are used to reflect the operational characteristics of the machine. This includes a simple "Pick and Go" task and a lifting task in which the telescopic functionality is used.

3.1.1. Pick and Go

The “Pick and Go” test cycle involves a defined load being moved in a stacking and transport operation. An illustration of this is shown in the left part of Figure 4. Starting from Position 1 (P1), located at a vertical height of 3 m above ground level, the telescopic loader picks up the load. In the second step, the load is transferred to Position 2 (P2), which is located at ground level, covering a horizontal distance of 20 m. This operation represents a downward transfer from elevation level 1 to level 2.

In the third phase of the cycle, the telescopic loader drives backward in an unloaded state. In steps 4 and 5, the load is picked up again at P2 and transported back to P1, followed by a return to the starting position.

3.1.2. Lifting

To take the far-reaching telescopic functionality into consideration, a second load cycle is introduced. Thereby, the load is transferred to a position that is elevated to 6 m above ground level. This extended reach scenario is meant to represent typical use cases in e.g. construction and logistics where elevated placement of materials is required

In the first step of the cycle, the telescopic handler picks up the load at P1, which lies at ground level. This is followed by a transport manoeuvre that includes a 180-degree rotation of the machine. In the third step, the load is elevated to the height level of the target position P2. The horizontal distance between both points P1 and P2 is 20 m. Upon reaching the target position, the load is lowered back to ground level in step 4. In step 5, the telescopic handler traverses the transport path back to P1. Finally, in step 6, the load is placed back at the initial location. An illustration of this can be seen in the right part of Figure 4.

	Conventional System	**IM System**
#1	800 kg | 23 l/min | Pick & Go	800 kg | 23 l/min | Pick & Go
#2	800 kg | 40 l/min | Pick & Go	800 kg | 40 l/min | Pick & Go
#3	800 kg | 100 l/min | Pick & Go	800 kg | 100 l/min | Pick & Go
#4	1500 kg | 23 l/min | Pick & Go	1500 kg | 23 l/min | Pick & Go
#5	1500 kg | 40 l/min | Pick & Go	1500 kg | 40 l/min | Pick & Go
#6	1500 kg | 100 l/min | Pick & Go	1500 kg | 100 l/min | Pick & Go
#7	800 kg | 23 l/min | Lifting	800 kg | 23 l/min | Lifting
#8	800 kg | 40 l/min | Lifting	800 kg | 40 l/min | Lifting
#9	800 kg | 100 l/min | Lifting	800 kg | 100 l/min | Lifting
#10	1500 kg | 23 l/min | Lifting	1500 kg | 23 l/min | Lifting
#11	1500 kg | 40 l/min | Lifting	1500 kg | 40 l/min | Lifting
#12	1500 kg | 100 l/min | Lifting	1500 kg | 100 l/min | Lifting

Table 2: List of all experiments conducted for this work.

3.2. Load and Pump Conditions

Both test cycles are conducted under two distinct load conditions. In the first scenario, a transport weight of 800 kg is selected. This corresponds to 25% of the machine's nominal lifting capacity, which would be 3200 kg. In the second scenario, a heavier load of 1500 kg is used to evaluate system performance under increased mechanical stress.

Due to the constant-flow hydraulic architecture, the test cycle is executed at discrete volumetric flow rates. In regular operation, the pump speed is adjusted according to the flow demand of each subtask. However, to ensure comparability across measurements, each test run is performed at fixed flow rates. Three discrete flow settings are selected: 23 l/min (motor speed of 1550 rpm and pump speed of 516 rpm), 40 l/min (2550 rpm | 850 rpm), and 100 l/min (6600 rpm | 2200 rpm). Each flow rate is tested in a separate experimental run.

Overall, this results in a total of 12 experiments. For the sake of clarity, an overview of all the experiments and their different settings can be found in Table 2. Each experiment listed in Table 2 contains of two runs, where one is conducted with the conventional system and the other one is conducted with the IM system.

3.3. Evaluation Metrics

To evaluate the machine's energy consumption, measurement data is recorded during operation. Energy usage is captured across multiple reference boundaries (see Figure 5) in order to better understand the impact of different consumers on the overall energy balance.

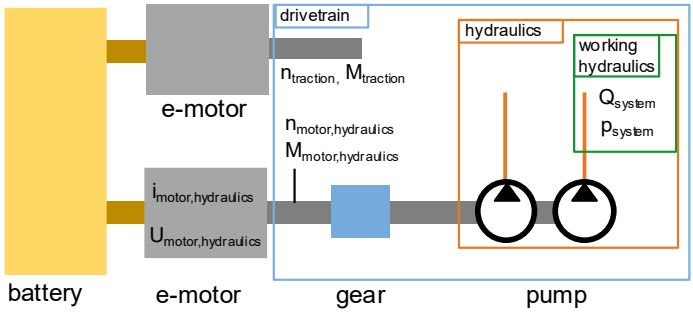


Figure 5: Schematic view of the energetic consumers of the telescopic loader including the reference boundaries.

The power input to the drivetrain encompasses the power for both the traction system and all hydraulic systems, the working hydraulics and auxiliary drives. So, the power of the whole drivetrain is calculated as:

$$P_{Drivetrain} = P_{Traction} + P_{Hydraulics}. \quad (3.1)$$

Hereby, $P_{Drivetrain}$, $P_{Traction}$, and $P_{Hydraulics}$ are powers in W. Moreover, the power of the traction system is:

$$P_{traction} = \omega_{traction} * M_{traction} = 2\pi * n_{traction} * M_{traction} \quad (3.2)$$

with $n_{traction}$ as a rotational speed in 1/s and $M_{traction}$ as a torque in Nm. Moreover, $P_{Hydraulics}$ is defined as:

$$P_{Hydraulics} = i_{Motor,Hydraulics} * u_{Motor,Hydraulics} \tag{3.3}$$

with $i_{Motor,Hydraulics}$ as a current in A and $u_{Motor,Hydraulics}$ as a voltage in V. It is to note that here, both the power consumption for the main and the secondary pump are included. However, this study focuses specifically on the working hydraulics.

Thus, the power output of the main pump is measured in addition to $P_{Hydraulics}$. For this, the system pressure, namely p_{System}, in the working circuit as well as the volumetric flow rate, Q_{System}, is measured. So, $P_{Hydraulics,Working}$ is calculated as:

$$P_{Hydraulics,Working} = p_{System} * Q_{System} \tag{3.4}$$

It is to note that losses introduced by the flow sensor are compensated analytically during post-processing to ensure accurate energy assessments.

Considering the different reference systems, two metrics are used to quantify the energy savings achieved using the IM system. The first one focuses just on the working hydraulics:

$$\delta_{Hydraulics,Working} = 1 - \frac{\int_{t_{start}}^{t_{end}} P_{Hydraulics,Working,IM} dt}{\int_{t_{start}}^{t_{end}} P_{Hydraulics,Working,Conventional} dt}. \tag{3.5}$$

Here, t_{start} and t_{end} is the total working cycle duration in s. The second one focuses on the whole drivetrain:

$$\delta_{Drivetrain} = 1 - \frac{\int_{t_{start}}^{t_{end}} P_{Drivetrain,IM} dt}{\int_{t_{start}}^{t_{end}} P_{Drivetrain,Conventional} dt}. \tag{3.6}$$

Beyond energy-related data, the position of the hydraulic cylinders and the chamber pressures are also recorded. Moreover, the operational state of the drive system is monitored via the rotational speed of the traction drive. Based on the collected measurement data, a comprehensive evaluation of the conductive hydraulic systems is performed with respect to energy consumption and overall system suitability.

4. EXPERIMENTAL EVALUATION

The experimental evaluation of both hydraulic systems is conducted using the setting shown in the previous section. In the following, the results are introduced and thoroughly discussed. For this, the time series data of two exemplary working cycles is presented in detail. Afterwards, the energy consumptions of both hydraulic systems are shown and compared.

4.1. Working Cycles

Figure 6 shows an excerpt of the sensor signals of experiment #2 (see Table 2). Here, the volume flow is set to 40 l/min and the weight of the load is 800 kg. Moreover, the Pick & Go task is performed. The diagram illustrates two repetitions of two representative test runs:

the conventional baseline system is shown on the left in blue, while the proposed IM architecture is depicted on the right in orange.

In the first step of the work cycle, indicated by (I) in the graph, the telescopic handler needs to pick up the load at position P1 on the test track. During this phase, the boom arm is moved to an elevated position without load. Under identical boundary conditions, the IM valve system exhibits a system pressure approximately 10 bar lower than that of the conventional system.

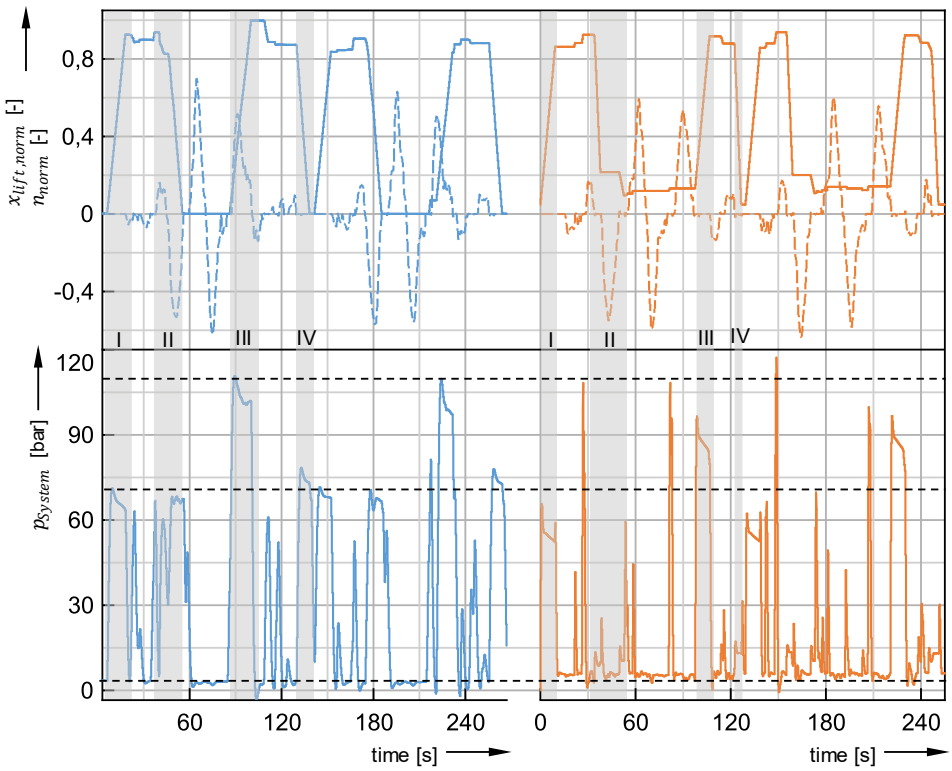


Figure 6: Excerpt of sensor signals from experiment #2 (800 kg | 40 l/min | Pick & Go).

In the second step, indicated by (II), the load is transferred to position P2. This involves lowering the boom arm under load from P1 to ground level. In the conventional system, a system pressure of 65 bar is observed, primarily required to actuate the load-holding valve. In contrast, the IM system operates at a significantly reduced pressure level of approximately 10 bar, which is the system's standby pressure. This comes from the gravity lowering functionality that the IM system offers since no load holding valves are required because of the decentralized mounting.

In the third step, (III), the load is lifted from position P2 back to P1, representing actuator movement under load. Similar to the behavior in (I), the IM system demonstrates a lower system pressure compared to the conventional design. In step four, (IV), the boom is lowered

without load. The pressure behavior mirrors that of (II), with the IM system enabling gravity-assisted lowering without the need for additional pressure.

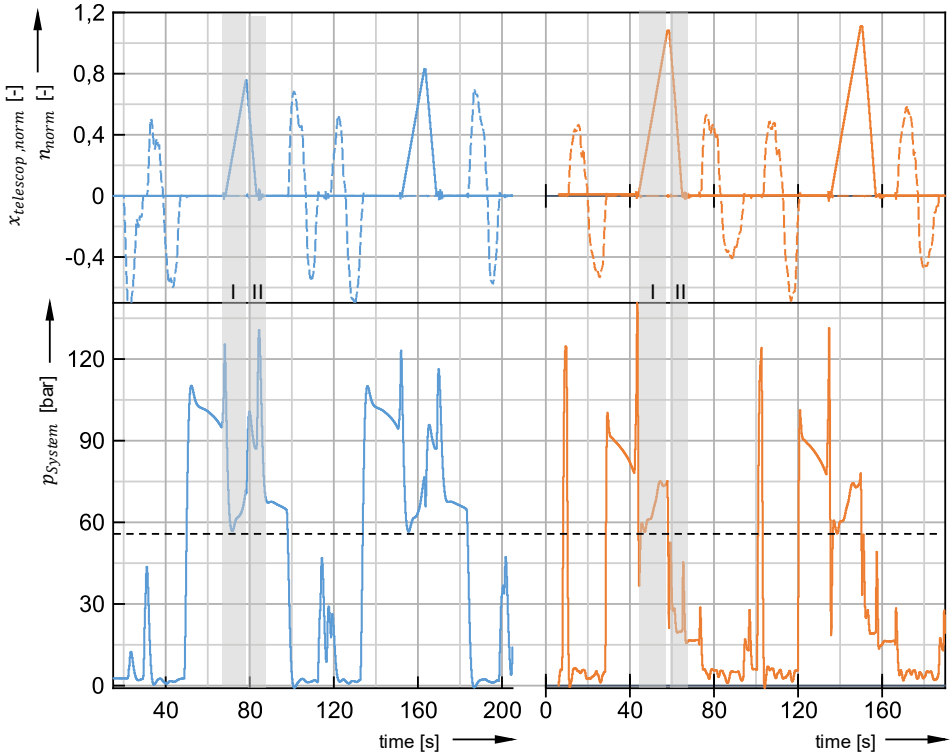


Figure 7: Excerpt of sensor signals from experiment #8 (800 kg | 40 l/min | Lifting).

Figure 7 shows an excerpt of sensor signals from two repetitions of experiment #8 (see Table 2). Again, the volume flow is set to 40 l/min and the weight of the load is 800 kg and the lifting task is performed.

As shown in Figure 7, the segments involving active drive mode without engagement of the working hydraulics exhibit recirculation behavior identical to that observed in experiment #2. Moreover, the lifting arm actuation remains as in the Pick & Go task. Thus, only the segments in which the telescopic cylinder is actuated, are marked in Figure 7.

Here, segment (I) of the cycle involves lifting the load to the target elevation. In this phase, the pressure profiles of both the conventional and IM systems show similar characteristics. In the subsequent lowering operation in segment (II), the conventional system requires a system pressure of 100 bar, primarily driven by the actuation of the load-holding valve. In contrast, the IM system operates at a reduced pressure level of 25 bar, which is sufficient to enable a stable and controlled movement. This pressure requirement is attributed to the static friction within the linear guide mechanism.

The data from both experiments demonstrate that the IM system requires significantly lower operating pressures within the working hydraulics. This reduction is primarily attributed to

minimized flow losses resulting from direct piping to the valves and the elimination of hydraulic components, i.e. the load holding valves.

Furthermore, the IM system exhibits substantially lower working pressures during cylinder retraction phases. In these scenarios, the observed pressure corresponds directly to the mechanical requirements of the task, rather than being influenced by auxiliary valve actuation. Due to the system's design, energy savings of up to 98% are achievable in lowering operations where no active hydraulic energy input is required. It is to note that the recirculating flow, inherent to the constant-flow architecture, is included in the overall energy balance and contributes to the system's efficiency profile.

Both working cycles, Pick & Go as well as the lifting task, show a clear tendency when comparing the pressure signatures from the conventional and the IM system. Since the first working cycle corresponds to a more realistic Pick & Go task, based on the unloading of the load on both positions, the following evaluation of the energy usage is going to focus on this particular working cycle with different load conditions as well as flow settings.

4.2. Energy Consumption

Figure 8 illustrates the energy balance of the working hydraulics, highlighting the energy savings achieved by the IM system relative to the conventional architecture as shown in equation (3.5). It is to note that the results are averaged over five runs with each system to ensure repeatability. Moreover, it ensured that the time taken to complete the working cycle remains constant over each run. This analysis, focused exclusively on the working hydraulic subsystem, reveals a clear dependency between load magnitude, volumetric flow rate, and achievable energy savings.

For the Pick and Go task, the IM system demonstrates energy savings ranging from 25% to 40%, depending on the operating point. The observed decline in energy savings with increasing flow rate is attributed to the diminishing influence of gravity-assisted lowering, as flow-induced losses become more dominant within the system.

The load dependency is primarily governed by the behavior of the load holding valves. Their actuation is based on a force equilibrium across the valve ports, with the working port contributing to the balance. Under high load conditions, the resulting cylinder pressure supports the process of valve opening, thereby reducing the pressure differential between the conventional and IM systems.

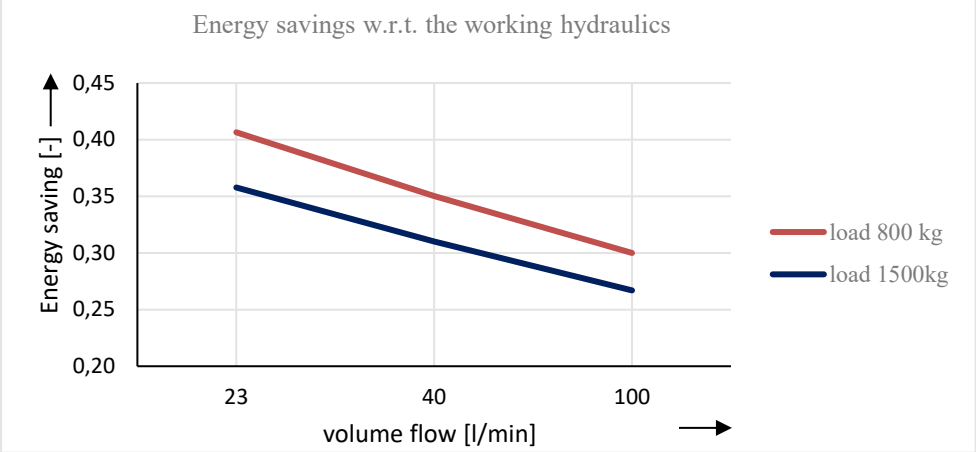


Figure 8: Energy savings $\delta_{Hydraulics,Working}$ of the IM system in the Pick & Go task w.r.t. the working hydraulics.

When extending the energy balance to the entire drivetrain, where the hydraulic consumers as well as the traction drive is included, the Pick & Go task reveals overall energy savings ranging from 15% to 30%. This behavior can be seen in Figure 9. A declining trend with increasing volumetric flow rate is again observed, driven by the same phenomenon of increasingly dominant flow losses.

In contrast to the behavior shown in Figure 8, the drivetrain-level analysis shows an increasing energy savings trend with rising load weight. As the load increases, the proportion of power demand attributed to the hydraulic subsystem grows relative to the electric drive. Consequently, the efficiency gains achieved within the hydraulic domain have a more pronounced impact on the overall energy balance.

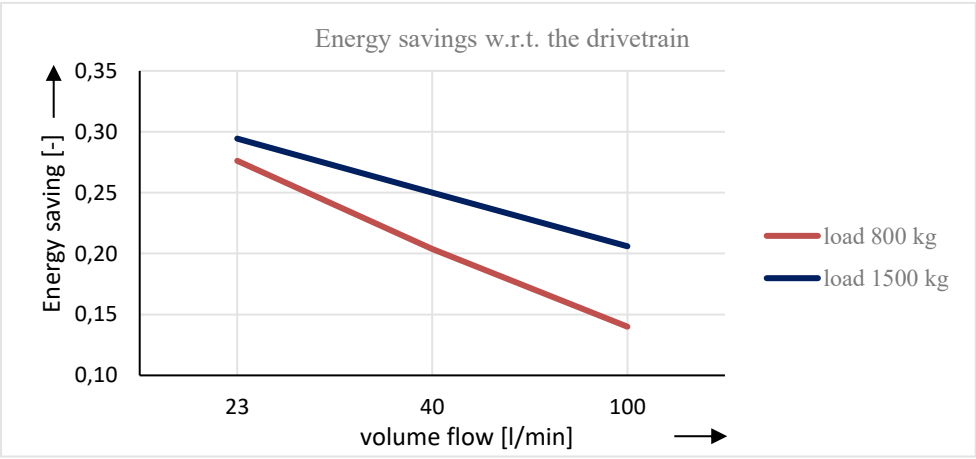


Figure 9: Energy savings $\delta_{Drivetrain}$ of the IM system in the Pick & Go task w.r.t. the drivetrain.

4.3. Practical Considerations

To emphasize the impact of the calculated energy savings, the potential impact of the telescopic loader's overall design on practical considerations is discussed. As listed in Table 1, the battery capacity of the machine used in this work is 63 kWh, which is just enough to use the machine for a working day of 8 hours with one loading cycle.

As stated before, the IM approach saves between 15% and 30% of energy, which can be seen as a range extension between 1.2 hours and 2.4 hours. This allows for more flexible machine usage, especially in areas where flexible shift models are required. Moreover, recharging is faster with a higher remaining battery capacity.

From a financial perspective, reducing the overall battery capacity up to 30% can significantly reduce the cost of the machine. Considering a retail price for off-highway applications of at least 300 € per kWh, the potential cost reduction lies in an area from 2835€ to 5670€ [12]. Additionally, the lower power usage might enable a redesign of the motor-pump unit and inverters, which can lower the price even more.

Overall, it is to mention that the energy gains of the IM approach can trigger further innovation and redesign of the whole machine. This underlines also the importance of taking a holistic approach on the design of modern mobile machinery, since hydraulic system, drive system, and the battery system significantly influence each other.

5. CONCLUSION

This work presents a comprehensive case study, highlighting the energy savings that a modern decentralized independent metering (IM) approach can be realized in comparison to a conventional hydraulic system design. To ensure practical relevancy, the accompanying experiments are conducted on an electrified telescopic loader, a machine which is heavily used in fields as e.g. agriculture and construction. Moreover, test cycles with a clear reference to real-life scenarios are used in order to come to a representative conclusion.

The case study clearly indicates that the decentralized IM approach can significantly reduce the power consumption of the machine's hydraulic circuit up to 40%. Related to the whole drivetrain, there is still a gain of up to 30% possible. Especially the gravity lowering functionality, which is possible because load holding valves are spared, makes nearly energy-free load lowering possible and accounts for a great amount of the savings. However, the pressure levels needed for cylinder extraction tasks are significantly lower using the IM approach.

This shows that redesigning conventional hydraulic circuits can be a great factor in making future machines more efficient. This is a benefit especially for electrified machines, since energy savings directly reduce the battery capacity needed. This reduces the space requirements of the electrical system, the weight of the machine, and the overall invest cost as well as the end users total cost of ownership. However, also mobile machinery using internal combustion engines can benefit from the presented IM approach, since a significant reduction in energy consumption reduces both fuel costs and emissions.

Despite the promising results, there is much room for future work on decentralized as well as centralized IM systems. So, stable control of IM systems is an ongoing challenge and requires modern control techniques. Moreover, the usage of the proposed architecture offers

possibilities like regeneration and recuperation, which can be further explored and might even further improve machine efficiency. Overall, it is to state that modern hydraulic architectures such as IM offer a great possibility to design future-proof efficient and sustainable mobile machines.

6. REFERENCES

[1] D. Padovani, P. Dimitriou, T. Minav, 'Challenges and solutions for designing Energy-Efficient and Low-Pollutant Machines in Off-Road hydraulics.', Energy Conversion and Management: X, Volume 1, 2024.

[2] A. Rankka, A. Dell'Amico, 'Procedure of developing more energy efficient and electrified hydraulic systems for loader cranes', Frontiers in Mechanical Engineering, Volume 10, 2004.

[3] K. Abuowda, I. Okhotnikov, S. Noroozi, P. Godfrey, M. Dupac, 'A review of electrohydraulic independent metering technology', ISA Transactions, Volume 98, pp. 364-381, 2020.

[4] A. Opgenoorth, C. Haas, K. Frischkorn, K. Schmitz, 'Hydraulische Mehrdrucksysteme für mobile Arbeitsmaschinen mit elektrischen Antrieben', 8. Fachtagung Hybride und energieeffiziente Antriebe für mobile Arbeitsmaschinen, 2021.

[5] T.-H. Nguyen, T.-C. Do, K.-K. Ahn, 'A Study on a New Independent Metering Valve for Hydraulic Boom Excavator', Applied Sciences, Volume 12, Issue 2, 2022.

[6] A. Hansen, H. Pedersen, T. Andersen, L. Wachmann, 'Investigation of energy saving separate meter-in separate meter-out control strategies', Proceedings of the 12th Scandinavian International Conference on Fluid Power (SICFP'11), 2011.

[7] F. Kreutmayr, M. Imlauer, 'Application of machine learning to improve the performance of a pressure-controlled system', Proceedings of the 12th International Fluid Power Conference (12th IFK), Volume 1, pp. 77-86, Dresden, 2020.

[8] F. Kreutmayr, C. Ament, 'Automated Tuning of Gain-Scheduled PI Controllers Based on SANARX Models', IFAC-PapersOnLine (PID 2024), Volume 58, Issue 7, pp. 120-125, 2024.

[9] J. Schwarz, B. Lohmann, 'Robust identification and control of mobile hydraulic systems using a decentralized valve structure', Robust Control Engineering Practice, Volume 151, 2024.

[10] H. Deiters, 'Standardisierung von Lastzyklen zur Beurteilung der Effizienz mobiler Arbeitsmaschinen', Dissertation, TU Braunschweig, 2009.

[11] A. Huber, 'Ermittlung von prozessabhängigen Lastkollektiven eines hydrostatischen Fahrantriebsstrangs am Beispiel eines Teleskopladers, Dissertation', KIT, 2010.

[12] O. Fakunle, 'Slow shift to electrification for ~$10bn off-highway powertrain market as diesel continues to dominate', Interact Analysis, Press Release, 2024.

Biographies



Fabian Kreutmayr received the bachelor's degree in electrical engineering from Technical University Munich (TUM) in 2016, the master's degree in electrical engineering from TUM in 2019 and is pursuing his PhD in control engineering at the University of Augsburg. He is currently working as an Expert Machine Learning at HAWE Hydraulik SE. His areas of interest include machine learning, intelligent control systems, optimal control, and hydraulic system design.



Lucas Caye received the bachelor's degree in mechanical engineering from RWTH Aachen in 2021 and the master's degree in mechanical engineering from RWTH Aachen University in 2023. He is currently working as a Hydraulic Systems Engineer at HAWE Hydraulik SE. His areas of interest include mobile hydraulic working machines, hydraulic system design and energy efficiency in hydraulic systems and components.



Markus Imlauer received the Dipl. Ing. (FH) degree in mechatronics from University Mittweida in 2010. He is currently working as a team leader at the R&D Department for the area “PowerPacks & SmartControl" at HAWE Hydraulik SE. His areas of interest include working machines for mobile and industrial applications, compact hydraulic power units, and energy efficient hydraulic architectures.

Holistic system simulation method for the analysis and evaluation of thermal management of electrified construction machinery using inverse hydraulic modelling

Dr.-Ing. Marcel Neu, Dr.-Ing. Fiona Tichi, Dr.-Ing. Martin Piechnick, Lars Holländer

marpitec GmbH, Marcel.Neu@marpitec.com

Abstract.

The electrification of construction machinery comes with new opportunities and challenges for the system architecture, leading to varying approaches and subsystem designs. Missing experiences from preceding product generations prevent an efficient comparison of system architectures early in the product development process. Thus, product development needs a tool for evaluating designs towards efficiency and energy consumption, taking functionality and environmental impacts into account. Holistic numerical system simulations enable quantitative analysis and allow comparison of concepts, without the need for tangible component designs. [1]

Electrified construction machinery is mainly described by a combination of mechanical components as well as hydraulic-, electric- and thermal circuits. A holistic system simulation describes the interaction between these domains from which two conflicting objectives arise. First, while the mechanic and hydraulic interaction is highly dynamic and takes place on short time scales, the thermal behaviour unfolds in the long-term, which leads to long periods of simulation time with high temporal resolution. Second, the complete, forward, coupling of a mechanical load simulation with the hydraulics has a high simulation complexity which comes down to extensive computational costs. [2]

As a research gap the need for an efficient holistic simulation method arises. The method shall represent all couplings relevant for the design while keeping the computational costs low, so that a quick evaluation of concepts becomes accessible. To do so, the main conflict between highly dynamic hydraulics and the long-term thermal behaviour needs to be resolved.

This contribution proposes an innovative simulation method to resolve the outlined conflict for holistic system simulations, enabling its use in the early stages of development. The main advancement is the reduction of the full bidirectional coupling of mechanics and hydraulics to a unidirectional coupling and a separation of the mechanic-hydraulic domains from the

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

thermal simulation. The mechanics is calculated with a multibody simulation. Standard tasks and work patterns are calculated to approximate realistic forces, torques and displacements as input for the hydraulic simulation. The hydraulic simulation is implemented inversely, meaning the valve positions, pressures, and volumetric flow rates required to fulfil the mechanical movement pattern are calculated incrementally, omitting the solution of the pressure buildup equation. To comply with the required long-term analysis for thermal behaviour the output of several sequential mechanic-hydraulic work patterns is concatenated and input into the thermal simulation. This simulation method allows the efficient comparison of different system architectures and variations of subsystems and enables the initial dimensioning of the hydraulic system according to the requirements.

The usage of this holistic simulation approach is conducted exemplary on an excavator. Slewing, boom motion and driving patterns are considered providing the load and displacement input for the inverse hydraulics. The thermal simulation got the outputs of the mechanic-hydraulic simulation comprising the required electrical quantities, thermo-hydraulic losses and environmental influences to determine the required thermal regulation for the individual subsystems. This information is used to assess different designs of the cooling circuits and showcasing the added value of efficient holistic system simulations via inverse hydraulic modelling at the same time.

Keywords. electrified construction machinery, inverse kinematics, system simulation, coupled thermal-kinematic-hydraulic simulation.

1. INTRODUCTION

Electrification of construction machinery gained significant importance lately, since regulations for pollutant emissions became stricter not only for on-road vehicles but also for off-road heavy-duty vehicles. [3, pp. 780, 781] [4, pp. 1, 2] The requirements deduced from the amended regulations led to a change in the drive technology towards electric and hybrid drive concepts. These drive concepts on one hand allow to meet the requirements for pollutant emissions [5, p. 1] but furthermore offer advantages regarding noise emission, energy efficiency by utilizing recuperation of kinetic and thermal energy and optimisation of subsystem's operating point and increased productivity. [3, pp. 780, 787, 788, 797–799]

On the other hand, new challenges arise from the electrification of construction machinery. The crucial design decision is the system architecture around the electric drive system including the electronical periphery. Besides the changed drive system recuperation and electrical storage systems are required, beyond others. [3, pp. 782, 783]

One important aspect to evaluate the different system architectures is the thermal management within the machine. This includes the electrical subsystems as well as temperature control of the cabin and hydraulic subsystems. As electrified construction machines are not only a new generation of machines but involve a new drive train and energy supply concept, there is limited experience available on preceding product generations and therefore on sufficient criteria for the evaluation of the system variant's performance and optimum system design [6, p. 3]. Hence, comparing the variants of system architecture and subsystem design require analysing them during the design phase. For the evaluation of the thermal management using simulation results is feasible to reduce experimental effort and verify development considerations, furthermore, accelerating the development phase by

identifying subsystems with the most potential for optimisation. To describe a system of electrified construction machinery properly, the interaction of kinematics, hydraulics and thermal behaviour within the system need to be considered in a holistic simulation approach. [7] Therefore this contribution follows the trend of coupling multiple physical domains and subsystems into a single simulation with its challenges and opportunities.[8]

2. FUNDAMENTALS

The characterization of a mechanical system with actuators, sensory and a control loop is known as a mechatronic system. As the aim of the examinations in this contribution is the thermodynamic characteristics, for evaluation the thermal management, the mechatronic system is extended to include the thermal domain, which then describes a multi physics or multi domain system. To simulate this multi physics system, each domain needs its individual modelling. Initial objective of holistic system simulations is the identification of suitable numerical tools to capture the system's characteristics. Subsequently the exchange of information and the calculation process need to be described. In the following chapter the fundamental principles for the holistic numerical approach utilized in this work to couple the afore mentioned physical domains efficiently are introduced, following the idea of assessing the thermal management for the whole system. [9]

Computational Dynamics (also referred to as multi-Body Dynamics) serves the purpose of assisting engineers during the analysis of kinematic and dynamic behaviour of complex mechanical systems comprising numerous moving bodies connected by joints. Ideally Computational Dynamics allow the prediction of the component's position, velocity and acceleration based on the external forces acting upon the individual mechanical parts. It utilises available computing power to reduce the effort for analytically/manually calculating the interaction between multiple bodies in an acceptable time. Computational Dynamics describes the mechanical domain, and its application is usually divided into rigid body and flexible body analysis, as well as kinematic determined and dynamic determined systems. While rigid body and kinematic analysis produce algebraic equations, the dynamic analysis induces second order partial differential equations to solve. The regular computational dynamics determines the accelerations, velocities and positions of the system under consideration based on the forces acting upon it. [10]

In certain cases, if the accelerations, velocities and positions for the system in question are known, it is possible to calculate the external forces in reverse. This approach is called inverse kinematic simulation. Consequently, the partial differential equations of the dynamic approach collapse to differential-algebraic equations which simplifies the solution procedure significantly. The comparison of forward and inverse simulation is depicted in Figure 2.1.

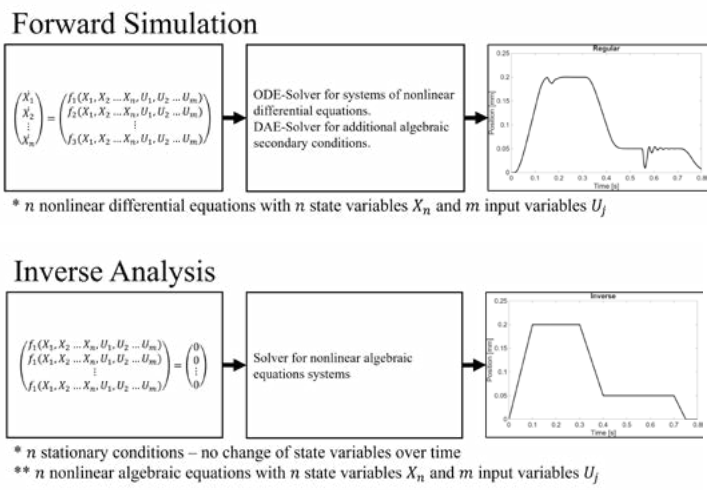


Figure 2.1: Forward Simulation a) in comparison to Inverse Analysis b) due to the inversion the equation system is simplified, while dynamic effects get eliminated from the analysis.

The approach of inverse kinematics is used frequently to analyse the forces needed to complete a functional task inside a technical system. [9,10,11,12]

Though Computational Dynamics only achieves a degree of accuracy compared to finite element calculations, which are much more accurate at a much higher computational cost, the granularity of Computational Dynamics meets the requirements for modelling complex mechanical systems. The same gradation in simulation tools is found in hydraulics, thermodynamics and electro dynamics. While for all three physical domains a continuum representation exists (i.e. CFD, CHT and CEM) the calculation with lumped parameter models, focussing on the interaction between components is used frequently for less effortful system analysis. [13,14]

Thermal simulations are primarily based on the Fourier's heat conduction equation to describe the heat flux through solid bodies. This form of heat transfer is dominating the heat development in a technical system. In addition, heat radiation and thermal convection are relevant depending on the system's physics and its boundary conditions. The main boundary conditions to be considered are the initial temperature of the system as well as its immediate environment and heat sources usually described by an input heat flux. [15, p. 265]

By reducing the hydraulic, electric and thermodynamic components to system level ordinary differential equations that describe the correlations, for example between volumetric flow rate into the cylinder and its stroke, the calculation of system dynamics is much more accessible than modelling the hydraulic cylinder using a continuous flow representation. The continuous flow representation is not only time-dependent but also a function of space and therefore leads to a complex system of computational expensive partial differential equations.

In hydraulic system simulations pressure or volumetric flow rates from fluid energy machines like pumps are specified along with control signals for the valves to analyse the

hydraulic system's response comprising piston or motor accelerations, velocities and positions. In combination with control circuits that feed those positions, velocities and accelerations into a control loop modifying the input variables, specific motions of the hydraulic system are realizable.

Analogue to inverse kinematics an inverse hydraulic analysis is applicable. Instead of specifying pressure, volumetric flow rates and valve positions, the desired state of motion and position of the hydraulic system is specified, and the needed pressures, volumetric flow rates and valve positions are deduced. As with inverse kinematic inverse hydraulic reduces the computational effort at the expense of the system dynamic behaviour, which is not representable anymore.

The lumped parametric calculation approach entails an increased demand of fitting the actual system components to the, mostly generic, component substitutes of the simulation tool, while allowing longer real-world time spans to be calculated.

But, although the simulation time for each individual domain is reduced drastically, the time span in which processes happen in the individual domains differs significantly, becoming particularly notable when comparing hydraulic and thermodynamic processes. The period required in the respective physical domains for a steady state to be reached is describable by its characteristic times that originate from an analogy to electrodynamic oscillatory circuits [15]. The characteristic time is calculable from the basic differential equation of the domain regarded. The analogy between hydraulics and electrodynamics is depicted in Figure 2.2, taken from [16].

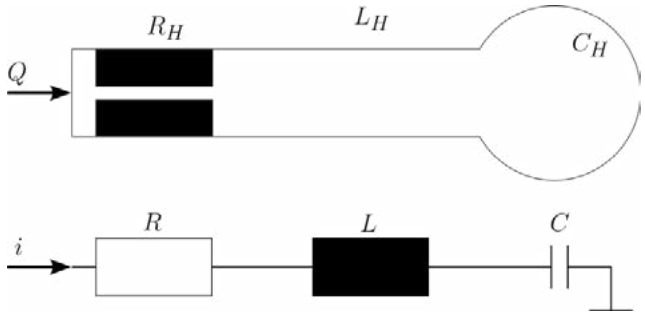


Figure 2.2: Principal illustration of the analogy between hydraulics and electrodynamics [16]

The circuit diagram can be transferred into the differential equation (1) and equation (2).

$$u = R \cdot i + L \cdot \frac{di}{dt} + \frac{1}{C} \cdot \int i dt \tag{1}$$

$$\Delta p = Q \cdot R_{\mathrm{H}} + L_{\mathrm{H}} \cdot \frac{dQ}{dt} + \frac{1}{C_{\mathrm{H}}} \cdot \int Q dt \tag{2}$$

Depending on the regarded component the equations vary. By transformation to the frequency domain, characteristic time constants are derived which indicate required period for a change in state of the component described. The analogies between electrodynamics circuits and the relevant domains for the analysis in this contribution are listed in Table 2.1.

Table 2.1: Electrodynamic Network Analogy for Hydraulics, Thermodynamics and Mechanical Systems

				Characteristic Times
Electrodynamics	Capacitor C	Inductor I	Resistor R	$\tau = RC, \frac{L}{R}, \sqrt{LC}$
Hydraulics	Hydraulic Capacitor $C_{\mathrm{H}} = \frac{V_0}{K_{\mathrm{Fl}}} = V_0\beta$	Hydraulic inertia $L_{\mathrm{H}} = l \cdot \frac{\rho}{A}$	Cylindrical pipe (Hagen-Poiseuille) $R_{\mathrm{H}} = \frac{12 \cdot \eta \cdot l}{\pi \cdot r^4}$	$\tau = RC, \frac{L}{R}, \sqrt{LC}$
Thermodynamics	Heat Capacity $C_{\mathrm{TH}} = C_{\mathrm{p}}, C_{\mathrm{v}} \left[\frac{\mathrm{J}}{\mathrm{K}}\right]$	-	Thermal Resistance $R_{\mathrm{TH}} = \lambda \cdot l \left[\frac{\mathrm{K}}{\mathrm{W}}\right]$	$\tau = \lambda \cdot l \cdot C_{\mathrm{p}}$
Mechanics	Spring $C_{\mathrm{m}} = k$	Inertia $L_{\mathrm{h}} = m$	Friction $R_{\mathrm{m}} = D$	$\tau = \sqrt{\frac{m}{c}}$

From this analogy the different characteristic times for the lumped parameter systems are calculable, which allows an estimation of the overall simulation time needed. A hydraulic network, typically consisting of valves, pistons and tanks, reaches its stationary state in the order of seconds. To capture the hydraulics dynamic behaviour, simulation time step sizes below $< 1 \cdot 10^{-3}$[s] are necessary. Based on experience a thermodynamic system reaches a stationary state in the order of minutes or hours, depending on its size, which leads to time step sizes bigger than 1 [s] to capture the systems dynamics. In many cases in Multiphysics, a problem emerges, if those domains are considered in combination within a single simulation approach, since this forces the calculation of thermodynamic effects on a hydrodynamic timescale if no measures are taken. A combined hydraulic and thermal simulation ultimately leads to a significant overhead in computational costs, which is why an innovative approach is needed to bring those two domains together.

3. RESEARCH OBJECTIVE AND APPROACH

The evaluation of newly developed electrified construction machinery and its subsystems, accordingly, regarding the thermal management is sensibly conducted using simulation tools for the system description. The relevant subsystems include mechanical kinematics, hydraulics and thermal processes, which raises the obstacle of conflicting time scales between the hydro-kinematic and the thermal characteristics of the system as derived in chapter 2. While the characteristic of hydraulic and kinematic subsystem is highly dynamic, the thermal behaviour is long-term in comparison. A holistic simulation of these domains, reconciling the significantly different time scales, requires innovative simulation methods, which outlines the research gap.

The objective of this contribution is to introduce a holistic simulation methodology, that enables coupling kinematics, hydraulics and thermodynamics while concurrently reducing simulation time and effort to a minimum.

The proposed solution provides for a reduction of the coupling between mechanic-hydraulic and thermal behaviour from bidirectional to unidirectional. This allows the simulations to be separated.

The mechanical part of the simulation is modelled as a multi body system, depicting the bodies' kinematics. Kinematics and hydraulics both are simulated inversely and unidirectionally coupled using the inverse kinematics' output as the inverse hydraulics' input. The hydraulics output combines the results of the hydraulic-mechanical coupling and is input of a subsequent forward thermodynamic simulation. To match the longer time span of thermodynamic phenomena the short-term hydro-mechanical simulation output is concatenated over time. [12]

Using this simulation methodology highly dynamic phenomena, mainly settling-in behaviour, in hydraulics are omitted to make the simulation less CPU-intensive. Furthermore, the hydro-mechanical simulation must be calculated only once, instead of repeatedly over the whole period the thermal behaviour takes up. Thereby, the thermal simulation becomes substantially more efficient and less effortful.

4. SIMULATION METHODOLOGY

The simulation methodology for modelling electrified construction machinery and similar applications according to thermal behaviour taking electric and hydraulic losses and external thermal influences like thermal radiation into account, bases on the idea of a holistic and simultaneously efficient simulation. The approach for a holistic model needs to handle the time scale difference between mechanical and hydraulic phenomena and thermal phenomena. This challenge is addressed by subdividing the simulation according to the time scale. The hydraulic-mechanical simulation is separated from the thermal simulation by decoupling the two simulations in one direction, so that the unidirectional coupling from the hydraulic-mechanical part to the thermal simulation remains.

The desire for efficiency furthermore demands low effort simulation methods for both, the highly dynamic mechanical and hydraulic part and the long-term thermal behaviour. For this reason, the mechanical and the hydraulic simulation are unidirectionally coupled, as well. This way it is possible to implement inverse simulations to the mechanics as well as the hydraulics. In comparison, the conventional forward simulation uses the input of manipulated variables e.g. force in mechanics or valve positions in hydraulics and simulates the causal system behaviour, outputting dependent variables like movements or pressure build-up, principally depicted in Figure 2.1 a). Whereas the inverse simulation illustrated in Figure 2.1 b) predefines the dependent variables, e.g. a position of components in a mechanical system and calculates backwards to determine the input variables' values required to achieve the predefined dependent variables. The inverse simulation is advantageous regarding computational effort and thus simulation time due to the omission of transient phenomena, especially oscillations, which moreover reduces the partial differential equation system to a nonlinear ordinary differential equation system.

Mechanical Simulation

The overall simulation methodology is presented in Figure 4.2. As described the mechanical simulation is implemented inversely as a multi body system (MBS) simulation, that is generally able to provide approximately realistic forces, torques and displacements for

standard tasks and work patterns of systems. The inverse mechanical simulation is reduced to a mere kinematic simulation, excluding the transient effects considered in a forward simulation, as mentioned in chapter 2. The highly dynamic transient effects are computationally expensive but their impact on the thermal behaviour of the system is insignificant, allowing for them to be neglected in a thermal analysis.

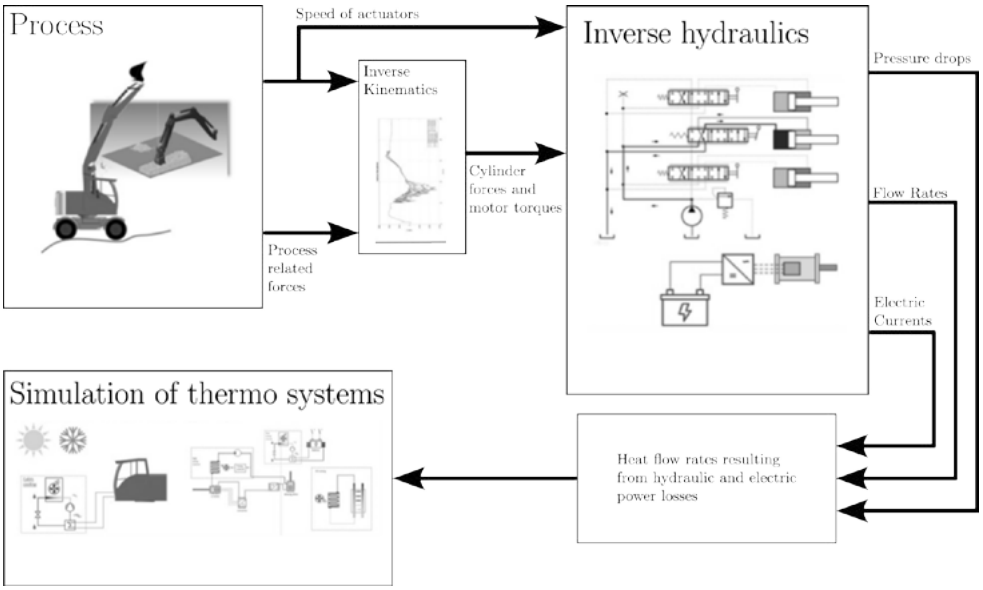


Figure 4.1: Overview of the simulation methodology.

For the inverse kinematic simulation of construction machinery, first the movements of all relevant components of the machine are defined, determining the free degrees of freedom (DoF) and considering restrictions of the component's connections in form of bearings and joints. This results in internal and external forces and torques. The external forces and torques arise from gravity and the interaction of the construction machine with the environment and are forwarded into the system where they cause internal forces and torques in the joints. The joints comprise bearings, swivel and thrust joints as well as the hydraulic cylinders also considered as thrust joints. In the unconstrained DoFs of the bearings and conventional joints, the movements are simply performed, since there is no counterforce or torque, while the forces and torques in the constrained DoFs are supported by the bearings. The hydraulic cylinders experience forces along the direction of piston movement since they can apply a counterforce. This counterforce is determined by the inverse kinematic simulation for each cylinder and describes the required actuator force to achieve the predefined movement and displacements, respectively. The determined cylinder forces subsequently serve as input for the hydraulic simulation. Compared to complex inverse kinematic problems like robotics, where iterative methods are necessary to reach a solution with inverse kinematics since a multivariant problem is looked at, the motion of an excavator or similar piece of equipment is quite simple with a small solution space. Therefore, the inverse kinematic solution is simpler and therefore more accurate. [17]

Hydraulic Simulation

The hydraulic subsystem is modelled using mechatronic system simulations with specified hydraulic and mechatronic components, e.g. cylinders with given Pivot-point. Like for the mechanic simulation inverse hydraulic simulation is used to exclude highly dynamic effects of a real hydraulic control system that are likewise neglectable for the overall result of the thermal system behaviour.

The inverse hydraulic simulation resorts to the inversely calculated required forces from the kinematic simulation. The inverse hydraulic uses the equation system, equivalent to the equation system in Figure 2.1 b) to calculate the required pressures and therefore volume flows in the cylinders, parameterised by their cross section and stroke, and valve positions as well as electric control signals for the system control for providing the desired actuator forces in the mechanical system.

The perfused valves, the piston movements and the electric currents for the control signals are subject to thermal losses. These losses represent the relevant thermal influence of the hydraulic subsystem on the overall machine and is consequently input to the thermal simulation. The simulated period is short, though, compared to the time span of thermal phenomena. And while the other thermal influences are nearly stationary the thermal influence of the mechanic and hydraulic system varies during operating of the construction machine and depending on the operating condition. To depict an average of the operating conditions, weighed with their assumed average duration, it is recommended to generate and simulate operation cycles in the mechanical and hydraulic subsystems. The determined thermal losses during the operation cycle are concatenated over time and used for the thermal simulation. The repeated operation cycles, provide a realistic approximation of the system's thermal behaviour in a real application, while in contrast a single operation only provides information about the thermal impact of this operation and neglects relevant further influences. The estimation is feasible for the initial system design and component dimensioning required to fulfil the requirements in terms of temperature management of the overall system.

Thermal Simulation

The thermal simulation is conducted conventionally forward. The concatenated heat input from the mechatronic and hydraulic simulation, arising from electric losses in motors, the inner resistance of the battery, inverters and hydraulic losses due to drops in pressure within the valves, transient losses in the compressor and exhaust heat in the converter are used as an input modelled either using the efficiency factor or if available characteristic curves. Furthermore, external heat sources like the environmental temperature and solar radiation on the cabin of the machine and internal heat sources like the air conditioning including the cooling cycle. The heat flux in between the subsystems and domains is calculated using dedicated connector elements describing the thermal transfer behaviour based on the fundamental thermodynamic equations as described in chapter 2 and local temperatures. The temperature changes and heat flux are provided as a result. From these outputs information about the temperature control or the overall thermal energy loss can be deduced. Further examples of potential interpreted results are presented in the following chapter, where the methodology is exemplary applied to an excavator.

5. Exemplary Application by Means of an Electrified Excavator

To demonstrate its operating principle and benefits the presented methodology is applied to a case study. The electrified excavator presented in the following is a generic design without a real-world equivalent but provides a level of detail close to a real-world system. For the mechanical domain the drivetrain and sewing drive as well as the excavator arm are modelled as multibody systems. The individual hydraulic cylinders for the excavator arm are supplied by a hydraulic circuit, depicted in Figure 5.2. The drivetrain, sewing drive, oil pump and coolant pumps are all driven by alternating current electric motors, which are supplied from power electronics, converting the direct current from the battery block to appropriate alternating currents for the electric motors. The multiple thermal circuits in the model compensate for the efficiency losses caused by transformation losses or friction, which produces heat. The oil of the hydraulic circuit is passed through a heat exchanger, modelled as a single fluid circuit giving off heat to a thermal mass, to compensate for the hydraulic losses. A low temperature coolant circuit is modelled for cooling the power electronics and the electric drives operating the pumps and drivetrain, comprising two cooling circuits with liquid ethylene glycol in one and coolant R134A in the other circuit, exchanging heat via contact to a chiller. Another coolant circuit with exchange of heat between fluid circuit and coolant circuit is deployed to regulate the battery system. Contrary to the oil cooling the coolant circuits for the power electronics and battery are temperature controlled since the electric components are only operable close to standard conditions. Finally, cabin air conditioning is included in the thermal domain analysis, which ultimately represents a temperature control giving off heat to a thermal mass, too. Separating the different cooling circuits allows to individually regulate each temperature in the most efficient working point of the component or subsystem. Aside from the cooling circuits heat radiation from the cabin towards the environment is considered as a heat sink as a counterpart to the solar radiation into the cabin.

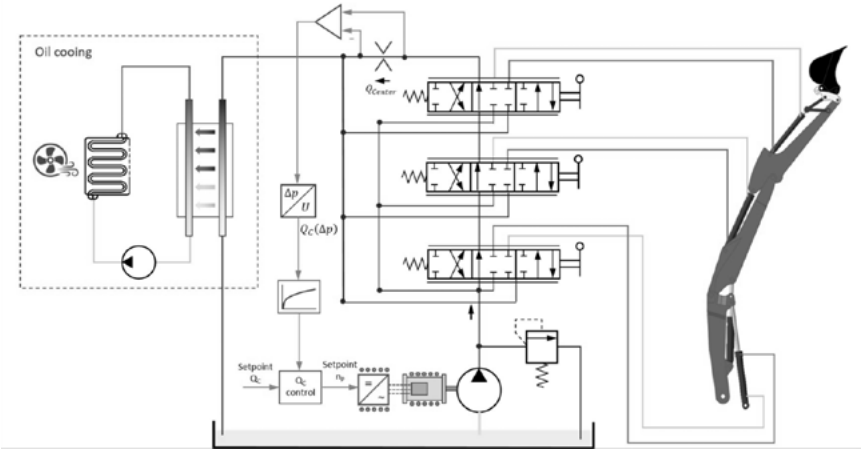


Figure 5.1: Schematic hydraulic circuit diagram of the excavator arm

In total the model consists of a mechanical, electrical, thermal and hydraulic dimension with, multiple, partly independent, circuits. A schematic overview of the system, including mechanical, hydraulic and thermal coupling of the excavator arm, is given in Figure 5.1.

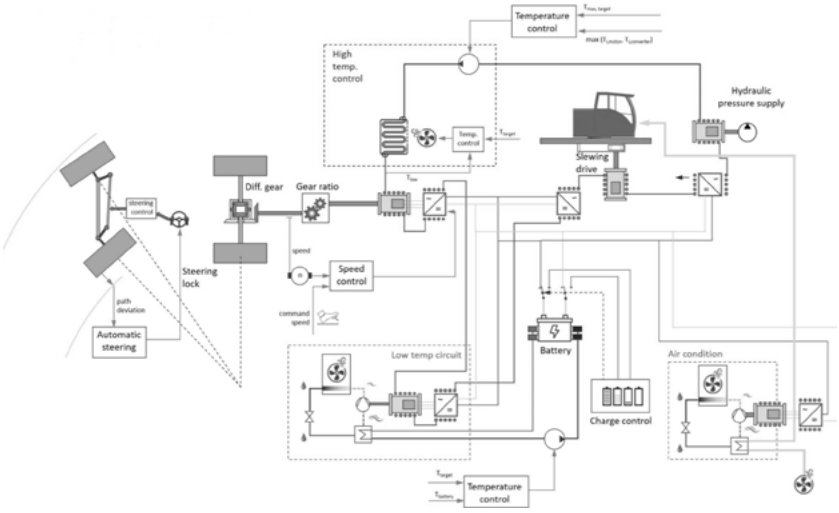


Figure 5.2: Electric actuators and cooling circuits of the exemplary excavator arm

To conduct a holistic system simulation and analyse the thermal management, which was implemented as described, a load cycle and environmental conditions are needed. The load cycle needs to represent realistic work pattern for the construction machine including slewing, drive patterns and boom motion. For the simulation the following work cycle is assumed:

- 5x Moving earth – based on a particle Simulation
- 1x acceleration/deceleration
- 1x uphill and downhill drive
- 1x idle slewing
- 1x idle boom motion

A single duty cycle takes four minutes and is looped 10 times for a single simulation to reach stationary conditions for each physical domain, which leads to a depicted real time of 40 minutes.

Figure 5.3 shows the drivetrain torque for the working cycle compared to the temperature development in the cooling circuit of the power electronics in Figure 5.4. The plots show how the thermal domain develops over multiple cycles while the drivetrain reaches its stationary states immediately. These results are consistent with the known discrepancy in the timescale between long-term thermal phenomena and considerably shorter-term mechanical, electrical and hydraulic phenomena. It also highlights the need to concatenate load cycles multiple times to obtain reliable information about the system behaviour.

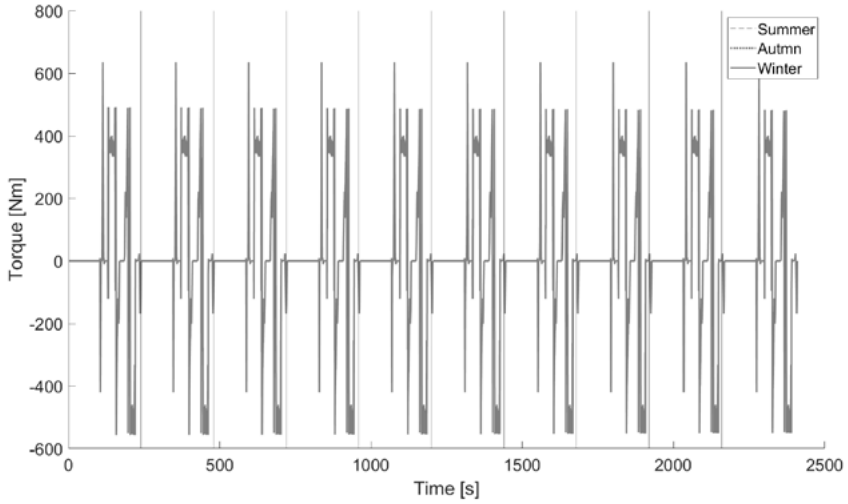


Figure 5.3: Torque progression over time of the drivetrain

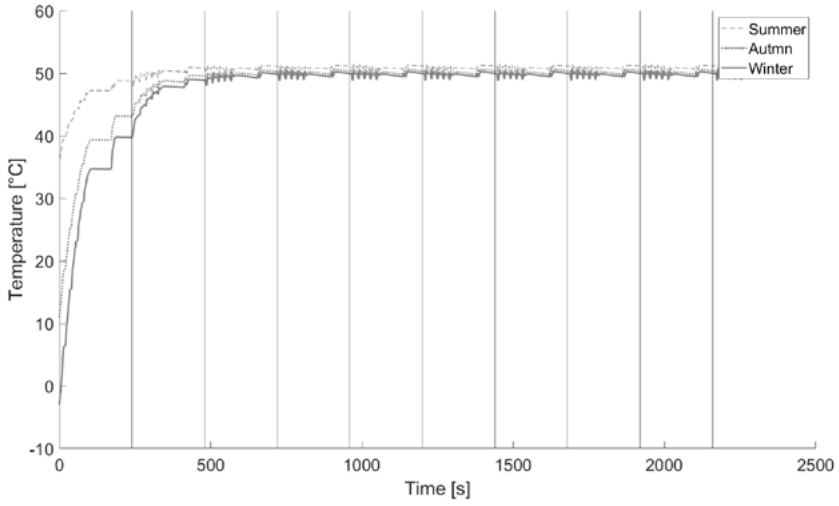


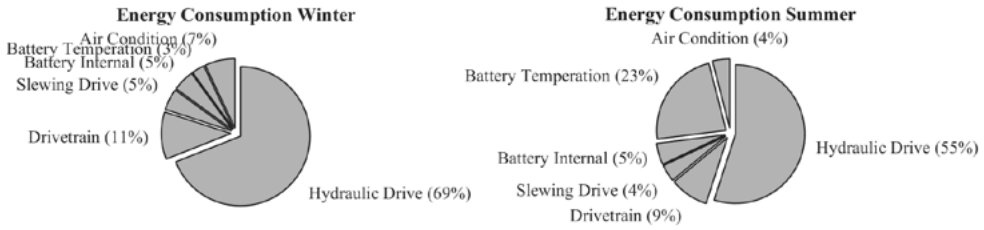
Figure 5.4: Temperature development over time of the power electronics coolant

Figure 5.4 also contains multiple plots for different environmental conditions that are tested to identify the dependency of the thermal management system on the environment. Therefore different, central European, seasons were taken as a basis:

- Summer: Temperature: 35 °C; Humidity: 78 %
- Autumn: Temperature: 11 °C; Humidity: 98 %
- Winter: Temperature: -3 °C; Humidity: 85 %

This analysis for different environmental conditions allows the evaluation of the relative energy consumption of the tested design. Figure 5.5 shows a comparison between summer

and winter according to the proportional energy consumptions arising after the system reached its stationary thermal operating conditions.



These evaluations are only possible because the system is considered holistically. It allows, aside from analysis of the total energy consumption, to identify design areas that are most promising for improvement. Furthermore, this example already emphasizes the cruciality of the environmental conditions for the design of thermal management systems. The battery cooling system for example has a significantly higher impact in the summer than the winter, compared to the other considered subsystems in this study. An analysis that ignores these differences could lead to flawed design.

6. DISCUSSION

The proposed simulation methodology serves as an analysis tool during the system development phase with the aim to enable a quantified evaluation and comparison of system concepts regarding the thermal aspects. Prerequisite for the use of the simulation methodology is, as with simulations in general, a sufficient level of details of the design to have the necessary information for a calculation available. This allows the methodology's application during the late concept and draft phase of the product development process. In this phase the simulation following the presented methodology allows a data-based comparison of different system architectures and an initial dimensioning of the components. Besides information about interrelations within the system and the impact of present influencing factors become known and understandable.

The simulation methodology omits the transient and highly dynamic processes of the mechanical and the hydraulic subsystem, as their impact on the long-term thermal development within the system is insignificant and therefore feasible to neglect. Furthermore, transient effects like friction and leakage are only represented to a limited extend by efficiency factors. These simplifications lead to a limited accuracy of the resulting values but still provides useful information about the main interrelations within the system and the main influencing factors as well as an estimation of the required subsystem and component dimensions. Another limitation of the inverse kinematics is that the exact and physically possible moving path of the components is not calculated. This is acceptable for the initial concept simulation because the exact moving path' do not influence the results for the forces substantially. For cases that require a detailed analysis of these effects as well as for the further steps of the product development process, an adaption of the methodology is recommended, resorting to transient forward simulations. As this approach revokes the decisive advantage of reduces computational and time effort, it is only sensible during later development phases or in separate simulations.

7. CONCLUSION AND OUTLOOK

Electrified construction machinery offers a variety of advantages, especially the compliance with the latest emission regulations. Using the simplifications of inverse kinematics and hydraulics, as well as the unidirectional coupling of mechanics to hydraulics and hydraulics to thermodynamics, the presented methodology enables a holistic and efficient simulation and evaluation of the thermal properties of a mechanic-hydraulic system, particularly designated for electrified construction machinery. This methodology allows for system understanding, system concept comparison and initial dimensioning during the early product development phases.

The example of the excavator demonstrates the applicability of the proposed simulation methodology on an academically level and provides the prospect on an application in a real system. As an outlook the verification of the methodology by a comparison to a fully coupled simulation and measurement data of a real example of an excavator are of interest to examine the accuracy of the simulation. To furthermore validate the application and the advantages of the simulation methodology the application in a practical development process is pending.

8. REFERENCES

[1] J. Feldhusen and K.-H. Grote, *Pahl/Beitz Konstruktionslehre - Methoden und Anwendung erfolgreicher Produktentwicklung*, 8th ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-29569-0.

[2] D. E. Keyes *et al.*, "Multiphysics simulations - Challenges and opportunities," *Int J High Perform Comput Appl*, vol. 27, no. 1, pp. 4–83, Feb. 2013, doi: 10.1177/1094342012468181.

[3] F. Un-Noor *et al.*, "Off-Road Construction and Agricultural Equipment Electrification: Review, Challenges, and Opportunities," Sep. 01, 2022, *MDPI*. doi: 10.3390/vehicles4030044.

[4] E. Scolaro, M. Beligoj, M. P. Estevez, L. Alberti, M. Renzi, and M. Mattetti, "Electrification of Agricultural Machinery: A Review," 2021, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2021.3135037.

[5] X. Huang, Q. Huang, H. Cao, W. Yan, L. Cao, and Q. Zhang, "Optimal design for improving operation performance of electric construction machinery collaborative system: Method and application," *Energy*, vol. 263, p. 125629, Jan. 2023, doi: 10.1016/J.ENERGY.2022.125629.

[6] A. Albers and N. Bursac, "Produktgenerationsentwicklung-Bedeutung und Herausforderungen aus einer entwicklungsmethodischen Perspektive," 2015. https://www.researchgate.net/publication/279539363

[7] J. Ponn and U. Lindemann, *Konzeptentwicklung und Gestaltung technischer Produkte*. Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-20580-4.

[8] M. O. Faruque, V. Dinavahi, M. Steurer, A. Monti, K. Strunz, J. A. Martinez, G. W. Chang, J. Jatskevich, R. Iravani, and A. Davoudi, *Interfacing Issues in Multi-Domain Simulation Tools.* IEEE TRANSACTIONS ON POWER DELIVERY, VOL. 27, NO. 1, JANUARY 2012

[9] A. A. . Shabana, *Computational dynamics*. John Wiley, 2001.

[10] Garcia de Jalon, Javier, and Eduardo Bayo. *Kinematic and Dynamic Simulation of Multibody Systems : The Real-Time Challenge*. 1st ed. 1994. New York, NY: Springer New York, 1994.

[11] Bruno Siciliano, Luigi Villani, *An inverse kinematics algorithm for interaction control of a flexible arm with a compliant surface*, Control Engineering Practice, Volume 9, Issue 2, 2001, Pages 191-198, ISSN 0967-0661

[12] Korkealaakso, P. (2009). *Real-time simulation of mobile and industrial machine using the multibody simulation approach*. Lappeenranta University of Technology.

[13] Rapetti L, Tirupachuri Y, Darvish K, Dafarra S, Nava G, Latella C, Pucci D. Model-Based Real-Time Motion Tracking Using Dynamical Inverse Kinematics. *Algorithms*. 2020; 13(10):266.

[14] Luostarinen LO, Åman R, Handroos H. Development of control interface for HIL simulation of electro-hydraulic energy converter. International Review on Modelling and Simulations. 2014;7(4):653-660. p.99

[15] J. R. . Melcher, *Continuum electromechanics*. MIT Press, 1981.

[16] K. Schmitz, *Fluidtechnik - Systeme und Komponenten*, vol. 9. Düren: Shaker Verlag, 2022.

[17] Aristidou, Andreas, and Joan Lasenby. "Inverse kinematics: a review of existing techniques and introduction of a new fast iterative solver." (2009).

Biographies



Dr.-Ing. Marcel Neu received the bachelor's degree and master's degree in mechanical engineering from Technical University of Darmstadt in 2014 and 2017, and the philosophy of doctorate degree in Mechanical Engineering from Technical University of Darmstadt in 2022, respectively. He is currently working as a simulation engineer and head of continuum simulation group at marpitec GmbH. His research areas include computational fluid dynamics, multiphase flows, supersonic flows and fluid structure interactions.



Dr.-Ing. Fiona Tichi received the bachelor's degree and master's degree in mechanical engineering from Technical University of Darmstadt in 2015 and 2018, and the philosophy of doctorate degree in Mechanical Engineering from Technical University of Darmstadt in 2023, respectively. She is currently working as a simulation engineer in the continuum simulation group at marpitec GmbH. Her research areas include Finite Elements Method, product development and Additive Manufacturing.

On Displacement Controlled Drive Networks

[1]Lasse Schmidt & [2]Mikkel van Binsbergen-Galán

[1] *AAU Energy, Aalborg University, Denmark, lsc@energy.aau.dk*

[2] *AAU Energy, Aalborg University, Denmark, mbg@energy.aau.dk*

Abstract.

In the efforts to improve the energy efficiency of hydraulic systems, a key initiative may be to eliminate the throttle losses induced by hydraulic control valves. Existing solutions targeting this aspect include standalone electro-hydraulic variable-speed drives for control of hydraulic actuators such as cylinders and motors. A main challenge of this approach may be the electric component sizes, hence their cost, required to meet both the maximum force/torque and speed of a cylinder/motor, even though not concurrently. In case of systems with multiple hydraulic cylinders/motors to be actuated, the idea of realizing highly integrated systemwide drives in terms of so-called electro-hydraulic variable-speed drive networks have emerged in recent years. As opposed to standalone electro-hydraulic variable-speed drives, such drive networks enable power sharing not only on the electric side, but also on the hydraulic side. A consequence of this is that components may be downsized considerably in many applications, while maintaining a similar efficiency. However, the cost may still be prohibitively high in many industry segments, especially due to the number of electric machines and drives. This provides the incitement to investigate hydraulic drive networks enabling control via variable displacement machines rather than via multiple variable-speed electric machines. This approach may be feasible due to the presumed less cost of variable displacement machines compared to electric machines/drives. This paper analyses the feasibility of hydraulic drive networks using a single electric machine powering multiple variable displacement machines, organized in network configurations. Especially, emphasis is placed on the component sizing in such displacement-controlled drive networks when applied in a crane application. Furthermore, the resulting displacement-controlled drive networks are compared to existing individual displacement-controlled drive approaches, where it is found that significant reductions in variable displacement machine sizes may be enabled by drive network structures.

Keywords. Displacement-controlled Drive Networks, Displacement-controlled Drives Electro-hydraulic Drive Networks, Energy Efficiency, Crane Actuation, Component Sizing.

1. INTRODUCTION

Facing the current climate crisis, efforts to reduce carbon emissions are pursued throughout societies, including various industries. In this endeavour, machinery relying on hydraulic

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

systems as means of power transmission is especially relevant to consider. This is partly because actuation in such machinery commonly is associated with large forces and power, and partly because traditional hydraulic systems rely on throttle valves to be controlled. That is, relying on control via loss adjustment. Consequently, traditional hydraulic machinery is inherently energy inefficient and associated with substantial energy losses and related emissions [1]. Hydraulic machinery may be stationary or mobile, and in the former case these are often already electrically powered, whereas the latter case remains powered by internal combustion engines in most cases. Trends suggests, however, that mobile hydraulic machinery are increasingly electrified by replacing internal combustion engines with battery powered electric machines. To enable an actual decarbonization of such machinery, the electric energy used must be provided by renewable energy sources, and furthermore, preferably only use as much energy as needed to minimize stress on the electric grid in a world where electrification of society is a key priority. Here, the energy inefficiency of traditional hydraulic systems poses a great challenge due to their high energy losses. As a first approach to reduce losses, variable-speed fixed or variable displacement pumps may be applied, allowing to reduce energy losses of hydraulic power supplies by adjusting shaft speeds to flow requirements, more optimal operating points etc. [2,3]. However, a substantial part of the losses in traditional hydraulic systems are the principal losses associated with throttle control, and to enable a real impact in terms of reduced energy consumption and emissions, these principal losses must be significantly reduced.

Approaches to achieve such reductions in energy consumption involves avoiding or limiting the use of control valves, and here electro-hydraulic standalone drives for actuation of e.g. hydraulic cylinders may be applied [4-8]. In systems containing two or more cylinders, however, it may be more feasible to consider more dedicated multi-axis electro-hydraulic drives such as the so-called HHEA [9,10] or so-called electro-hydraulic variable-speed drive networks [11-15]. Even though these electro-hydraulic systems suggest high energy efficiencies, the use of one or more electric machines per cylinder or motor to be actuated, may cause drive cost to be prohibitively high in an industrial setting. An approach to reduce the number of required electric machines may be so-called displacement control, where the control is shifted from speed variability to displacement variability. A recent displacement control approach suggests this to be combined with traditional valves for low power functions [16,17]. The topic of displacement-controlled drives has been considered extensively, see e.g. [18-21]. However, a one-pump per actuator requirement may also here cause cost levels to be prohibitively high [21]. To overcome this challenge the integration of on/off valves was proposed in [22], allowing to control individual actuators in a sequential way [21]. A drawback of this approach, however, is that concurrent control of all actuators is not possible.

An alternative approach mitigating the abovementioned challenges with displacement control, may be to organize multiple variable pumps on a common shaft in network configurations, similar to electro-hydraulic variable-speed drive networks. That is, having the same offset with a common shaft with multiple pumps as e.g. [18-22], but implement *appropriate* short circuit connections between cylinder chambers and *appropriate* hydraulic connections between pump and actuator ports in a network kind of way, forming *displacement-controlled drive networks* (DCDNs). This paper presents an analysis of displacement-controlled drive networks and their design in the context of a crane application. It is shown that the total required pump displacement may be significantly

reduced compared to *conventional* displacement-controlled multi-actuator drive systems, while allowing concurrent control of all actuators. The nomenclature used throughout the paper is explained in the Appendix.

2. CRANE APPLICATION & DESIGN REQUIREMENTS

The drive designs considered in the following are based on the load imposed by the two-axis knuckle boom crane depicted in Figure 1.1. Any drive system considered must enable a payload mass of up to $m_L = 5000$ [kg], moving at 200 [mm/s] both vertically and horizontally throughout the operating space, i.e. $\dot{x}_L = \pm$ 200 [mm/s], $\dot{y}_L = \pm$ 200 [mm/s]. The following serves as a case study the load and speed requirements have been chosen to provide reasonable flow and pressure levels and are not related to specific application tasks.

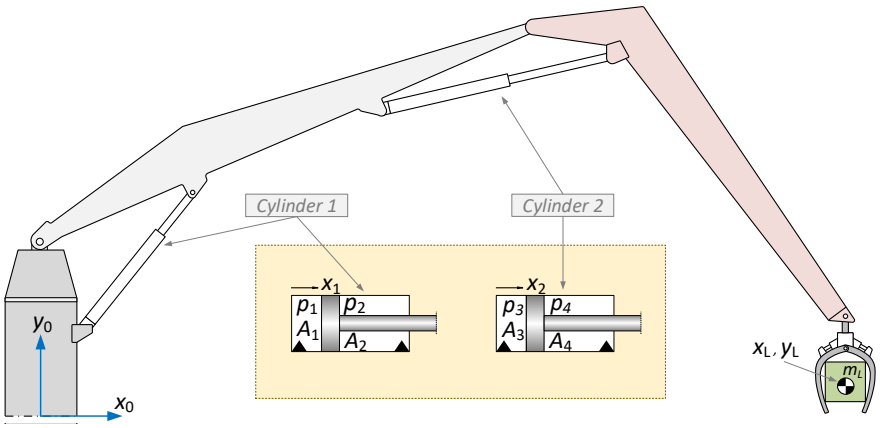


Figure 1.1. Two-axis knuckle boom crane application for investigations.

The hydraulic piston forces for the two cylinders are given by Eqs. (1.1).

$$F_{L1} = A_1 p_1 - A_2 p_2 \quad , \qquad F_{L2} = A_3 p_3 - A_4 p_4 \tag{1.1}$$

Using the crane parameters considered in [11] and a payload of 5000 [kg], the piston forces appear as depicted in Fig. 1.2. when considering the operating space spanned by the cylinder pistons.

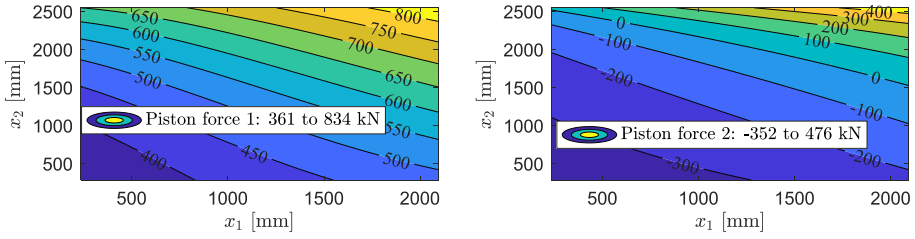


Figure 1.2. Cylinder forces ([kN]) of knuckle boom crane at payload of 5000 [kg].

The piston speeds evaluated at these extreme payload speeds, $\dot{x}_L = \pm$ 200 [mm/s], $\dot{y}_L = \pm$ 200 [mm/s], are depicted in Fig. 1.3., accounting for the gearings induced by the crane mechanism itself.

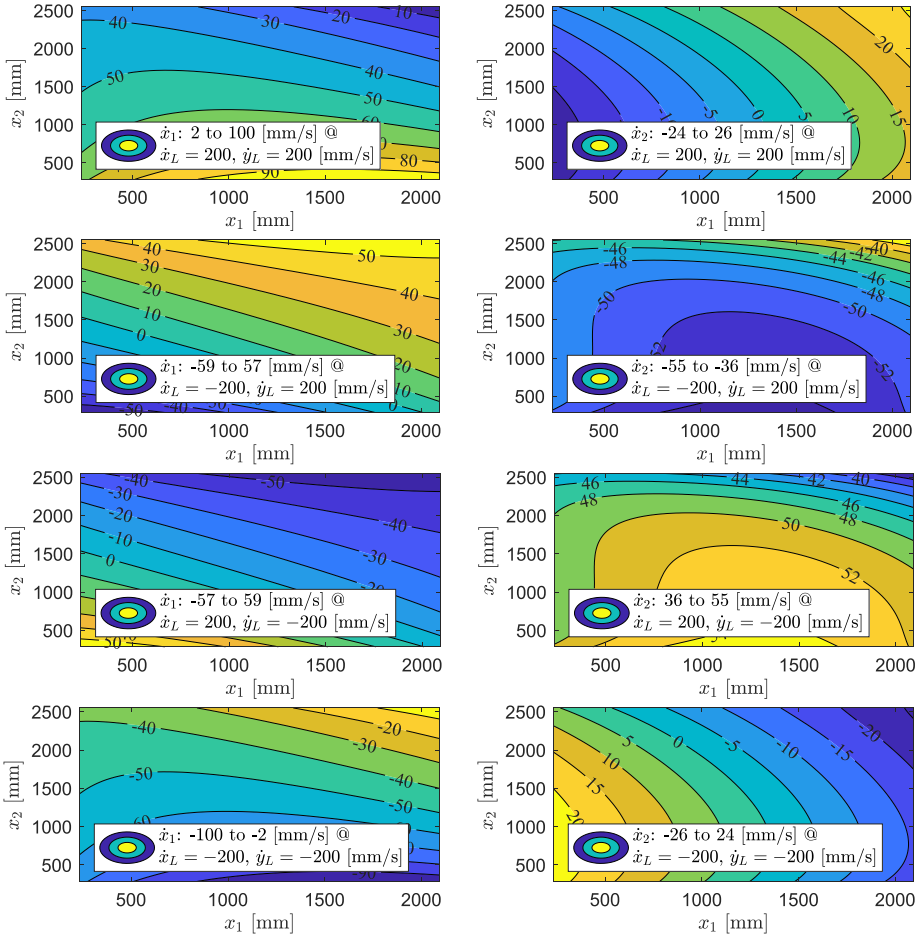


Figure 1.3. Cylinder piston speed at max./min. payload speeds and piston positions.

The piston forces and piston velocities depicted in Figures 1.2. and 1.3. serves as design requirements in the following sections.

Note: The following disregards the pilot supplies for the swivel control mechanisms and other flow losses as well as torque losses. Furthermore, all drives considered are assumed to operate at a constant shaft speed of 3000 [rpm].

3. INDIVIDUAL DISPLACEMENT CONTROLLED DRIVES

As mentioned in the introduction, individual displacement-controlled (IDC) drives have been proposed in several versions and for several applications [18-22]. Here the case of a multi-actuator IDC drive without switching valves are considered as a benchmark for displacement-controlled drive networks (DCDNs), as this allows for individual control of both cylinder pistons simultaneously, which is also the objective for the DCDNs. For the crane application the dual cylinder IDC drive depicted in Fig. 2.1 is considered.

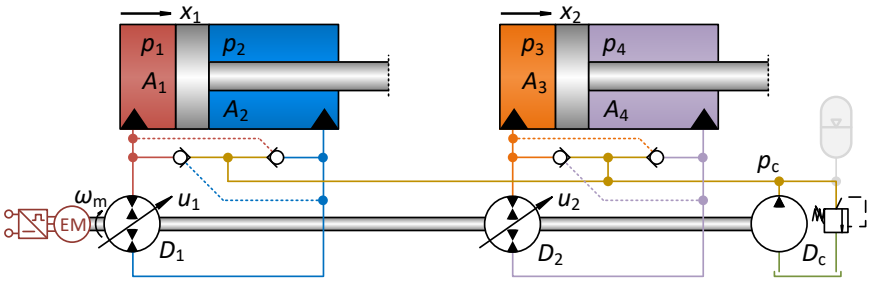


Figure 2.1. Individual displacement-controlled drive for crane application.

In literature IDC drives are generally equipped with an accumulator attached to the charge pump outlet, forming a charge line having the purpose of compensating the cylinder differential flows, flow losses and fluid exchange related to thermal management and filtering, and to supply the swivel control valves [22]. Hence, if no fluid exchange functionality is implemented as is the case in Fig. 2.1, the charge pump displacement size may be rather small. On the other hand, if no accumulator is desired, the charge pump must be sized to handle the cylinder differential flows (and in practice also flow losses and pilot flows). Considering traditional transmission pumps, these are equipped with a build-in charge pump. From these considerations, three different cases may be established.

Case 1: IDC without accumulator, but with charge pump

Case 2: IDC with accumulator and charge pump

Case 3: IDC with accumulator, but without charge pump (utopia)

Here Case 1 and Case 2 may be applied in practical use, and Case 2 is a common implementation. Case 3 is highly idealized and could work only if flow losses were absent and is thus a utopia. However, this case serves as an idealized benchmark. The theoretically required displacement sizes for all three cases are considered in the following. It should be noted that for all three cases the maximum steady state pressure here is 190 [bar].

3.1. Case 1 sizing – IDC with Charge Pump & No Accumulator

Considering Fig. 2.1 and the Case 1 conditions, the displacement machine flows in a steady state are given by Eqs. (3.1), (3.2).

$$Q_1 = D_1 u_1 \omega_m = \begin{cases} A_1 \dot{x}_1, & p_1 \geq p_2 \\ A_2 \dot{x}_1, & p_1 < p_2 \end{cases} \quad , \quad Q_2 = D_2 u_2 \omega_m = \begin{cases} A_3 \dot{x}_2, & p_3 \geq p_4 \\ A_4 \dot{x}_2, & p_3 < p_4 \end{cases} \tag{3.1}$$

$$Q_c = D_c \omega_m = (A_1 - A_2)\dot{x}_1 + (A_3 - A_4)\dot{x}_2 \tag{3.2}$$

From the cylinder piston forces and speed ranges of Figures 1.2 and 1.3, the required individual displacements and the total displacement are obtained as Eqs. (3.3).

$$D_1 = 123\ [\text{ccm}]\ ,\ D_2 = 48\ [\text{ccm}]\ ,\ D_\text{c} = 70\ [\text{ccm}] \rightarrow D_\text{tot} = 241\ [\text{ccm}] \tag{3.3}$$

3.2. Case 2 sizing – IDC with Charge Pump & Accumulator

Due to the application of both a charge pump and an accumulator, the charge pump could be chosen to only compensate flow losses (and in reality, also the swivel control valve pilot flow) at the specified shaft speed. Here, the sizing takes offset in charge pump sizing applied in the Bosch Rexroth A4VG series [23], with relevant information outlined in Table 1.

Table 1: Charge pump vs. nominal pump sizes for A4VG [23].

A4VG	**28**	**40**	**56**	**71**	**90**	**125**
Nominal size in [ccm]	28	40	56	71	90	125
Charge pump size [ccm]	6.1	8.6	11.6	19.6	19.6	28.3
Charge pump size in % of nominal size	22	22	21	28	22	24

For the A4VG [23], the charge pumps are averagely sized at 23% of the nominal pump size. To be conservative (sizing in favour of Case 2), the charge pump size is here assumed to be 20% of the total nominal pump size required. In this case, the displacement sizes appear as Eqs. (3.4).

$$D_1 = 123\ [\text{ccm}]\ ,\ D_2 = 48\ [\text{ccm}]\ ,\ D_\text{c} = \frac{D_1+D_2}{5} = 34\ [\text{ccm}] \rightarrow D_\text{tot} = 205\ [\text{ccm}] \tag{3.4}$$

3.3. Case 3 sizing – IDC without Charge Pump

As Case 3 assumes that no charge pump is required, total displacement simply corresponds to the sum of the nominal displacement, also required in Case 1 and Case 2, and is hence given by Eqs. (3.5).

$$D_1 = 123\ [\text{ccm}]\ ,\ D_2 = 48\ [\text{ccm}]\ ,\ D_\text{c} = 0\ [\text{ccm}] \rightarrow D_\text{tot} = 171\ [\text{ccm}] \tag{3.5}$$

3.4. IDC Shaft Torque Requirements

In evaluating the resulting torque applied to the common drive shaft, the chamber pressures are estimated as follows with cylinder forces $F_{\text{L1s}} = (A_1 - A_2)p_\text{c}$, $F_{\text{L2s}} = (A_3 - A_4)p_\text{c}$.

$$p_1 = \begin{cases} \dfrac{A_2 p_c + F_{L1}}{A_1} & , \quad F_{\text{L1}s} < F_{\text{L1}} \\ p_c & , \quad F_{\text{L1}s} \geq F_{\text{L1}} \end{cases} \quad , \quad p_2 = \begin{cases} p_c & , \quad F_{L1s} < F_{L1} \\ \dfrac{A_1 p_c - F_{\text{L1}}}{A_2} & , \quad F_{L1s} \geq F_{L1} \end{cases} \tag{3.6}$$

$$p_3 = \begin{cases} \dfrac{A_4 p_c + F_{\text{L2}}}{A_3} & , \quad F_{\text{L2}s} < F_{\text{L2}} \\ p_c & , \quad F_{\text{L2}s} \geq F_{\text{L2}} \end{cases} \quad , \quad p_4 = \begin{cases} p_c & , \quad F_{\text{L2}s} < F_{\text{L2}} \\ \dfrac{A_3 p_c - F_{\text{L2}}}{A_4} & , \quad F_{\text{L2}s} \geq F_{\text{L2}} \end{cases} \tag{3.7}$$

The common shaft torque for the IDC drives is then given by Eq. (3.8).

$$\tau_\text{IDC} = D_1 u_1 (p_1 - p_2) + D_2 u_2 (p_3 - p_4) + D_\text{c} p_\text{c} \tag{3.8}$$

Here, the factors D_1u_1 and D_2u_2 are obtained using Eqs. (3.1) and Eqs. (3.9), noting that the shaft speed ω_m is constant.

$$D_1u_1 = \frac{Q_1}{\omega_\mathrm{m}} \quad , \quad D_2u_2 = \frac{Q_2}{\omega_\mathrm{m}} \tag{3.9}$$

Evaluating Eq. (3.8) with Eqs. (3.9) for the cylinder force maps Fig. 1.2. and the cylinder velocity maps Fig. 1.3., the maximum shaft torque for the three cases is given by Eqs. (3.10), (3.11) and (3.12).

$$\text{Case 1: } \tau_\mathrm{IDC}|_\mathrm{max} = 139 \text{ [Nm]} \tag{3.10}$$

$$\text{Case 2: } \tau_\mathrm{IDC}|_\mathrm{max} = 127 \text{ [Nm]} \tag{3.11}$$

$$\text{Case 3: } \tau_\mathrm{IDC}|_\mathrm{max} = 116 \text{ [Nm]} \tag{3.12}$$

The reason for the reduced maximum shaft torque of benchmark case 2 and 3 compared to benchmark case 1 is attributed the reduced charge pump sizes.

4. DISPLACEMENT CONTROLLED DRIVE NETWORKS

Displacement controlled drive networks (DCDNs) for dual cylinder systems may basically appear as depicted in Fig. 4.1. It is desired to use as few displacement machines as possible, while maintaining controllability of both cylinder pistons and the system sum pressure as in [11]. Hence, a single short circuit connection between two chambers should be deployed, necessitating three displacement machines to achieve the desired level of controllability. This, in turn, provides 36 unique ways of realizing the drive connections, which should be considered in context of the design requirements discussed in Section 2.

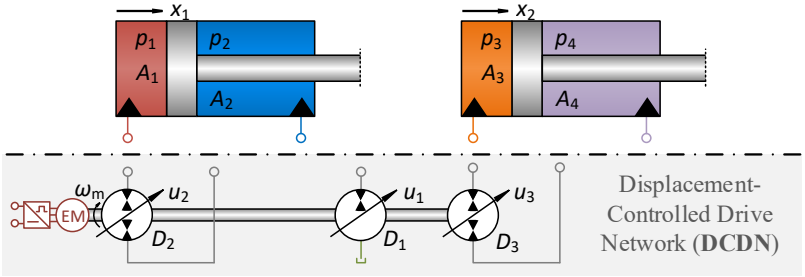


Figure 4.1. Displacement-controlled drive network (DCDN) for crane application.

4.1. DCDN Sizing Example

As an example on component sizing of one of the 36 unique configurations, consider the DCDN in Fig. 4.2. Here, the rod-side chamber of cylinder 1 is short circuited with the bore side chamber of cylinder 2, forming the combined volume *23*. Furthermore, the bore side chamber of cylinder 1 is connected to the common volume 23 via pump 2 and volume 23 is also connected to the rod side volume of cylinder 2 via pump 3 (notation 1 ↔ 23 ↔ 4). Finally, the tank volume 0 is connected to volume 23 via pump 1 (notation 0 ↔ 23).

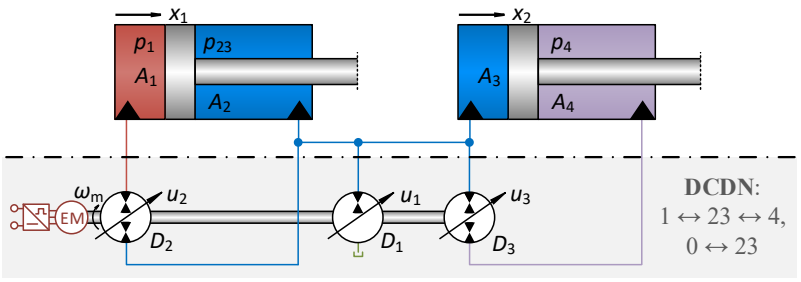


Figure 4.2. Displacement-controlled drive network 1 ↔ 23 ↔ 4, 0 ↔ 23.

The displacement machine flows in a steady state, assuming flow losses absent, are obtained considering flow continuity and are given by Eqs. (4.1)-(4.3).

$$Q_1 = D_1 u_1 \omega_\mathrm{m} = (A_1 - A_2)\dot{x}_1 + (A_3 - A_4)\dot{x}_2 \tag{4.1}$$

$$Q_2 = D_2 u_2 \omega_\mathrm{m} = A_1 \dot{x}_1 \tag{4.2}$$

$$Q_3 = D_3 u_3 \omega_\mathrm{m} = A_4 \dot{x}_2 \tag{4.3}$$

When considering the cylinder piston force and speed ranges in Figures 1.2. and 1.3., the required displacement sizes in Eq. (4.4) are obtained.

$$D_1 = 70\ [\mathrm{ccm}]\,,\ D_2 = 123\ [\mathrm{ccm}]\,,\ D_3 = 26\ [\mathrm{ccm}] \rightarrow D_\mathrm{tot} = 219\ [\mathrm{ccm}] \tag{4.4}$$

This is comparable to the IDC drive, Case 1, as no accumulator is used. It is notable that total required displacement for the IDC of Case 1 is 241 [ccm], whereas it is only 219 [ccm] for the DCDN example considered here.

4.2. Sizing Results for all 36 DCDNs

Applying the same sizing method for all 36 unique DCDNs, the total required displacements presented in Table 2 are obtained. It should be noted that the maximum steady state pressure depends on the short circuit applied, hence this is also specified in Table 2. The DCDNs requiring the lowest and highest displacement are depicted in Fig. 4.3 to the right and to the left, respectively.

From the sizing results in Table 2 it is evident that, even though requiring 9% less displacement compared to the IDC in Case 1, the DCDN example of section 4.1 is not the one requiring the lowest amount of displacement.

The sizing results emphasize the importance of architecture selection, and its influence on required displacement when considered for a certain application. The DCDN denoted 1 ↔ 23 ↔ 4, 0 ↔ 1 is the architecture that requires the lowest amount of displacement, 155 [ccm], whereas the DCDN denoted 13 ↔ 4 ↔ 2, 0 ↔ 2 requires the highest displacement with 337 [ccm] which is more than twice that of DCDN 1 ↔ 23 ↔ 4, 0 ↔ 1.

Table 2: Sizing results for all 36 unique DCDNs.

Configuration	Max. pressure	0 ↔ 13	0 ↔ 2	0 ↔ 4
13 ↔ 2 ↔ 4	294 [bar]	163 [ccm]	234 [ccm]	294 [ccm]
13 ↔ 4 ↔ 2	294 [bar]	197 [ccm]	**337 [ccm]**	268 [ccm]
2 ↔ 13 ↔ 4	294 [bar]	156 [ccm]	225 [ccm]	216 [ccm]
Configuration	Max. pressure	0 ↔ 14	0 ↔ 2	0 ↔ 3
14 ↔ 2 ↔ 3	269 [bar]	199 [ccm]	244 [ccm]	251 [ccm]
14 ↔ 3 ↔ 2	269 [bar]	205 [ccm]	319 [ccm]	250 [ccm]
2 ↔ 14 ↔ 3	269 [bar]	184 [ccm]	253 [ccm]	191 [ccm]
Configuration	Max. pressure	0 ↔ 23	0 ↔ 1	0 ↔ 4
23 ↔ 1 ↔ 4	190 [bar]	216 [ccm]	171 [ccm]	231 [ccm]
23 ↔ 4 ↔ 1	190 [bar]	313 [ccm]	204 [ccm]	268 [ccm]
1 ↔ 23 ↔ 4	190 [bar]	219 [ccm]	**155 [ccm]**	279 [ccm]
Configuration	Max. pressure	0 ↔ 24	0 ↔ 1	0 ↔ 3
24 ↔ 1 ↔ 3	190 [bar]	262 [ccm]	191 [ccm]	198 [ccm]
24 ↔ 3 ↔ 1	190 [bar]	331 [ccm]	196 [ccm]	260 [ccm]
1 ↔ 24 ↔ 3	190 [bar]	247 [ccm]	183 [ccm]	254 [ccm]

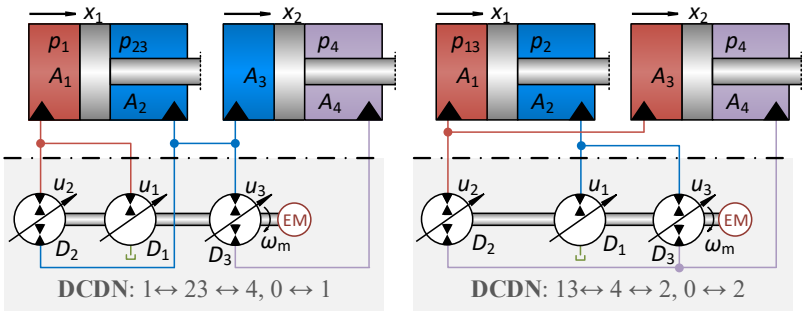


Figure 4.3. DCDNs with min. (left) and max. (right) required total displacement.

4.1. DCDN Shaft Torque Requirements

Numbering the 36 unique DCDN architectures from $i = 1 \ldots 36$, the instantaneous power of the i^{th} DCDN architecture is given by Eq. (4.5) where $\tau_{\mathrm{DCDC},i}$ is the total torque acting on the common shaft.

$$\tau_{\mathrm{DCDC},i}\omega_{\mathrm{m}} = \left(D_{1,i}u_{1,i}\Delta p_{1,i} + D_2 u_2 \Delta p_{2,i} + D_{3,i}u_{3,i}\Delta p_{3,i}\right)\omega_{\mathrm{m}} \tag{4.5}$$

Indeed, when the shaft speed ω_{m} is constant, the swivel angles $u_{1,i}, u_{2,i}, u_{3,i}$ must inevitably be adjusted such that the total shaft torque of any of the 36 DCDNs is identical for any identical operating point and given by Eq. (4.6), whereas the individual displacement machine torques may differ among the different DCDNs.

$$\tau_{\mathrm{DCDC},i} = D_{1,i}u_{1,i}\Delta p_{1,i} + D_2 u_2 \Delta p_{2,i} + D_{3,i}u_{3,i}\Delta p_{3,i} \tag{4.6}$$

The instantaneous displacements may be obtained as Eq. (4.7).

$$D_{1,i}u_{1,i} = \frac{Q_{1,i}}{\omega_{\mathrm{m}}} \quad , \quad D_{2,i}u_{2,i} = \frac{Q_{2,i}}{\omega_{\mathrm{m}}} \quad , \quad D_{3,i}u_{3,i} = \frac{Q_{2,i}}{\omega_{\mathrm{m}}} \tag{4.7}$$

Evaluating Eq. (4.6) with Eqs. (4.7) for the cylinder force and velocity maps in Fig. 1.2. and Fig. 1.3, the maximum shaft torque for all 36 DCDNs is given by Eq. (4.8).

$$\tau_{\mathrm{DCDN}}|_{\max} = 139\ [\mathrm{Nm}] \tag{4.8}$$

This is coincident with the IDC Case 1 and is a consequence of not including any hydraulic energy recuperation abilities, the specified shaft speed and not considering torque losses.

5. Comparison of Required Displacements

To provide an overview of what can be gained in terms of component downsizing using a DCDN compared to IDC drives, all sizing results are depicted in Figure 5.1 for comparison.

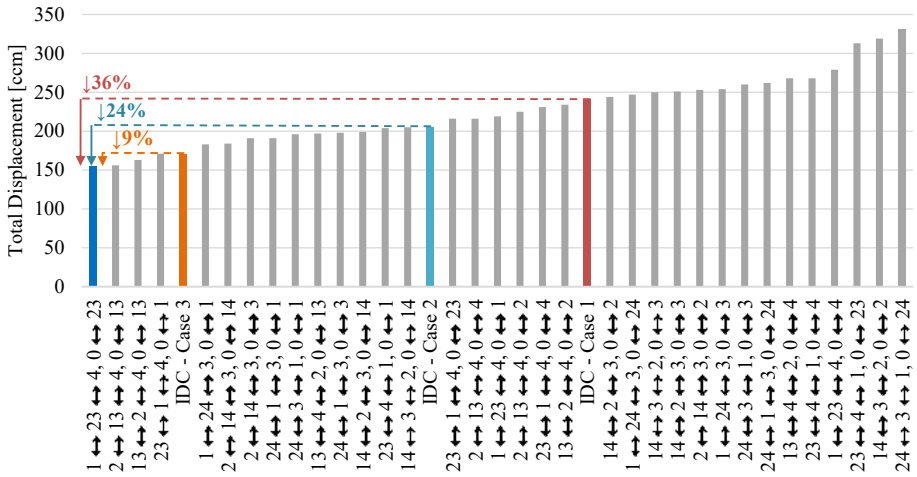


Figure 5.1. Ideal DCDN and IDC drive sizing results for the crane application.

Considering IDC Case 1, which is the case most comparable with the DCDNs as no accumulators are used, the DCDN subject to the lowest amount of displacement allows a total displacement reduction of 36%. Compared to the more conventional IDC Case 2, the same DCDN allow for a displacement reduction of 24%, and 9% when compared to the utopian IDC Case 3. The highest DCDN displacement requirement, on the other hand, by far exceed those of all IDC cases. These results emphasize the downsizing perspectives of DCDNs compared IDCs implying that cost may be significantly reduced.

6. SUMMARY & OUTLOOK

This paper considers the possible feasibility of displacement-controlled drive networks when compared to individual displacement-controlled drives applied to a dual axis crane application. The presented study takes offset in idealized cases, not considering losses such as torque and flow losses. Special attention is given to displacement requirements, and the possibility to reduce installed displacement using displacement-controlled drive networks. As benchmarks to the displacement-controlled drive networks, three cases for individual displacement-controlled drives are considered, namely being a case with a charge pump but no accumulator, a case with both a charge pump and an accumulator and a utopian case with no charge pump but with an accumulator. For the crane application considered, the use of a displacement-controlled drive network allows for a reduction in required displacement of 36% compared to the individual displacement-controlled drive without accumulator and a reduction of 24% compared to the individual displacement-controlled drive with accumulator. These results suggest that the cost of using displacement-control may be significantly reduced compared to earlier reports in literature, potentially rendering such drives commercially feasible.

In the endeavour to further develop displacement-controlled drive network technology, multiple topics should be addressed. The work presented in this paper takes offset in the static loads of the chosen crane application, however, future design developments should also consider the impact of transient loads. Especially, loads resulting from acceleration and rapid motion reversal may significantly influence the common shaft torque as well as the flow balance of a displacement-controlled drive network, and really any displacement-controlled drive system, and should be taken into consideration.

From a cost point of view, it is preferred to use as small component sizes as possible. However, transient loads may cause flow and torque/current saturation preventing a system to perform as intended, and hence peak demand conditions should be considered with great care. Such challenges may be mitigated by the system level controls via the pressure and flow coupling introduced in displacement-controlled drive networks, allowing to distribute hydraulic energy in alternative ways compared to individual controlled drive networks. Furthermore, the bandwidths of displacement control mechanisms and electric machine/inverter drive assemblies should be considered to ensure that these allow the system level controls to maintain proper behaviour of a system, both under rapid changes in motion commands and as well as under impact loads.

To this end, the use of charge pumps in displacement-controlled drive networks may also have a perspective and should be considered in future research. This approach is not considered in the work presented in this paper, where the use of three variable displacement machines with two cylinders allows the control of the lower pressure level in the system. A further advantage of the presented approach is, that excess energy of a fixed charge pump is not dissipated in the tank, hence the energy losses are anticipated to be less compared to individual displacement-controlled drives. This, however, depends on the energy consumption of the additional swash plate control mechanism introduced with the network solutions. Other losses introduced in displacement-controlled drive networks as well as individual displacement-controlled drives, are associated with idle conditions when the cylinders are inactive. These idle losses may be mitigated by adjusting the shaft speed command to the instantaneous flow command.

7. REFERENCES

[1] Estimating the Impact (Energy, Emissions and Economics) of the U.S. Fluid Power Industry, Oak Ridge National Laboratory, 2012. https://info.ornl.gov/sites/publications/files/Pub28014.pdf

[2] Willkomm, J., Wahler, M., Weber, J., Quadratic Programming to Optimize Energy Efficiency of Speed- and Displacement-Variable Pumps, Proceedings of the 8th FPNI Ph.D Symposium on Fluid Power, 2014. https://doi.org/10.1115/FPNI2014-7802

[3] Willkomm, J., Wahler, M., Weber, J., Potentials of Speed and Displacement Variable Pumps in Hydraulic Applications, 10th International Fluid Power Conference, Dresden, Germany, 2016.

[4] Jensen KJ, Ebbesen MK, Hansen MR (2021) Novel Concept for Electro-Hydrostatic Actuators for Motion Control of Hydraulic Manipulators. Energies 14:6566. https://doi.org/10.3390/en14206566

[5] Padovani D, Ketelsen S, Hagen D, Schmidt L (2019) A Self-Contained Electro-Hydraulic Cylinder with Passive Load-Holding Capability. Energies 12:292. https://doi.org/10.3390/en12020292

[6] Casoli P, Scolari F, Minav T, Rundo M (2020) Comparative Energy Analysis of a Load Sensing System and a Zonal Hydraulics for a 9-Tonne Excavator. Actuators 9:39. https://doi.org/10.3390/act9020039

[7] Schmidt, L, Ketelsen, S, Padovani, D, & Mortensen, KA (2019) Improving the Efficiency and Dynamic Properties of a Flow Control Unit in a Self-Locking Compact Electro-Hydraulic Cylinder Drive. Proceedings of the ASME/BATH Symposium on Fluid Power and Motion Control. https://doi.org/10.1115/FPMC2019-1671

[8] Ketelsen S, Michel S, Andersen TO, et al (2021) Thermo-Hydraulic Modelling and Experimental Validation of an Electro-Hydraulic Compact Drive. Energies 14:2375. https://doi.org/10.3390/en14092375

[9] Siefert J, Li PY (2022) Optimal Control of the Energy-Saving Hybrid Hydraulic-Electric Architecture (HHEA) for Off-Highway Mobile Machines. IEEE Transactions on Control Systems Technology 30:5. https://doi.org/10.1109/tcst.2021.3131435

[10] Khandekar AA, Siefert J, Li PY (2021) Co-Design of a Fully Electric Hybrid Hydraulic-Electric Architecture (FE-HHEA) for Off-Road Mobile Machines. Proceedings of the American Control Conference. https://doi.org/10.23919/acc50511.2021.9482615

[11] Schmidt L, Hansen KV (2022) Electro-Hydraulic Variable-Speed Drive Networks—Idea, Perspectives, and Energy Saving Potentials. Energies 15:1228. https://doi.org/10.3390/en15031228

[12] Schmidt, L., van Binsbergen-Galán, M., Knöll, R., Riedmann, M., Schneider, B., & Heemskerk, E. (2024). Energy Efficient Excavator Implement by Electro-Hydraulic/Mechanical Drive Network. International Journal of Fluid Power, 25(4), 413-438. https://doi.org/10.13052/ijfp1439-9776.2541

[13] Schmidt, L., & van Binsbergen-Galán, M. (2024). Electro-Hydraulic Variable-Speed Drive Network Technology: First Experimental Validation. Energies, 17(13), 1-15. https://doi.org/10.3390/en17133192

[14] van Binsbergen-Galán, M., Videbæk, B., Hansen, K. V., & Schmidt, L. (2024). Experimental Investigation of Hydraulic Power Sharing Potential in a Dual Cylinder Electro-Hydraulic Variable-Speed Drive Network. In Proceedings of the BATH/ASME 2024 Symposium on Fluid Power and Motion Control (FPMC2024) (pp. 1-9). The American Society of Mechanical Engineers (ASME). https://doi.org/10.1115/FPMC2024-140322

[15] Schmidt, L., Hansen, K. V., Videbæk, B., & van Binsbergen-Galán, M. (2024). Experimental Validation of a State Decoupling Method Applied to a Dual Cylinder Electro-Hydraulic Variable-Speed Drive Network. In Proceedings of BATH/ASME 2024 Symposium on Fluid Power and Motion Control, FPMC 2024 (pp. 1-10). The American Society of Mechanical Engineers (ASME). https://doi.org/10.1115/FPMC2024-140149

[16] Fassbender D, Zakharov V, Minav T (2021) Utilization of electric primemovers in hydraulic heavy-duty-mobile-machine implement systems. Automation in Construction 132. https://doi.org/10.1016/j.autcon.2021.103964

[17] Fassbender D, Brach C, Minav T (2023) Experimental Investigations of Partially Valve-, Partially Displacement-Controlled Electrified Telehandler Implements. Actuators 12:50. https://doi.org/10.3390/act12020050

[18] Rahmfeld, R., & Ivantysynova, M. (1998). Energy Saving Hydraulic Actuators for Mobile Machines. Proceedings of 1st Bratislavian Fluid Power Sysmposium, (pp. 47-57). Častá – Píla, Slovakia

[19] Busquets, E., & Ivantysynova, M. (2013). Temperature Prediction of Displacement Controlled Multi-Actuator Machines. International Journal of Fluid Power, 25 - 36.

[20] Williamson, C., Zimmerman, J., & Ivantysynova, M. (2008). Efficiency Study of an Excavator Hydraulic System Based on Displacement-Controlled Actuators. Proceedings of the Bath ASME Symposium on Fluid Power and Motion Control. Bath, UK.

[21] Busquets, E., (2016) Advanced control algorithms for compact and highly efficient displacement-controlled multi-actuator and hydraulic hybrid systems. Purdue University: PhD Thesis. https://docs.lib.purdue.edu/dissertations/AAI10170559/

[22] Zimmerman, J.D., (2012) Toward optimal multi-actuator displacement controlled mobile hydraulic systems. Purdue University: PhD Thesis. https://www.proquest.com/docview/1238001343?sourcetype=Dissertations%20&%20Theses

[23] Bosch Rexroth AG, (2023) Axial piston variable pump A4VG series 32, RE-E 92003/2023-10-30.

APPENDIX

The nomenclature used in this paper is outlined in Table 3 below.

Table 3: Nomenclature used.

Symbol	Description
A_1, A_2, A_3, A_4	Cylinder piston areas
D_1, D_2, D_3	Max. displacement of displacement machine
D_c	Displacement of charge pump
$F_{\mathrm{L}1}, F_{\mathrm{L}2}$	Piston pressure forces
m_L	Crane payload
p_1, p_2, p_3, p_4	Cylinder piston pressures
p_c	Charge pressure
$\Delta p_1, \Delta p_2, \Delta p_3$	Displacement machine pressure differences
Q_1, Q_2, Q_3	Displacement machine flows
Q_c	Charge pump flow
u_1, u_2, u_3	Normalized swivel angles
$\dot{x}_L, \dot{y}_L$	Payload speed components
$\dot{x}_1, \dot{x}_2$	Cylinder piston speeds
ω_m	Shaft speed
$\tau_{\mathrm{IDC}}, \tau_{\mathrm{DCDC}}$	Common shaft torques

Biographies



Lasse Schmidt received the M.Sc. degree in engineering (mechatronics) from Aalborg University, Denmark, in 2008. From 2008 he was with the application engineering group of Bosch Rexroth A/S, Denmark, and from 2010 an industrial Ph.D. fellow also associated with Aalborg University. He received the Ph.D. degree in robust control of hydraulic cylinder drives in 2014. Subsequently, he has been a postdoctoral researcher at AAU Energy while concurrently being with Bosch Rexroth AG. Hereafter he became an Assistant Professor with AAU Energy. He is currently an Associate Professor with AAU Energy and heading research activities related to electro-hydraulic drive network technology, a field in which he is the founder of the fundamental design and control principles. He is the main author or co-author of nearly 70 scientific peer-reviewed publications, most of them on topics related hydraulic drives and systems control. Lasses current research interests are in design and control of electro-hydraulic drive networks and their integration into both mobile working machines and industrial systems.



Mikkel van Binsbergen-Galán received the B.Sc. and M.Sc. degrees, both from Aalborg University, Denmark, in 2020 and 2022 respectively, where he is currently working towards the Ph.D. degree at AAU Energy. His research interests are related to design and control of electro-hydraulic drives, actuators and systems, with special focus on electro-hydraulic variable-speed drive technology.

Development of a flow-optimised slip-in servo cartridge valve for high-performance applications

Kristof Schlemmer[1], Jörg Wagner[1], Felix Lang[2]

[1]AXXERON Hydrolux S.à r.l., k.schlemmer@axxeron-hydrolux.com
[2]SMS group GmbH, felix.lang@sms-group.com

Abstract.

Hydraulic drive technology remains essential in applications demanding high force, power density, and precise dynamic control. Systems such as die-casting machines require extremely high flow rates to manage large cylinder volumes, rapid motion sequences, and fast pressure build-up and release. This paper presents the development methodology of a new, flow-optimized generation of hydraulic slip-in servo cartridge valves tailored to such high-performance requirements. The process includes the derivation of performance criteria from demanding use cases, the identification of key influencing parameters, and the formulation of a suitable design approach. State-of-the-art simulation methods and extensive experimental investigations are employed to evaluate design variants and optimize flow efficiency, dynamic behaviour, hysteresis, repeatability, and positioning accuracy, while enabling digital control and sensor integration. The result is benchmarked against relevant performance indicators. The study demonstrates how combining high dynamics with high flow capability provides a technical advantage in modern hydraulic systems and supports industrial efforts to reduce carbon footprints. A case study illustrates system integration and quantifies the resulting performance benefits.

Keywords. Hydraulic servo cartridge valve; slip-in valve; CFD; FEA; high-flow hydraulics; dynamic response; open-die forging press.

1. INTRODUCTION

Hydraulic drive technology remains essential in industrial applications that require high force density, compact installation space and reliable control behaviour. Many systems operate with large cylinder volumes and high operating pressures while demanding fast motion profiles and accurate pressure regulation. As a result, substantial hydraulic flow must be controlled with precision, particularly during rapid pressurisation and equally fast release or decompression phases. Applications such as die casting machines, open die forging presses and large injection moulding units exemplify these combined requirements and place correspondingly high demands on the components used.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Slip-in servo cartridge valves provide a suitable basis for such environments because they combine high flow capacity with the controllability of servo-actuated designs. Their architecture supports efficient flow guidance, low pressure losses and stable closed loop behaviour across a wide range of dynamic operating conditions. By virtue of their fundamental characteristics, valves of this class thus provide specific advantages in high performance hydraulic systems.

The present work discusses the development of a new flow optimised generation of such valves, named X700 [1]. The approach follows a structured methodology that integrates application driven requirement analysis, numerical design and simulation, and iterative experimental refinement. By combining mechanical, hydraulic and electronic design aspects within a coordinated development process, a new valve series has been created that offers significantly improved flow efficiency and dynamic control behaviour.

The resulting design is evaluated using representative performance metrics and is examined in a system context through a case study of an open die forging press. This example illustrates how the characteristics of the new valve generation influence manifold architecture, system dynamics and energy efficiency, and how the systematic development approach supports tangible improvements in demanding hydraulic applications.

2. DEVELOPMENT METHODOLOGY

The development of high-performance hydraulic servo cartridge valves at Axxeron Hydrolux follows a structured and repeatable methodology that combines virtual analysis and physical testing within iterative loops. This approach ensures that functional, structural and dynamic requirements are met with a high degree of reliability before a new valve design is released for field use.

2.1. Requirements Analysis and Definition of Performance Targets

The process begins with a systematic requirement analysis for the relevant classes of applications. Flow demands are derived from cylinder dimensions, volumetric stroke requirements and acceptable speed profiles, whereas upper pressure limits are determined from structural load cases of the hydraulic system. Requirements concerning dynamic response, hysteresis and control precision arise from the motion control tasks of the target machine. Energy related aspects, such as permissible pilot stage leakage and acceptable pressure losses, further influence the performance targets. The result of this phase is a consolidated and quantified specification that governs all subsequent development stages.

2.2. Virtual Development Cycle

The first development loop takes place in the virtual domain. Three-dimensional CAD models of sleeve, poppet, cover and pilot interface form the foundation for numerical analyses. Strength evaluation is performed using finite element methods to assess static and cyclic loading up to the intended nominal operating pressure. This ensures that critical elements such as flow windows, seat transitions and connecting threads provide sufficient safety margins. The methodology includes both static load analyses and pressure cycling evaluations, with results systematically reviewed against design rules.

Flow behaviour is evaluated using computational fluid dynamics. Throughout the entire stroke range, simulations provide detailed information on pressure distributions, velocity

fields, potential cavitation and flow induced hydraulic forces. The numerical development workflow, illustrated in Figure 2.1, supports the comparison of alternative geometries under identical boundary conditions and assists in identifying variants that combine low pressure losses with favourable dynamic characteristics. By iterating CAD design, FEA and CFD within the virtual loop, only the most promising concepts are transferred into physical prototypes.

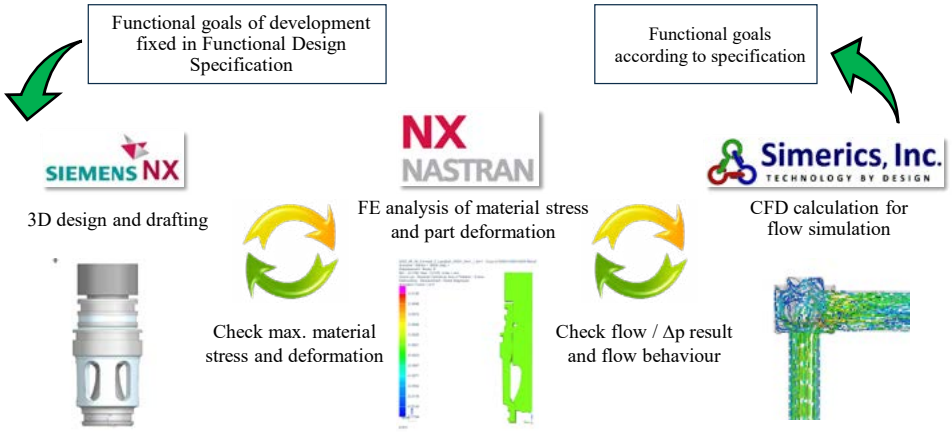


Figure 2.1. Iterative numerical development in continuous refinement loop.

2.3. Experimental Development Cycle

After virtual optimisation, selected variants are manufactured as prototypes and subjected to a standardised functional testing programme. This includes the measurement of flow characteristics in both flow directions at controlled pressure drops, typically ranging from a few bars to double digit values, depending on the intended application scenario. Dynamic evaluation involves step response measurements at pilot pressures up to 210 bar and frequency response tests across a representative bandwidth. Threshold, hysteresis, and drift behaviours are measured at different temperatures to assess control accuracy under varying environmental conditions. The complete experimental workflow is depicted in Figure 2.2.

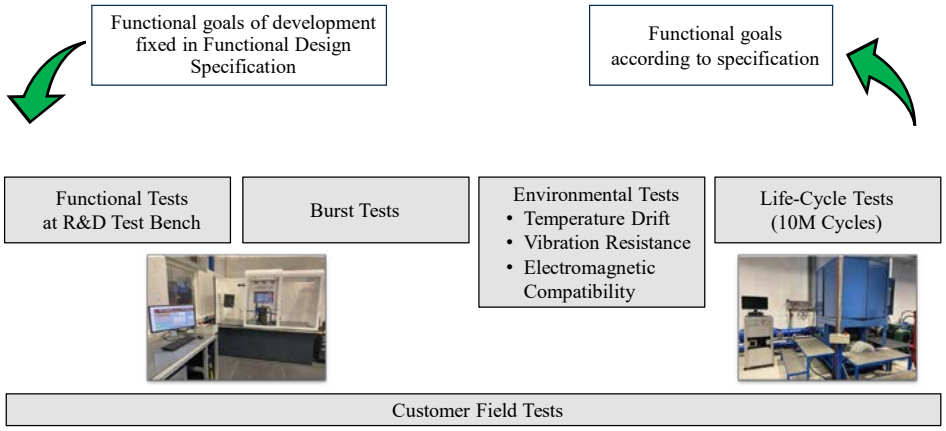


Figure 2.2. Iterative experimental development involving various types of testing.

These experiments validate the simulation results and provide insight into frictional behaviour, guiding tolerances and poppet position stability. Any deviation between measured and predicted behaviour triggers an additional virtual development loop, ensuring alignment between numerical predictions and hardware performance.

2.4. Static and Dynamic Validation Testing

Once a prototype demonstrates stable operation under functional conditions, it enters the structural verification phase. This includes pressure cycling tests with amplitudes up to the intended nominal operating pressure and endurance tests amounting to several million switching cycles. These tests assess the long-term stability of the main stage, potential wear mechanisms of guiding surfaces and drift behaviour of measurement and pilot components. Static burst pressure tests to several times the nominal operating pressure serve as an additional safety measure. Together, these assessments demonstrate that the valve design withstands the mechanical loads expected throughout its life cycle.

2.5. Environmental Qualification Testing

Environmental qualification ensures that the valve performs reliably under demanding industrial conditions. Vibration tests with multi axis excitation up to 30 g and shock tests up to 50 g evaluate mechanical robustness. Thermal cycling and temperature dependent measurements assess the stability of materials, electronics and hydraulic components across the intended ambient conditions. Electromagnetic compatibility testing confirms the robustness of the integrated electronics with respect to both emitted interference and immunity to external disturbances.

2.6. Consolidation and Release for Field Testing

Only when virtual, functional, structural and environmental evaluations all converge does a design enter field testing. Pre series valves are installed in selected customer machines, enabling long term observation under real operating conditions. These tests confirm dynamic behaviour, leakage stability and robustness against contamination and installation tolerances. The feedback from this stage closes the development loop and supports the final release for series production.

3. APPLICATION OF THE DEVELOPMENT METHODOLOGY TO THE NEW VALVE SERIES

The methodology described in Section 2 was applied to the development of the X700 series, a new generation of two-way slip-in servo cartridge valves intended for applications with exceptionally high flow demand and stringent dynamic requirements. While the individual development steps follow the generic process, several aspects of the X700 project required particular attention due to the unusually high flow-to-size ratio and the intended operating pressure of 420 bar across the entire flow range from NG32 to NG100.

3.1. Derivation of Application Specific Requirements

The requirement analysis was carried out with a focus on die casting machines and metal forming presses, as these applications represent some of the most demanding use cases for servo cartridge valves. A comparative assessment of design drivers, shown schematically in Figure 3.1, highlighted flow capacity, operating pressure and dynamic response as primary

influencing parameters for both application classes. Press systems additionally emphasised failsafe behaviour and the need for compact manifolds, while die casting machines placed stronger emphasis on control precision during fast injection and decompression phases. Based on these boundary conditions, the X700 development targets included significantly higher flow performance compared to earlier cartridge generations, reduced pressure losses through the main stage, and improved dynamic behaviour through the use of direct drive pilot valves and integrated digital electronics.

Criteria	Die Casting	Presses
Flow capacity	++	++
Operating pressure	++	++
Dynamics	++	+
Control	++	+
Size and space	+	++
Failsafe functionality	○	++
Digital and IoT features	+	+
Robustness	++	+



Figure 3.1. Application-driven requirement matrix for die-casting and press applications.

3.2. Virtual Development and Design Evolution

The virtual development stage started from a baseline design as used in conventional servo cartridge valves. Early variants followed established geometries with drilled radial bores and cylindrical poppets, which provided a useful reference but proved insufficient to achieve the targeted flow levels. Subsequent design iterations explored alternative sleeve window shapes and modified poppet contours. The systematic comparison of these variants through FEM and CFD simulations led to the identification of a geometry that allowed a substantial increase in effective flow cross section while maintaining acceptable stress levels. This evolution is illustrated in Figure 3.2, which shows how the transition from drilled bores to shaped flow windows enabled larger port diameters and lower local pressure losses.

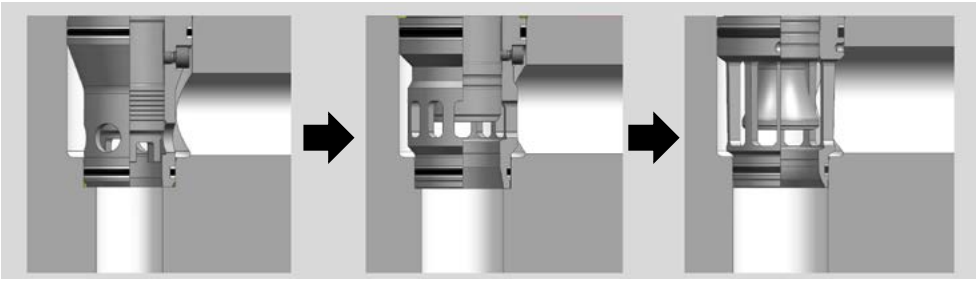


Figure 3.2. Evolution of main stage sleeve and poppet geometry.

Structural analysis confirmed that the resulting geometry provided sufficient safety margins under both static loading at 420 bar and dynamic pressure cycling. CFD simulations demonstrated improved flow characteristics and a favourable force balance over the stroke range, which was essential for ensuring stable poppet motion under high flow rates. Figure

3.3 and Figure 3.4 summarise typical FEA and CFD results processed during the design evaluation.

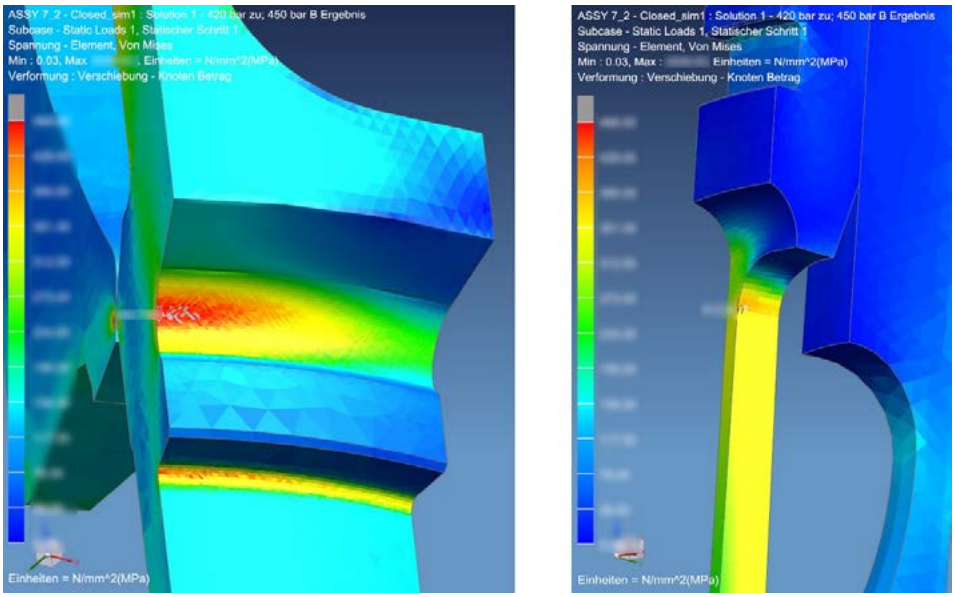


Figure 3.3. FEA-based stress distribution used to verify structural integrity.

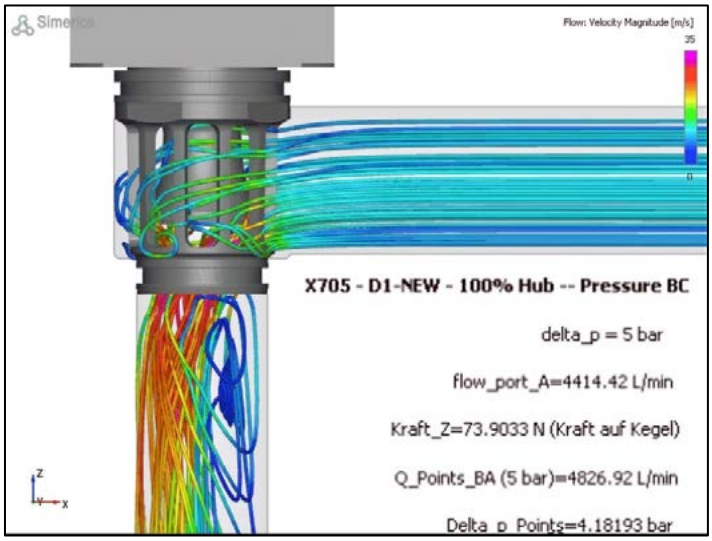


Figure 3.4. CFD-based evaluation of velocity field and flow induced forces.

3.3. Experimental Verification and Iterative Refinement

The experimental development phase follows the methodology outlined in Section 2 and typically reveals aspects of the valve behaviour that cannot be predicted with full precision using numerical tools alone. This is an inherent characteristic of hydraulic systems operating at very high flow rates, where small geometric deviations or local pressure effects may

influence the dynamic response. In the case of the X700 development, such an iterative refinement step occurred during functional testing of early prototypes. While the overall flow behaviour and pressure loss characteristics closely matched the CFD predictions, one prototype variant showed a lightly oscillatory poppet motion in a narrow operating range at high flow demand. This observation was not treated as an anomaly but as a natural part of the iterative process, in which simulation and physical testing complement each other.

A combined analysis of measurement data and targeted CFD studies revealed that, in this particular geometry, the hydraulic force balance at certain intermediate stroke positions was sensitive to small variations in the surrounding pressure field. As shown by result in Figure 3.5, the pressure acting on the poppet surfaces could shift slightly under these conditions, producing alternating net forces that caused a lightly damped dynamic response. Since such behaviours are difficult to capture fully in simplified simulation models, this finding demonstrated the value of the iterative test and refine loop that is integral to the applied development methodology. Several modified poppet geometries were subsequently analysed and manufactured. Among these, one variant achieved a consistently stable force balance across the full stroke range while maintaining the targeted flow performance. The refinement confirmed that virtual models and physical testing converge most effectively when such iterative corrections are incorporated into the development process as planned and expected steps.

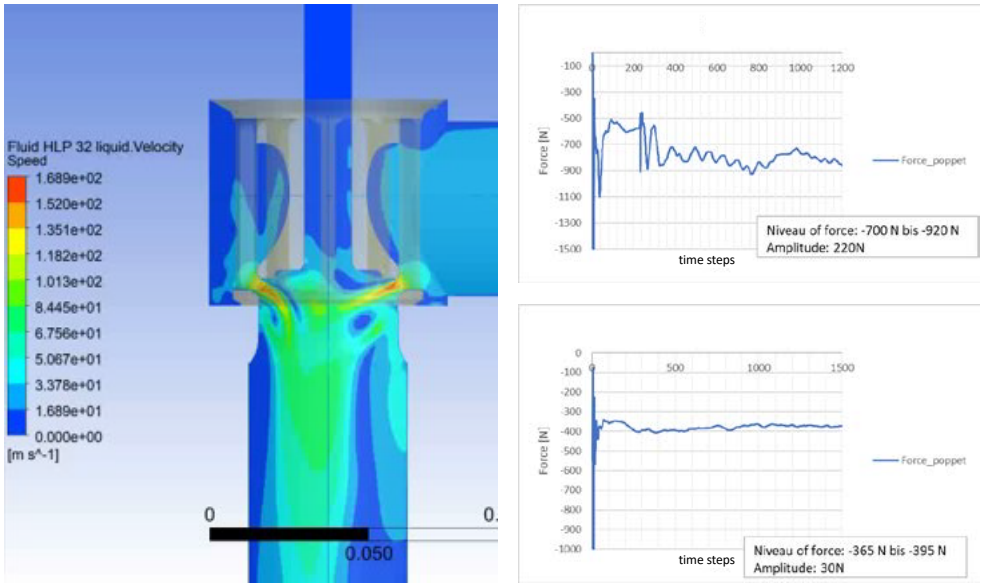


Figure 3.5. Flow-induced force imbalance as identified during prototype testing.

3.4. Qualification and Validation for Series Release

After resolving all functional and dynamic issues, the X700 prototypes were subjected to the standard qualification programme described in Section 2. This included extended endurance testing with up to ten million switching cycles at elevated pressure levels, verifying fatigue behaviour, wear and long-term stability of the main stage mechanics, the pilot valve and the displacement transducer. Structural safety was confirmed through burst pressure tests

conducted up to a multiple of the rated pressure. Environmental testing demonstrated robustness against vibration, shock loads and thermal cycling, while EMC qualification ensured the immunity of the integrated electronics against electromagnetic disturbances. These results validated the valve for field installation. Pre series units of the X700 were subsequently tested in selected industrial machines to verify performance in operational conditions. The field tests confirmed the high flow capacity, excellent dynamic behaviour and low leakage levels predicted in simulation and demonstrated in laboratory tests.

4. DEVELOPMENT RESULTS

The X700 series represents the consolidated outcome of the virtual and experimental development loops described in Sections 2 and 3. The resulting valve family covers ISO sizes 32 to 100 and is designed for an operating pressure of 420 bar across all variants. Its performance characteristics reflect the accumulated effects of the flow optimised main stage geometry, the direct drive pilot concept and the iterative refinement of dynamic behaviour and stability margins. Figure 4.1 provides an overview of the final valve design and its structural elements.



Figure 4.1. Final design of the X700 servo cartridge valve series.

Across the entire size range, the valves exhibit a consistently high flow efficiency. The shaped sleeve windows and coordinated poppet contour minimise pressure losses in both flow directions, so that the valves reach required flow levels at comparatively low command signals. This behaviour enhances the achievable dynamic response, because the cylinder velocity rises more quickly after a change in the command value. The measured dynamic flow characteristics and excitation responses shown in Figure 4.2 illustrate this relationship for representative size and cavity configurations. The curves confirm that the valves translate command signals into valve opening and subsequently into stable and repeatable flow behaviour across the full operating range.

The dynamic response of the valves is supported by the direct drive pilot stage, which provides rapid actuation without the pilot oil consumption associated with traditional two-

stage concepts. The integrated electronics ensure a stable closed-loop relation between command value and poppet position, even under varying pilot pressures and temperatures. This contributes to the low hysteresis and high repeatability observed in testing, which remain within the targets established during the requirement analysis. The low internal leakage of the pilot stage enhances energy efficiency, particularly in cyclic operation where leakage-related losses accumulate significantly over time.

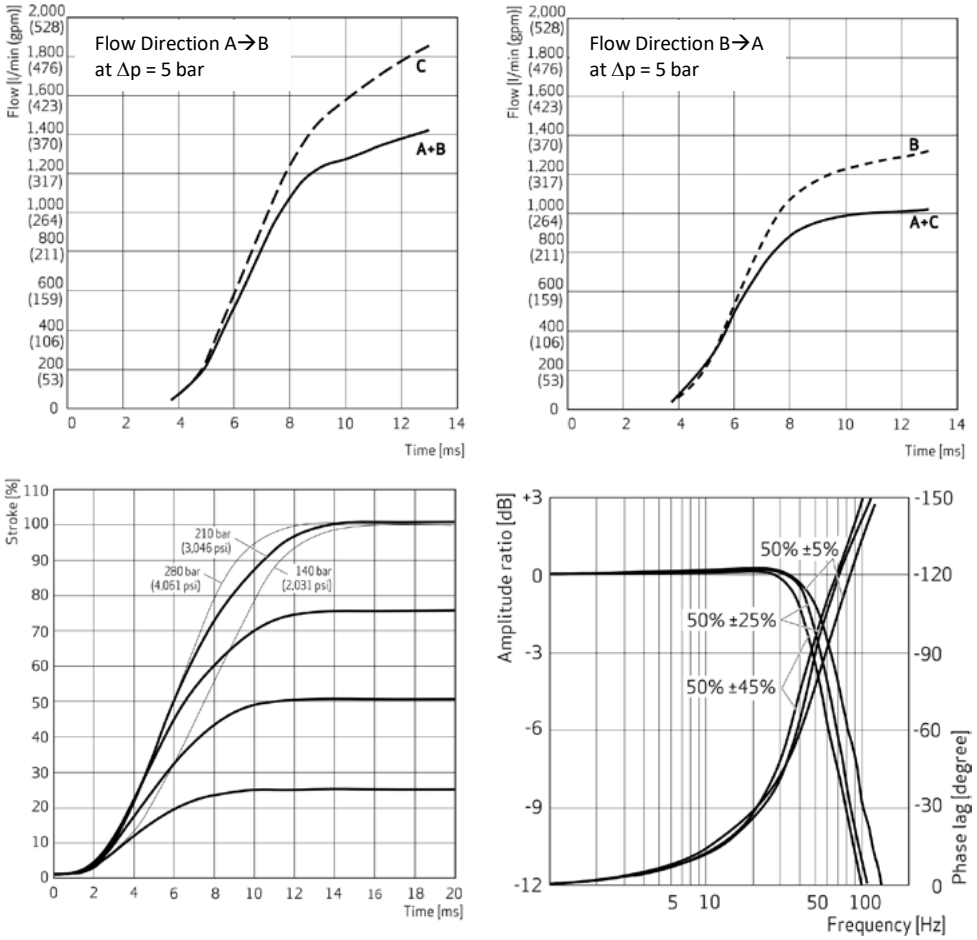


Figure 4.2. Characteristic dynamic flow response, step response, and frequency response for size 32. [2]

When assessing the achieved performance relative to predecessor products and competing valve families, the X700 series demonstrates a clear step forward. For a given valve size, the flow-optimised geometry provides noticeably higher flow capacity than previous slip-in servo cartridge designs, and the direct-drive pilot concept delivers a more favourable combination of dynamic response and control precision. In this regard, the essential improvement lies in the enhanced capability to quickly, but in a stable and controlled fashion ramp up volumetric flow, i.e. to provide a superior flow rate increase. This is shown exemplarily for one size in the comparison in Figure 4.3. Notwithstanding, the available

flow performance exceeds that of comparable servo cartridges in many size classes by a substantial margin, enabling downsizing of one or more valve sizes in typical installations. As an example, Figure 4.4 exhibits the increased dynamic response of cylinder velocity in an existing die-casting machine, where the previously installed size-50 valve was outperformed by the newly developed size-40 valve. This improvement is not only a result of geometric changes but a direct consequence of the systematic application of the development methodology, in which simulation, testing and iterative refinement were closely aligned.

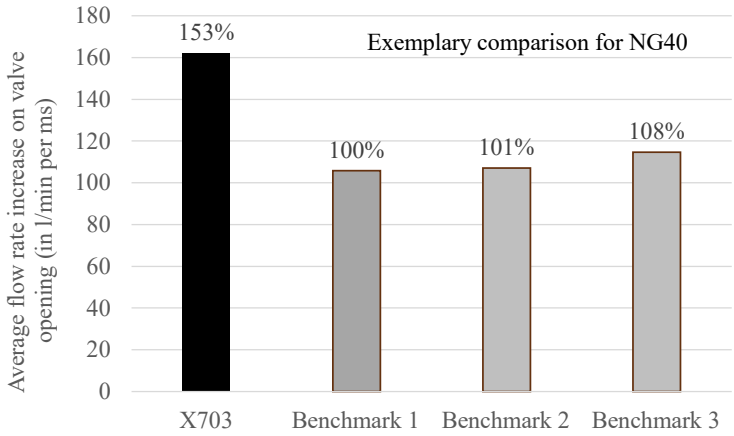


Figure 4.3. Flow performance indicator for exemplary Size 40.

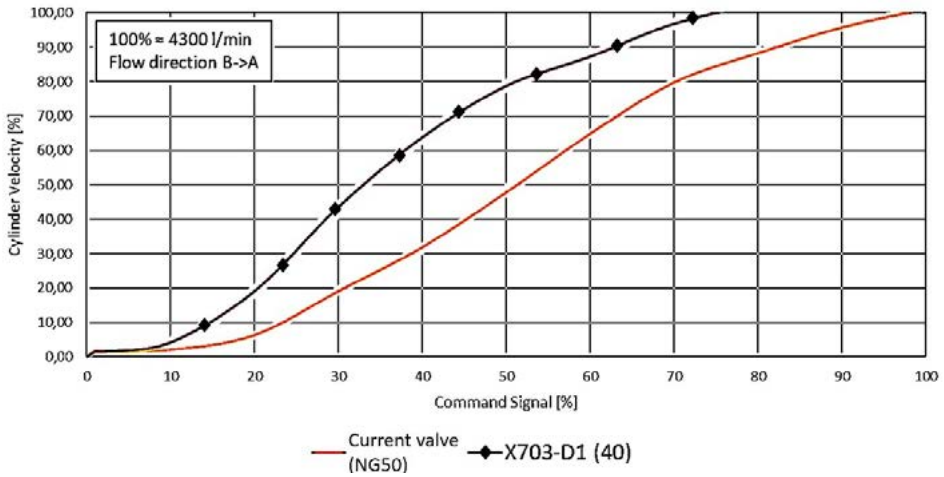


Figure 4.4. Increased dynamic performance at downsizing in a die-casting machine.

The X700 series therefore establishes a new level of performance within its class. The valves combine high flow capacity with stable dynamic behaviour and robust closed loop control. Their characteristics support the implementation of compact manifold layouts with reduced pressure losses and allow machine builders to achieve faster cycle times and higher productivity. The series is well suited for applications in which both high volumetric flow and precise motion control are required, as demonstrated in the case study of Section 5.

5. CASE STUDY: INTEGRATION IN AN OPEN-DIE FORGING PRESS

The improvements achieved in the development of the X700 series become particularly evident when the valves are integrated into a demanding hydraulic system. A representative example is an open die forging press in the 50–55 MN class, which combines large cylinder volumes, high operating pressures and stringent requirements for cycle time and repeatability [3]. Figure 5.1 depicts an exemplary press implementation. The hydraulic system must provide fast approach motions, accurately controlled forging strokes and rapid decompression, while executing a high number of cycles over the machine's lifetime. Figure 5.2 illustrates the typical functional layout and working cycle of such a press and the role of servo cartridge valves in the inlet and outlet flow paths (all pictures courtesy of SMS group GmbH, Mönchengladbach).



Figure 5.1. Exemplary open-die forging press.

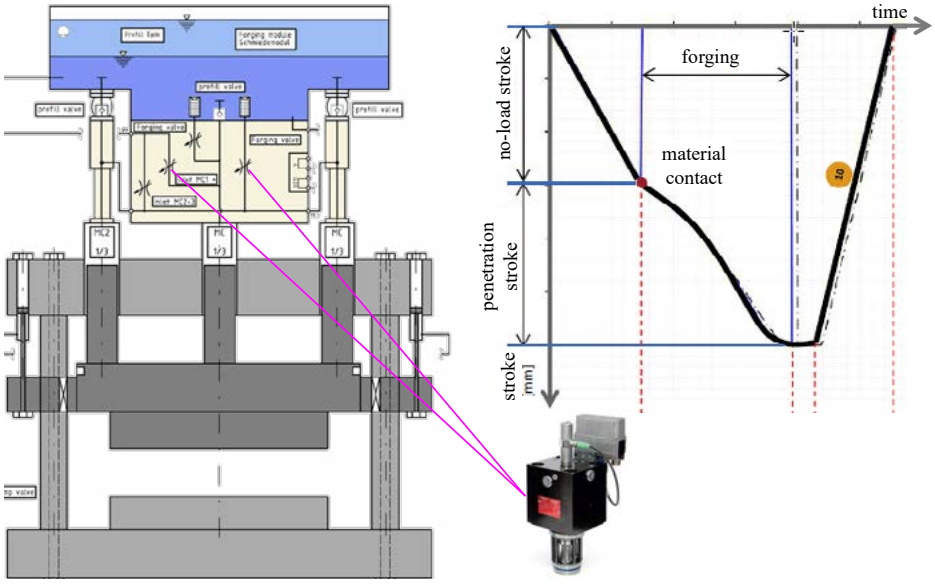


Figure 5.2. Simplified layout and working cycle of the open-die forging press.

5.1. Integration of the Valves into the Forging Manifold

In the hydraulic manifold originally used for this press class, several large servo cartridges are installed in parallel to achieve the required pressurisation and decompression flows. The configuration includes multiple valves of nominal size 125, complemented by size 80 units, and results in a structurally robust but comparatively heavy and voluminous manifold. With the introduction of the X700 series, the same functional capacity can be achieved using smaller nominal sizes, because each valve provides higher flow capability relative to its ISO size. This enables a simplified configuration based exclusively on size 100 and size 80 valves and leads to a reduction in manifold mass of about 35 %. A direct comparison of the two configurations is given in Table 5.1. Figure 5.3 visualises the achieved result in the revised, compact manifold.

Table 5.1: Valve configurations before and after introducing the new valve series.

Configuration	Inlet Valves	Outlet Valves	Manifold Mass
Before	1 × size 125 + 1 × size 80	3 × size 125	~16 200 kg
After (X700)	2 × size 100	2 × size 100 + 2 × size 80	~12 000 kg

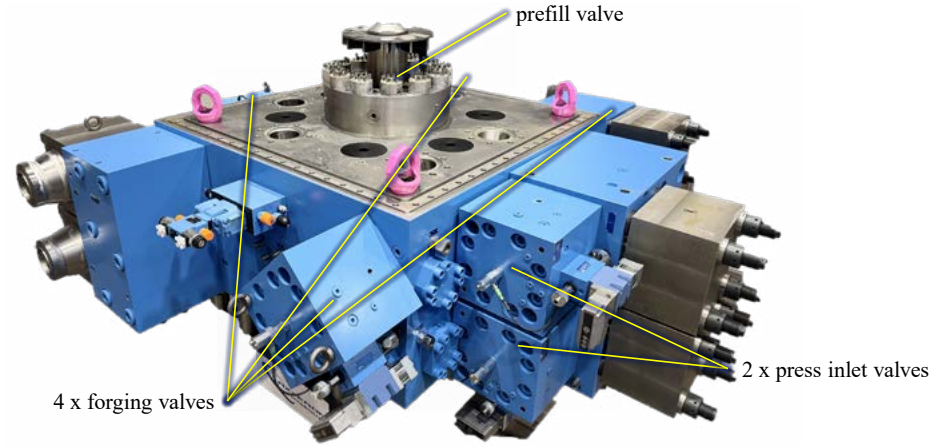


Figure 5.3. Improved manifold configuration and layout using X700 valves.

5.2. Effect on System Dynamics and Cycle Performance

The influence of the new valve configuration becomes clear when examining typical forging cycles. The flow optimised main stage of the X700 series enables the valves to reach high volumetric flows at comparatively low command signals. For example, size 100 valves of the series provide rated flows clearly above 10 000 l/min at a pressure drop of 5 bar, with recommended maximum flows significantly higher when using optimised cavity designs. This behaviour shortens the time required for pressure build up and exhaust. During the decompression phase, the updated manifold can release the hydraulic fluid from the main cylinder more rapidly than the previous design, despite using smaller nominated valve sizes. The corresponding reduction in non-productive time within each cycle contributes to a higher throughput of the press. At the same time, the stable closed loop characteristics of the X700 series support precise control of the forging stroke, so that the higher dynamics do not come at the expense of positioning accuracy or repeatability.

5.3. Energy Related Effects and System Efficiency

Beyond the dynamic aspects, the introduction of the new valve series also influences the energy demand of the forging system. The reduced pressure losses across the main stage, achieved through the flow optimised geometry, lower the hydraulic power needed for a given motion profile. In addition, the direct drive pilot valves used in the series exhibit significantly reduced internal leakage compared to traditional pilot operated concepts. For size ranges up to size 100, the pilot leakage remains below approximately 1.4 l/min at 100 bar, which is substantially lower than the values associated with earlier two-stage solutions.



Figure 5.4. Integration of the forging manifold into the machine structure.

In a forging environment with thousands of cycles per shift, such differences accumulate measurably. Lower leakage reduces the continuous energy losses in the system, while improved flow efficiency decreases the peak power required from the hydraulic power unit. The downsized manifold, with its 35 % lower mass and reduced volume, and thus improved integration into the overall machine structure as shown in Figure 5.4, further supports efficient operation by reducing the total amount of oil in circulation and associated thermal loads. Together, these effects help operators reduce operating costs and support current efforts to lower the carbon footprint of industrial processes.

5.4. Summary of Case Study Findings

The open die forging press illustrates how the X700 series translates component level improvements into system level benefits. Increased flow capability per valve enables a more compact manifold with fewer very large cartridges and a significantly lower structural mass. Dynamic response and control precision are improved, leading to shorter non-productive phases and higher throughput. The reduced pressure losses and low pilot leakage contribute to lower energy demand. Together, these findings show how the systematic development methodology supports the design of more efficient and competitive hydraulic systems.

6. CONCLUSIONS AND OUTLOOK

The development of the X700 series shows that a systematic combination of numerical simulation, experimental refinement and structured qualification enables tangible advances

in hydraulic servo cartridge technology. Building on this foundation, several extensions of the concept have already emerged that broaden its applicability.

One development path concerns variants with moderately reduced performance levels. These valves follow the same design principles as the X700 series but are adapted for applications with lower dynamic requirements or reduced operating pressures. By simplifying selected elements of the main stage and electronics, these variants offer a cost-efficient option for less demanding industrial environments while retaining the characteristic stability and controllability of the base design.

At the upper end of the performance range, versions designed for higher operating pressures have also been realised. These valves extend the pressure capability from the standard 420 bar to 450 bar and are intended for applications that require increased load capacity or additional structural safety margins. Their design and qualification follow the same methodological steps as the standard series, ensuring consistent behaviour across the extended operating range.

In parallel, work continues on new control concepts derived from the same architecture. A digitally pressure compensated flow control valve is currently being developed, combining the mechanical robustness of the slip in cartridge format with electronically controlled flow characteristics. This concept enables more flexible integration into digital machine control systems and supports advanced functions such as remote parameterisation, integrated diagnostic information and coordinated motion control.

Together, these developments demonstrate that the underlying methodology not only supports high performance solutions such as the X700 series but also provides a scalable basis for future hydraulic product families.

7. REFERENCES

[1] Wagner, J., Wiegandt, M. and Jost, F.: (2023): Weniger ist mehr – Servocartridges auf einem neuen Leistungsniveau, in: O+P Fluidtechnik 2023/04, pp. 30-33.

[2] Axxeron Hydrolux S.à r.l.: X700 Series 2-Way Servo Cartridge Valves – Product Catalogue, Rev. E, March 2025.

[3] K. Schlemmer, F. Lang: Leistung trifft Präzision – Servo-Einbauventiltechnik für High-Performance-Anwendungen, IFAS Industrie-Kolloquium, 16 May 2025.

Biographies



Kristof Schlemmer received his Dipl.-Ing. degree in Mechanical Engineering from RWTH Aachen University in 2004, and his Dr.-Ing. degree in the field of fluid power from the same institution in 2012. He currently works as Engineering Manager at AXXERON Hydrolux S.à r.l. in Bettembourg, Luxembourg, and acts as a guest lecturer at IFAS of RWTH Aachen. His areas of activity are focussed on advanced products, systems and applications in industrial hydraulics.



Jörg Wagner received the degree in Mechanical Engineering from Balthasar Neumann Technikum Trier in 2008. He currently works as Team Leader Engineering Products at AXXERON Hydrolux S.à r.l. in Bettembourg, Luxembourg. He is responsible for Product Development focused on cartridge valves and cartridge valve applications.



Felix Lang received his Dipl.-Ing. degree in Mechanical Engineering from RWTH Aachen University in 2010 and subsequently joined SMS group GmbH in Mönchengladbach as Project Engineer in the field of forging technology. Since 2018, he has been leading a team of specialised engineers as Head of Hydraulics Extrusion & Forging Presses.

Modelling and Simulation of Damping Orifices for Pressure Control Valves （2nd Report)

Seiei Masuda*, Yasushi Matsunaga*, Fumio Shimizu, Kazuhiro Tanaka****

*IHI Corporation, Control Systems Engineering Department, 3-2-1 Daikanyama, Akishima City, Tokyo, Japan**

*Kyushu Institute of Technology, Graduate School of Computer Science and Systems Engineering, 680-4 Kawazu, Iizuka-shi, Fukuoka, JAPAN***

masuda4361@ihi-g.com, matsunaga4536@ihi-g.com, shimizu@phys.kyutech.ac.jp, tanakafa@fka.att.ne.jp

Abstract.

Because alternating flow passes through the damping orifice in a pressure control valve, the size and location of the contraction region are not well understood. However, it is known that the flow coefficient of the orifice is significantly affected by the contraction coefficient of the jet. If the flow coefficient of the damping orifice fluctuates depending on the frequency of the alternating flow, a 40% change in the flow coefficient poses a risk of a 6 dB increase or decrease in the gain margin of the pressure control system. This research was motivated by this risk, predicted from a 1D model of the pressure control valve. There is a high need to reduce the risk to stability of the pressure control system inside an aircraft jet engine fuel control device, whose operating range varies significantly depending on flight conditions. The authors previously presented the above risk in reference [1]. When considering the dynamic characteristics of an orifice, they approximate it as an LRC circuit. In this case, L in the LRC circuit is corrected for the fluid inertia of the short choke tube, and the R element can be approximated as the orifice opening area, and the C element can be approximated as the fluid compressibility upstream of the orifice. Comparison with CFD frequency response demonstrated its validity. The remaining issues include the effect of the damping orifice delay on the differential pressure control valve, understanding the effect of the damping orifice amplitude, deriving the correction value for the *L* element, and confirming the flow pattern that the alternating flow has on the orifice. This paper reports on the progress of the results of these investigations and also mentions the remaining issues.

Keywords. Pressure Control Valve, Damping Orifice, Alternative Flow, Flow Pattern, Flow Coefficient, Bond graph, LRC Circuit, Frequency Response Diagram, CFD

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Most jet engine fuel metering units today combine a fuel metering valve, which can control the metering port opening area to a desired value through feedback control using a digital controller, with a differential pressure control valve, a type of pressure control valve, which maintains a constant differential pressure across the fuel metering valve. Because the flow velocity of the fuel metering valve is proportional to the square root of the differential pressure and regulated to constant, the fuel metering flow is proportional to the metering port opening area.[11][12] As noted in [1], the differential pressure control valves, which is a type of pressure control valve used in the fuel control units of aircraft jet engines must achieve both stability and responsiveness over a wide operating range. Among these valves, the dynamic characteristics of the differential pressure control valve, which maintains a constant differential pressure across the fuel metering valve, are particularly important because they directly affect the stability of the fuel flow rate itself. On the other hand, pressure disturbances are directly applied to the differential pressure across the metering valve; however, the control of the metering valve opening area is performed by feedback control. By designing the bandwidth of the fuel metering valve to be low, the system becomes robust against high-frequency disturbances, and the effects of pressure disturbances can be more easily compensated by the feedback control. The configuration of a differential pressure control valve is shown in Figure 1.

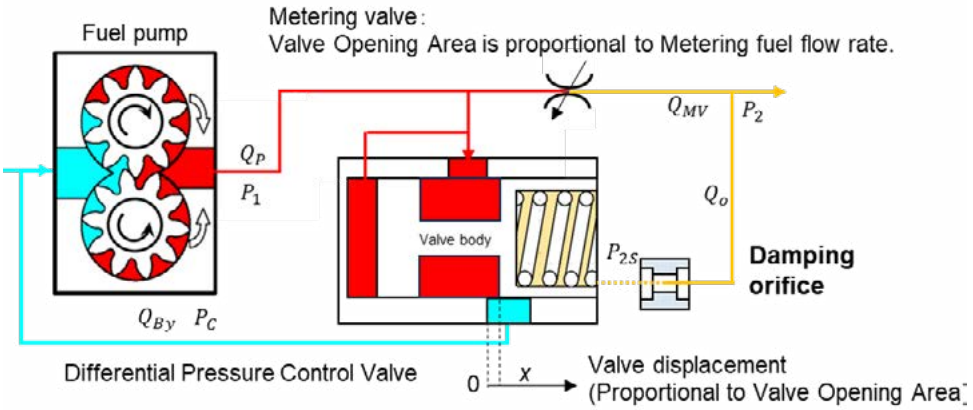


Figure 1 Schematic Diagram of a Pressure Control Valve [1]

Typically, the differential pressure sensing valve operates by multiplying the differential pressure between the control pressure and the reference pressure by the valve's pressure-receiving area to generate a fluid force. This force balances against a preloaded spring force to determine the valve's position. Additionally, a reaction force from the valve's damping resistance is applied, imparting dynamic characteristics to the spring-mass system. Differential pressure control valves must respond sufficiently quickly to changes in fuel flow and maintain adequate stability margin across the entire operating range. This prevents unexpected fuel flow overshoot. To achieve this, the damping ratio of the valve body's spring-mass system is increased beyond the limit damping ratio using a damping orifice, reducing the rate of change of phase margin with frequency. The proportional gain is set so that the crossover frequency satisfies the lower limit of the responsiveness requirement, while ensuring sufficient stability margin is always maintained under the valve spool geometry and operating conditions.

In the pressure control valve's spring-mass system, a damping orifice is placed in series with the valve spool and is used to provide damping force ideally proportional to valve velocity. Typically, the damping force exhibits a characteristic proportional to the square of the flow rate through the orifice, resulting in zero damping force at valve standstill and significant nonlinearity. [2] [6] [7] In the damping orifice, the flow pattern through the orifice, such as the direction of flow rotation, is considered to exhibit symmetry in both positive and negative directions. Furthermore, the effect of changes in flow pattern on the orifice's Pena-Contractor section is expected to vary significantly with frequency.

However, there are few studies on the frequency response, flow pattern, contraction coefficient, and flow coefficient of the damping orifice in hydraulic pressure control valves. Oshima et al. [2] and Miyashita [3] reported experimental results on the frequency response of damping orifices, but did not mention flow patterns. Furthermore, research on pulsating flow by Nakano et al. [4] has been conducted, sharing many commonalities from a modelling perspective. However, in alternating flow, the average flow velocity is zero. Since the damping force has polarity and is proportional to the square of the flow velocity, the damping force becomes zero in the valve's steady state and varies significantly with velocity amplitude. If the damping force were zero, the valve's spring-mass system would not be damped and would continue oscillating at a constant amplitude. When the valve vibrates, the damping force acts, reducing the amplitude. A method to estimate the valve amplitude in this state has not yet been established.

On the other hand, Iimura et al. [10] studied the equivalent electrical circuit model and transfer function of the Pitot tube orifice in air pressure control applications. The Pitot tube is a type of air speed indicator and a pressure measurement device, so its input is pressure and its output is pressure. While similarities exist in the shape of the transfer function and other aspects, differences between aircraft and liquid systems, as well as the disparity between pressure and flow rate as inputs, necessitate an evaluation of this similarity.

2. PURPOSE

This study reports on further investigations into the results obtained by modelling the frequency characteristics of the damping orifice in a pressure control valve using unsteady CFD analysis and a Bond graph model, aiming to clarify the dynamic characteristics of the damping orifice. Specifically, it reports on the effect of damping orifice delay on differential pressure control valves, understanding the influence of damping orifice amplitude, deriving correction values for L-elements, and confirming flow patterns affecting orifices in alternating flow.

3. 1D MODELLING AND SIMULATION

3.1. Equations of the Pressure Control valve and the Damping orifice

We derive the linearized transfer function of the differential pressure control valve receiving flow from the flow source. The derivation of the differential pressure control valve transfer function below is based on Reference [5][6][7][8]. Furthermore, regarding the dynamic characteristics of the damping orifice, we add the differential pressure caused by fluid inertia occurring in the short pipe section of the orifice restriction, as described in Reference [1]. The objective is to verify changes in crossover frequency and phase margin in response to variations in the damping force or damping ratio within the differential pressure control

valve. However, the purpose of this paper is to outline the impact of changes in the damping orifice on the entire differential pressure control system. NOMENCLATURE lists the symbols used in deriving the transfer function.

Equation (3.1) shows the equilibrium of forces acting on the valve body of a differential pressure control valve. For simplification, the fluid viscous resistance and friction force of the spool valve are approximated as zero, and the viscous resistance is only the damping force from the damping orifice. Similarly, the fluid force is also assumed to be negligible. These simplifications do not compromise the generality for high-precision spool valves. The motion equation of the valve body is balanced by the fluid dynamic force due to the differential pressure acting on the valve's pressure-receiving area and the spring force. Here, x denotes the valve displacement of the differential pressure control valve, M is the mass of the spool valve, C is the viscous resistance coefficient, k is the spring constant of the pressure control valve, A_R is the receiving area of the spool, P_1 is the pressure upstream of the pressure control valve, and P_2 is the pressure downstream of the metering valve.

$$M \cdot \ddot{x} + C \cdot \dot{x} + k \cdot x = A_R \cdot (P_1 - P_2) \tag{3.1}$$

The damping force generated by the damping orifice of the differential pressure control valve is expressed by equation (3.2). Here, P_{2s} is the pressure in the spring chamber obtained from P_2 after passing through the damping orifice.

$$F_o = A_R \cdot (P_{2s} - P_2) \tag{3.2}$$

The damping orifice is modelled below. The differential pressure across the damping orifice is defined by equation (3.3). Here, P_O is the differential pressure of the damping orifice, and PL is the pressure due to fluid inertia as it passes through the short pipe section of the damping orifice. In addition to the pressure loss due to the orifice restriction, the pressure changes originating from the inertial force generated in the short pipe section of the restriction is also formulated. The pressure downstream of the metering valve is governed by the dynamic characteristics of the jet engine and changes sufficiently slowly, so its time variation can be considered negligible, resulting in equation (3.4).

$$P_{2s} - P_O - PL = P_2 \tag{3.3}$$

$$\frac{d}{dt} P_2 \cong 0 \tag{3.4}$$

The relationship between flow rate and pressure loss for the damping orifice is shown in Equation (3.5). Here, ρ represents the fluid density, Q_o represents the flow rate through the orifice, A_o represents the orifice opening area, and C_d represents Coefficient of discharge flow r*ate* of the orifice.

$$P_O = \frac{\rho}{2} \left(\frac{Q_O}{C_d \cdot A_o} \right)^2 \tag{3.5}$$

Fluid inertia is shown in Equations (3.6) and (3.7). This paper considers the pressure due to the inertial effect occurring in the short pipe of the orifice.[6] [7] Here, L represents the length of the pipe of the damping orifice.

$$PL = IL \frac{d}{dt} Q_o \tag{3.6}$$

$$IL = \frac{\rho \cdot L}{A_o} \tag{3.7}$$

Equation (3.8) represents the flow rate applied to the damping orifice from the differential pressure control valve.

$$Q_{In} = A_R \cdot \dot{x} \tag{3.8}$$

Substituting the equations for each element into equation (3.3) yields equation (3.9). Linearizing equation (3.9) results in equations (3.10) and (3.11).

$$P_{2s} = \frac{\rho}{2}\left(\frac{Q_O}{C_d \cdot A_O}\right)^2 + IL \cdot \frac{d}{dt}Q_O \tag{3.9}$$

$$\Delta P_{2s} = R \cdot \Delta Q_O + IL \cdot \frac{d}{dt}\Delta Q_O \tag{3.10}$$

$$= \rho \frac{(\Delta P_O)_0}{(Q_O)_0} \tag{3.11}$$

The upstream pressure at the dumping orifice is calculated by integrating the difference between the inflow rate and the flow rate through the orifice using Equation (3.8), as given by Equation (3.12). [6] [7] B represents Bulk modulus of fluid. V_s represents the volume upstream of Spring room. Here, Equation (3.13) represents the spring effect of the hydraulic pressure.

$$P_{2S} = \int \frac{B}{V_s}(A_R \cdot \dot{x} - Q_O)dt + P_{2S0} \tag{3.12}$$

$$C_I = \frac{B}{V_s} \tag{3.13}$$

Taking the Laplace transform of equation (3.12) and substituting equation (3.8) yields equation (3.14). Here, Δ indicates a prefix of variable for micro-time perturbation and C_I represents the C element of damping orifice Volume. And s represents Laplace operator. Q_{In} represents the input flow rate of damping orifice.

$$\Delta P_{2S} = \frac{1}{s} C_I (Q_{In} - Q_O) \tag{3.14}$$

Substituting equation (3.9) into equation (3.14) gives the relationship between Q_{In} and Q_o as equation (3.15).

$$\frac{\rho}{2}\left(\frac{Q_O}{C_d \cdot A_O}\right)^2 + IL \cdot \frac{d}{dt}Q_O = \frac{1}{s} \cdot C_I \cdot (Q_{In} - Q_O) \tag{3.15}$$

The transfer function from linearized Q_{In} to Q_o is derived from equations (3.16) to (3.17). Here, R represents the R element of orifice restriction.

$$R \cdot \Delta Q_O + IL \cdot \frac{d}{dt}\Delta Q_O = \frac{1}{s} \cdot C_I \cdot (\Delta Q_{In} - \Delta Q_O) \tag{3.16}$$

$$Q_O = \frac{C_I}{IL \cdot s^2 + R \cdot s + \cdot C_I} Q_{In} \tag{3.17}$$

$$\Delta P_{2s} = \frac{C_I \cdot (R + IL \cdot s)}{IL \cdot s^2 + R \cdot s + \cdot C_I} Q_{In} \tag{3.18}$$

Furthermore, from equations (3.1) and (3.8), the damping coefficient C is obtained as shown in equation (3.19). Since the characteristics of the damping orifice vary significantly with amplitude, applying a sinusoidal frequency sweep input to the nonlinear state equations and performing a Fourier transform on the input and output yields a transfer function nearly equivalent to a linear transfer function. That is, equation (3.20) shows the second-order nonlinear state equation. By applying a frequency-swept input to Q_{In} while varying the amplitude, the frequency response curve for each amplitude can be obtained.

$$C \cdot \dot{x} = A_R \cdot \Delta P_{2s} = \frac{{A_R}^2 \cdot C_I \cdot (R + IL \cdot s)}{IL \cdot s^2 + R \cdot s + \cdot C_I} \dot{x} \tag{3.19}$$

$$\begin{bmatrix} \dot{P_{2s}} \\ Q_O \end{bmatrix} = \begin{bmatrix} C_I(Q_{In} - Q_O) \\ \frac{1}{IL}\left(\Delta P_{2s} - \frac{\rho}{2}\left(\frac{Q_O}{C_d \cdot A_o}\right)^2\right) \end{bmatrix} \tag{3.20}$$

In the motion equation of the valve body for a pressure control valve, damping resistance is dominated by the damping orifice. For precision valves, friction and fluid viscous resistance can be considered negligible. The control pressure for a differential pressure control valve varies due to the compressibility of the fluid and the bypass flow rate of the differential pressure control valve. Equation (3.21) relates the upstream pressure of a differential pressure control valve to the inflow rate into its volume. Here V represents the volume upstream of metering valve. Q_P, Q_{By} and Q_{MV} represent the input flow rate of damping orifice, the metering fuel flow rate and the flow source discharge flow rate.

$$P_1 = \int \frac{B}{V} \left(Q_P - Q_{By} - Q_{MV}\right) dt + P_{10}) \tag{3.21}$$

Equation (3.22) relates the orifice area of a pressure control valve to its flow rate based on Bernoulli's law. [6][7] Here A_{By} represents Pressure control valve opening area. k_V represents the pressure control valve opening area per displacement.

$$Q_{By} = C_d \cdot A_{By} \sqrt{\frac{2 \cdot (P_1 - P_c)}{\rho}} = C_d \cdot k_V \cdot x \sqrt{\frac{2 \cdot (P_1 - P_c)}{\rho}} \tag{3.22}$$

Equation (3.23) shows the linearized transfer function between the upstream pressure of a differential pressure control valve and its bypass valve orifice area.

$$\Delta P_1 = \frac{-2 \cdot (P_1 - P_c)_0}{Q_P} \cdot \left(\frac{Q_{By}}{A_{By}}\right)_0 \cdot \frac{1}{1 + \frac{V}{B} \cdot \frac{2 \cdot (P_1 - P_c)_0}{Q_P} s} \cdot k_V \cdot \Delta x \tag{3.23}$$

Figure 3.1 shows a block diagram of the overall linearized transfer function of the differential pressure control valve. In the open-loop transfer function depicted in this block diagram, the crossover frequency and the phase delay at that frequency can be used as evaluation metrics for responsiveness and stability. By satisfying each required value, it is possible to determine whether responsiveness and stability can be achieved simultaneously.

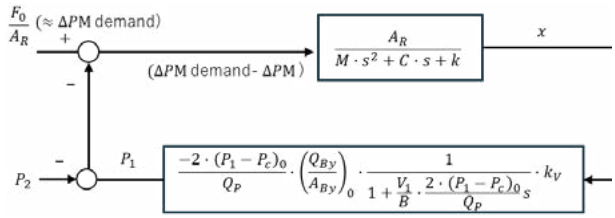


Figure 3.1 Block Diagram of Differential Pressure Control Valve

3.2. Frequency Response Results for Damping Orifice

Frequency response simulations were performed using a sinusoidal frequency sweep input to vary the amplitude in the state equation for the damped orifice (Equation 3.20). Frequency Response Simulation Conditions is shown Table 3.1. Figure 3.2 shows the time series of the waveform of a sinusoidal sweep of the orifice input flow rate. The input waveform is the same, but the flow rate amplitude is different. Figure 3.3 shows the time series response of the normalized orifice differential pressure. Figure 3.4 also shows the normalized response when the internal inertance of the orifice is changed. Each frequency response diagram is shown in Figures 3.5 and 3.6. When the input amplitude and the response is small, as shown in Figure 3.3 a) or b), the amplitude increases with frequency relative to the input, exhibiting a peak frequency. However, as the input amplitude increases, as seen in Figure 3.3, the peak amplitude tends to decrease. The frequency response diagrams with varying amplitude in Figure 3.5 show a first order/second-order response at small amplitudes. As the amplitude increases, the open-loop gain of the orifice's LCR circuit decreases, phase margin increases, and the response approaches a first-order lag shape. For the short pipe section of the orifice, it was confirmed that the frequency changes with pipe length, like the linear analysis.[1]
In addition, Figure 3.6 shows the change in the Bode diagram due to changes in inertance. The results are almost the same as those obtained by Fourier transform analysis. [1]

able 3.1 Frequency Response Simulation Conditions

Simulation Condition	***Dimension***	***Parameter***
Flow Amplitude Range	*mL/min*	a):6.3, b): 31.5, c):63.0, d):314.5, e):630.0
Inertance	*kg/m⁴*	1):1.739E+6, 2):4.288 E+6
Sine Sweep Range	*rad/sec*	From 0.628 to 3000·6.28
Frequency Sweep Time	*s*	0.5
Time Step	*s*	0.5E-06
Integration Method	-	Eulerian Method
ΔP₂ₛ	*kPa*	Differential Pressure of orifice

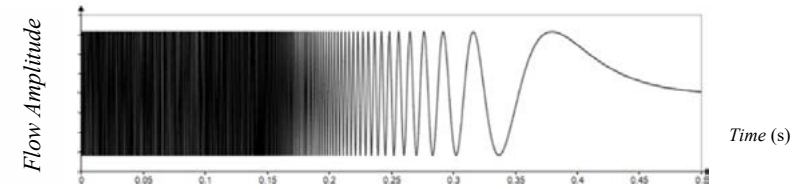


Figure 3.2 Q_{In} sine wave sweep input (from a) to e) have same waveform)

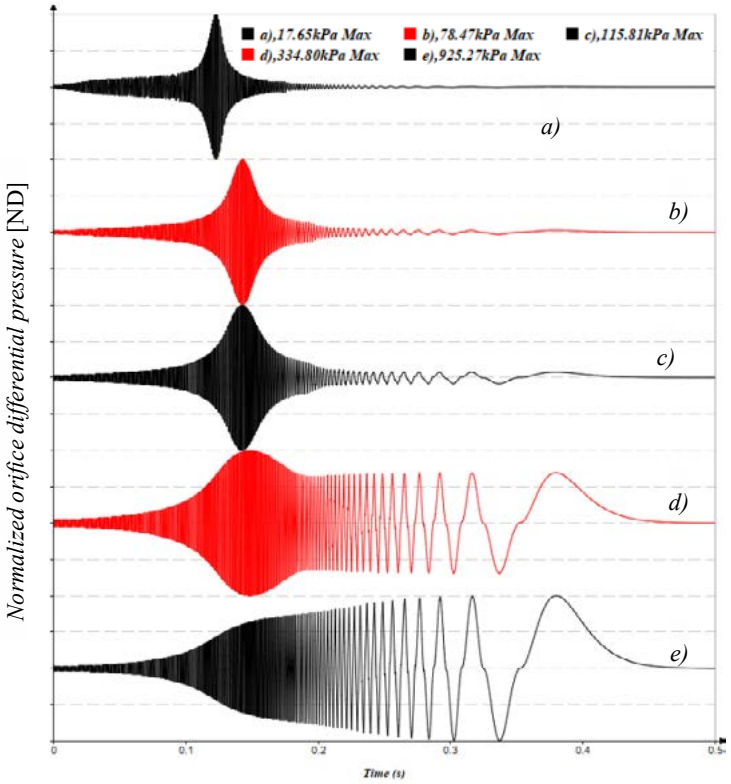


Figure 3.3 Response to Q_{In} input （Normalized ΔP_{2S} Response, Input amplitude change）

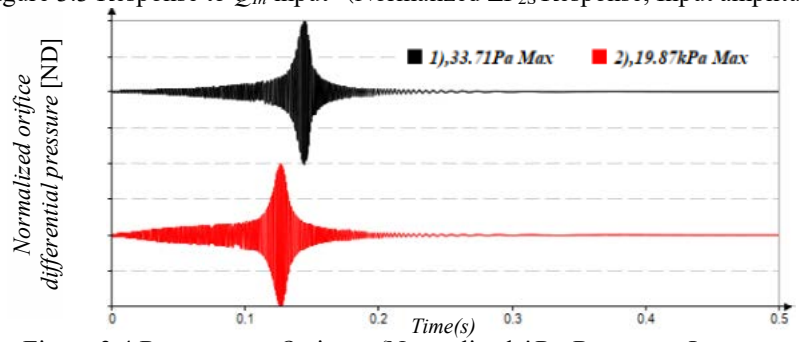


Figure 3.4 Response to Q_{In} input (Normalized ΔP_{2S} Response, Inertance change)

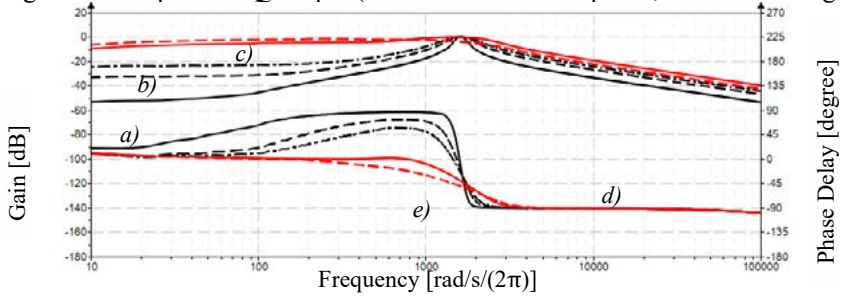


Figure 3.5 Normalized Frequency Response Diagram of ΔP_{2S}/ Q_{In} (Input amplitude change)

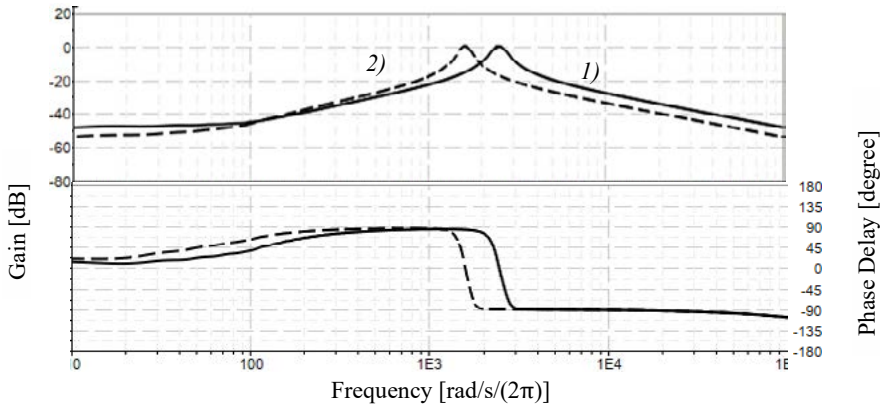
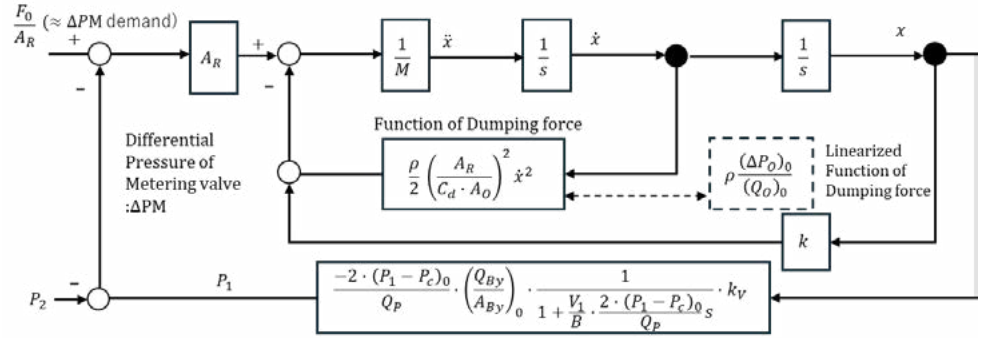


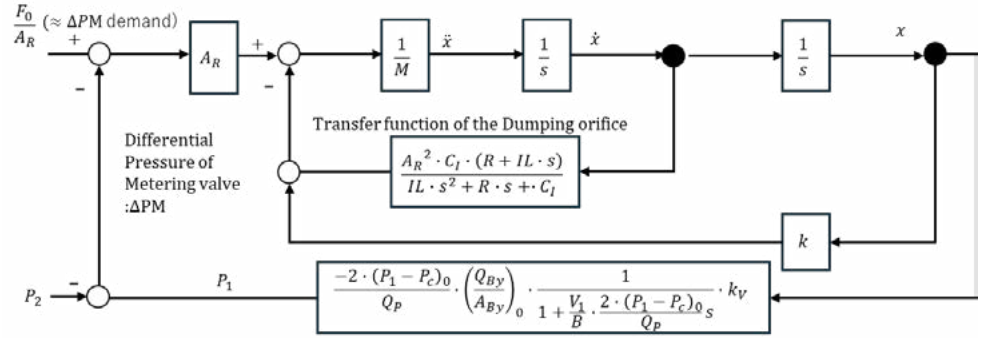
Figure 3.6 Normalized Frequency Response Diagram of $\Delta P_{2S}/ Q_{In}$ (Input Inertance change)

3.3. Effect of Damping Orifice Transfer Function on the Response of a Differential Pressure Control System

We investigated whether the transfer function of Equation (3.18) affects the frequency response diagram of a differential pressure control system. The transfer function of the damping coefficient in Equation (3.18) was substituted for the damping coefficient in Equation (3.1). Figure 3.7 shows the block diagram before and after the substitution.



(a) Differential Pressure Control Valve Block Diagram (No Orifice Delay)



(b) Differential Pressure Control Valve Block Diagram (With Orifice Delay)

Figure 3.7 block diagrams of the Pressure control valve

The transfer function of the pressure control valve with and without orifice delay was calculated by varying the valve amplitude under the conditions listed in Table 3.2. Figure

3.8 shows the Bode diagram. Table 3.2 lists the pressure control valve parameters. The orifice delay has a resonance point at approximately 10000 rad/s (1.6 kHz), which was previously thought to have a small impact due to a frequency band sufficiently different from the bandwidth of the pressure control valve. However, due to factors such as the phase lead element, it can affect the stability and responsiveness of the pressure control valve.

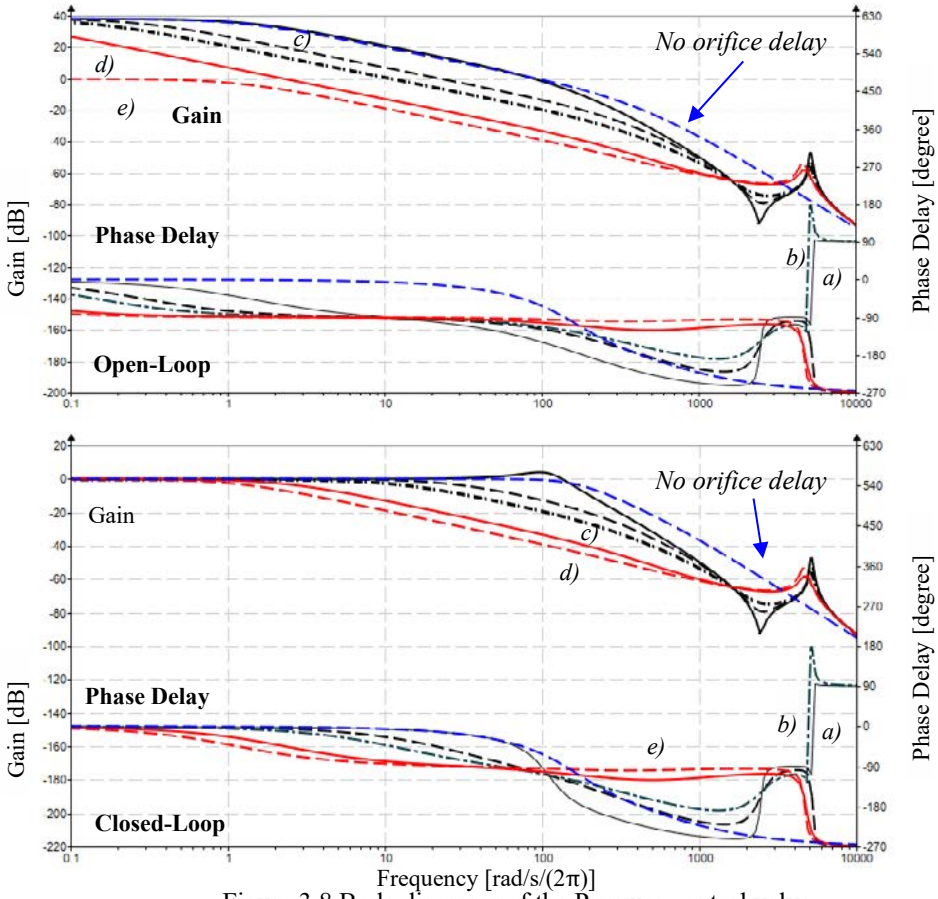


Figure 3.8 Bode diagrams of the Pressure control valve

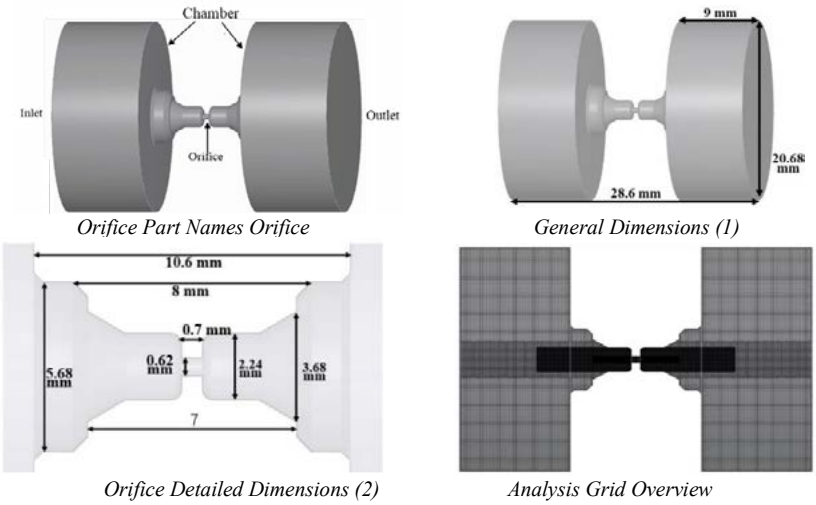
Table 3.2 Parameter of Pressure Control Valve and simulation conditions

Simulation Condition	***Dimension***	***Parameter***
Inertance of orifice	*kg/m⁴*	1.739E+6,
R of orifice	*kg/(s·m4)*	a):1.728e+9, b):8.639E+9, c):1.728E+10, d): 8.639E+10, e): 1.728e+11,
Inertance of orifice	*kg/m⁴*	1.739E+6, 2)
C of orifice	*kg/(s²·m4)*	4.067E+14
Resonant Frequency of The Valve	*rad/sec*	176.14
Mass of Valve	*kg*	0.05
Resistance of Valve	*N/mm/s*	0.2
Open-loop Gain	*ND*	80

It was found that greater stability is achieved when the proportional element or orifice flow rate is large, and resistance: R is high. Because the purpose of this paper is to examine the dynamic characteristics of the orifice, the operating conditions are not yet clear. It was found that even in systems with fast orifice dynamics, the pressure control valve is affected by a frequency band that differs by more than one order of magnitude. A more detailed study will be conducted separately after determining the specifications of the pressure control valve.

4. CFD ANALYSIS

The CFD model used in this study is shown in Figure 4.1. In examining the flow patterns, a simpler orifice model was also considered. However, the unique flow characteristics of the orifice's alternating flow were also reproduced in the model described in [1], so that model is omitted for simplicity's sake. The flow patterns were examined at the timing of changes in flow rate. The 3D model and simulation conditions are shown in Table 4.1, and the simulation model is shown in Figure 4.1. Figure 4.2 shows the velocity diagram near the orifice's short-view and the velocity distribution at the short-view cross section for each simulation. Figure 4.3 also shows the pressure distribution near the orifice. These analyses revealed no contraction within the orifice's short-pipe passage. Furthermore, during the alternating flow transition, a cylindrical flow remained at the orifice exit without changing flow direction, and no fluid flow entered from the orifice's axis when the flow reversed. Therefore, fluid is thought to be entering from the radial direction. Furthermore, the orifice contraction occurs at the orifice entrance and immediately shifts to the opposite side when the flow alternates. Table 4.2 shows the flow coefficient for each orifice frequency at flow amplitude rate:6.3mL/min. Although there is almost no contraction in the flow pattern, the flow coefficient is decreasing. It is thought that the flow rate at the throttle changes because of compressibility and inertia, which is why the flow coefficient changed. However, it is thought that it will be necessary to separate the effects of compressibility and inertia in the future.



Orifice Part Names Orifice *General Dimensions (1)*

Orifice Detailed Dimensions (2) *Analysis Grid Overview*

Figure 4.1 3D model of CFD Analyses [1]

Table 4.1 Condition of CFD Analyses [1]

Item	Analysis conditions
Software used	Simerics-MP+® [9]
Analysis type	Transient
Turbulence model	Standard k-ε
Fluid density	750 kg/m
Fluid viscosity	9.5×10^{-7} m^2/s
Number of grid points	2,488,000
Outlet boundary condition	4.2 MPa
Inlet boundary condition	$Q_{In} = A_R \cdot X \cdot \sin(2\pi \cdot \omega \cdot t)$

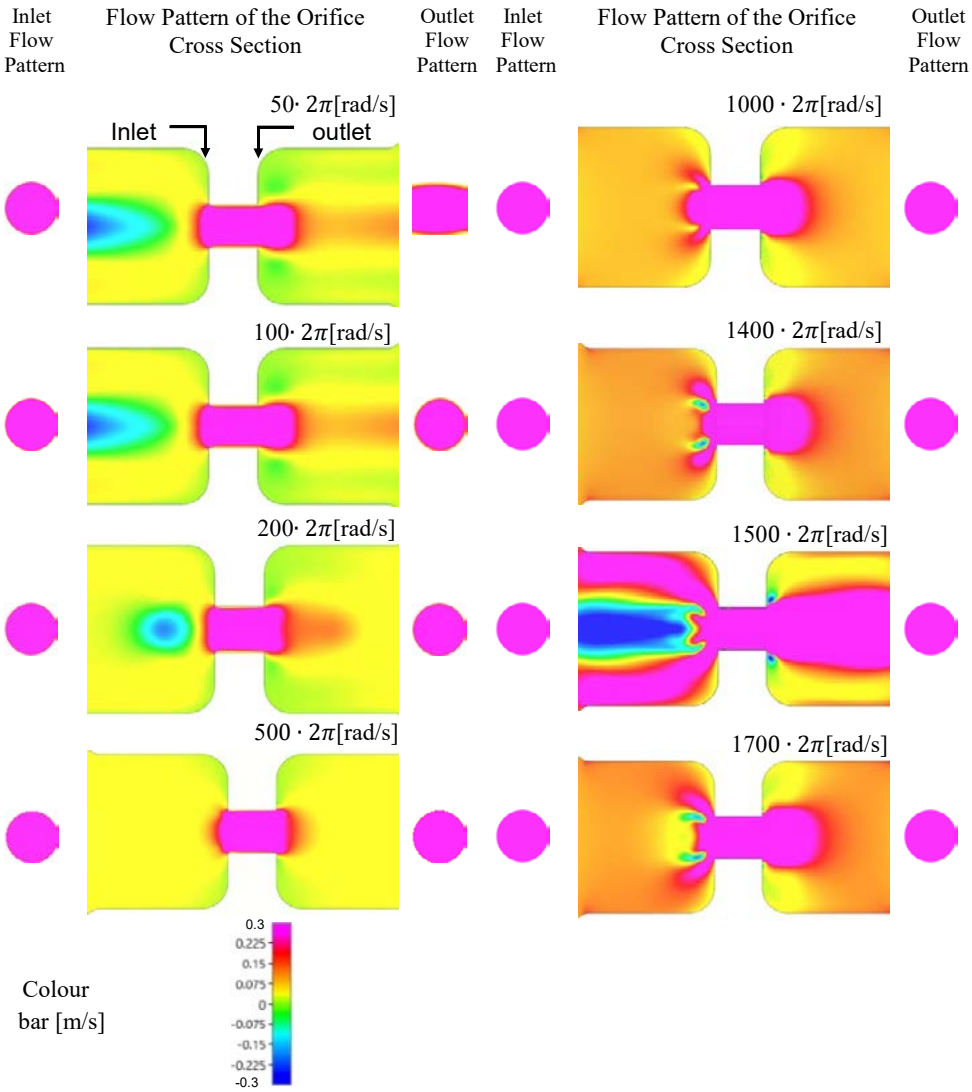
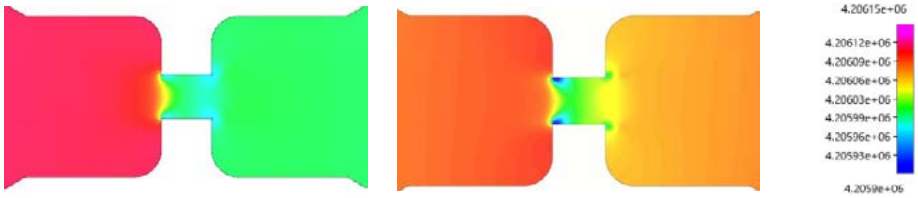


Figure 4.2 Flow patterns of flow velocity near the orifice at each frequency



Pressure Distribution at Maximum Flow Rate (500·2π rad/s) Pressure Distribution at Maximum Flow Rate (1600·2π rad/s) [Pa]

Figure.4.3 Orifice Pressure Distribution

Table 4.2 flow coefficient of orifices at Flow Amplitude Range:6.3mL/min

Frequency $[rad/s]$	*The flow coefficient*	*Frequency* $[rad/s]$	*The flow coefficient*
50· 2π	0.5	1000· 2π	0.3
100· 2π	0.4	1500· 2π	0.1
200· 2π	0.3	1600· 2π	0.3
500· 2π	0.2	-	-

5. DISCUSSION

5.1. Considerations on the Shift in Orifice Resonance Frequency

In our previous paper, we used transient CFD frequency responses to derive the orifice resonant frequency (the frequency at which the sinusoidal response of the orifice differential pressure to the applied flow rate lags approximately 90 degrees in phase). However, the reason for the approximately 63% decrease in resonant frequency in the 1D modelling was unclear. Based on the flow patterns in the CFD analysis, the inertial length: L, which is proportional to the inertial resistance $\rho \cdot L/A_O$ in the 1D model, appears to be 2.5 times larger in the CFD model. Velocity contours show that a uniform flow from the orifice extends beyond the length of the short pipe. In this paper, we examine the flow patterns and conclude that the increase in inertial length is due to two factors. One factor is additive mass. As shown in [11], when a cylindrical object moves through a fluid, it entrains and simultaneously moves a mass of fluid $3/8 \cdot \rho \cdot (D_O/2)^3$. This increases the inertial length of the fluid by $3 \cdot \pi \cdot D_O/4$. Since $D_O = 6 \cdot L/7$ for this orifice, L is expected to be extended by approximately 0.75 times. Furthermore, attachment is expected to occur at both ends of the fluid passing through the short pipe. While additional mass typically occurs at the end face of a spool valve, in this flow, it is expected to attach to the cylindrical fluid in the short pipe. The second factor is that the fluid flow entering and exiting the orifice is similar to the flow field passing through the curtain area of a nozzle/flapper valve, and thus simultaneously moves the cylindrical fluid at the height of the curtain area. Flow along the orifice axis creates a flow pattern that obstructs the flow at both the inlet and outlet of the orifice. Because the flow velocity is slower outside the orifice, the cylindrical flow passing through the orifice continues to flow away from the orifice. When the flow direction is reversed in the orifice's alternating flow, the axial flow near the orifice is obstructed, resulting in radial flow. From a macroscopic perspective, the 90-degree flow bend is like the flow within the curtain area of a nozzle/flapper. This is thought to increase the inertial length, which is equivalent to the curtain area height d/2 ($\cong 3/7L$). Figure 5.1 shows a conceptual diagram of inertial length correction around the orifice. Due to the above two factors, the

inertial length is thought to increase by approximately 2.2 times. If a shape factor that approximates the radial flow near the orifice to the curtain area can be identified in the future,

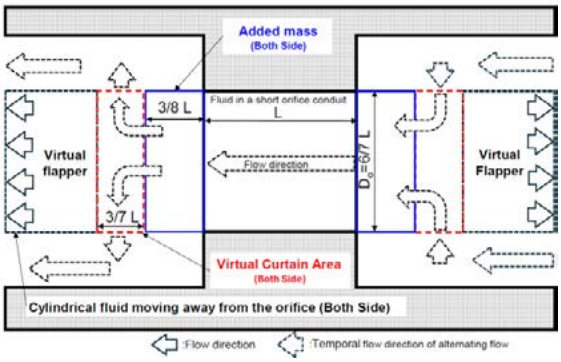


Figure 5.1 Simplified flow pattern near an orifice

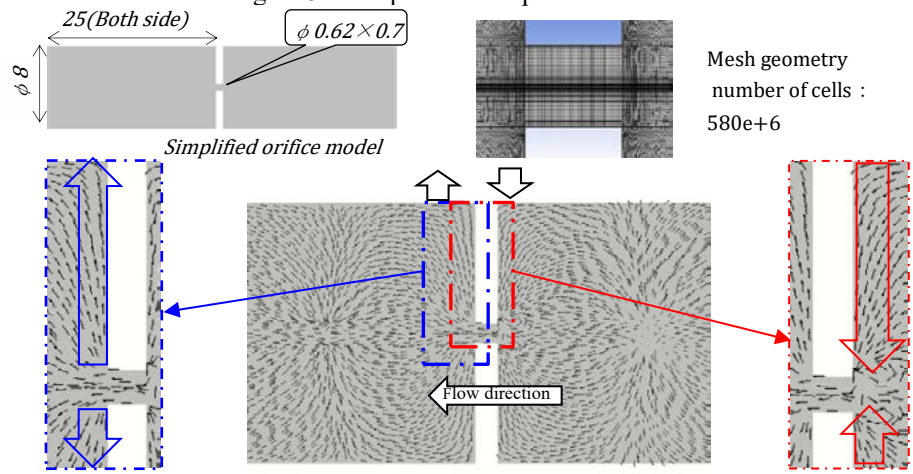


Figure 5.2 Unsteady CFD analysis results for orifice frequency response (vector diagram)

it is expected that the accuracy of deriving the inertial length will improve. Figure 5.2 shows the results of the orifice frequency response analysis using a simplified geometry model with ANSYS CFX, presented as a vector diagram.[13] The frequency is set at $400 \cdot 2\pi[rad/s]$, and the flow rate amplitude is equivalent to that in Table 4.1. ANSYS CFX was used to easily obtain the vector diagrams. As indicated by the arrows in Figure 5.2, the flow pattern on both sides of the orifice exhibits radial inflow (inside of red hollow arrow) and outflow (inside of blue hollow arrow). This flow field is characteristic of the curtain area and occurs just before the direction of the alternating flow changes. It is considered that the radial outflow is generated as the flow from the previous cycle is overtaken and collided with by the flow of the subsequent cycle.

5.2. Considerations on the Shift of the Orifice Natural Vibration

Confirmation of the local discharge coefficient using CFD suggests that no contraction occurs within the orifice. Although contraction does occur before and after the orifice, due to flow symmetry, the location of contraction shifts instantly to the orifice inlet, making the effect of alternating frequency on contraction small. On the other hand, the delay in high

frequencies through the orifice is likely to have a significant impact on the flow difficulty. First, we must further investigate whether orifice impedance has any effect on reducing the discharge coefficient. This will require defining the orifice's impedance characteristics using a 1DCAE approach rather than a flow field approach and correlating them with the discharge coefficient as necessary.

6. CONCLUSION

The following was revealed through 1D and 3D orifice simulations.

- In the LCR circuit model of orifice alternating flow, the inertance increased by approximately 2.5 times in the unsteady CFD analysis, and the co-movement frequency decreased by approximately 63%. This is thought to be mainly due to the added mass and the fluid within the curtain area of the virtual nozzle flapper moving simultaneously with the fluid within the pipe. The virtual curtain area is formed by the end face of the added mass, the orifice outlet flow rate, and the surface (virtual flapper) where the axial velocity becomes zero when the flow direction inside the orifice reverses.
- The orifice discharge coefficient is thought to be largely unaffected by the frequency of the alternating flow. Backflow does not occur in short pipes, and the contracted flow that occurs at the orifice inlet quickly moves to the orifice outlet (the inlet after the alternating flow) due to the alternating flow.
- As shown in the example above, the effect of the alternating flow frequency on the discharge coefficient of the steady-state orifice restriction equation is thought to be small. However, the impedance of an LCR circuit is generally affected by the frequency of the alternating flow, and this effect is thought to be dominant.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to Dr. Takeshi Yamaguchi and Ms. Akari Takekawa, a second-year master's student at the Graduate School of Information Engineering, Kyushu Institute of Technology, for their valuable advice on this research.

7. REFERENCES

[1] S. Masuda, Y. Matsunaga, A. Takegawa, F. Shimizu, K. Tanaka, Modelling and simulation of damping orifices for pressure control valves, Proc. ISFP 2024 JFPS, 1A2-03, https://doi.org/10.5739/jfps.52.8

[2] S. Oshima, T. Ichikawa, Frequency Characteristics of Cylindrical Short Chokes for Hydraulic Valves (In case of using general hydraulic oil), J. Hydraul. Pneumatics (Japanese) Vol.15, No.2, 1984, pp.157-165, https://doi.org/10.5739/jfps1970.2.72

[3] S. Miyagawa, Frequency Characteristics of Cylindrical Short Chokes, J. JSME (Japanese) Vol.42, No.355, 1971, pp.847-856, https://doi.org/10.1299/kikai1938.42.847

[4] K. Nakano, S. Yokota, Instantaneous Flowmeter for Hydraulic Systems Utilizing the Dynamic Characteristics of Cylindrical Chokes, J. Hydraul. Pneumatics (Japanese)Vol.17, No.4, 1986, pp.311-317, https://doi.org/10.5739/jfps1970.17.311

[5] S. Masuda, Simulation of the Pressure Control Valve Improving Responsiveness and Stability by Variable Restrict Orifice, Pro. ISFP2017, 2C17

[6] K. Tanaka, Y. Sakurai, System Dynamics and Bondgraphs (2) (Basics of Bondgraph), Turbomachinery Vol.50, No.6, 2022, pp365-375
[7] K. Tanaka, Y. Sakurai, System Dynamics and Bondgraphs (3) (Bondgraph Modelling of Fluid Systems), Turbomachinery Vol.50, No.8, 2022, pp494-505
[8] A. Takegawa, F. Shimizu, S. Masuda, Y. Matsunaga, K. Tanaka, Unsteady Characteristics of Orifice Wake Flow, Proc. JSME Kyushu, Student Conf. 2024
[9] Simerics: Simerics MP+® Manual V5.0 (2019).
[10] I. Iimura, N. Goto, H. Hatanaka, Impedance Characteristics of Orifices, Transactions of the Society of Instrument and Control Engineers, Volume 23 Issue 1 Pages 21-26, 1987
[11] L. C. Jaw, J. D. Mattingly, Aircraft Engine Controls, Aiaa, 2002, pp50-56, https://doi.org/10.2514/4.867057
[12] K. Nishio, M. Endo, Y. Matsuda, The Control of Turbofan Engines, Technical Report of NAL (Japanese), TR-428, 1975, https://jaxa.repo.nii.ac.jp/records/44678
[13] ANSYS: CFX-Pre User's Guide Release 2025R1 (2025)

NOMENCLATURE

Symbol	***Explanation of variable symbols***
A_M	*Metering valve opening area*
C_1	*C element of piping volume*
C_2	*C element of pressure control valve and piping end volume*
D_o	*Diameter of the Dumping orifice*
F_S	*Spring preload*
F_O	*Dumping force of orifice*
F_{DP}	*Force due to differential pressure*
IL	*Inertance of Fluid inertia in orifice pipe*
k_v	*Pressure Control Valve Opening Area per displacement*
t	*Time*
P_{2s}	*Pressure of Spring room*
P_C	*Pressure downstream of Pressure Control valve*
PL	*Effort (pressure) due to fluid inertia*
Po	*Differential pressure of Damping orifice*
R_1	*Pipe pressure loss (pipe resistance) R_1*
R_2	*Pressure control valve restriction R_2*
R_3	*Metering valve restriction R3*
V_1	*Volume upstream of metering valve*
V_S	*Volume upstream of Spring room*
ρ	*Fluid density*
$(X)_0$	*Subscript indicating the steady-state value at time: 0 for variable: X*
ΔPM	*Differential Pressure of Metering Valve*
ΔQ	*Fluid fluctuation in piping volume*
μ	*Viscosity*
ζ	*Damping ratio*

Biographies



Seiei Masuda graduated from the Master course at the Graduate School of Science and Technology, Sophia University, and joined Ishikawajima-Harima Heavy Industries Co., Ltd. (currently IHI Corporation) in 1990. Since then, he has been working in the Mechanical Control Engineering Group, Control Systems Engineering Department, Aerospace and Business Area, mainly developing jet engine fuel control equipment. He holds a Professional Engineer, Japan (Mechanical Engineering), a Professional Engineer, Japan (Engineering Management), and a Doctor of Engineering (Kyushu Institute of Technology, 2020).



Yasushi Matsunaga graduated from the Master's program at the Graduate School of Engineering, Tohoku University in 1986 and joined Ishikawajima-Harima Heavy Industries Co., Ltd. (now IHI Corporation) in the same year. Since then, he has been engaged in research and development of jet engine FADEC systems and rocket turbopumps in the aerospace and business fields.



Fumio Shimizu received his Doctor of Engineering from Tokyo University of Agriculture and Technology in 1995 and subsequently worked in the Department of Physics and Information Technology at the Kyushu Institute of Technology Faculty of Computer Science and Systems Engineering. His research interests include Computational Fluid Dynamics, Gas/Liquid Two-Phase Flow, Acoustic Noise, and Fluid Dynamics.



Kazuhiro Tanaka received his Doctor of Engineering from the University of Tokyo in 1983. He went on to Sophia University and the University of Tokyo and is currently a Professor Emeritus at the Kyushu Institute of Technology.

Research Interests: Thermal/Fluid Engineering, CFD, Fluid Power Systems (Oil-Hydraulics, Pneumatics), Bond Graph Modelling & Simulation

The optimal design of hydraulic pump performance and reliability through lubrication, mechanism, and structural coupling analysis

Keitaro Kamata*, Hiroyuki Nomaguchi*, Toshikazu Hayashi*, and Atsushi Yuge*

*Mitsubishi Heavy Industries, Ltd., 5-717-1, Fukahori-machi, Nagasaki, 851-0392, Japan**

E-mail: keitaro.kamata.tk@mhi.com

Abstract.

This study proposes a design method for interacting sliding parts in hydraulic pumps that optimizes both performance and reliability within practical computation time. Conventional design analyzes each part individually using EHL (Elasto-Hydrodynamic Lubrication), overlooking interactions and causing redesign delays. The new approach integrates lubrication-mechanism-structure coupled analysis, Taguchi-based sensitivity analysis, and response surface optimization to determine optimal shapes. Validation using an elastic body model showed only a 7% error compared to the rigid model. Compared to the current design, total loss decreased by about 30%, improving overall efficiency by 4%. This technology enables faster development without compromising reliability.

Keywords. Hydraulic pump efficiency, Sliding part design, Coupled analysis, Response surface Optimization.

1. INTRODUCTION

Figure 1.1. shows the before-after design of the sliding part. In the conventional sliding part design, EHL (Elasto-Hydrodynamic Lubrication) analysis of each sliding part is performed, and the shape is determined individually. Therefore, the interaction between multiple sliding parts is overlooked, and redesign and verification are required to achieve both performance and reliability, resulting in an increase in development time.

In this study, a complex sliding part design method that can determine the optimum shape that achieves both reliability and performance in practical calculation time was constructed by simultaneously designing multiple complex sliding parts by the strong coupled analysis of mechanism and lubrication, extracting design parameters of sliding parts by the sensitivity analysis using the Taguchi method, and performing the optimum solution calculation using the response surface.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

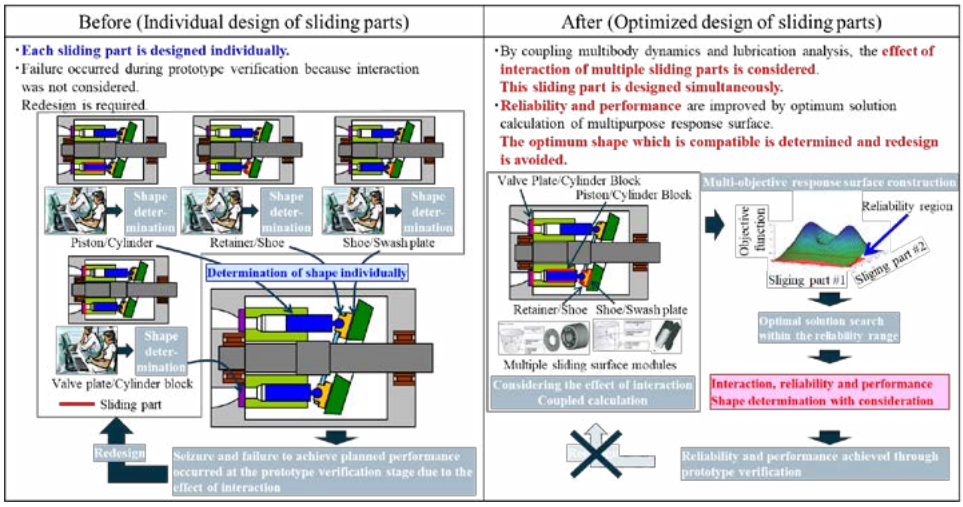


Figure 1.1. Before and After of Designing of Interacting Sliding Parts

2. OPTIMIZATION DESIGN METHOD FOR THE INTERACTION OF MULTIPLE SLIDING PARTS

2.1. Analysis Object

Figure 2.1.1. shows the analysis object. In this study, a simultaneous optimization design method of multiple sliding parts for a swash plate hydraulic pump was verified. In order to achieve both reliability and performance, it is necessary to ensure the integrity of the sliding surface and reduce the power loss. The sliding parts of a swash plate type hydraulic pump are (1) valve plate and cylinder block, (2) piston and cylinder, and (3) shoe and swash plate. In order to improve the output of a hydraulic pump, it is necessary to design the clearance of each sliding part in consideration of the increase of friction loss and seizure of the sliding surface due to insufficient clearance.

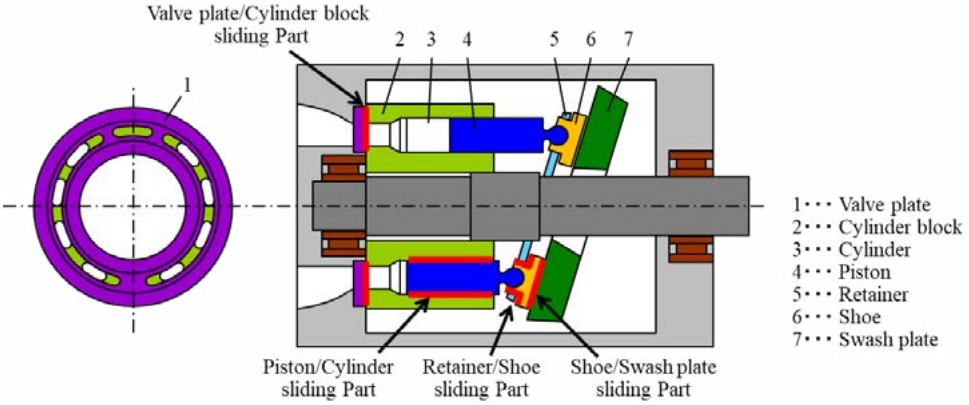
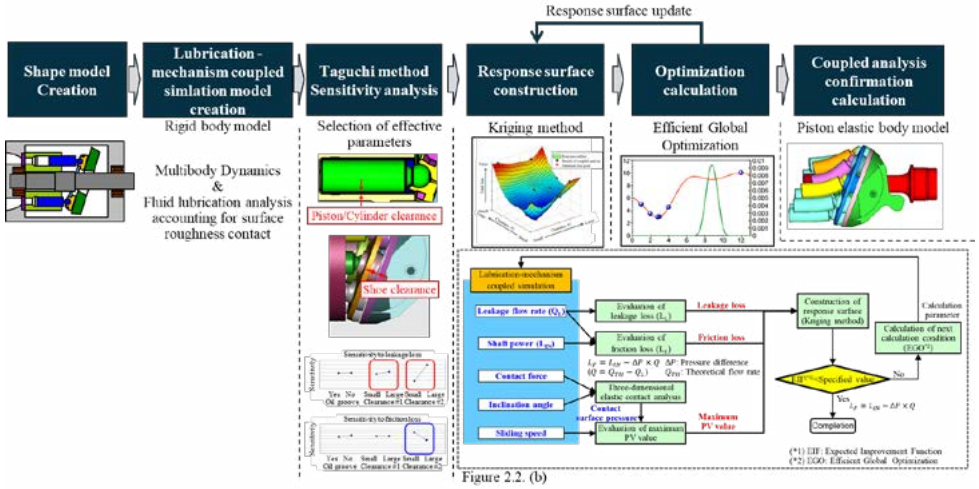


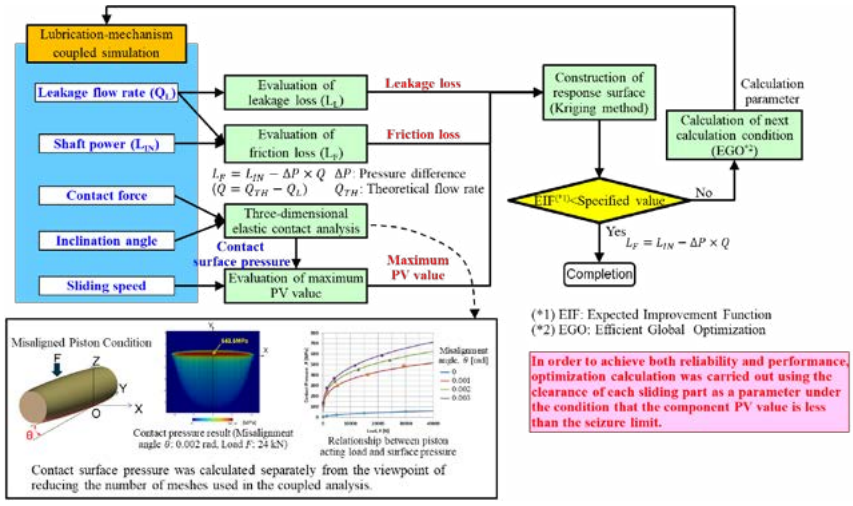
Figure 1.1.1. Analysis Object

2.2. *Design flow for optimization of sliding part*

Figure 2.2.1. shows the design flow for simultaneous optimization of sliding part. The calculation load of the elastic body model is more than 5 times as high as that of the rigid body model (calculation time 20hr per 90deg of axial rotation angle), and it takes time to obtain the optimum design dimensions. Therefore, the rigid body model was used for the coupled analysis in the sensitivity analysis and optimization calculation, which require calculation of multiple cases, and the calculation time was shortened by separately considering the elastic effect in contact with the map data of the 3D elastic contact analysis. For the final verification, the coupled analysis of the elastic body model was confirmed with the optimum solution obtained by the rigid body model.



(a) Overall design flow



(b) Details of Optimization Calculation

Figure 2.2.1. Flow Chart of Simultaneous Optimization Design of Sliding Parts

2.3. Lubrication-Mechanism-Structure Coupled Analysis simulation

The lubrication-mechanism-structure coupled analysis code FIRST developed by IST in Germany was used to construct an analysis model for multiple sliding surfaces. FIRST can couple the mechanism analysis by multibody dynamics with the modified Reynolds equation and the mixed lubrication analysis considering the roughness by the Greenwood-Tripp model, and the analysis considering the interaction of each sliding surface is possible. In addition, the analysis considering the elastic deformation of the parts is also possible by using the structural model of the degenerate parts by the Structural Analysis Tool (ANSYS).

In this study, the main sliding surfaces, (1) valve plate and cylinder block, (2) piston and cylinder, (3) shoe and swash plate, were modeled as sliding surfaces, and the other parts were modeled by connecting spring-damper elements.

As an input condition, the cylinder pressure fluctuation curve obtained by one-dimensional simulation was used as a pressure load on the piston and a pressure boundary condition on the sliding surface.

Using this model, the power loss and efficiency were evaluated based on the driving torque when the pump shaft was forced to rotate and the amount of leakage from each sliding surface.

2.4. Selection of High-Sensitivity Design Factors

In order to select design factors to be used in the optimization calculation, sensitivity evaluation of leakage loss and friction loss for each sliding part was carried out by the Taguchi method. The control factor for the sensitivity analysis is shown in Figure 2.4.1. Three factors (Presence or absence of cylinder oil groove, piston/cylinder clearance, shoe/swash plate clearance) were selected to carry out the sensitivity analysis. Among these three factors, the shoe has a pressure pocket at the bottom where the oil pressure from the cylinder is introduced, and the oil pressure balance is set to float freely. Therefore, the shoe floats from the swash plate during the compression stroke, and leakage occurs through the gap between the shoe and the swash plate, which is considered to have a large effect on the leakage loss. In addition to the leakage loss, the piston/cylinder gap has an effect on the contact pressure at the contact point, which greatly affects the contact PV value of the sliding surface. The cylinder oil groove is provided to improve the loading capacity of the oil film by introducing the high-pressure oil to the side surface of the piston.

Figure 2.4.2. shows the results of the sensitivity evaluation. The shoe/swash plate clearance showed the highest sensitivity for both leakage loss and friction loss, suggesting that it is an important parameter to minimize the total loss. The piston/cylinder clearance showed the highest sensitivity for leakage loss.

Based on the results of the sensitivity analysis, the shoe/ swash plate clearance and the piston/cylinder clearance were set as two design variables for optimization.

Sensitivity parameter	Set value
Cylinder oil groove	Yes, No
Piston/cylinder clearance	Small, Large
Shoe/Swash Plate clearance	Small, Large

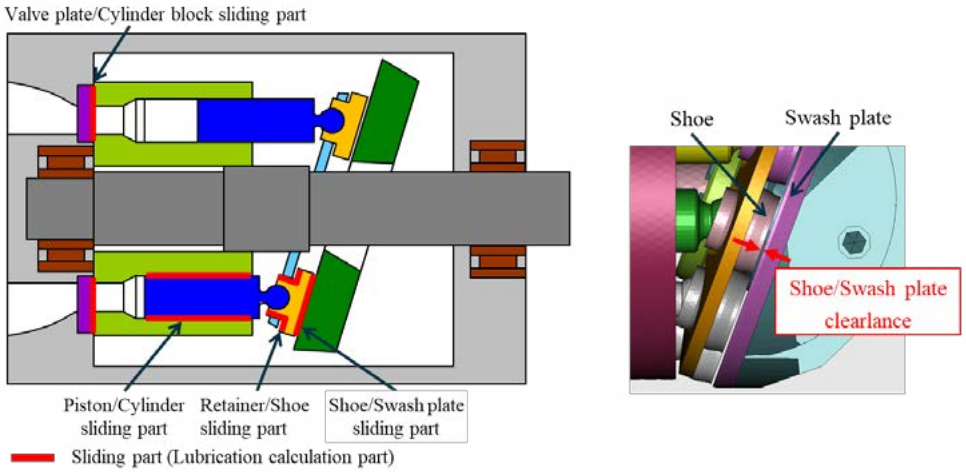
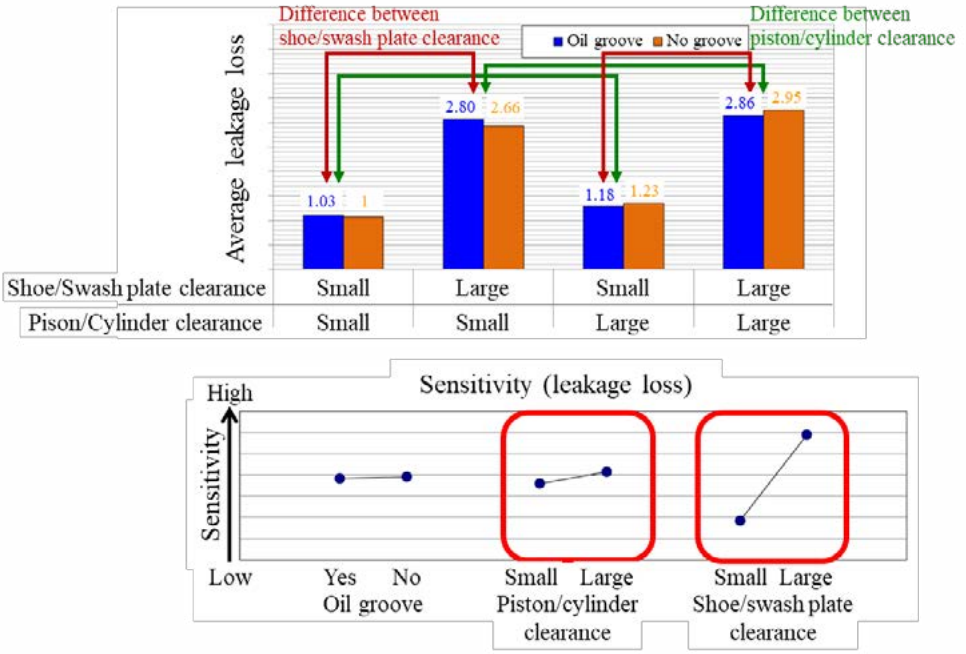
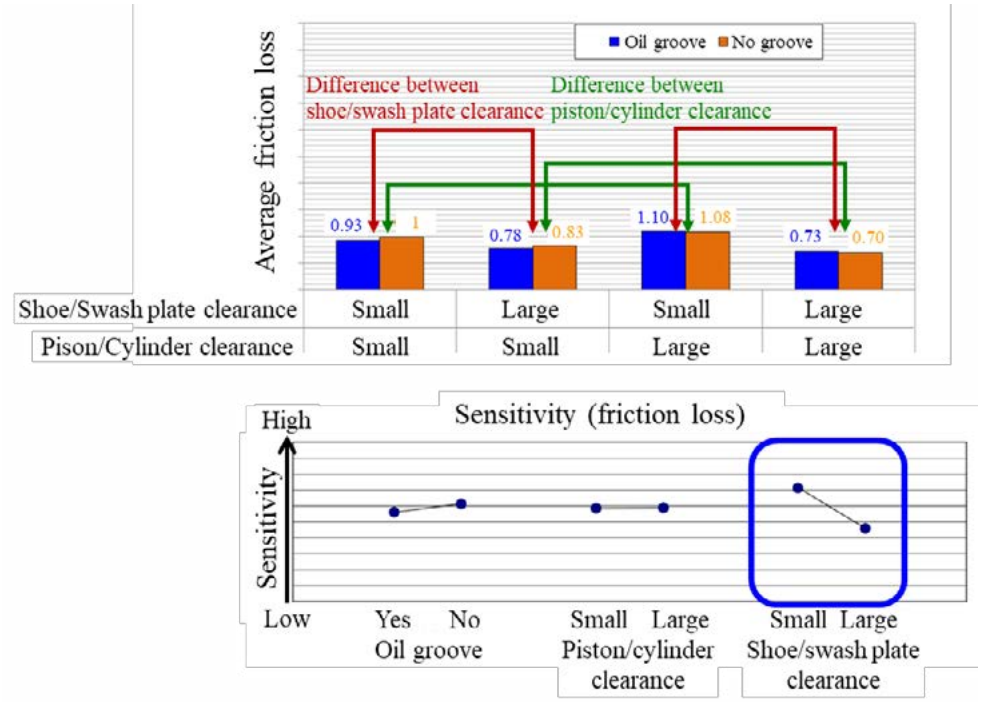


Figure 2.4.1. Control factors for sensitivity analysis



(a) Leakage loss



(b) Friction loss

Figure 2.4.2. Sensitivity evaluation result of the interaction of sliding surfaces

2.5. Optimal Solution Calculation Using Nonlinear Approximation Surface

2.5.1. Calculation method

The lubrication- mechanism-structure coupled analysis is a large-scale numerical analysis, and the calculation time is several hours to several days per case. By adopting the optimization method using the response surface, the number of analysis cases to obtain appropriate design dimensions can be reduced. In this study, the Kriging method, which is suitable for approximating nonlinear functions, was used as a response surface construction method. The response surface by the Kriging method includes not only the distribution of the function value itself but also the uncertainty (approximation error) information that can be included in the estimated function value. Efficient Global Optimization (EGO), which is an optimization algorithm to improve the accuracy of the response surface, was adopted with this information. In EGO, the expected improvement (EI), which is an index expressing how much the estimation accuracy can be improved when a new sample is added to a certain point and the response surface is reconstructed, is evaluated. Then, the location where the EI value is maximum is searched, and a new sample point is added to reconstruct the response surface. By repeating these operations, the improvement of the approximation accuracy of the response surface and the search for the global optimum solution can be achieved simultaneously. Compared with the conventional optimization method (genetic

algorithm), it is reported that the number of calculations required to obtain a solution with the same accuracy is less than 1/20 [1]. Therefore, it is expected that the solution can be obtained in a practical time. In this study, the response surface was generated using Dakota, an open source optimization tool developed by Sandia National Laboratories, USA, and the optimum solution was searched by EGO.

In the optimum solution search, a constrained nonlinear optimization was performed to minimize the total loss by keeping the PV value below the criterion value within the range of design variables (Piston/cylinder clearance, shoe/swash plate clearance) specified in the optimization problem.

In the coupled analysis model used for the optimization calculation, all parts were treated as rigid bodies and the analysis mesh of the sliding surface was made relatively coarse to reduce the calculation time. Considering the periodicity of the pump behavior, the analysis range was limited to 1/4 rotation of the main shaft. As a result, the calculation time per 1 case was about 16 hr, and the optimum solution could be obtained within a practical period. However, the analysis accuracy of the contact pressure could not be guaranteed because the sliding surface analysis mesh is coarse and the number of contact nodes per piston piece may be 1 to several points. Therefore, the contact pressure of the piston was separately calculated based on the three-dimensional elastic contact model from the contact load of each contact region (the upper end of the piston and the small mouth of the cylinder) and the inclination angle of the piston with respect to the cylinder obtained from the analysis result of the coupled analysis.

2.5.2. Optimization calculation result

Figure 2.5.1. shows the response surface at steps 0, 3, and 10 of optimization, and Figure 2.5.2. shows the final response surface and analysis points.

An initial response surface was created based on the analysis result of 4 points used in the sensitivity analysis, and the optimization calculation was executed. As a result, the amount of movement of the search point fell below the criteria at step 10, and it was judged to be convergence. The initial response surface showed a monotonic decreasing trend for each design variable with low estimation accuracy. However, the optimum point appeared at the tip of the bowl-shaped region at step 10.

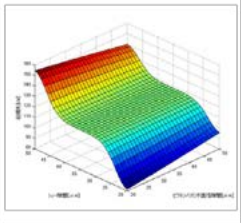
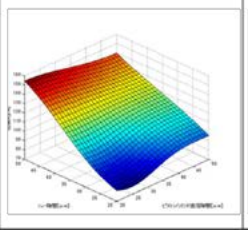
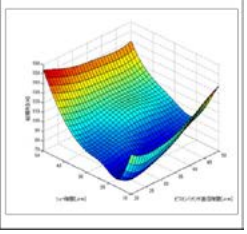
Number of optimization steps	#0 (Initial)	#3	#10 (Convergence)
Response surface			

Figure 2.5.1. Optimization steps

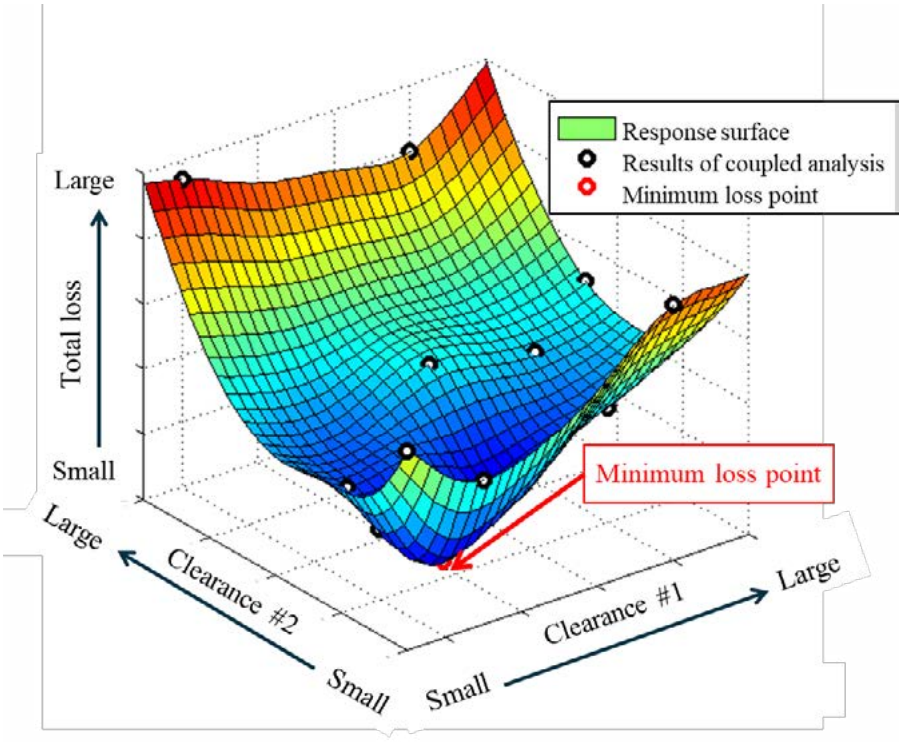


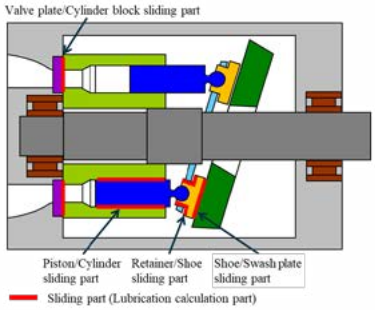
Figure 2.5.2. Optimal solution and kriging surface after convergence

2.5.3. Confirmation of Optimum Solution by Elastic Body Model

Figure 2.5.3. shows the comparison result between the rigid body model and elastic body model. In the optimization calculation, the rigid body model was used to reduce the calculation time. In order to confirm the validity of the optimization result using the rigid body model, an analysis was performed considering the elastic behavior of the piston at the optimum point. Comparing the rigid body model with the elastic body model, the error of the calculation result of the total loss was about 7%, and there was no significant difference in the distribution of the oil film thickness, oil film pressure, and contact pressure (the calculation time of the elastic body model was 4-10 times longer than that of the rigid body model).

Comparison of Loss Calculation Results

	Total loss (-)	Error (%)
Elastic body model	1.07	-
Rigid body model	1	7



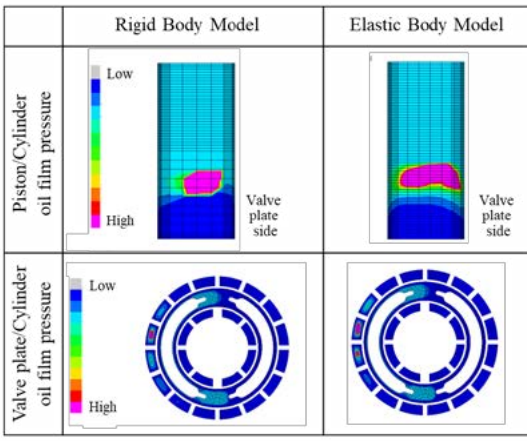


Figure 2.5.3. Comparison of Elastic and Rigid Analysis Result

2.5.4. Comparison of the Optimization Result and the Current Design Result

Figure 2.5.4. and Figure2.5.5. show the comparison between the analysis result and the optimization result and the current design result. Comparing the current design point with the optimization result, it was confirmed that the total loss was reduced by about 30%, and the corresponding improvement in total efficiency was 4%.

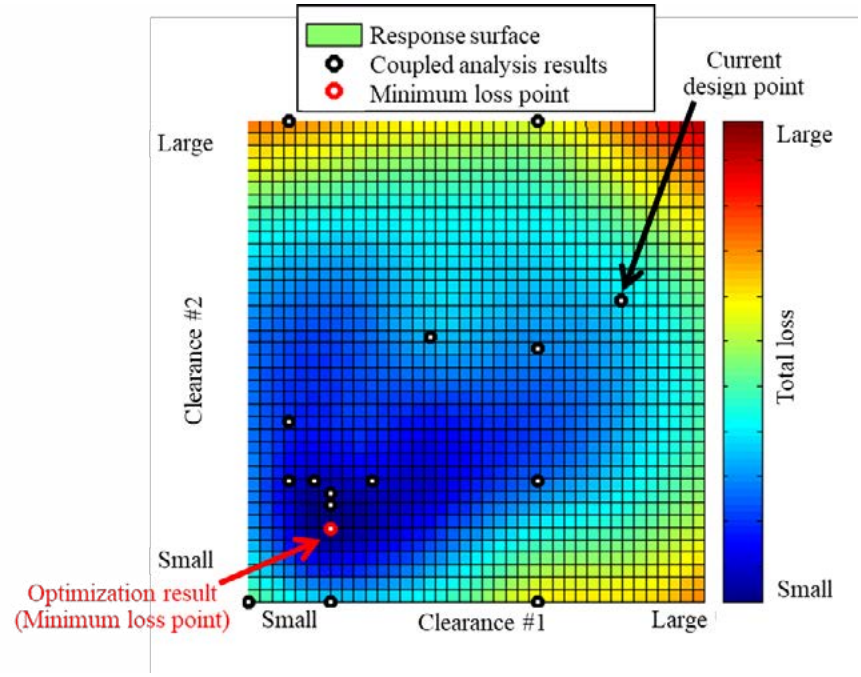


Figure 2.5.4. Response Surface Contour of total loss and tolerance zone of clearance

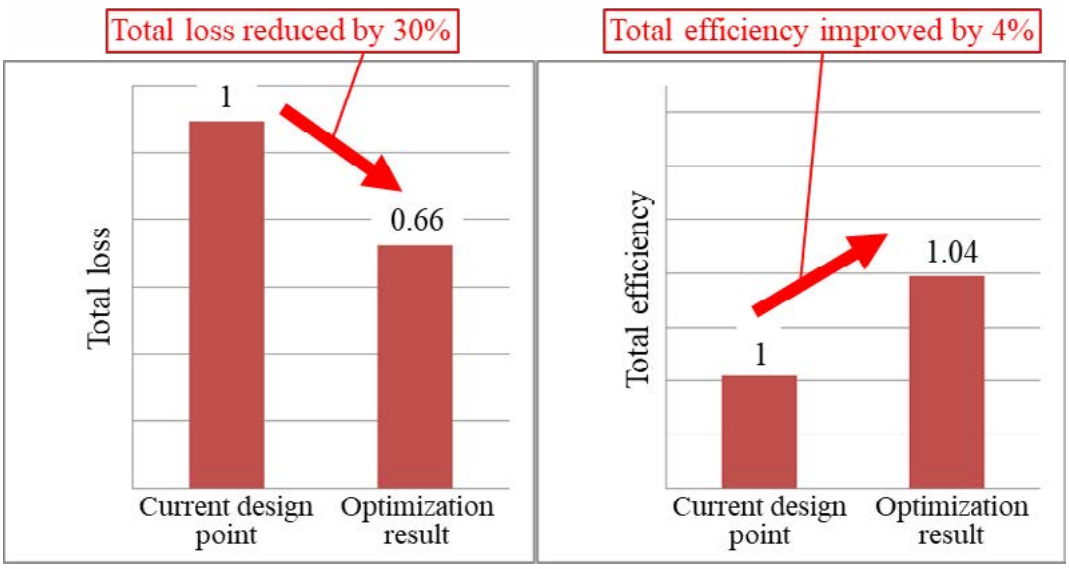


Figure 2.5.5. Comparison between the optimization result and the current design result

3. SUMMARY

By constructing a sliding part design technology that combines lubrication-mechanism-structure coupled analysis and nonlinear optimization technology, it is possible to design a sliding part shape that achieves both performance and reliability of a hydraulic pump in a short period of time, and it is expected to avoid an increase in the development period.

4. REFERENCES

[1] R. Aror, et. al., 'Design Technique for Mechanism Analysis Using Nonlinear Response Surface by Applying High-speed, High-precision Optimisation Method', Mitsubishi Heavy Industries Technical Review Vol. 53 No. 4, Dec., 2016.

[2] Y. Chuman, et. al., 'Innovation of Design and Development Process using Structure Optimization Technology', Mitsubishi Heavy Industries Technical Review Vol. 52 No. 4, Dec., 2015.

Performance Improvement of Pressure Compensated Flow Control Valve Using Virtual Engineering

Harsh Maniar[1], Ujas Patel[1] and Nirav Patel[1]

[1]Bosch Rexroth (I) Pvt Ltd, harsh.maniar@boschrexroth.co.in; ujasp.patel@boschrexroth.co.in; niravkumardahyalal.patel@boschrexroth.co.in

Abstract.

This paper addresses the performance challenges associated with pressure compensated flow control valves, particularly in maintaining effective pressure compensation at higher flow rates. The primary objective is to investigate how various design parameters influence the valve ability to maintain consistent flow under pressure variations. Given the complexities and resource demands of physical testing across numerous design variants, a more efficient approach is required. This study uses virtual engineering techniques to create a simulation model of the valve, enabling a comprehensive analysis of the impact of different design parameters on its performance.

Keywords. Pressure compensation, Flow control valve, Flow force, CFD, Mathematical model.

1. INTRODUCTION

Hydraulic systems are widely used in both industrial and mobile applications because of their high-power density and ability to generate large output forces [1]. Actuator motion control is required in most hydraulic applications, which can be achieved by regulating the flow rate of hydraulic oil. Table 1 summarizes the different components used to control flow rate in hydraulic systems.

Components	Remarks
Fixed displacement pump	- Requires a mechanism to regulate pump speed. - Cannot provide independent speed control for multiple actuators.
Variable displacement pump	- More expensive compared to fixed displacement pumps. - Cannot provide independent speed control for multiple actuators.
Flow control valves	- Allow independent speed control of multiple actuators.
Closed loop control with servo system	- Provides precise control of flow rate. - Involves higher complexity and cost.

Table 1: Overview of different components for flow rate control.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Flow control valves regulate the flow rate by varying the throttle area. Such valves are commonly referred to as throttle valves. They are available from several manufacturers, including F valves (Bosch Rexroth), T valves (Yuken), and Q valves (Atos). The main limitation of throttle valves is that the regulated flow rate depends on the pressure drop across the valve. To maintain a constant flow independent of pressure variations, pressure compensation is employed in valves. Pressure compensated flow control valves (referred as PCFCV here) are offered by several manufacturers, including FRM valves (Bosch Rexroth), F valves (Yuken), QV valves (Atos).

A new design of a pressure-compensated flow control valve for water hydraulics was developed [2]. The valve dimensions were determined by solving the simplified equations in MATLAB. Experimental investigations were then carried out on the prototype valve, which revealed the following key findings:

- The pressure–flow characteristics indicated that an increase in pressure led to a slight decrease in flow rate.
- Incorporation of flow-force compensation reduced the rate of flow reduction with increasing pressure.
- Flow-force compensation using an O-ring further improved the pressure–flow characteristics; however, it also introduced a negative impact on hysteresis.

[3] analysed such a valve using (i) steady-state analysis with a nonlinear model and (ii) dynamic response analysis with a linearized model. The nonlinear model was used to extract parameters for the frequency response model, from which flow-gain and pressure-coefficient transfer functions were derived. The study also examined the effect of various design parameters on valve dynamics, including critically compensated, over-compensated, and under-compensated conditions. Similar investigations have been reported in [4], [5], [6], focusing on the analysis and performance evaluation of pressure-compensated flow control valves.

Performance improvement of pressure compensated flow control valve can be done by improving flow control accuracy, effective pressure compensation and output flow stability. In reference [7] shows influence of main orifice shape on accuracy of flow control. They identified two shapes (triangle and circular) for proportional flow control valve. They created characteristic curves for valve using CFD data in conjunction with MATLAB/SIMULINK. [8] performed dynamic analysis of two stage poppet valve for flow control. The effect of flow forces on force equilibrium to maintain constant outflow is studied.

The present study aims to develop a methodology for predicting the performance of pressure-compensated flow control valves (PCFCVs) using CFD. For this purpose, 2FRM6 valve from Bosch Rexroth is used. Furthermore, the study seeks to enhance valve performance by investigating the influence of various design parameters on its dynamic and steady-state behaviour

2. FLOW CONTROL VALVE: 2FRM6

2.1. Working of 2FRM6

The 2FRM6 is a pressure- and temperature-compensated two-way flow control valve. Figure 2.1 illustrates its schematic layout. The valve housing (1) accommodates a main sleeve (3) and a compensator sleeve (9). The main sleeve contains a main orifice (5) of triangular shape, the opening of which is regulated by the throttle screw (8). This screw moves axially within the main sleeve and is externally adjusted through the control knob (2).

The compensator spool (4), located within the compensator sleeve (9), governs the pressure-compensation mechanism. Flow from the main orifice discharges through circular holes in both the compensator spool and sleeve. The compensator spool is free to move axially to an equilibrium position determined by the balance of forces acting on it. Specifically, fluid pressure acts on its lower face via the nozzle (7) and on its upper face via the main orifice (5). The resulting position of the spool adjusts the throttling area, thereby compensating for pressure variations. Additionally, the spring (6) applies a restoring force on the compensator spool, which varies with the relative positions of the throttle screw and spool.

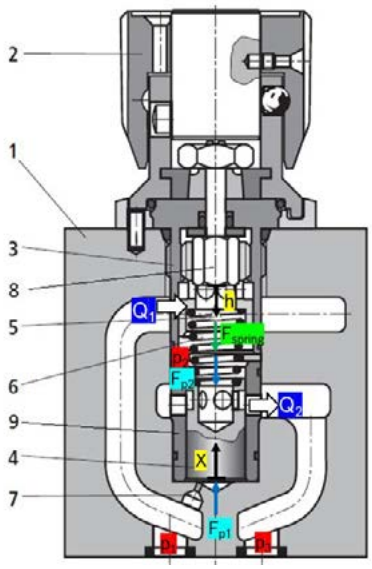


Figure 2.1: Schematic representation of 2FRM6: 1-Housing, 2-Adjustment knob, 3-Main sleeve, 4- Compensator spool, 5- Main orifice, 6- Spring, 7- Nozzle, 8- Throttle screw, 9-Compensator sleeve.

Subsequent section presents the approach used to estimate the performance of the 2FRM6 valve using CFD simulations. Before detailing the CFD methodology, a mathematical description of the valve operation is provided to illustrate how the equilibrium position of the compensator spool governs the flow characteristics. This formulation establishes the foundation for the subsequent numerical analysis by linking hydraulic forces and spring forces to the valve's operating points.

2.2. Mathematical Description of 2FRM6

Flow through main orifice is described by standard orifice equation as:

$$Q_1 = c_{d1} A_{main} \sqrt{\frac{2(p_1 - p_2)}{\rho}} \tag{2.1}$$

Similarly, the flow through compensator orifice can be described as:

$$Q_2 = c_{d2} A_{compensator} \sqrt{\frac{2(p_2 - p_3)}{\rho}} \tag{2.2}$$

where, c_{d1} and c_{d2} are discharge coefficients of main and compensator orifices, respectively, and A_{main} and $A_{compensator}$ are the effective flow areas of main orifice and compensator orifice. For simplicity, the downstream pressure p_3 can be assumed equal to tank pressure, which is taken as zero.

In valve performance evaluation, the main throttle area A_{main} is fixed according to the control knob setting. The discharge coefficients c_{d1}, c_{d2} are generally unknown and can be determined experimentally or through CFD simulations.

At steady state, the flow through both orifices must be equal ($Q_1 = Q_2$). By equating Eqs. (2.1) and (2.2), the intermediate pressure p_2 can be expressed as:

$$p_2 = \frac{c^2}{1 + c^2} p_1 \tag{2.3}$$

where, $c = \frac{c_{d1} A_{main}}{c_{d2} A_{compensator}}$.

The position of the compensator spool is determined from the equilibrium between the hydraulic and spring forces:

$$F_{spring} = F_{hyd} \tag{2.4}$$

The spring (6) is compressed between the throttle screw and compensator spool, and its force can be written as:

$$F_{spring} = F_0 + kx \tag{2.5}$$

where F_0 is the initial compression force dependent on the throttle screw position ($F_0 = f(h)$), k is the spring stiffness, and x is the axial displacement of the compensator spool. F_{hyd} consists of all type of forces acting on compensator, i.e., force due to pressure, flow, and friction. Considering only pressure force for simplicity one can express it as:

$$F_0 + kx = \frac{p_1 A}{1 + c^2} \tag{2.6}$$

A is area of compensator where pressure acts. Since c depends on $A_{compensator}$, which in turn depends on the spool displacement x, Eq. (2.6) is nonlinear and can be solved iteratively to determine the compensator position for a given inlet pressure p_1. Once the equilibrium position is known, the corresponding flow rate can be calculated using Eq. (2.1).

3. PERFORMANCE EVALUATION OF 2FRM6 USING CFD

The mathematical model provides a framework for understanding the steady-state behaviour of the 2FRM6. However, the model has inherent limitations. The discharge coefficients of the main and compensator orifices (c_{d1} and c_{d2}) are generally unknown and cannot be predicted analytically accurately without detailed flow information. Furthermore, while the model accounts for the pressure forces acting on the compensator spool, it neglects the effects of flow-induced forces and other secondary effects such as friction. These simplifications limit the predictive capability of the mathematical model for precise performance evaluation and hence, CFD simulations are employed. By combining the mathematical understanding of spool force balance with CFD-predicted pressure distributions, the equilibrium position of the compensator spool and the corresponding flow–pressure characteristics of the valve can be reliably evaluated.

3.1. Simulation Strategy

For each simulation, the main throttle screw was fixed at a prescribed position, while multiple axial positions of the compensator spool were considered to capture the variation in pressure and flow. This procedure can be repeated for different main spool positions to cover the full operational range of the valve. In the present study, the main spool position, h, is set to 2.8 mm, and five different compensator positions are considered.

For each compensator spool configuration, CFD simulations were performed to obtain the detailed pressure and velocity distributions. The resulting pressure fields were then combined with the spring force acting on the compensator spool to evaluate the equilibrium position using the force-balance relation. This procedure allowed the determination of the operating point corresponding to each spool configuration, from which the flow–pressure characteristics of the valve were obtained.

The computational domain was discretized using a hybrid mesh generated with commercial software. Mesh was refined in the vicinity of the orifices and the compensator spool to accurately resolve the pressure gradients. A representative mesh snapshot is shown in Figure 3.1, illustrating the resolution in critical regions.

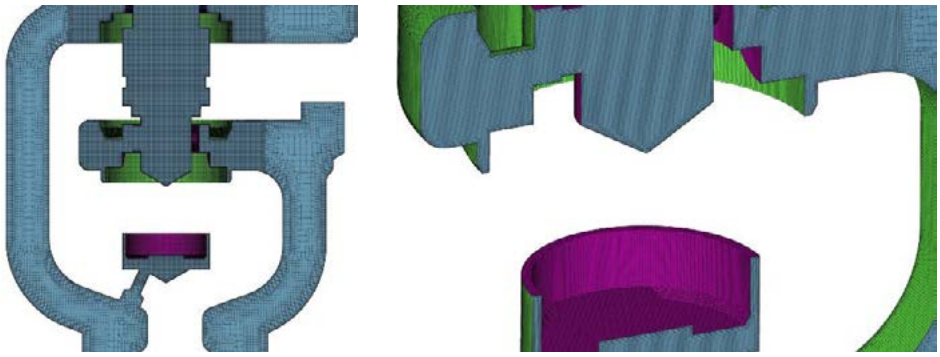


Figure 3.2: Snapshot of discretized domain.

3.2. CFD Model Setup and Parameters

Steady-state CFD simulations were performed using commercial software to resolve the pressure and velocity fields within the 2FRM6 valve. The fluid was treated as incompressible and isothermal, with properties corresponding to HLP46 hydraulic oil at 40 °C. Boundary conditions were applied as follows: a prescribed pressure at the inlet (upstream of the main orifice) and zero-gauge pressure at the outlet (tank). No-slip conditions were applied at all solid surfaces, including the valve housing and spool.

Software provides a wide range of turbulence modelling options. Among them realizable $k-\epsilon$ model [9] was selected, as it is widely regarded as a suitable choice for simulating turbulent flow in hydraulic valves and has been successfully employed in several previous studies [10], [11]. The boundary values of turbulent kinetic energy (k) and dissipation rate (ϵ) were specified using standard relations:

$$k = \frac{3}{2}\left(\frac{V}{TI}\right)^2 \tag{2.7}$$

$$\epsilon = k^{3/2}/L \tag{2.8}$$

Here, V is free stream velocity estimated using orifice equation and using a discharge coefficient of 0.7. TI is turbulence intensity assumed as 3%. L is the turbulence length scale taken as port radius.

4. RESULTS

This section presents the results obtained from the CFD simulations described in the previous section. The subsequent section will focus on the validation of these CFD results through comparison with experimental data.

4.1. Experimental Validation

Figure 4.1 shows the experimental test bench used to measure the performance of the 2FRM6 valve. The throttle screw of the valve was set to the fully open condition (scale position 10), which corresponds to the maximum flow capacity. CFD simulations were carried out for the same configuration, and the figure also presents a comparison between experimental measurements and CFD predictions. The results demonstrate that the CFD predictions are in very good agreement with the experimental data. Next section, provide in detail discussion on valve performance.

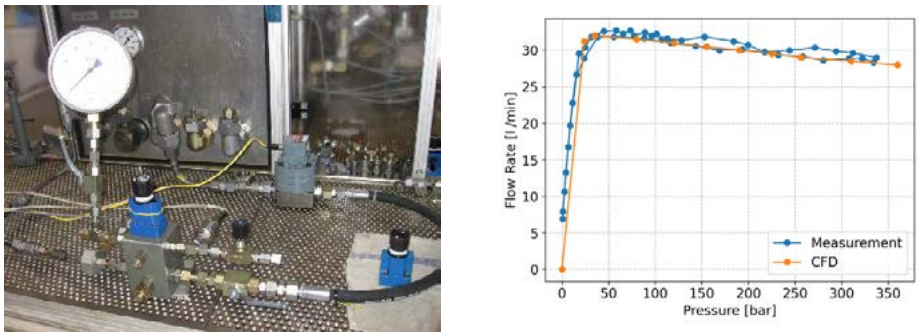


Figure 4.1: Experimental test bench and comparison of CFD and measurement data.

4.2. Discussion on Valve Performance

In a pressure-compensated flow control valve, the performance curve exhibits a distinct characteristic of maintaining nearly constant flow irrespective of pressure variations. At low flow rates, the pressure drop increases gradually, governed primarily by the resistance of the valve geometry. As the flow continues to rise, the compensator mechanism becomes active, adjusting its position in response to the pressure differential across it. Once the compensator engages, any further increase in flow is restricted because the compensator movement effectively reduces the available flow area. This behaviour is represented graphically by a plateau region in the performance curve, where the flow rate remains constant despite variations in pressure. After the plateau region, where the valve maintains nearly constant flow, the performance curve may exhibit a slight fall in the flow rate. This decline occurs due to nonlinear effects that influence the valve dynamics. Maximum flow capacity of 2FRM6 valve is 32 l/min, with the decline in flow rate being highest at this maximum and lower at reduced rates. The force balance can be disturbed by factors like friction in the compensator mechanism and the flow forces acting on the spool. It is important to note that when the inlet pressure is the same, this flow force is higher at increased flow rates and lower at reduced flow rates. These nonlinearities reduce the effective control of the compensator, causing a gradual decrease in flow rate even though the pressure differential remains high. The width of the plateau depends on flow rate and the extent of nonlinearities.

CFD simulations predict the decline in flow even without modeling effects such as friction, indicating that flow forces are the dominant source of nonlinearity in this case. Figure 4.2 illustrates the forces acting on the compensator at a flow rate of 32 l/min, along with the corresponding spring force. As the stroke increases (i.e., as the flow area decreases), the flow force rises sharply compared to the spring force. This rapid increase in flow force causes the compensator to move further than required for ideal pressure compensation, resulting in an excessive reduction in flow area. Consequently, the flow rate decreases at higher pressures.

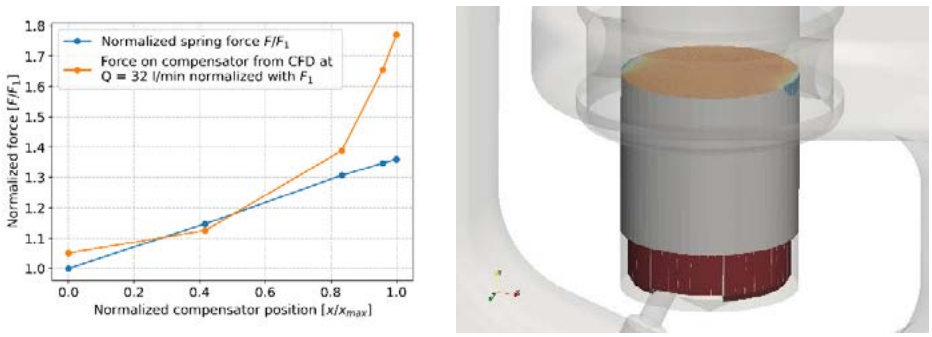


Figure 4.2: Flow force from CFD and pressure profile on compensator surface.

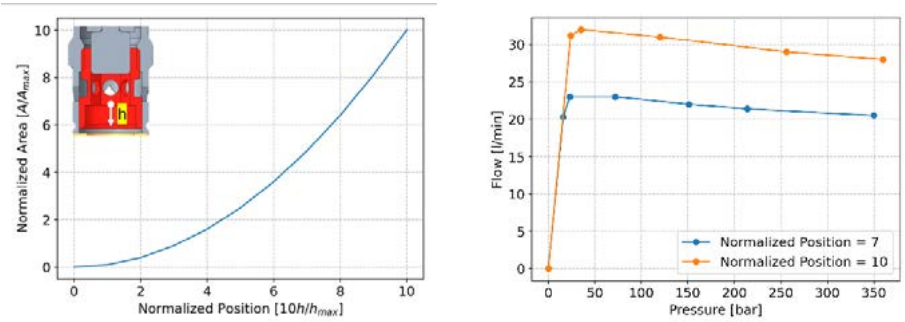
4.3. Influence of Design Parameters

The performance of a pressure-compensated valve is strongly affected by its design parameters. Key elements such as the main orifice, spring, compensator design determine how the valve responds to flow and pressure variations. Understanding the influence of these parameters is essential for optimizing valve behaviour, achieving the desired flow control,

and ensuring stable operation under different operating conditions. This section presents the influence of these design parameters on the valve performance.

4.3.1. Main Orifice Area

The main orifice (5) shown in Figure 2.1 plays a crucial role in the valve operation. While the orifice area determines the set flow rate through the valve, the orifice shape controls how this area changes with the movement of the throttle screw. Triangular-shaped orifices are commonly preferred to achieve a gradual increase in flow area. Figure 4.3 illustrates the variation of orifice area with different throttle screw positions and presents the characteristic curves of the valve for two set positions. It can be observed that changing the throttle screw position affects only the set flow, without altering the overall valve behaviour.

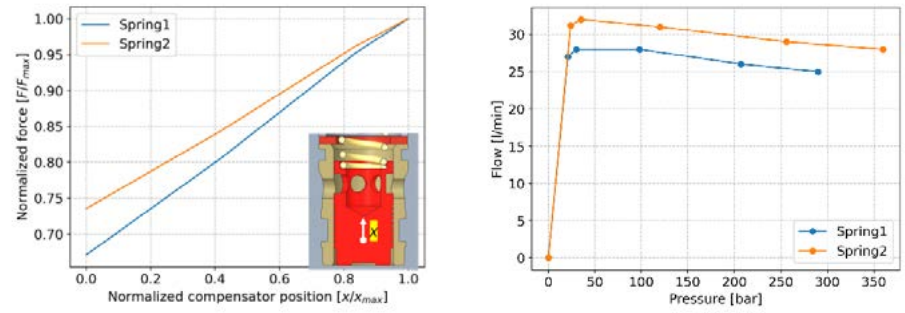


(a) (b)

Figure 4.3: (a) Variation of the main orifice area with the throttle screw position, (b) Performance curves corresponding to two different set positions of the throttle screw.

4.3.2. Influence of Spring

The spring in a pressure-compensated valve plays a crucial role in determining its operational characteristics. The regulation of flow rate under varying pressure conditions is primarily governed by the force balance between the spring and the compensator. The spring's initial compression force (F_1), its rate, and any nonlinearity directly influence the valve's response.

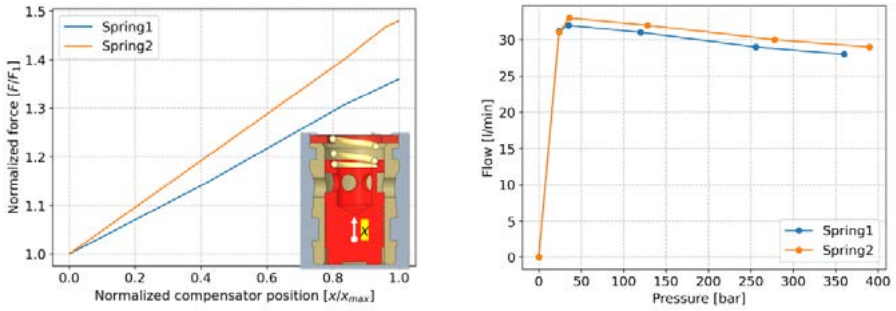


(a) (b)

Figure 4.4: (a) Normalized spring force with normalized compensator position, (b) Performance curves corresponding to two different springs.

Figure 4.4 illustrates the influence of the initial spring compression force on valve performance. Figure 4.4(a) shows the normalized spring force plotted against the compensator position. In this study, the initial compression of the same spring was increased while maintaining the same spring rate. As the spring rate remains unchanged, the force at $x/x_{max} = 1$ also increases, and this value is used for normalization. Two characteristic curves corresponding to different initial forces are presented. Increasing the initial compression force raises the valve's set pressure. As shown in the figure, with spring 1 the flow is set to 27 l/min at 21 bar, whereas with spring 2 it increases to 31 l/min at 24 bar.

Figure 4.5 shows influence of spring rate on performance curve. Figure 4.5(a) shows normalized force for spring1 and spring2 for different compensator position. Spring force is normalized by initial compression of the spring. It can be observed that with changing spring rate there is minor change in set flow, however, it does not influence the remaining characteristic of the flow.



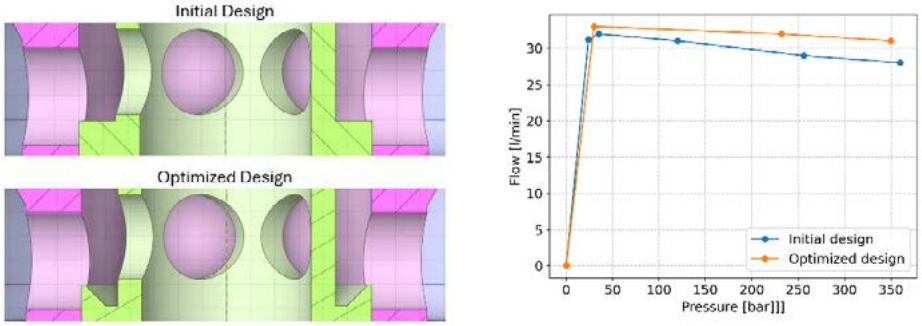
(a) (b)

Figure 4.5: (a) Normalized spring force with normalized compensator position, (b) Performance curves corresponding to two different springs.

The variation of flow rate with inlet pressure is governed by the force balance between the compensator and the spring. As discussed earlier (see Figure 4.2), the nonlinear nature of the flow-induced force acting on the compensator plays a decisive role in defining the valve's pressure–flow characteristics. To achieve a stable response, the spring must generate a corresponding nonlinear force. This can be realized through a conical spring with variable stiffness.

4.3.3. Compensator Design

The design of the compensator plays an important role in determining the overall performance of a pressure-compensated flow control valve, as the flow-induced forces act directly on its surface. By optimizing the compensator geometry—particularly the profile of the surface exposed to flow—the magnitude of these forces can be effectively reduced. Incorporating sharper edges can limit the area over which the pressure distribution acts, thereby decreasing the effect of flow forces on compensator. Such geometric refinements improve the valve's ability to maintain a consistent flow rate under varying pressure conditions.



(a) (b)

Figure 4.6: (a) Design changes of compensator, (b) Performance curves corresponding to two different compensator designs.

5. CONCLUSION

A comprehensive methodology has been developed for predicting the performance of a pressure-compensated flow control valve. The approach enables accurate estimation of pressure–flow characteristics and internal forces under different operating conditions. The predicted results have been validated against experimental measurements, showing good agreement and demonstrating the reliability of the developed simulation methodology. Using this validated model, the influence of key design parameters—such as the main orifice geometry, spring characteristics, and compensator land dimensions—has been systematically investigated. The study highlights how each parameter governs the valve's set pressure and compensating behaviour. The findings emphasize the importance of proper selection and optimization of these design parameters to achieve desired performance and ensure consistent pressure compensation across the operating range.

6. REFERENCES

[1] M. G. El-Din and M. Rabi, *Fluid power engineering*. McGraw-Hill Education, 2009.

[2] K. Suzuki and E. Urata, "Development of a water hydraulic pressure-compensated flow control valve," *Int. J. Fluid Power*, vol. 9, no. 3, pp. 25–33, 2008.

[3] D. Wu, R. Burton, G. Schoenau, and D. Bitner, "Analysis of a pressure-compensated flow control valve," 2007.

[4] J. Ruan, R. Burton, and P. Ukrainetz, "An investigation into the characteristics of a two dimensional '2D' flow control valve," *J Dyn Sys Meas Control*, vol. 124, no. 1, pp. 214–220, 2002.

[5] H. Handroos and M. Vilenius, "Flexible semi-empirical models for hydraulic flow control valves," 1991.

[6] H. Pettersson, P. Krus, A. Jansson, and J.-O. Palmberg, "The design of pressure compensators for load sensing hydraulic systems," in *UKACC International Conference on Control'96 (Conf. Publ. No. 427)*, 1996, vol. 2, pp. 1456–1461.

[7] E. Lisowski and G. Filo, "CFD analysis of the characteristics of a proportional flow control valve with an innovative opening shape," *Energy Convers. Manag.*, vol. 123, pp. 15–28, 2016, doi: https://doi.org/10.1016/j.enconman.2016.06.025.

[8] M. T. Muller and R. C. Fales, “Design and analysis of a two-stage poppet valve for flow control,” *Int. J. Fluid Power*, vol. 9, no. 1, pp. 17–26, 2008.

[9] U. Goldberg, O. Peroomian, and S. Chakravarthy, “A wall-distance-free k-ε model with enhanced near-wall treatment,” 1998.

[10] X. Pan, G. Wang, and Z. Lu, “Flow field simulation and a flow model of servo-valve spool valve orifice,” *Energy Convers. Manag.*, vol. 52, no. 10, pp. 3249–3256, 2011.

[11] G. D. Vescovo and A. Lippolis, “Three-Dimensional Analysis of Flow Forces on Directional Control Valves,” *Int. J. Fluid Power*, vol. 4, no. 2, pp. 15–24, 2003, doi: 10.1080/14399776.2003.10781162

Biographies



Harsh Maniar received the bachelor's degree in mechanical engineering from Gujarat Technological University in 2013, the master's degree in mechanical engineering from Indian Institute of Technology, Dhanbad in 2017, respectively. He is currently working as Computational Fluid Dynamics engineer at the Bosch Rexroth India. He has broad experience in macroscopic particle method, multiphase flow and transient simulation. His research interests include flow-induced forces and pressure drop analysis.



Ujas Patel received the bachelor's degree in mechanical engineering from Gujarat Technological University in 2015, the master's degree in mechanical engineering from Indian Institute of Technology, Madras in 2019, respectively. He is currently working as Computational Fluid Dynamics engineer at the Bosch Rexroth India. He has broad experience in fluid systems, encompassing gas turbine combustors, turbomachinery, hydraulic systems, and water transport and treatment. His work focuses on the study and optimization of complex flow behaviour



Nirav Patel received the bachelor's degree in mechanical engineering from Gujarat Technological University in 2015, the master's degree in mechanical engineering from Pandit Deendayal Energy University in 2018, respectively. He is currently working design engineer at the Bosch Rexroth India. He has broad experience in screw compressor design for gas application, pneumatic and hydraulic systems. His focus study on optimisation of fatigue strength, pressure drop and induced flow forces.

Fluid Solutions Innovative and sustainable hydraulic fluid concepts

Wolfgang Bock, Romina Ambrosi, Gerit Dornacher

Fuchs Lubricants Germany GmbH

Abstract.

Hydraulic fluids are an important design element in stationary and mobile hydraulic equipment. Based on different sustainable base oil concept, FUCHS has developed new hydraulic fluid concepts which in addition to offering "hydraulic performance" have a low carbon footprint value. The product carbon footprint (PCF) of the relevant ACT (Advanced Circular Technology) fluids is significantly lower than vigor oils. Additionally, the C02 footprint of different fluid concepts during the application phase of hydraulic fluids will be discussed, based on a life cycle assessment (LCA) study. This presentation will provide an overview of the latest developments in hydraulic fluid technology. Energy-saving potential, lifetime aspects and performance levels will also be discussed.

1. INTRODUCTION

Hydraulic fluids are a significant group of lubricants in Germany. The fluid group includes in general mineral oil based hydraulic fluids, fire resistant, biodegradable, food-grade and multigrade hydraulic fluids as well as other specialties. The diagram shows the total market development of hydraulic fluids in Germany. Mineral based hydraulic fluids can be classified according to API base oil groups: Group 1, Group II and Group III. Conventional hydraulic fluids are based on Group I base oils, also known as solvent neutral (SN) base oils. New innovative hydraulic fluids are based on Group II/Group III, hydrotreated or hydrocracked which have a lot of technical and sustainable advantages (high lifetime, no aromatics, low sulphur content, etc.).

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

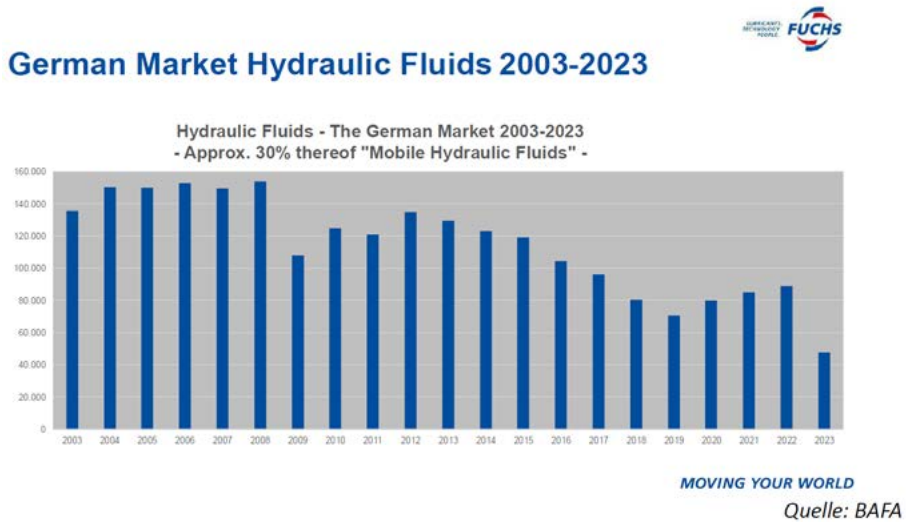


Figure 1.1. German Market Hydraulic Fluids 2003-2023

2. NEW INNOVATIVE MINERAL OIL BASED HYDRAULIC FLUIDS

Today, more than approximately 30% of mineral oil based hydraulic fluids are based on hydraulic fluids made from hydrotreated, hydrocracked Group II/ Group Ill. These high-performance fluids have a very long service life (2-3 times longer than comparable conventional fluids based on Group I), a low sulphur content, no aromatics, and a slightly higher viscosity index.

They also have low/ no varnish potential due to their extremely high oxidation stability. The hydrotreated fluids of the new generation of base oils are also increasingly important worldwide for industrial lubricants (e.g. Asia, America, Europe).

Diagram 2.1. shows the extended service life of these high-performance fluids.

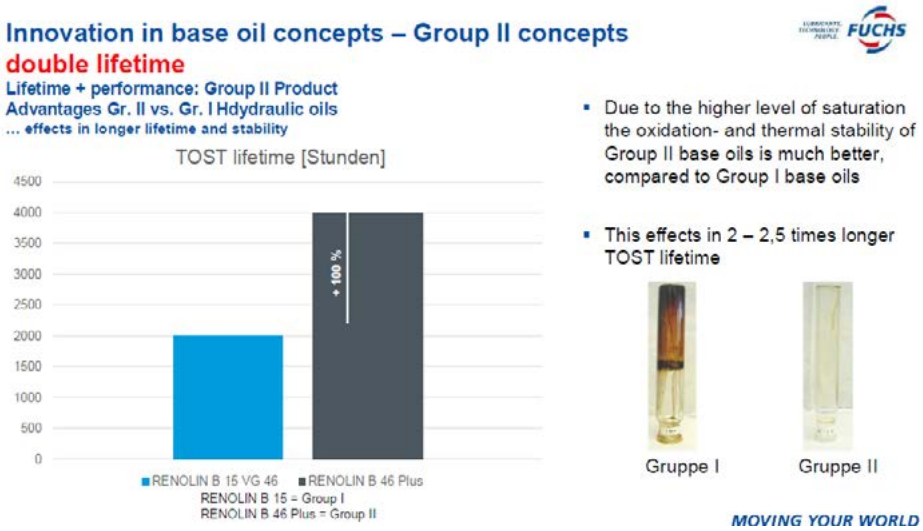


Diagram 2.1. Innovation in base oil concepts – Group II concepts

3. SUSTAINABILITY ASPECTS PCF VALUES, TRENDS

In general, sustainability, energy-saving and additive labeling aspects must be considered when formulating hydraulic fluids.

In addition, new fluid types must also meet economic criteria.

An important sustainability value is the Product Carbon Footprint (PCF) of hydraulic fluids, the PCF value. FUCHS evaluates the PCF value of the products within the cradle-to-gate boundaries.

Through the targeted use of sustainable base oils (e.g. re-refined base oils: Group I, Group II and Group III), the PCF value of sustainable products can be specifically reduced significantly (PCF reduction of up to 70%).

Diagram 3.1. shows different PCF values of virgin and sustainable base oil based hydraulic fluids.

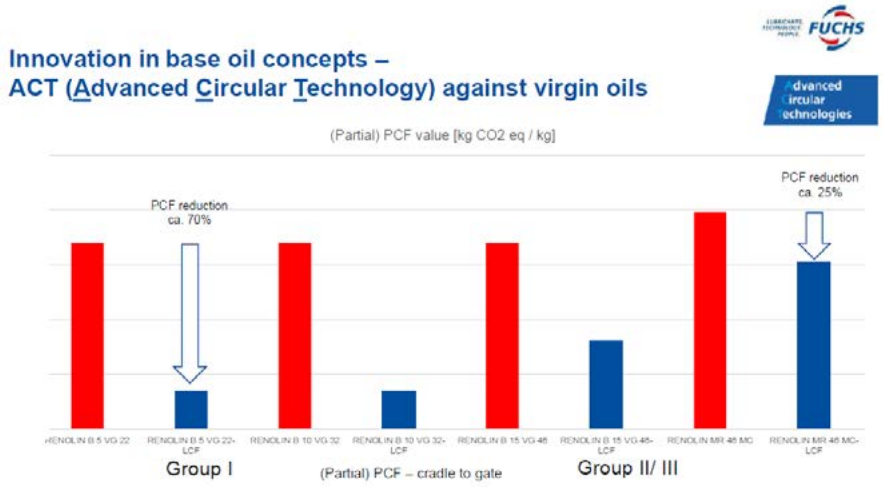


Diagram 3.1. Innovation in base oil concepts – ACT against virgin oils

4. SUSTAINABILITY PERFORMANCE PARAMETER, LIFETIME, ENERGY SAVING, LCA-STUDY, CUSTOMER STUDY

The sustainability of fluids can be assessed from various perspectives. The specific PCF of a product/ (per year of use) fluid, the overall lifespan, the use of re-refined base oi ls (RRBO), the use of renewable raw materials (for bio hydraulic fluids with the EU eco-label, > 25%), and additive systems that are not labelled acc. To CLP regulation.

In addition, life cycle assessment studies (LCA) - (comparison of mono- to multigrade hydraulic fluids) can provide information on the CO2 footprint of the products used during the application phase.

Figures 4.1. to 4.3. show the different aspects of sustainability in hydraulic fluids and the results of a LCA study.

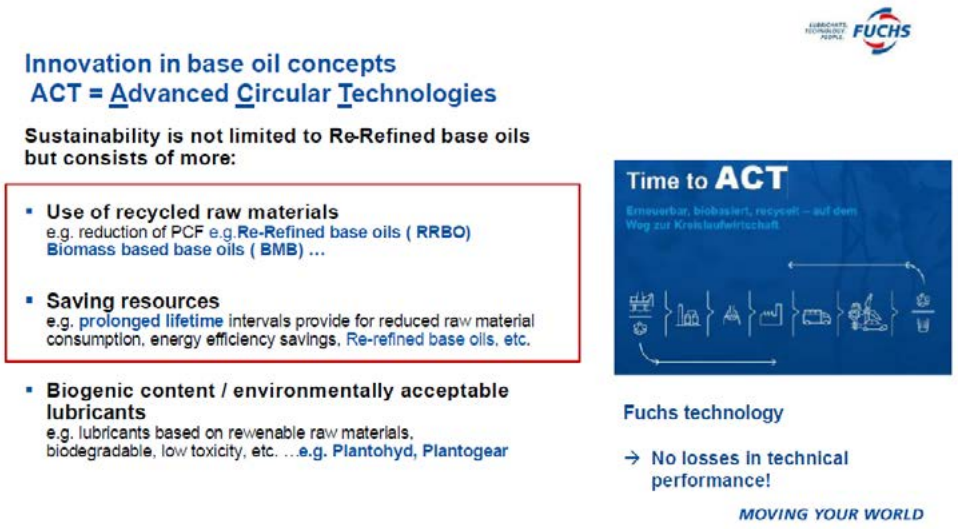


Figure 4.1. Innovation in base oil concepts

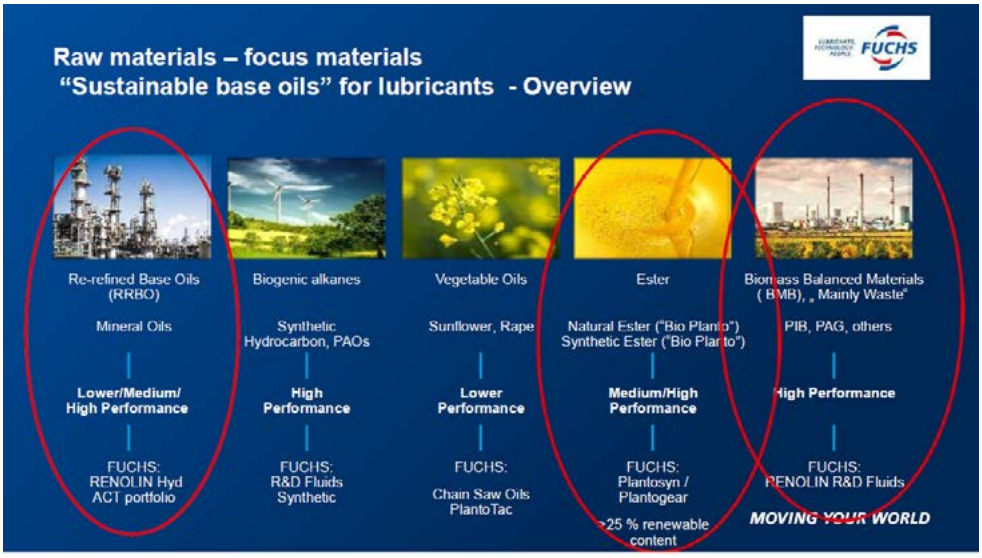


Figure 4.2. Raw materials – focus materials

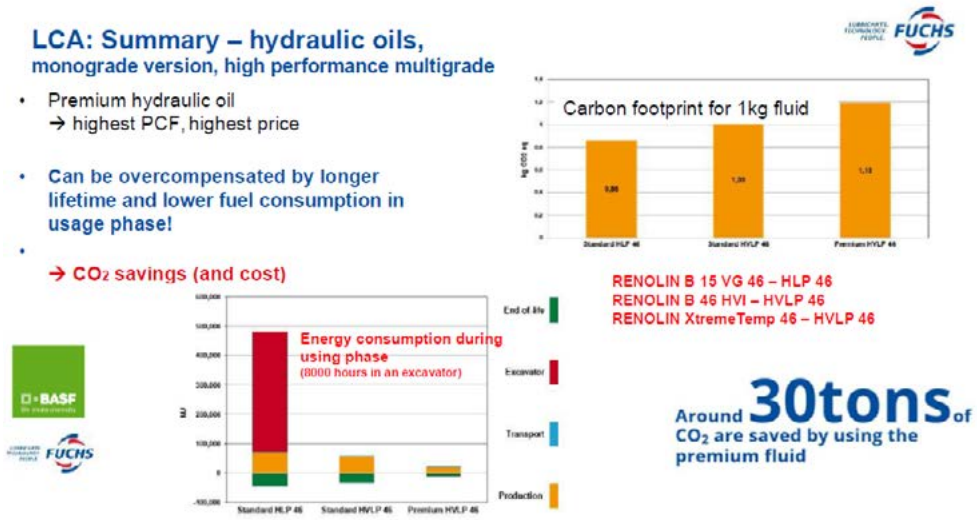


Figure 4.3. LCA: Summary – hydraulic oils, monograde version, high performance multigrade

5. CONCLUSION

Sustainability aspects should be considered when selecting hydraulic fluids. The product carbon footprint value (cradle to gate) is a meaningful parameter for assessing the carbon footprint of a product.

Other aspects such as energy efficiency, flu id lifetime, LCA considerations, and the labeling of additives must be considered and will become very important for products and users of fluids in the future.

Biographies



Wolfgang Bock is a Mechanical Engineer (Dipl.-Ing.) who graduated from the University of Applied Sciences Mannheim. After gaining experience in various professional roles, he joined Fuchs in May 1991. Since then, he has held several positions within the Fuchs Group, including Application Engineer, Product Manager, and Project Management Lead (PML). Over the years, he has worked for Fuchs Mineralöle, Fuchs DEA, Fuchs Petrolub, Fuchs Europe, Fuchs Lubricants, and Fuchs SE. Within PM4, his main areas of responsibility include a broad product portfolio with a focus on GPM-Senior topics, refrigerants, and turbine applications. He is an active member of several professional associations, including DIN, FVA, VDMA, FKT, and VGB.

Rating Hydraulic Fluids for Performance and Sustainability: Bosch Rexroth's Approach

Klaus Ellenrieder[1]*, Karl-Heinz Blum[2]

[1]Bosch Rexroth AG, Glockeraustraße 2, 89275 Elchingen, Germany
[2]Bosch Rexroth AG, An den Kelterwiesen 14, 72160 Horb, Germany

**Corresponding author: klaus.ellenrieder@boschrexroth.de*

Abstract

The increasing power density of hydraulic systems and growing demands for environmental compatibility, resource conservation, and energy efficiency are putting traditional mineral oil-based hydraulic fluids under increasing pressure. This encourages the development of new fluid types such as zinc- and ash-free formulations, biodegradable and bio-based hydraulic fluids, and re-refined base oils (RRBO). In view of this development, Bosch Rexroth has expanded its "Fluid Rating 90235" [1], established in 2015, which originally focused on wear behaviour and component lifespan. Since 2026, the rating includes a practical oil aging test for precise evaluation of fluid lifetime. In the future, the Fluid Rating will be supplemented by a comprehensive sustainability block, integrating criteria such as energy efficiency, environmental compatibility, and aspects of circular economy, including life cycle assessments (LCA) and product carbon footprints (PCF). This expanded assessment approach enables a standardized and transparent evaluation of hydraulic fluids, considering modern technical and regulatory requirements, and supports the transformation towards more sustainable fluid solutions.

Keywords. Fluid Rating, sustainability, energy efficiency, oil aging, performance, regulatory requirements

1. INTRODUCTION

In recent years, the performance of hydraulic systems has once again significantly increased. Consequently, the demands on the hydraulic fluids used are also rising steadily, as they can, for example, decisively influence the wear properties of hydraulic components. Relevant standards only define minimum requirements for hydraulic fluids. They do not reflect technical advancements and do not allow for differentiation between fluids. Bosch Rexroth offers lubricant and additive manufacturers a standardized evaluation procedure based on scientific findings, which takes these aspects into account and tests hydraulic fluids for suitability under realistic operating conditions.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non-Commercial 4.0 International

The increasing power density, growing demands for environmental compatibility, resource conservation, energy efficiency, occupational safety, and rising regulatory requirements from countries and authorities pose new challenges for the development of hydraulic fluids. Traditional mineral oil-based fluids, which currently make up the majority of the market, are increasingly under pressure as the mentioned aspects become crucial criteria (Figure 1).

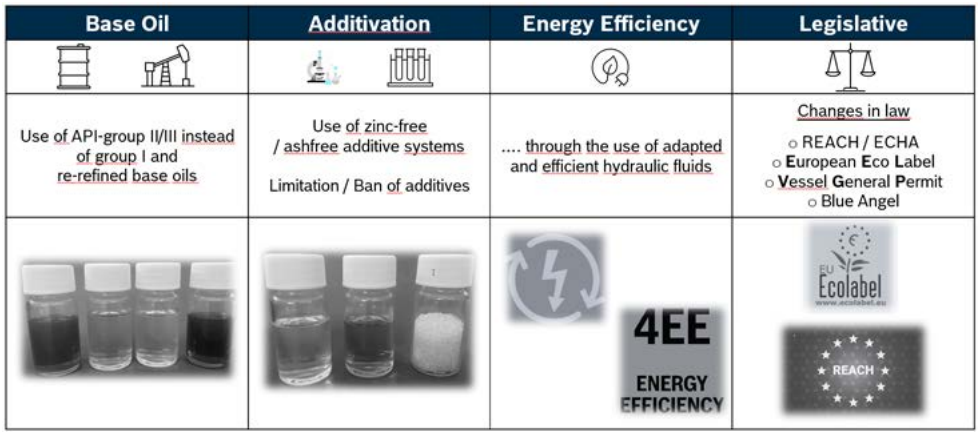


Figure 1: Challenges for future high performance hydraulic fluids

This development promotes the emergence and use of new fluid types, including zinc- and ash-free formulations, biodegradable hydraulic fluids, bio-based products, re-refined base oils (RRBO), and fire-resistant fluids. This change is no longer driven solely by environmental awareness but significantly by stricter regulations, the pursuit of carbon neutrality, and technical innovations. For example, biodegradable lubricants are increasingly used in environmentally sensitive areas such as transport, marine, food processing, and forestry and agriculture.

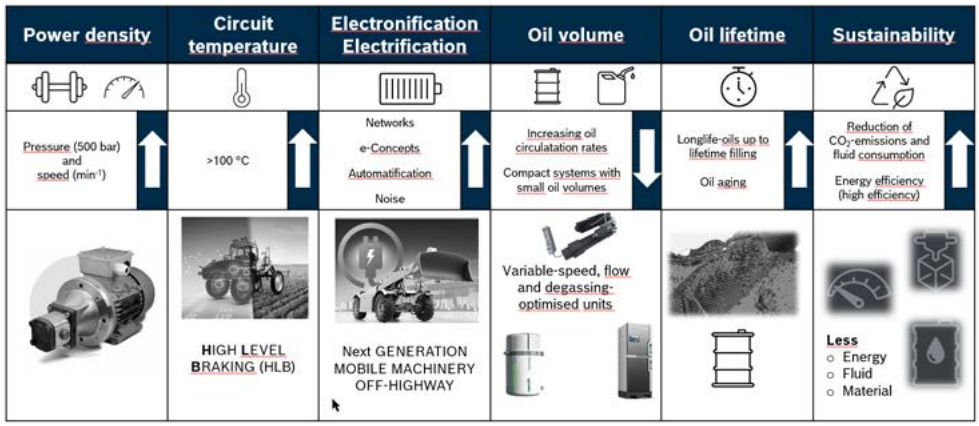


Figure 2: Changes and trends in hydraulic systems and components

At the same time, hydraulic applications are also being further developed through electronification and electrification. The replacement of internal combustion engine-driven systems by electric drives, power-on-demand systems with electrically controlled pumps,

and decentralized hydraulic systems with small oil volumes and high-power density require correspondingly adapted fluid properties (Figure 2).

2. STATE OF THE ART AND TRENDS

Hydraulic fluids based on mineral oils are the most used group, accounting for approximately 85%. About 15% of the market share is accounted for by synthetic base oils, such as esters, polyalphaolefins (PAO), and polyalkylene glycols (PAG). Among mineral oil-based base oils, the trend is moving away from paraffinic API (American Petroleum Institute) Group I towards hydrotreated paraffins of API Group II (hydrotreated) and Group III (hydrocracked) (Figure 3).

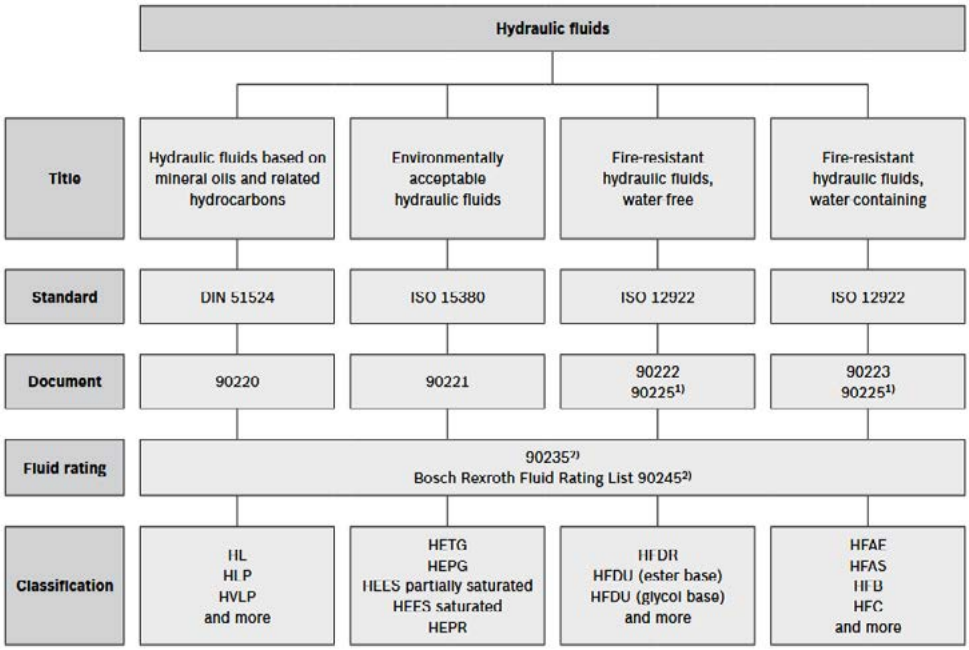


Figure 3: Overview of classification and relevant standards for hydraulic fluids [1]

To address this development, Bosch Rexroth has already established a standardized rating procedure for hydraulic fluids – the Bosch Rexroth Fluid Rating 90235 (Figure 4) [1]. Since then, this has become a recognized industry standard and is used by many fluid and additive manufacturers for the development of fluid formulations and hydraulic fluids. The primary focus has been and continues to be mainly on the wear behaviour and lifetime of hydraulic components.

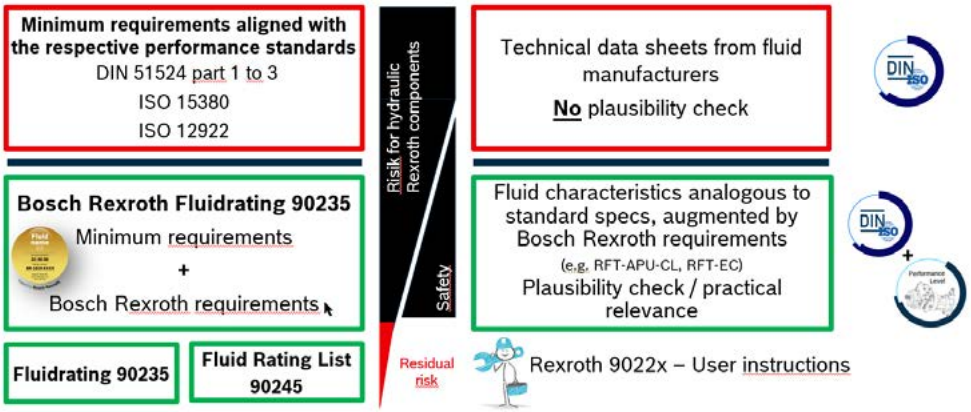


Figure 4: Bosch Rexroth Fluidrating: actual structure

The central part of the Fluid Rating, in addition to stricter requirements based on relevant performance standards and a seal and oil aging test, is a comprehensive fluid test based on the use of hydraulic components such as swash plate axial piston pumps and bent axis axial piston motors (Figure 5).

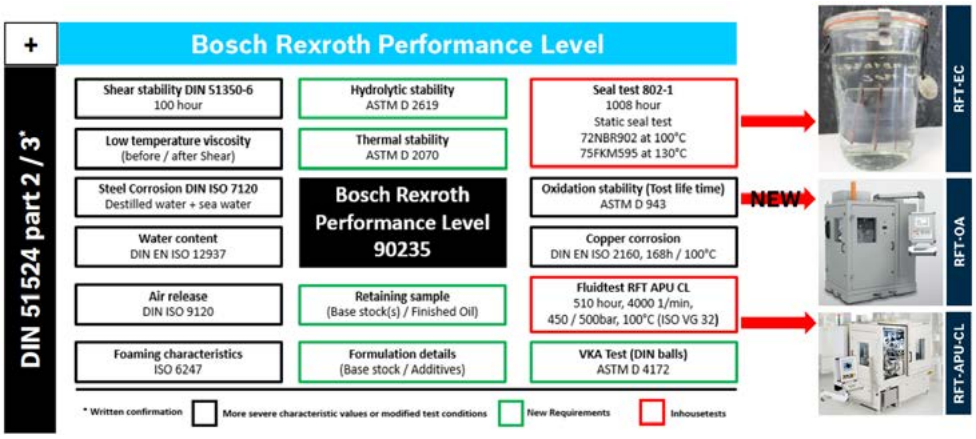
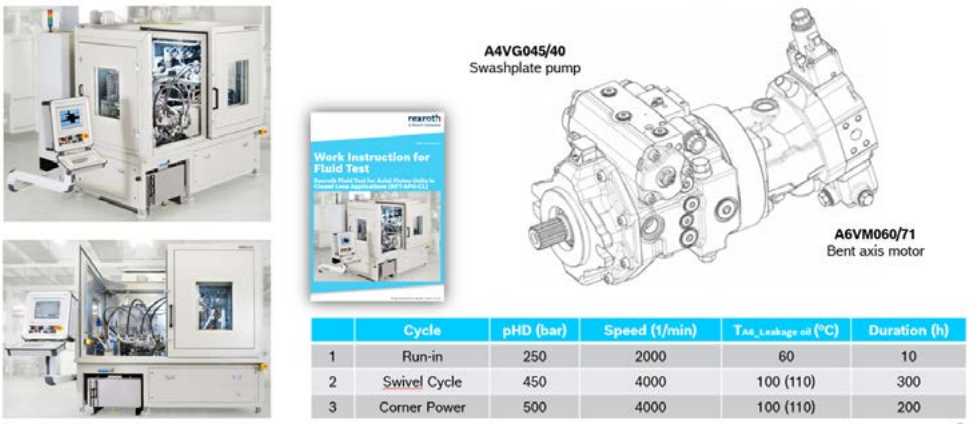


Figure 5: Bosch Rexroth Performance Level for mineral oils H(V)LP

These units simulate a typical hydrostatic drive in a closed circuit and, due to their high-performance parameters (pressure, speed, temperature), also cover the requirements of units in open circuits regarding wear behaviour and material compatibility. The drive components used are subject to strict quality control, are precisely measured and documented to minimize influences from test specimens. Such a fluid test lasts a total of 510 hours and includes phases such as run-in, swivel cycle, and corner power operation (Figure 6). After the test run, a detailed evaluation of the relevant components is carried out.

To reduce resources, the test units for the fluid tests are not completely replaced by new ones but reconditioned, and only relevant components, such as the rotary group and seals, are replaced – already a corresponding contribution to the sustainability claim.



	Cycle	pHD (bar)	Speed (1/min)	$T_{A6_Leakage\ oil}$ (°C)	Duration (h)
1	Run-in	250	2000	60	10
2	Swivel Cycle	450	4000	100 (110)	300
3	Corner Power	500	4000	100 (110)	200

Figure 6: Bosch Rexroth Fluid Test RFT-APU-CL (Rexroth Fluid Test Axial Piston Unit Closed Loop)

The Fluid Rating currently covers the different fluid categories: mineral oils, environmentally acceptable hydraulic fluids, UTTO oils, and hydraulic fluids based on re-refined base oils (RRBO).

3. OBJECTIVE AND APPROACH

Given the necessity to transparently rate new fluids (e.g., RRBO, biodegradable fluids, fire-resistant fluids) and the associated requirements regarding environmental compatibility, efficiency, and resource conservation, an expansion of the rating to include a comprehensive sustainability block is sought.Criteria such as fluid lifetime, energy efficiency, environmental compatibility in the context of the entire lifespan, and the CO2 footprint are to be integrated into the Fluid Rating in the future (Figure 7).

A first step for this is the recently implemented inclusion of an oil aging test, to assess the lifetime of hydraulic fluids more practically than the TOST TEST (Total Oxidation Stability Test) currently required in standardization. This will be described in more detail in Section 4.1.

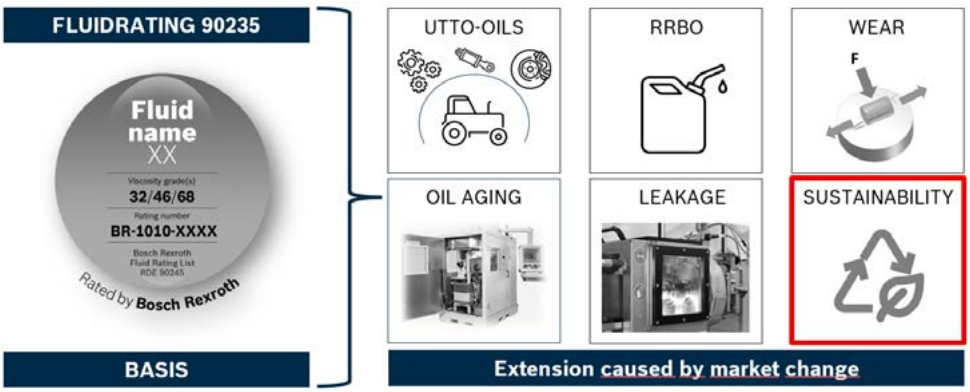


Figure 7: Extension of the fluid rating due to market change

The new requirements, in addition to the lifetime of hydraulic components which serves as the basis, are now to be integrated into the Fluid Rating as follows:
The primary objective is the development of a standardized and transparent evaluation procedure for additional sustainability aspects of hydraulic fluids. This essentially includes the following criteria (Figure 8):

- **Fluid Lifetime:** (Expanded in 2026 by an oil aging test)
- **Energy Efficiency:** Evaluation of the hydraulic fluid's influence on the system's energy consumption.
- **Environmental Compatibility:** Evaluation of ecological impacts, including biodegradability, toxicity, and water hazard class.
- **Circular Economy** (Life Cycle Assessment, LCA) **and CO2 Footprint** (Product Carbon Footprint, PCF): Evaluation of environmental impacts throughout the fluid's entire life cycle, from raw material extraction to disposal/recycling.

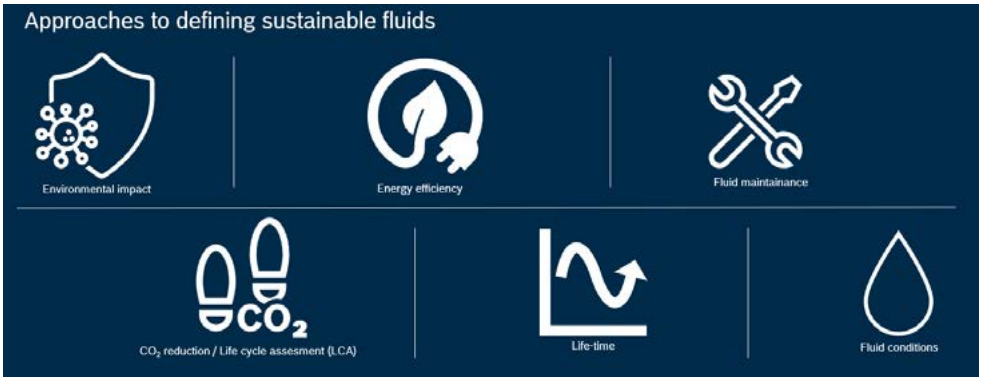


Figure 8: Sustainability aspects for hydraulic fluids

The approach envisages offering this expanded evaluation block successively in several expansion stages as a further module to the existing Fluid Rating. This requires the development of new test methods, the identification of relevant fluid parameters, and the creation of a transparent rating system that enables the comparison and classification of fluids regarding their sustainability performance (Figure 9).

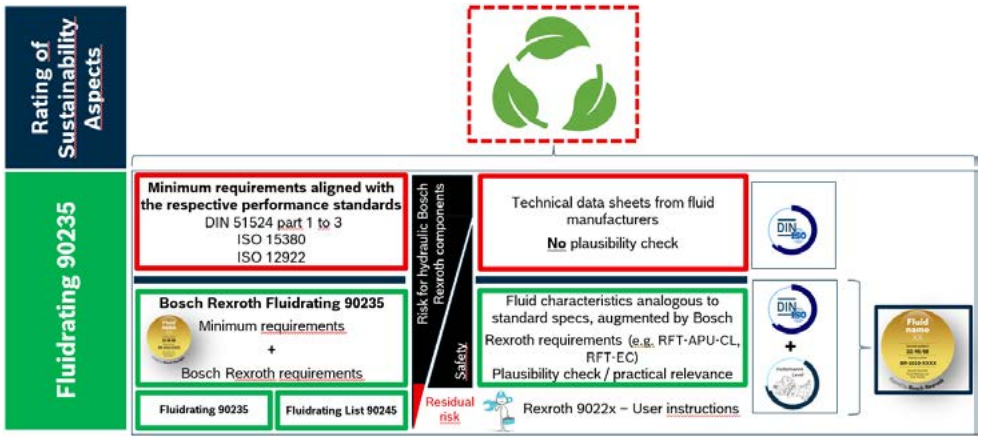


Figure 9: Sustainability aspects as a further rating block on top of the existing Bosch Rexroth Fluidrating

4. CURRENT STATUS AND ACHIEVED RESULTS

The expansion of the Fluid Rating is being carried out through targeted research and development efforts in the following areas:

4.1. Lifetime of Hydraulic Fluids

The determination and evaluation of the hydraulic fluid's lifetime are carried out using a specially developed oil aging test rig.

This fully automated test rig subjects the fluid to accelerated aging for over 300 hours at 90 °C with the addition of a copper catalyst and air. Automatically taken oil samples are regularly analyzed to comprehensively record and evaluate the aging behaviour (Figure 10). This test closes the gap for a practically relevant aging assessment that was not covered by the conventional TOST (Total Oxidation Stability Test).



Figure 10: Bosch Rexroth Oil Aging Test [2]

In preparation for inclusion in the Fluid Rating, tests were carried out with a reference oil (mineral oil HLP containing zinc), and the aging behaviour was determined over time based on NMR analyses (nuclear magnetic resonance analysis) and FTIR spectroscopy (Fourier-Transform Infrared Spectroscopy) and documented under different test parameters (speed, pressure, temperature).

4.2. Energy Efficiency (Development of a Systematic Approach)

The influence of hydraulic fluids on the energy efficiency of hydraulic systems is often underestimated. Currently, a uniform and standardized rating procedure is lacking. Bosch Rexroth plans, in cooperation with industrial partners and institutes, to identify relevant fluid parameters that influence the efficiency of various hydraulic components. Based on experimental data, a systematic approach will be developed and integrated into the Fluid Rating. This will enable a comparison and evaluation of hydraulic fluids regarding their contribution to energy efficiency in different applications. In the long term, this could include the development of a dedicated energy efficiency test rig and a standardized evaluation procedure to validly assess energy efficiency in various applications.

4.3. Environmental Compatibility (Integration of Labels and Criteria)

The environmental compatibility of hydraulic fluids is determined by their chemical composition, their behaviour in case of leaks, and their disposal. The assessment includes biodegradability, toxicity, and water hazard class. Transparent representation is achieved by integrating existing environmental labels and certifications such as the Blue Angel, VGP (Vessel General Permit), or the EU Ecolabel. These labels certify compliance with relevant environmental requirements. Additionally, ISO (International Organization for Standardization) certifications and manufacturer specifications, as well as fluid monitoring concepts, can be used to minimize environmental impact.

4.4. Circular Economy and CO2 Footprint (LCA/PCF Assessment)

For the evaluation of circular economy, the analysis includes the proportion of recycled or bio-based raw materials, the avoidance of scarce or poorly degradable raw materials, environmental impacts during raw material extraction and processing, reprocessability and recyclability, as well as corresponding service concepts and condition monitoring. The assessment of the CO2 footprint considers emissions from raw material extraction and production, behaviour during the use phase (energy efficiency, oil lifetime, leaks), and end-of-life management (emissions and energy consumption during transport, reprocessing, and disposal). For a neutral and comparable assessment, life cycle assessments (LCA) according to recognized standards are used, possibly in cooperation with independent institutions.

5. SUMMARY AND OUTLOOK

The increasing power density in hydraulic applications and global sustainability goals necessitate a fundamental adaptation of fluid formulations. The Bosch Rexroth Fluid Rating, a proven tool for rating the performance of hydraulic fluids, is being consistently expanded to meet these new requirements. By integrating aspects such as energy efficiency, oil aging, environmental compatibility, and holistic life cycle analyses (LCA/PCF), we are creating a comprehensive rating platform.

Achieved Results:

- Component lifetime is already established.
- The existing Fluid Rating has been supplemented since the beginning of the year by a practical oil aging test, which precisely evaluates the lifetime of fluids.
- The foundations for integrating energy efficiency parameters have been laid to make the influence of fluids on system efficiency measurable.
- A framework for evaluating environmental compatibility based on existing labels and other criteria is being established.
- The need for a comprehensive LCA/PCF assessment has been recognized, and the steps for implementation have been defined.

Outlook:
Bosch Rexroth will continue to actively support the transformation of the fluid industry. In the next steps, the detailed methodology for evaluating energy efficiency, environmental compatibility, and LCA/PCF aspects will be finalized and integrated into the Fluid Rating. This will be carried out gradually in several expansion stages.
Furthermore, we see potential for more intensive cooperation with the fluid industry to jointly develop sustainable hydraulic fluids that meet both the highest performance requirements and strict sustainability criteria. The possibility of offering fluid monitoring concepts in conjunction with the expanded rating criteria will also be evaluated. The goal is to provide users with transparent and reliable information for selecting the optimal fluid for their specific needs while promoting the sustainable development of hydraulic technology.

6. REFERENCES

[1] Bosch Rexroth data sheet ‘Rating of hydraulic fluids for Rexroth hydraulic components (pumps and motors)’, RE 90235/02.2015.
[2] Bosch Rexroth data sheet ‘Rexroth Oil aging test rig RFT-OA’, RE 98145/2025-06-11.

Biographies



Klaus Ellenrieder[1] received the Diplom Ingenieur degree in mechanical engineering 1995 at the Technical University in Munich and the MPhil degree 1996 at the Silsoe College / Cranfield University in UK. He was responsible for the testing department at Bosch Rexroth in Elchingen, Gemany, until end of 2023 and is currently working as the head of the Business Innovation Team Fluids at Bosch Rexroth.



Karl-Heinz Blum[2]**,** Mechanical engineer and business economist. Responsible for hydraulic fluid-related topics at Bosch Rexroth as well as for the Fluidrating 90235 and sustainable fluids. Located in Horb am Neckar, Germany. Currently working as chief expert fluids of the Business Innovation Team Fluids at Bosch Rexroth.

Influence of Fluid Mechanical Properties of Hydraulic Oils with Different Chemical Structures on the Efficiency of Hydraulic System Components

Sebastian Deuster, Marius Hofmeister, Faras Brumand-Poor, Katharina Schmitz

RWTH Aachen University - Institute for Fluid Power Drives and Systems (ifas), sebastian.deuster@ifas.rwth-aachen.de

Abstract.

This paper aims to describe the influence of synthetic fluids from Group 5 on various system boundaries and components of hydraulic systems, and to compare them with conventional hydraulic mineral oil. The focus is on evaluating the fluid mechanical properties. For this purpose, losses in pumps and pipes, as well as in valves, are analysed. Initially, these components are investigated and modelled mathematically to obtain a fundamental understanding. Experimental tests were conducted on a real hydraulic system of a mobile working machine using two different types of hydraulic fluids. These fluids include a mineral oil-based HLP (Group 2), a synthetic ester HEES (Group 5). The fluids are compared based on system boundaries losses in hydraulic systems components. The influence of their different chemical and physical properties is discussed.
The paper concludes with an analysis of the influence of the different losses when using alternative hydraulic fluids compared to mineral oils in hydraulic systems and potential future research activities in this field.

Keywords. Alternative hydraulic fluids, loss mechanism, energy efficiency, mobile machinery.

1. INTRODUCTION

In addition to the mechanical components of a hydraulic system, the hydraulic fluid is a key component that plays a crucial role in the efficiency and operating behavior of hydraulic machines. Hydraulic fluids consist mainly of a base oil and a small proportion of additives. Different base oils can result in varying chemical and physical properties. Base oils are classified according to the American Petroleum Institute (API). The API classifies base oils into five groups. Groups 1 to 3 include mineral oils of varying purity. Groups 4 and 5 comprise fully synthetic base oils. Most lubricants are based on mineral oils from Groups 1 to 3. Their market share is approximately 85%, with synthetic base oils accounting for 15 % [1]. Synthetic base oils, such as esters or polyalphaolefins, offer, for example, the advantage

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

of adapting physical properties, including viscosity-temperature behavior, to the respective conditions of the hydraulic system during operation. Due to higher acquisition costs and the lack of suitable application tools for designing hydraulic systems tailored to a specific system, market sales are low compared to conventional mineral oils.

However, the use of hydraulic fluids based on API group 4 to 5 base oils has increased significantly in recent years [2]. These synthetic base oils provide physical and chemical properties that differ significantly, in some cases, from those of conventional mineral oils (Group 2). Using these oils in hydraulic systems results in distinct fluid mechanical and thermodynamic properties during their operation.

Previous research focused on the performance characteristics of synthetic hydraulic fluids in hydraulic displacement units [3, 4], which showed that the use of synthetic hydraulic fluids results in different efficiency levels than those achieved with conventional mineral oils. The tribological behaviour of synthetic esters was investigated in greater depth in comparison with mineral oils by [5, 6]. It was found that friction losses and wear are higher when synthetic esters are used. Jakobs [7] demonstrated in his work that the use of synthetic esters results in differentiated switching behaviour compared to mineral oils.

Previous work by us [8] showed high deviations in the overall efficiency of a mobile machine when using water-based hydraulic fluids instead of mineral oils.

This paper aims to expand on this work by identifying the causes of efficiency differences and balancing energy losses in the components of the hydraulic system. Unlike previous work, the focus here is on comparing mineral oils and synthetic esters.

The work is structured as follows: First, the fluid properties of the hydraulic fluid under investigation, mineral oil and synthetic esters, are presented in a comparative manner. This is followed by a presentation of the mathematical principles relating to the loss mechanisms that occur during the operation of a mobile machine. Particular attention is paid to the influence of different physical properties between mineral oils and synthetic esters. This is followed by validation of the mathematical assumptions using experimental data from a test excavator. The system is divided into different boundaries. Furthermore, the total losses during a dig and dump cycle of the machine, operating with different hydraulic fluids, are presented. Finally, a summary is provided in which the question should be answered: Which losses in the system are responsible for the different energy requirements associated with the use of different fluids in a mobile working machine?

2. FLUID PROPERTIES

The use of alternative lubricants results in variations in physical properties, which affect their operating behaviour when used in components and machines. Compared to conventional hydraulic fluids based on mineral oils, it is primarily the chemical structure of the base fluids that is responsible for differences in physical properties such as viscosity behaviour or density. Additives for hydraulic applications are used in nearly all base oil groups, featuring similar chemical components. These correlations can only be used qualitatively in the development of lubricants in the form of predictions and cannot be applied in a targeted manner by now.

The hydraulic fluids considered for the experimental investigations in this work differ fundamentally in their base fluid structure. This chapter examines the chemical structure of various base oils and explains the resulting physical properties.

2.1. Rheology and chemical structure

Viscosity generally describes the thickness of a fluid. It is used as a measure of a liquid's internal friction. This internal friction arises between the molecules of a flowing medium [9]. Physically, viscosity can be derived using the relationships shown in **Figure 2.1**. A fluid element located between a wall and a plate parallel to it is displaced by a force F applied on the plate. This creates a velocity gradient $d\dot{x}$ to the fluid between the moving plate and the stationary wall. The resulting shear stress τ depends on the area A of the moving plate.

The dynamic viscosity η is defined for Newtonian fluids as the proportionality constant between the shear rate $\dot{\gamma}$ and the shear stress τ acting on the fluid.

Figure 2.1 b) shows the atomic relationships for the graphical representation of the viscosity of lubricants in general. Molecular chains of hydrocarbons slide over each other due to an initiating force. The positively charged nuclei of the molecules repel each other due to their identical charge. For molecules to move from a given location to a vacant space, they must overcome a potential barrier. This requires activation energy. The potential barrier described defines the viscosity of the fluid. Additionally, mechanical resistance is caused by the branching of molecular chains and by branching structures within the chains. [10]

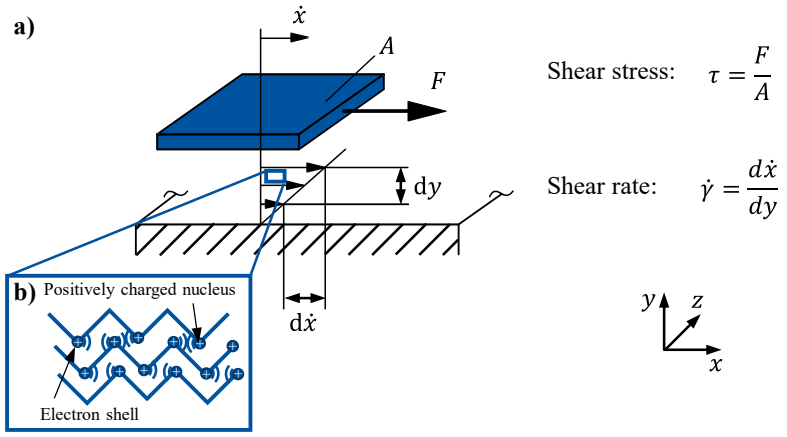


Figure 2.1. (a) Derivation of dynamic viscosity and (b) Intermolecular interactions between hydrocarbon chains in lubricants to illustrate viscosity at the atomic level

The base oils of the hydraulic fluids considered in this paper differ in their chemical structure. The differences in structure between mineral oils and esters primarily lie in the extent of branching and side chains. The chain length of the hydrocarbon chains, on the other hand, is similar for both types. Esters usually have long linear aliphatic molecular chains, whereas mineral oils have inflexible structural elements in the form of side chains. This difference in chemical structure influences the viscosity behavior of hydraulic fluids depending on temperature and pressure. Furthermore, this means that the carbon chains in ester oils are located nearer to each other, resulting in a higher packing capacity and thus a

higher density of ester oils compared to mineral oils. These relationships will be examined in more detail in the following chapters.

2.2. Temperature-viscosity behaviour

The viscosity of hydraulic fluids is highly dependent on temperature. As the temperature rises, the viscosity of the fluid decreases. At the microscopic level, this behavior can be explained as follows: Temperature affects the vibration and changing speeds of the atoms and molecules in the lubricant. The resulting vibration-energy reduces the activation energy required to overcome the potential barrier (see Figure 2.1). Corresponding mathematical descriptions of this behavior take the form of the Arrhenius equation [10]. Viscosity-temperature behavior depends on the chemical structure of the base fluid. Flexible, long linear molecular chains, such as esters, exhibit the lowest temperature dependence of viscosity. Inflexible structural elements, such as branches or sidechains, like in mineral oils, increase the dependence of viscosity on temperature accordingly [11]. Branching causes the hydrocarbon chains to separate further apart spatially, thereby reducing their influence compared to closely linked chains. The degree of dependence of viscosity on temperature is described using the viscosity index. As the viscosity index (VI) increases, the dependence of viscosity on temperature decreases. The calculation basis for determining the VI is described in DIN ISO 2909 [12]. While the VI of mineral oils is typically around 100, ester-based oils achieve values exceeding 180 [13].

2.3. Pressure-viscosity behaviour

The pressure-viscosity behavior is a fluid-specific characteristic that depends mainly on the composition of the base fluid. The addition of additives has a negligible influence. The pressure-viscosity behavior generally increases with the degree of branching, the number of cyclic structures, the presence of aromatic compounds, and the length of the side chains of the base oil (see **Fehler! Verweisquelle konnte nicht gefunden werden.**). Both naphthenic and paraffinic-based mineral oils have pronounced branched chain structures. These structures intertwine under increasing pressure, which in turn increases the internal friction within the fluid. Ester molecules have a significantly flatter structure with few side structures. This results in a lower pressure dependence on the viscosity. [14]

There are several approaches to calculating viscosity-pressure behavior. The first and still current approach is that, according to Barus (Eq. 2.1) [15]. Here, the dynamic viscosity η_0 at a specific pair of values for temperature and atmospheric pressure is the initial value. For a pressure p, the pressure-dependent dynamic viscosity $\eta(p)$ is determined. α_η is the pressure-viscosity coefficient, which is determined experimentally. The pressure-viscosity coefficient depends on the fluid and as described above, is largely determined by the chemical structure of the base oil. The coefficient for ester-based oils is therefore lower than for naphthenic mineral oils.

$$A_0 = \frac{\mu_0 \omega I_0}{4\pi} = 10^{-7} \omega I_0 \eta_0 \cdot e^{(\alpha_\eta \cdot p)} \quad (2.1)$$

Due to the exponential relationship, the pressure-viscosity coefficient (pV-Coefficient) and, thus, the use of different base oils have a significant influence on the viscosity of the lubricant in highly loaded tribological contacts. Studies in [16] show that the influence of the pressure-viscosity behavior is decisive for the formation of a load-bearing lubricating film and is only influenced to a small extent by the temperature-viscosity behavior.

2.4. Fluids investigated

The hydraulic fluids used in this study are a naphthenic mineral oil (HLP 46) and a synthetic ester (HEES 46) with a minimum content of 25 % renewable raw materials. It is a biobased lubricant in accordance with DIN 16807 [17]. Both fluids are commercially available hydraulic fluids containing additives. Table **2.1** shows the most important physical properties of the hydraulic fluids considered in this paper.

Table 2.1. Physical properties of the investigated hydraulic fluids

Fluid	Viscosity index	Density (15 °C) [kg/m³]	Spec. heat capacity [kJ/(kg·K)]	Therm. conductivity [W/(m·K)]	pV-coefficient [10^{-3}bar^{-1}]
HLP 46	113	858	2.0	0.135	2.0
HEES 46	184	944	2.3	0.14	1.5

3. MATHEMATICAL FUNDAMENTALS OF LOSS MECHANISMS IN HYDRAULIC SYSTEMS

This section presents a mathematical analysis of various common hydraulic components in terms of their losses resulting from fluid properties, which are eventually utilized for the examination of the experimental investigations.

3.1. Flow Friction

The hydrostatic and hydrodynamic pressure drops of the pipe and the orifice, as well as the tribological contacts in a pump, are investigated. Regarding the pipe, an advanced description of the relationship between laminar volumetric flow rate and pressure drop is provided by computing both steady-state and transient relations. The steady relation is determined by the Hagen-Poiseuille law:

$$\mathrm{A}_0 = \frac{\mu_0 \omega \mathrm{I}_0}{4\pi} = 10^{-7} \omega \mathrm{I}_0 \frac{\boldsymbol{\pi R^4}}{\boldsymbol{8\eta L}} \boldsymbol{\Delta p} = \frac{\boldsymbol{\Delta p}}{\boldsymbol{R_H}}, \qquad (3.1)$$

with the dynamic viscosity η, the pipe radius R, the pipe length L, and the hydraulic resistance R_H. The transient phenomena can be determined by incorporating the frequency-dependent wall shear stress as a convolutional integral:

$$\mathrm{A}_0 = \frac{\mu_0 \omega \mathrm{I}_0}{4\pi} = 10^{-7} \omega \mathrm{I}_0 \int_0^{t_n} W(\tau) \frac{\Delta p(t_n - \tau_n)}{R_H} d\tau_n, \qquad (3.2)$$

with $W(\tau)$ being a weight function in the form of $\sum_0^k m_k e^{-n_k t_n}$, $m_k \in [5.53, 1.05, 0.427, 0.23, 0.143]$, $n_k \in [5.78, 30.74, 74.9, 139, 222]$, the normalized time $t_n = \frac{v}{R^2} t$ with the kinematic viscosity v and the time t and the normalized integral dummy variable τ_n [18]. Extending this mathematical formula to consider fluid compressibility would introduce the dissipation number $D_n = \frac{vL}{R^2 a}$, which describes the relation of the time span of the axial propagation of pressure waves to the radial diffusion of axial momentum [19]. Regarding these compressibility phenomena, the effect on hydraulic loss is minor compared to the incompressible effects and is therefore not further examined. Investigating these equations reveals that the steady volumetric flow rate is inversely proportional to η, thus resulting in a decreased pressure drop for a lower dynamic viscosity while maintaining a constant flow rate. This translates to a lower hydraulic loss, which is determined by the

product of pressure and flow rate. Investigating the Reynolds number Re = $\frac{\rho Q D}{\eta A}$, with the pipe diameter D and the pipe area A, which relates inertial and viscous forces, reveals that Re increases for a lower viscosity, thus resulting in a flow that is more dominated by inertial forces than viscous forces. Until the tipping point of Re < 2300, the pipe flow remains laminar; however, for higher Re due to a further decrease of the viscosity, the flow reaches the non-laminar regime, which is characterized by eddies, vortices, and flow instabilities and results in a higher hydraulic loss, specifically a square-root dependency of the pressure with regard to the flow rate. Regarding the transient flow rate, the normalized time t_n increases with a higher kinematic viscosity, which results in a faster steady state and thus a higher value for the hydraulic loss for a given dynamic flow rate.

3.2. Orifice Losses

Another commonly installed component in hydraulic systems is the orifice, which determines the following pressure-volumetric flow rate relation:

$$\mathrm{A_0} = \frac{\mu_0 \omega \mathrm{I_0}}{4\pi} = 10^{-7} \omega \mathrm{I_0} \boldsymbol{\Delta p} = \boldsymbol{L}\frac{\boldsymbol{dQ}}{\boldsymbol{dt}} + \frac{\boldsymbol{\rho}}{\boldsymbol{2\alpha_D^2 A^2}} \boldsymbol{|Q|Q}, \qquad (3.3)$$

with the hydraulic inertance $L = \frac{l_{eq}\rho}{A^2}$ and the discharge coefficient α_D [20]. The hydraulic inertance accounts for fluid inertia and thus incorporates dynamic phenomena occurring in orifice flows. The second term of the equation represents the stationary flow rate.

In the case of dig and dump, a defined trajectory of the excavator bucket is followed. To do this, the piston length of the cylinders is controlled. To provide the required volume flow $Q(t)$, a pressure difference Δp at the orifice is necessary. For a defined movement this pressure difference is called Δp_{ref}. By changing the fluid properties of the used oil, the pressure difference must be adjusted to achieve an identical cylinder movement. According to Eq. 3.4, the adjusted pressure Δp_i, only depends on the oil's density ρ_i when the transient term is neglected.

$$\mathrm{A_0} = \frac{\mu_0 \omega \mathrm{I_0}}{4\pi} = 10^{-7} \omega \mathrm{I_0} \frac{\boldsymbol{\Delta p_i}}{\boldsymbol{\Delta p_{ref}}} = \frac{\boldsymbol{\rho_i}}{\boldsymbol{\rho_{ref}}}, \qquad (3.4)$$

While the volume flow $Q(t)$ remains constant for a defined cylinder movement, the pressure difference Δp required to perform this movement increases with higher density. Consequently, the energy $E_{p\to A,B\to T}$ required also increases.

$$\mathrm{A_0} = \frac{\mu_0 \omega \mathrm{I_0}}{4\pi} = 10^{-7} \omega \mathrm{I_0} \boldsymbol{E_{p\to A,B\to T}} = \int_0^{\boldsymbol{t_n}} \boldsymbol{Q(t)} \cdot \underbrace{\boldsymbol{\Delta p_{ref}} \frac{\boldsymbol{\rho_i}}{\boldsymbol{\rho_{ref}}}}_{\boldsymbol{\Delta p}} \boldsymbol{dt}, \qquad (3.5)$$

In addition to the losses described above, there is also leakage on the tank side of the valve, which must be considered within the efficiency calculation. By rearranging equation 3.3 to $Q(t)$, the classic orifice equation from Eq. 3.5 is obtained. Eq. 3.4 establishes a relationship between the pressure p_i and p_{ref}. The resulting equation corresponds to the reference state and thus, under the given assumptions, the leakage volume flow is independent of the density of the fluid (Eq. 3.5).

$$\mathrm{A_0} = \frac{\mu_0 \omega \mathrm{I_0}}{4\pi} = 10^{-7} \omega \mathrm{I_0} \boldsymbol{Q} = \boldsymbol{\alpha} \cdot \boldsymbol{A}\sqrt{\frac{\boldsymbol{2\Delta p_i}}{\boldsymbol{\rho_i}}} = \boldsymbol{\alpha} \cdot \boldsymbol{A}\sqrt{\frac{\boldsymbol{2\Delta p_{ref}}\frac{\boldsymbol{\rho_i}}{\boldsymbol{\rho_{ref}}}}{\boldsymbol{\rho_i}}} = \boldsymbol{\alpha} \cdot \boldsymbol{A}\sqrt{\frac{\boldsymbol{2\Delta p_{ref}}}{\boldsymbol{\rho_{ref}}}}, \quad (3.6)$$

Finally, the power losses due to leakage can then be calculated using equation (3.7).

$$A_0 = \frac{\mu_0 \omega I_0}{4\pi} = 10^{-7} \omega I_0 \boldsymbol{E_{p \to T}} = \int_0^{t_n} \boldsymbol{Q_{leak}(t)} \cdot \Delta \boldsymbol{p_{ref}} \frac{\boldsymbol{\rho_i}}{\boldsymbol{\rho_{ref}}} \boldsymbol{dt}, \quad (3.7)$$

The power loss due to leakage increases similarly to the losses due to cylinder movement as the density of the oil increases, even though the actual leakage volume flow remains constant for a defined movement on the cylinder. The losses are therefore attributable to the higher pressure required for oils with higher density.

3.3. Hydraulic Pump (Tribological Contact)

Further losses occur in the tribological contacts of the pump (piston/bushing, slipper/swash plate, cylinder block/valve plate). In axial piston pumps, the greatest losses can be attributed to friction and leakage between the slipper and swash plate. For this reason, the relevant parameter influencing this tribological system should be examined more closely under simplified assumptions. Especially, the influence of pressure- and temperature-dependent viscosity behavior, which differs significantly for mineral oils and class 4 oils, on quantities like the average gap heights $\bar{h}$ shall be addressed. To do this, the contact between the slipper and swash plate should be considered in simplified form, as shown in **Figure 3.1**.

The pressure force F_p acting on the slide shoe is compensated hydrostatically by the inner bore and hydrodynamically by the inclination that occurs during operation. Friction losses in the contact made up of hydrodynamic friction and solid body friction. For small to average gap heights $\bar{h}$, the proportion of solid friction increases disproportionately, which also increases the total friction in contact.

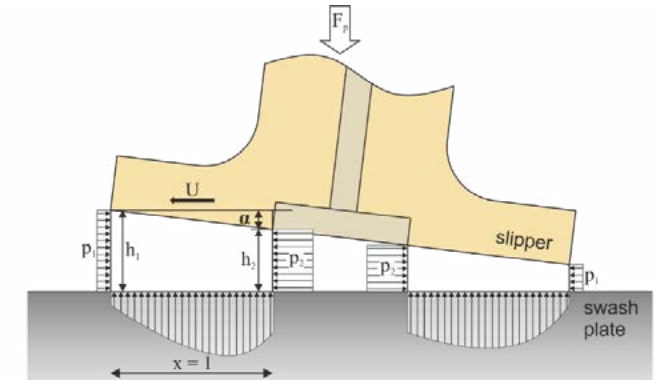


Figure 3.1. Schematic sketch of the slipper/swash plate contact

To estimate the qualitative influence of dynamic viscosity η on the average gap height $\bar{h}$, the Reynolds equation from Eq. 3.8 is used. The Reynolds equation describes the pressure distribution in small gaps for laminar flows and consists of the Couette term, the Poiseuille term, and a time-dependent term for pressure build-up. The relative movement U of the gap surfaces leads to a flow that is summarized in the Couette term. The flow driven by the pressure gradient $\partial p/\partial x$ is referred to as the Poiseuille term. Transient effects are represented in the unsteady case by the pressure build-up term.

$$\underbrace{\frac{\boldsymbol{U}}{\boldsymbol{2}} \frac{\boldsymbol{\partial}}{\boldsymbol{\partial x}} (\boldsymbol{\rho h})}_{\boldsymbol{Couette\ flow}} - \underbrace{\frac{\boldsymbol{1}}{\boldsymbol{12\eta}} \frac{\boldsymbol{\partial}}{\boldsymbol{\partial x}} \left(\boldsymbol{\rho h^3} \frac{\boldsymbol{\partial p}}{\boldsymbol{\partial x}} \right)}_{\boldsymbol{Poiseuille\ flow}} + \underbrace{\frac{\boldsymbol{\partial}}{\boldsymbol{\partial t}} (\boldsymbol{\rho h})}_{\boldsymbol{pressure\ build-up}} = \boldsymbol{0}, \quad (3.8)$$

The Reynolds equation can be solved analytically for the boundary conditions shown in **Figure 3.1** (p_1, p_2, h_1, h_2, U), assuming a constant inclination α and for the 1-dimensional steady-state case (Eq. 3.9).

$$\boldsymbol{p(x) = -\frac{6\eta U}{\alpha}\frac{1}{h(x)} - \frac{C_1}{2\alpha}\frac{1}{h(x)^2} + C_2}, \quad (3.9)$$

$$\boldsymbol{C_1 = \frac{12\eta U h_1 h_2 + 2(p_1 + p_2){h_1}^2{h_2}^2}{\alpha(h_1 + h_2)}}$$

$$\boldsymbol{C_2 = p_1 + \frac{6\eta U}{\alpha h_1} + \frac{C_1}{\alpha {h_1}^2}}$$

By integrating $p(x)$ with respect to x, a length-related force F_z is calculated, which depends on the average gap height $\bar{h}$, neglecting third order α terms $O(\alpha^3)$.

$$\boldsymbol{F_z = \int_0^1 p(x)\,dx} = \underbrace{\boldsymbol{\frac{p_1+p_1}{2} + \frac{\alpha}{4\bar{h}}(p_1 + p_1)}}_{\boldsymbol{Poiseuille\ fraction}} \underbrace{\boldsymbol{-\frac{\eta U \alpha}{2\bar{h}^3} + O(\alpha^3)}}_{\boldsymbol{Couette\ fraction}}, \quad (3.10)$$

The dynamic viscosity η is in the numerator of the Coutte fraction of the length-related force. If the viscosity decreases, this leads to a reduction in the force exerted on the slipper by the pressure field $p(x)$. To maintain a stable balance between the pressure field and F_p, the average gap height $\bar{h}$ decreases until a new equilibrium is reached. A reduction in the average gap height $\bar{h}$ simultaneously increases the proportion of solid contact and thus the friction between slipper and swash plate.

4. Experimental Investigation of the Loss Mechanisms During the Operation of a Mobile Working Machine

The experimental investigations were carried out on a test excavator. The crawler excavator belongs to the class of compact excavators, and it is driven by tracks. The hydraulic system of the test excavator is a closed-centre system with a downstream pressure compensator and a circulation valve in the form of a pressure relief valve. The displacement volume of the axial piston pump is controlled by a two-point controller, which swings back from full displacement to a minimum swing angle when a pressure of approx. 130 bar is reached. In addition to linear actuators, which are required for digging, there also are actuators for further operations which are not regarded in the investigations. The test excavator is controlled by electro-hydraulic pilot valves instead of the usually implemented hydraulic joysticks. Therefore, an automatic digging process can be performed by the control of the cylinder lengths over the cycle time to achieve identical output energy when using different fluids. For the investigations made a 90° dig and dump cycle (dig/swing/unload/swing) is performed. The excavated material is simulated due to a weight in the shovel of 25 kg. In **Figure 4.1** the bucket tip position is shown during the cycle for the two fluids investigated.

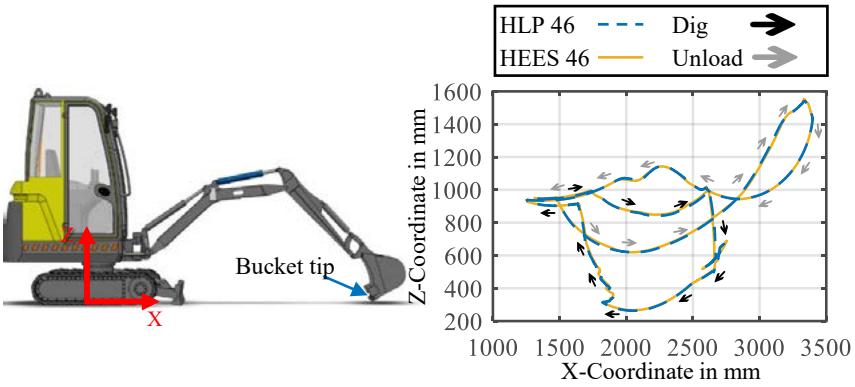


Figure 4.1. Position of the bucket tip during the dig and dump cycle at 40 °C tank temperature

The test operation is performed at three different tank temperature levels (25 °C, 40 °C and 50 °C). The allocation and assessment of losses is based on the mathematical principles set out in Chapter 3. To classify the losses, **Table 4.1** shows the dynamic viscosity of the hydraulic fluids investigated at the temperatures considered.

Table 4.1. Dynamic viscosity at investigated temperature stages

Hydraulic fluid	Dyn. Viscosity [mPas] (25 °C)	Dyn. Viscosity [mPas] (40 °C)	Dyn. Viscosity [mPas] (50 °C)
HLP 46	86.7	39.9	26.2
HEES 46	71.7	38.4	26.3

4.1. Flow Friction losses E_{Flow}

The flow friction losses $\boldsymbol{E_{Flow}}$ during the Dig and Dump cycle are calculated by integrating the power loss over cycle time t. The power loss caused by flow friction is determined by multiplying the pressure losses of a pipe section Δp by the volume flow Q_{Pump} (Eq.4.1).

Figure 4.2 shows the flow friction losses $\boldsymbol{E_{Flow}}$ of the hydraulic fluids considered at the three different temperature levels and the share of the losses on the total input energy. Both fluids investigated fit into the viscosity class of ISO VG 46 [21]. Due to the fact that the standard allows a deviation of ±10 % from the average kinematic viscosity (46 mm²/s), the kinematic viscosity of the same class can differ up to 20 %. This is the case with the examined: the HEES 46 has a viscosity of 41.4 mm²/s at 40 °C and the HLP 46 has a viscosity of 47.3 mm²/s. This means that the deviation in kinematic viscosity is approximately 12 %. This results in an intersection point of the dynamic viscosity of the two hydraulic oils at a temperature of 49 °C (Table **4.1**). Below this temperature, the dynamic viscosity of HEES 46 is lower than that of HLP 46. Due to the higher viscosity index, it lies above HLP 46 at higher temperature after the intersection point. The flow losses caused by friction are proportional to the dynamic viscosity according to Eq. 3.1.

$$E_{Flow} = \int_0^{t=end} Q_{Pump}(t) \cdot \Delta p \, dt, \qquad (4.1)$$

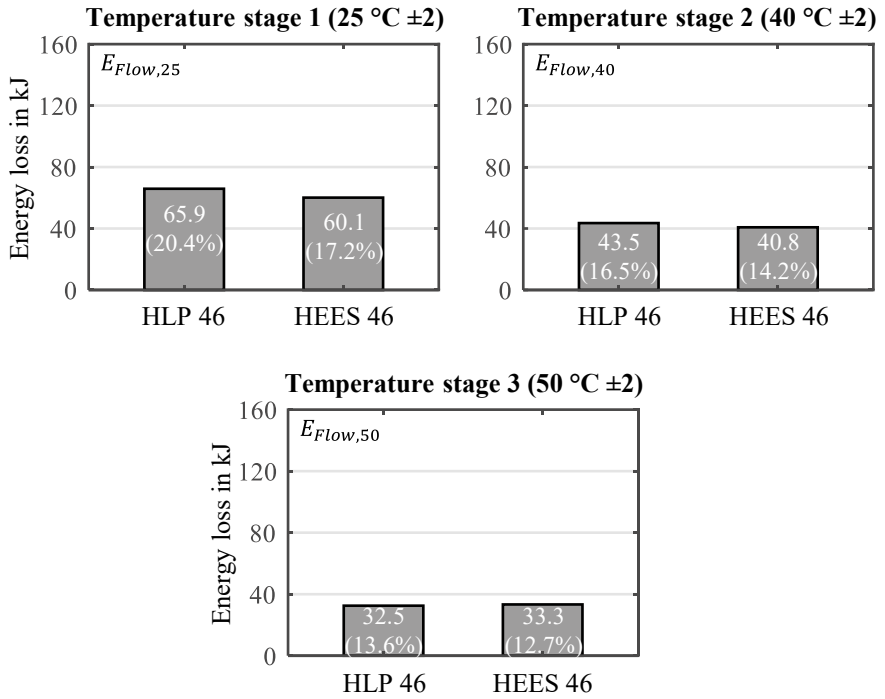


Figure 4.2. Flow friction energy losses $\boldsymbol{E_{Flow}}$ and their share on the input energy during dig and dump cycle dependent on tank temperature

4.2. Pump losses E_{Pump}

The hydraulic pump (axial piston pump) converts the incoming mechanical energy into hydraulic energy. This results in hydraulic-mechanical and volumetric losses. The total energy loss $\boldsymbol{E_{Pump}}$ is calculated from the difference between the measured mechanical input energy $\boldsymbol{E_{In}}$ measured at the drive shaft of the pump and the hydraulic output energy $\boldsymbol{E_{Out}}$ of the pump. The hydraulic output energy is calculated by integrating the product of the measured variables pressure at pump outlet $\boldsymbol{p_{Pump}}$and volume flow $\boldsymbol{Q_{Pump}}$ over cycle time $\boldsymbol{t}$ (**Eq. 4.2**).

$$E_{Pump} = E_{In} - E_{Out} = E_{In} - \int_0^{t=end} Q_{Pump}(t) \cdot p_{Pump}(t)\, dt, \qquad (4.2)$$

The comparison of the energy losses caused by the pump when using the two different hydraulic fluids at the different temperatures is shown in **Figure 4.3**. The pump losses are inversely proportional to the overall efficiency of the hydraulic pump. As the temperature rises, energy losses generally decrease. Except for HLP 46 between temperature stage 2 and 3. It can be assumed that the minimum proposed viscosity of the pump is reached at temperature stage 3. Due to the higher viscosity index of HEES 46 the behaviour of the energy losses is different. The overall efficiency is composed of the volumetric and hydraulic-mechanical efficiency. Volumetric losses are largely dependent on the dynamic viscosity of the fluid (Eq. 3.1). At higher dynamic viscosity, losses are minimized accordingly. Hydraulic mechanical efficiency, on the other hand, describes losses due to friction in the pump. Since the density of HEES 46 is higher than that of HLP 46, it can be assumed that the volumetric losses are higher. Since the total pump losses are higher when

using HEES 46, it can be expected that the hydraulic-mechanical losses and thus the friction losses are significantly higher. This assumption was confirmed in preliminary investigations and was attributed to the pressure-viscosity behaviour [5]. The influence of friction behavior in tribological contacts of pumps is described in detail in Chapter 3.3.

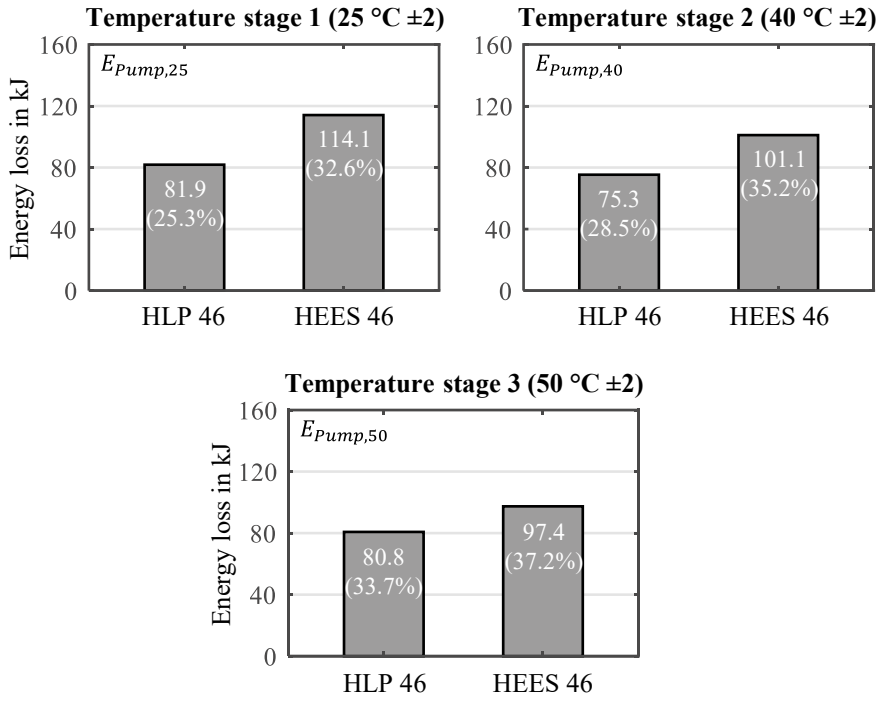


Figure 4.3. Energy losses caused by the pump $\boldsymbol{E_{Pump}}$ and their share on the input energy during dig and dump cycle dependent on tank temperature

4.3. System losses E_{Sys}

System losses $\boldsymbol{E_{Sys}}$ depend on the functionality and architecture of the hydraulic system. The excavator system under consideration is a closed-center system with a downstream pressure balance. The excess volume flow that is not used to fulfil the movement of the actuators is throttled back to the tank via a circulation valve. For this purpose, the volume flow requested by the activated consumer $\boldsymbol{Q_{Act}}$ is subtracted from the total volume flow provided by the pump $\boldsymbol{Q_{Pump}}$ and multiplied by the pump pressure $\boldsymbol{p_{Pump}}$. To calculate the energy loss $\boldsymbol{E_{Sys}}$, the measured values are integrated over the cycle time $\boldsymbol{t}$ (**Eq. 4.3**).

$$E_{Sys} = \int_0^{t=end} p_{Pump}(t) \cdot \left(Q_{Pump}(t) - \sum_{i=1}^{n} Q_{Act,i}(t)\right)\, dt, \quad (4.3)$$

When looking at the results of the system losses during the dig and dump cycle (**Figure 4.4**), only minor differences can be identified. The losses are almost identical, especially for the first two temperature levels (0.9%/0.38%). At the third temperature level, the difference is higher (4.8%).

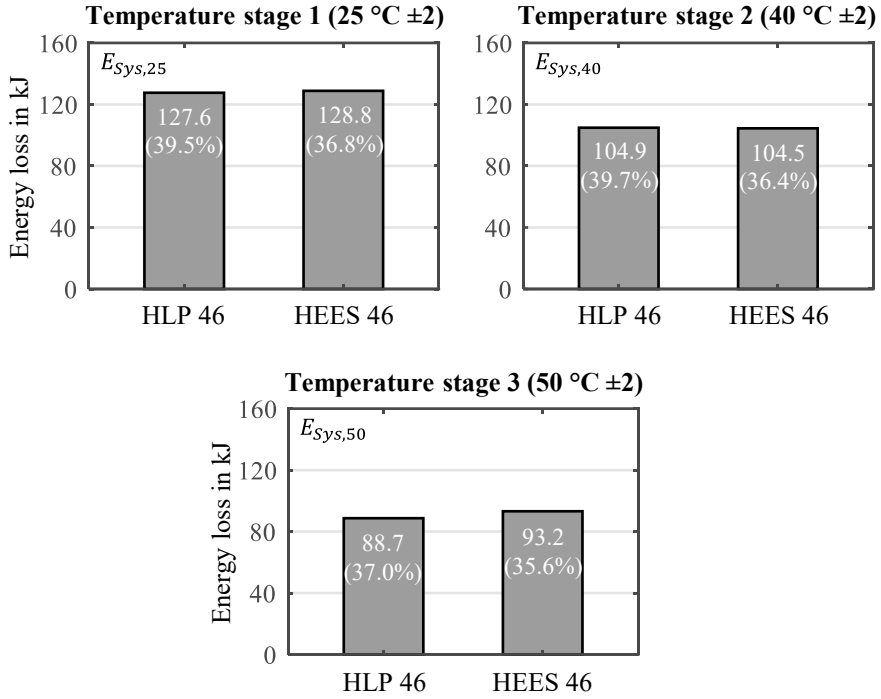


Figure 4.4. Energy losses caused by the system $\boldsymbol{E_{Sys}}$ and their share on the input energy during dig and dump cycle dependent on tank temperature

4.4. Valve losses E_{MCV}

When fluid passes through valves, several types of losses occur in varying magnitudes. The flow resistance depends on the type of closing element and the channel configuration. Furthermore, the geometric conditions depend on the opening position of the closing element. In case of the investigated excavator the valve sections for cylinders and swivel motor consist of spool valves combined into a main control valve. In spool valves, orifice losses and, to a lesser extent, throttling losses occur predominantly. The flow resistance of a valve is determined by the orifice area and the flow velocity. **Eq. 4.4** is used to determine the energy losses caused by the flow through the valves during the dig and dump cycle. This involves integrating the valve-related hydraulic power loss over cycle time $\boldsymbol{t}$. The power loss is determined by multiplying the pressure loss between the excavator´s pressure line near the pump $\boldsymbol{p_{Pump}}$ and the pressure right after the respective valve $\boldsymbol{p_{Act,MCV}}$ with the corresponding volume flow requested by the respective actuator $\boldsymbol{Q_{Act}}$. Losses in the valve due to leakage are not taken into account in this calculation.

$$E_{MCV} = \int_0^{t=end} \left(p_{Pump}(t) - p_{Act,MCV,i}(t)\right) \cdot Q_{Act,i}(t)\ dt, \quad (4.4)$$

Figure 4.5 shows the energy losses during the dig and dump cycle by operation with the different oils. The losses generated by the flow paths of the Main Control Valve (MCV) $\boldsymbol{E_{MCV}}$ generally show smaller differences between the various hydraulic fluids than the types of losses described before. This can be explained by the fact that the losses occurring at the valve can mainly be described as orifice-type resistance. This means that the losses are

largely independent of viscosity (see Eq. 3.3). Minor differences can be explained by the different densities of the fluids and sections with throttle losses of the inlet and outlet pipeline of the valves. Furthermore, the cross-section area can differ by using different fluids, while the cylinder lengths are regulated.

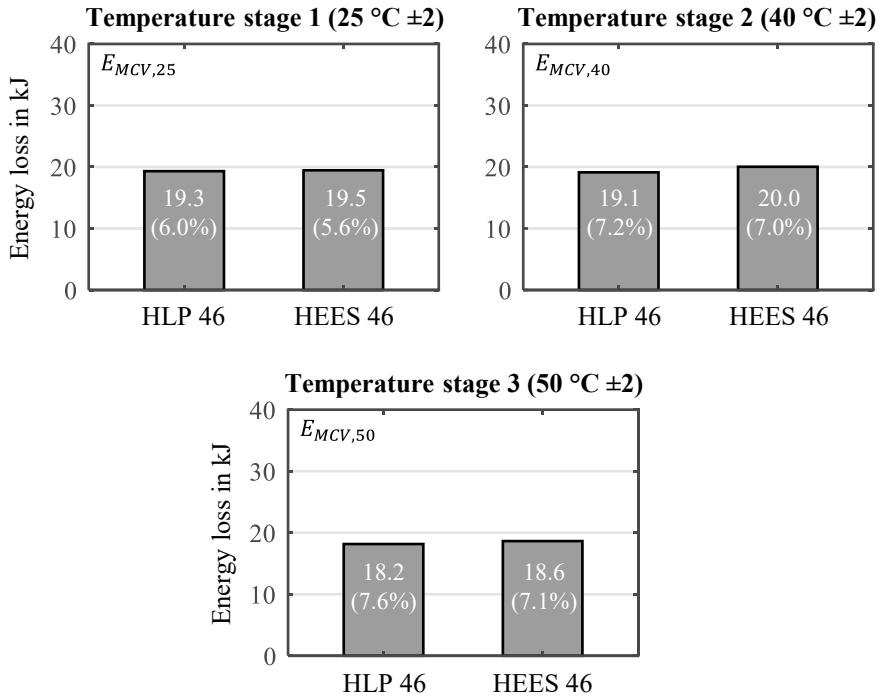


Figure 4.5. Energy losses caused by the main control valve $\boldsymbol{E_{MCV}}$ and their share on the input energy during dig and dump cycle dependent on tank temperature

5. DISCUSSION AND CONCLUSION

The individual losses described in previous chapters of course have an impact on the overall efficiency of the machine. **Figure 5.1** shows the energy requirement for the two hydraulic fluids over the tank temperature for the dig and dump cycle. The measured mechanical input energy E_{In} was integrated over the cycle time t. The highest measured energy requirement by the use the two fluids, which occurs at HEES 46 and 25 °C, serves as a reference. The losses when using HEES 46 are higher at all temperature stages than when performing the cycle with HLP 46. When considering the individual losses (Chapter 3), the higher losses caused by the hydraulic pump are mainly responsible for this. The further losses show only minor differences between the two hydraulic fluids investigated.

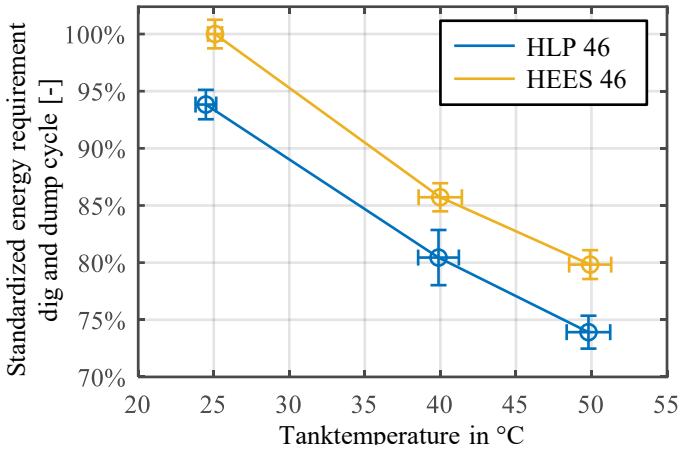


Figure 5.1. Standardized energy requirement on the highest measured energy consumption (HEES 46/25 °C) during dig and dump cycle

Since HLP 46 is one of the most used hydraulic fluids in the market, it was assumed that pump efficiency would be higher when using the mineral oil based fluid instead of HEES 46. In fact, the pumps and their tribological contacts were optimised based on the chemical and physical characteristics of HLP 46. In order to achieve comparable efficiencies with HEES 46, the geometry of the components needs to be adjusted by minimising the contact pressure. This can be achieved, for example, by increasing the hydrostatic relief in highly stressed tribological contacts. Furthermore, the research on ecological compatible additives for biobased lubricants, like extreme pressure (EP-) additives offer a high potential for further improvements of bio lubricants performance. These aspects offer further research potential for optimising both the resource efficiency and the ecological compatibility of hydraulic systems due to the use of biobased hydraulic fluids.

6. REFERENCES

[1] DYM Resources GmbH. Base oil market in 2023: challenges, trends, key market drivers. https://dymresources.com/de/news/base-oil_and_lubes/base-oil-market-in-2023-challenges-trends-key-market-drivers/ (Accessed August 1, 2025).

[2] Global Market Insights Inc. Base Oil Market - By Group, By Type, By Application & Forecast, 2024- 2032. https://www.gminsights.com/industry-analysis/base-oil-market (Accessed April 30, 2025).

[3] Chiavola, O.; Frattini, E.; Palmieri, F.; Fioravanti, A.; Marani, P. On the Efficiency of Mobile Hydraulic Power Packs Operating with New and Aged Eco-Friendly Fluids. *Energies*, **2023**, *16,* 5681.

[4] Otto, N.; Murrenhoff, H. *Öleinfluss auf den Wirkungsgrad von Hydraulikpumpen: Voruntersuchungen:* München, **2016**.

[5] Deuster, S.; Holzer, A.; Schmitz, K. Influence of the Pressure-Viscosity Behavior of Different Base Oils on the Formation of Lubricating Films in Tribological Contacts. In: *Global Fluid Power Society PhD Symposium*: Neapel, **2022**; pp. 1–15.

[6] Castro, R.d.M.; Curi, E.I.M.; Inácio, L.F.F.; Da Rocha, A.S. Analysis of the tribological performances of biodegradable hydraulic oils HEES and HEPR in the

sliding of Cu–Zn/WC–CoCr alloys using the Stribeck curve. *J Braz. Soc. Mech. Sci. Eng.*, **2020**, *42,* 1–20.

[7] Jakobsen, J.H. *Synthetic Esters in Hydraulic Valves;* Wittusen & Jensen, **2020**.

[8] Deuster, S.; Schmitz, K. In: *14th International Fluid Power Conference*, 14th International Fluid Power Conference (14. IFK), Dresden, Germany, 3/19/2024 - 3/21/2024; Weber, Prof. Dr.-Ing. Jürgen, Ed.; River Publishers: Denmark, **March 19, 2024 - March 21, 2024**; pp. 102–114.

[9] Newton, I. *Philosohiae Naturalis Principia Mathematica (Latin, 1687);* Benediction Classics, **2015**.

[10] Rütten, M. *Verallgemeinerte newtonsche Fluide: Thermische und viskose Strömungseigenschaften,* 1st ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2018**.

[11] Boyde, S. Esters. In: *Synthetics, Mineral Oils, and Bio-Based Lubricants*. Rudnick, L.R., Ed.; CRC Press, **2020**.

[12] DIN Deutsches Institut für Normung e.V. *Mineralölerzeugnisse Berechnung des Viskositätsindex aus der kinematischen Viskosität;* Beuth Verlag GmbH: Berlin, **2004 (Accessed** May 15, 2021).

[13] Mang, T.; Dresel, W., Eds. *Lubricants and lubrication;* Wiley-VCH: Weinheim, Germany, **2017**.

[14] Totten, G.E.; Shah, R.J.; Forester, D.R., Eds. *Fuels and Lubricants Handbook: Technology, Properties, Performance, and Testing, 2nd Edition;* ASTM International: West Conshohocken, **2019**.

[15] Barus, C. Isothermals, Isopiestics and isometrics relative to viscosity. *American Journal of Science*, **1893**.

[16] Murrenhoff, H. *Umweltverträgliche Tribosysteme: Die Vision einer umweltfreundlichen Werkzeugmaschine;* Springer-Verlag Berlin Heidelberg: Berlin, Heidelberg, **2010**.

[17] DIN Deutsches Institut für Normung e.V. *Flüssige Mineralöl-Erzeugnisse – Bio-Schmierstoffe – Kriterien und Anforderungen für Bio-Schmierstoffe und biobasierte Schmierstoffe (EN 16807);* Beuth Verlag: Berlin, **12.2016**.

[18] Brumand-Poor, F.; Kotte, T.; Pasquini, E.G.; Schmitz, K. Signal Processing for Transient Flow Rate Determination: An Analytical Soft Sensor Using Two Pressure Signals. *Signals*, **2024**, *5,* 812–840.

[19] Brumand-Poor, F.; Kotte, T.; Pasquini, E.; Kratschun, F.; Enking, J.; Schmitz, K. Unsteady flow rate in transient, incompressible pipe flow. *Z Angew Math Mech*, **2025**, *105*.

[20] Schmitz, K. *Fluidtechnik - Systeme und Komponenten: Umdruck zur Vorlesung,* 1st ed.; Shaker Verlag: Düren, **2022**.

[21] DIN Deutsches Institut für Normung e.V. *Flüssige Industrie-Schmierstoffe – ISO-Viskositätsklassifikation (ISO 3448:1992);* Beuth Verlag GmbH: Berlin, **2010**.

Biographies



Sebastian Deuster studied Mechanical Engineering at the University of Applied Sciences in Aachen and finished with a master's degree. Since 2018 he is part of the Institute for Fluid Power Drives and Systems at RWTH University as a research associate. His main research activity is to state out the performance abilities of biobased hydraulic fluids. Since 2023 Head of Sustainable Lubricants Committee (SLC). This is a working group consisting of fluid/additive manufacturers and OEMs to find solutions to further develop sustainable lubricants.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Advanced Hydraulic Fluids Make Mobile Machinery Not Only More Sustainable, but Also More Economical

Dmitriy Shakhvorostov, Michal Alibert, Thorsten Bartels, Rhishikesh Gokhale, Andreas Hees

Evonik Operations GmbH, Oil Additives, Presenting author: dmitriy.shakhvorostov@evonik.com

Abstract.

Europe is pursuing ambitious climate goals: by 2030, net greenhouse gas emissions are to be significantly reduced—by 65% in Germany compared to 1990. Globally, the pressure to cut emissions and promote sustainable technologies is also increasing. Mobile machinery such as excavators, wheel loaders, and other hydraulically operated equipment is coming into sharper focus, as these machines play a central role in construction, agriculture, and industry.

This technical study explores how advanced hydraulic fluids can contribute to achieving these goals. By using shear-stable viscosity index improvers based on polyalkylmethacrylates, fluids can be formulated to maintain high efficiency across a wide temperature range. This not only leads to energy savings of up to 20%, but also to a measurable reduction in CO_2 emissions.

In addition to the environmental impact, the economic perspective is also examined: despite higher initial costs, these fluids offer a positive long-term return on investment—through reduced energy consumption, extended component life, and lower maintenance costs.

The study is based on Evonik's many years of experience in developing and testing such high-performance fluids. Using technical data and case studies, it demonstrates how modern fluid technology can contribute to the sustainable transformation of mobile hydraulics.

Keywords: Hydraulic systems, Hydraulic fluids, Viscosity index improvers, Polyalkyl methacrylate (PAMA), Reduced energy consumption, Total cost of ownership (TCO)

1. INTRODUCTION

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

In mobile hydraulic machinery, the total cost of ownership has become one of the most critical competitive factors. For machine owners and operators, fuel or energy costs and maintenance requirements play a decisive role in daily operations.

Rising fuel prices, stricter emissions requirements, and uncertainties in the supply of energy and raw materials have intensified the pressure to reduce energy losses in hydraulic systems. As a result, hydraulic efficiency and fluid performance are no longer secondary topics — they directly determine operating cost, machine uptime, and competitive positioning in the mobile equipment market.

At the same time, mobile machinery contributes significantly to global energy consumption and CO_2 emissions. Even incremental improvements in hydraulic efficiency, fluid formulation, and component durability can therefore offer substantial potential for reducing operating costs and environmental impact across large fleets.

The reasons for this are high fuel and energy cost and uncertainties in the supply of energy [1, 2] and raw materials (Figure 1.1).

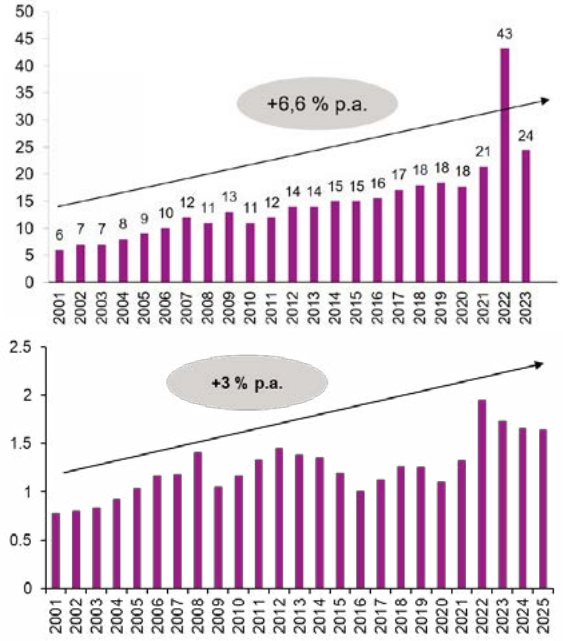


Figure 1.1. Electric energy price in €cent/kWh over time in the 21st century [3,4] and diesel retail price in €/l [5] in Germany.

Efficiency improvement potentials have been repeatedly demonstrated in the mobile machinery and industrial sector [6,7]. However, the implementation of energy efficiency measures is often slow, and several hurdles need to be overcome. The low acceptance, despite economic advantages, is attributed to a number of obstacles, which have been

investigated by numerous studies [8]. A well-known obstacle mentioned in several surveys is the lack of resources, time, and educated staff. Additionally, the effort required for the planning and implementation of energy efficiency measures represents a further barrier.

The optimization of construction sites in terms of up time process efficiency is not only a business requirement, but also a decisive step on the way to a climate-neutral society. A bundle of measures is available, which differ significantly in terms of investment costs and complexity [8]. A simple measure to achieve both, cost and emission reduction, is to increase the energy efficiency of the hydraulic system by choosing a suitable hydraulic lubricant. This approach does not require hardware design modifications and can therefore be easily applied to existing hydraulic driven machines. Since hydraulic oils are changed regularly anyway, the associated effort is also low.

2. HYDRAULIC OILS IN MOBILE MACHINES

Hydraulic fluids are widely used in industry. Their properties result from the choice of base oils and the selection and combination of suitable additives. Viscosity has a great influence. In principle, it is possible to reduce hydromechanical losses by using low-viscosity fluids and thus (theoretically) increase the efficiency of the machine, unless volumetric losses in hydraulic systems increase too much as a result of this measure. With very low viscosities there is a risk that no sufficient lubricating film will be formed in contact points that are subject to high temperatures leading to increase of mixed friction and increase of wear. This risk is avoided by reducing the temperature dependence of the viscosity of the lubricant – i.e. increasing the viscosity index (VI) – by using suitable VI improvers and/or base oils with high VI. An additional risk of simply lowering the viscosity is to increase the chance of an external leakage in older machines at elevated temperatures where fluids with low viscosity index will have dramatically reduced viscosity.

For hydraulic applications, it has been shown that both, viscosity index and shear stability of the fluid are decisive for the level of efficiency [6,7]. Figure 2.1 demonstrates the opposing characteristics of mechanical and volumetric efficiency in hydraulic equipment [6,7]. A low viscosity reduces internal friction and has a positive effect on the mechanical. The opposite is true for volumetric efficiency, which decreases as viscosity decreases. Overall efficiency is the combination of both efficiencies.

The aim is to minimize both, the "Total Cost of Ownership" (TCO) and the “GreenHouse Gas” (GHG) emissions, two parameters that are linked via the lifetime, reliability, and efficiency of equipment.

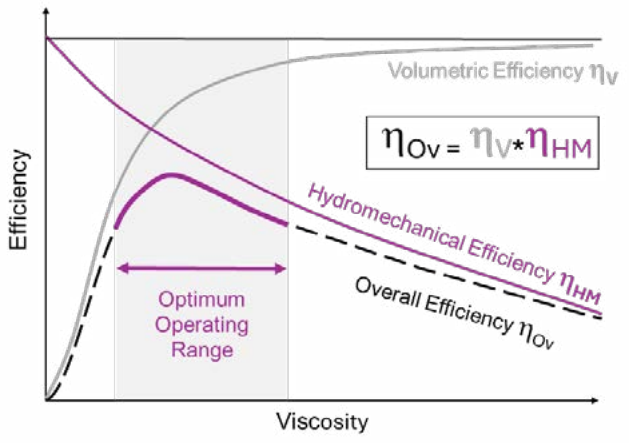


Figure 2.2. Correlation of efficiency and viscosity in a hydraulic pump [6,7]

3. VALIDATION OF FUEL AND ENERGY SAVINGS

Figure 2.1 describes a step-by-step approach to validate effects of an oil changes in hydraulic systems. An oil change can be an easy way to reduce overall cost in terms of savings on maintenance, energy consumption, and GHG emissions. The procedure to realize oil changes in mobile and stationary hydraulics was investigated and demonstrated for various systems (excavators, wheel loader, skid steer loader, tractor).

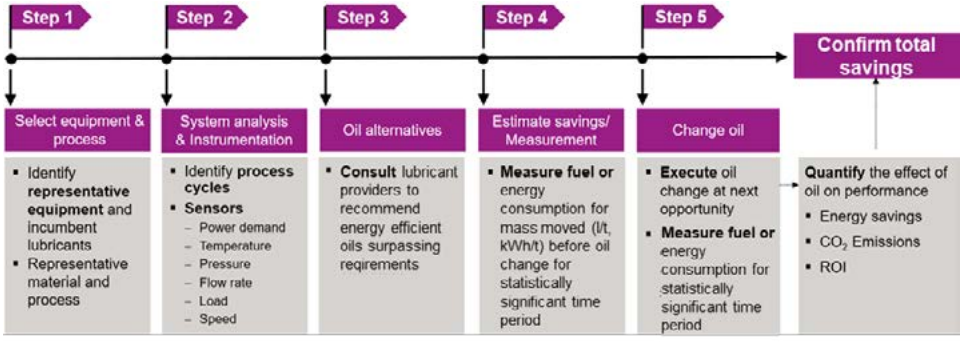


Figure 2.3. Universal step by step approach to execute oil changes and confirm fuel or energy savings.

In this article, the authors present an overview of observed effects depending on the fluid rheological properties and describe in more detail the effects of fluid in an excavator operation.

The field trial employed a medium-class hydraulic crawler excavator representative of machines in the 23–26 ton operating-weight range. According to published technical data for machines of this class, such excavators typically exhibit an operating weight between 23,400 and 26,050 kg and are powered by a diesel engine rated at approximately 129 kW (≈175 HP). The excavator used in the study incorporates a conventional hydraulic architecture consisting of three hydraulic cylinders controlling the boom, a single cylinder for the stick (arm), and an additional single cylinder for bucket articulation. Interchangeable digging buckets with nominal capacities of 0.75 t, 1.25 t, and 1.75 t were utilized to assess loading behavior and cycle efficiency under varying tool masses. This configuration is representative of typical quarry-class loading equipment.

To support the excavation cycles and evaluate downstream productivity effects, the test setup also employed a 25-ton class dump truck for material hauling and a wheeled loader for auxiliary material handling and site servicing. The integration of these machines enabled a controlled assessment of how changes in hydraulic fluid formulation influence the excavator's energy efficiency, loading dynamics, and overall productivity across a complete loading–hauling cycle.

The trial was conducted in a stone quarry environment in Gimbsheim, Germany, where conditions allowed repeatable and consistent measurement. An on-site truck scale facilitated accurate mass determinations for each truckload, enabling quantification of fill mass, bucket productivity, and cycle uniformity. The test material consisted of gravel stone with an approximate particle size of 1 inch, providing a homogeneous granular medium suitable for comparing hydraulic performance under realistic working conditions.

This instrumentation framework ensured a controlled, measurable, and representative evaluation of hydraulic system behavior in real-world quarry operations.

The experimental program consisted of seven test oils including reference fluid, each subjected to 31 operational test cycles per day, executed in three independent replicates to ensure statistical robustness. A reference oil was incorporated systematically throughout the campaign—at the beginning, midpoint, and end of each daily sequence—to monitor drift and establish baseline repeatability. Additional reference runs were scheduled immediately after the first test oil, after the machine's return to site, and at the conclusion of the final oil evaluation, providing a continuous calibration framework across the multi-day campaign. The order of oils was selected pragmatically to accommodate site logistics and thermal behavior, recognizing that operational temperatures naturally increased over the day, typically rising from 50 °C to 70 °C. Hydraulic oil temperature control was further refined by enforcing 100% cooling-fan duty via CAN-bus command, complemented by a deliberate reduction of the cold-air suction-hose cross-section, enabling tighter regulation of thermal boundary conditions during each test run.

Each oil was evaluated over a four-day operational block, followed by a dedicated operator test on day five, after which the machine underwent a one-day flushing and refilling procedure before transitioning to the next fluid. Two operators participated in the study, with both operators performing tests each day to reduce operator-specific bias. Operator sequencing was predefined and maintained consistently, applying the same order on days 1 and 3 as on days 2 and 4. For each oil, 16 test sequences per operator were executed, ensuring balanced exposure of each hydraulic fluid to varying operational styles and workloads. The 16 test sequences were distributed between moving, truck loading and digging cycles.

A stringent oil-flushing procedure was implemented between fluids to prevent cross-contamination. This process comprised four complete flushing cycles, performed with the excavator's boom and arm fully extended to maximize hydraulic-volume exchange, followed by one full refill of fresh test oil. The viscosity of the filled oil was measured at both the beginning and end of each test period to document potential degradation or shear-related viscosity loss over the operational interval. Collectively, this methodology provided a highly controlled and repeatable framework for quantifying the influence of hydraulic-fluid formulation on excavator performance, thermal stability, and cycle efficiency.

4. TESTED FLUIDS

The hydraulic fluids evaluated in this study consisted of formulations built on a standard hydraulic fluid anti-wear/oxidation inhibitor (DI) additive package, combined with base oils spanning API Group II, Group III, and Group IV quality levels. The product matrix was intentionally designed to see if the influence of base-oil quality and viscosity-index (VI) performance on hydraulic efficiency, including fluids blended with second-tier high-VI Group III stocks as well as maximum-VI formulations relying on Group IV (PAO) components can be elaborated. The set included conventional mono-grade fluids (ISO VG 46 and ISO VG 68) with moderate viscosity indices (VI 105 and VI 98, respectively), as well as a series of high-VI multigrade formulations engineered to deliver enhanced viscometric stability: VG 32 VI 189, VG 46 VI 112, VG 46 VI 155, VG 46 VI 167, and an ultra-high-VI VG 25 VI 285 fluid (see Figure 4.1). These exhibited viscosity indices ranging from 112 to 285, covering the practical performance envelope from improved cold-start viscosity control to advanced shear-stable elastomer-friendly multigrade behavior. Kinematic viscosities at 40 °C and 100 °C varied accordingly, from 26.0 to 71.2 cSt at 40 °C and 7.08 to 9.13 cSt at 100 °C, reflecting the tailored grade-dependent rheology. Shear stability, measured in a 40-minute ultra sonic test (ASTM D5621) viscosity loss, ranged from 0.45% in the mono-grade VG 68 to 15% in the ultra-high-VI VG 25, capturing differences in polymeric thickener architecture, treat rate, and mechanical shear sensitivity. Collectively, this hydraulic-fluid matrix provided a broad and controlled variation in base-oil structure, viscosity grade, and viscosity index improver properties, enabling systematic assessment of how fluid formulation influences excavator hydraulic efficiency, thermal behavior, and operational productivity.

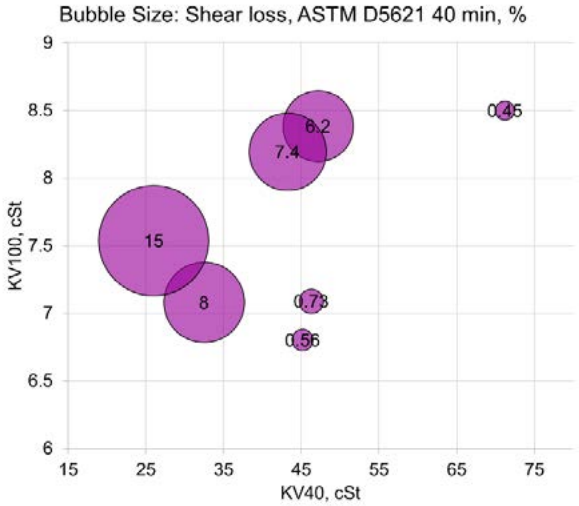


Figure 4.1. Overview of the rheological properties of seven tested fluids.

5. RESULTS

5.1. Performance Outcomes Across Viscosity Grades and Viscosity Indices

The weighted overall efficiency improvement (45% digging + 45% truck loading + 10% traveling) results (Figure 5.1) reveal a clear stratification of fluid performance driven primarily by viscosity grade (KV40) and viscosity index (VI). The reference monograde ISO VG 46 (KV40 = 45.16 cSt, VI = 105) establishes the 0% baseline. Deviations from this baseline strongly follow the expected physical mechanisms: reductions in viscosity grade paired with sufficiently high viscosity index produce substantial improvements in hydraulic efficiency, whereas increases in viscosity grade or insufficient VI lead to penalties. This is illustrated by the poorest-performing fluid, ISO VG 68 (KV40 = 71.2 cSt, VI = 98), which shows a −6.1% efficiency decrease due to elevated hydromechanical losses at higher viscosity. Conversely, the highest-performing fluids—ISO VG 22 with 285 VI and ISO VG 32 with 189 VI — demonstrate improvements of +13.6% and +10.7%, respectively, underscoring the strong advantage of pairing relatively low kinematic viscosity at 40°C with very high VI to minimize frictional losses while preserving pump leakage control at elevated temperatures.

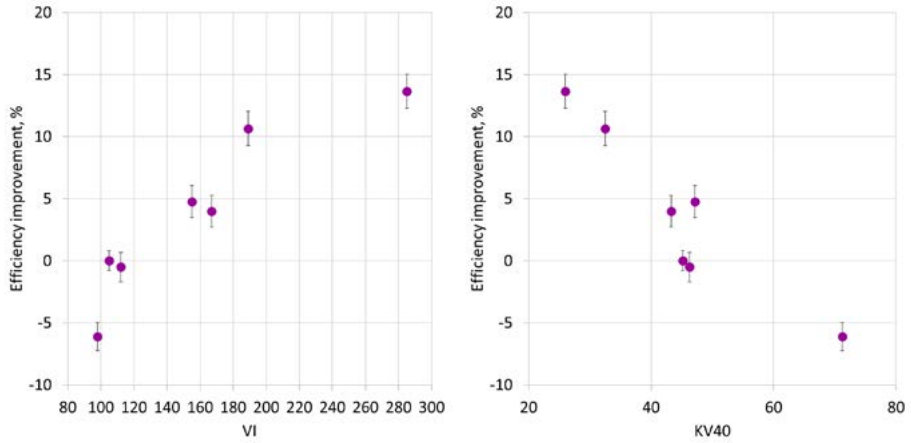


Figure 5.1. Efficiency improvement of test fluids with varying viscosity index (VI) and kinematic viscosity at 40°C (KV40) compared to the ISO VG 46 Monograde fluid in a combined cycle 45% digging + 45 % truck loading + 10 % traveling.

5.2. Influence of Viscosity Index (VI) on Efficiency

VI emerges as a decisive factor in the test results, exerting a strong positive influence on overall efficiency also when viscosity grade is held relatively constant. Among the ISO VG 46-based fluids, an increase in VI from 105-112 VI to 167 VI boosts efficiency from −0.5…0% baseline to +4.0…+5%, despite no significant differences in KV40. These results emphasize that VI governs the fluid's ability to maintain optimal viscosity as the temperature rises during operation. High-VI fluids prevent excessive thinning, thereby sustaining both lubrication, internal/external sealing effectiveness in pumps and actuators.

5.3. Effect of Viscosity Grade Reduction

The data strongly support viscosity-grade reduction—provided VI is sufficiently high. Fluids with markedly reduced KV40, such as 285 VI ISO VG22 (KV40 = 26 cSt) and 189 VI ISO VG 32 (KV40 = 32.5 cSt), consistently deliver the largest efficiency gains: +13.6% and +10.7%, respectively. These improvements are attributable to significantly lower hydrodynamic losses in pump, valve components and pipes, allowing hydraulic actuators to cycle faster and with reduced energy consumption. However, these gains only materialize because VI values are high (VI = 285 and VI = 189), preventing the fluid from becoming too thin during operation that it compromises volumetric efficiency or pump protection. In contrast, fluids with moderate KV40 but insufficient VI—such as 105…112 VI—experience efficiency losses or negligible gains due to poorer thermal viscosity retention.

5.4. Role of Shear Stability in Maintaining Volumetric Efficiency

Shear loss, measured via ultrasonic shear thinning (ASTM D5621, 40 min) for fluids, having the viscosity index improvers (VIIs) ranged from 6.2% to 15% across the fluids. This type

of shear loss correspond to the middle to high tier type of the VIIs used in hydraulic fluids. The results confirm that low to moderate shear loss supports efficient hydraulic operation by preventing excessive polymer breakdown and preserving the volumetric-efficiency advantage conferred by high VI. For example, 167 VI (shear loss = 7.4%) and 155_VI (6.2%) both maintain efficiency increases near 4–5%, however the fluid with 155 VI outperformes the fluid with 167 VI while even having slightly higher KV40, indicating that shear stability remains additional decisive factor to boost efficiency as already noted earlier [6]. The overall trend shows that lower shear loss helps sustain volumetric efficiency, in high-VI fluids where fluid temporary thinning or permanent breakdown would otherwise erode the sealing benefit.

6. IMPLICATIONS ON THE TOTAL COST OF OWNERSHIP AND AVOIDED EMISSIONS

Adopting a multigrade hydraulic fluid (ISO VG 46, VI 155) in place of a monograde ISO VG 46, VI 105 yields a measurable improvement in total cost of ownership and a material reduction in emissions over a single drain interval. Under the realistic assumption of a fleet of five machines, each consuming 26 L/h of diesel and filled with 900 L of hydraulic fluid, with a 2,500 h drain interval—the empirically observed 5% efficiency gain translates directly into lower fuel use during digging, loading, and travel. Priced at €1.60 per liter of diesel, the fleet-wide fuel saving over one interval amounts to €20,883. Against this benefit, the higher upfront cost of the multigrade—€4.50/L versus €2.00/L for the monograde—requires an additional €11,250 to fill the fleet (Δ€2.50/L × 900 L × 5 machines). Even after accounting for this premium, the upgrade produces €9,633 in net savings within the same service cycle. Typically high quality premium hydraulic fluids with high viscosity index bring an extended oil drain interval, which is not accounted for in the analysis here. Because fuel savings accrue with every operating hour, the investment reaches break-even well before the end of the interval—at roughly 1,350 operating hours per machine—after which all further savings are realized as pure financial gain. The environmental dividend is equally tangible: by avoiding the combustion of 13,050 L of diesel over the interval, the fleet prevents approximately 35 t of CO_2 emissions on a tank-to-wheel basis, contributing directly to Scope 1 reductions without any hardware changes or operational retraining. Mechanistically, these outcomes reflect the superior balance of mechanical and volumetric efficiency delivered by the higher-VI formulation: reduced viscous losses at operating temperature accelerate hydraulic cycles, while sufficient viscosity retention and controlled shear thinning preserve sealing and pump health. In practical terms, switching to ISO VG 46 VI 155 delivers a rapid-payback, low-risk pathway to lower operating costs and meaningful carbon abatement, realized entirely within a single drain interval.

7. BROADER FIELD EXPERIENCE AND GENERALIZATION OF RESULTS

The excavator study presented here represents only one example within a much broader body of long-term field evidence. Over more than 20 years of active field-trial work, Evonik has collaborated with equipment owners and OEM partners

to evaluate high-VI hydraulic fluids across a wide range of mobile machinery, operating environments, and hydraulic duty cycles. These long-duration studies—performed under real-world loading, digging, carrying, and travelling conditions—demonstrate that the qualitative efficiency- and productivity-enhancing mechanisms observed in the excavator trial are consistent, reproducible, and not machine-specific.

Across various equipment types—including skid steer loaders, compact and small-size excavators, mid-size excavators, and modern electronically controlled mid-size excavators—the introduction of high-VI, shear-stable multigrade fluids has resulted in clear operational gains. Depending on machine type and work mode, efficiency improvements have ranged from 6% up to 35%, and productivity improvements from 6% up to 25–35%. These effects have been observed under diverse hydraulic tasks: excavation, loading and carrying cycles, grading, travelling, and mixed operation. Importantly, the comparisons span a variety of baseline fluids—conventional ISO 46 HM monogrades, moderate-VI ISO 46 industrial fluids, and OEM 10W hydraulic formulations—demonstrating that the relative advantage of a higher-VI, thermally robust fluid holds regardless of the starting point.

A consistent pattern emerges from these long-term trials: fluids with higher viscosity index and carefully controlled shear stability maintain optimal viscosity under highly dynamic temperature and load conditions, thereby improving both mechanical and volumetric efficiency. This leads to faster hydraulic response, reduced pump losses, improved actuator speed, and more effective transmission of engine power into productive work. Crucially, these performance gains are achieved without any hardware modification. Machines require no changes to pumps, valves, filters, cooling systems, or control logic—operators continue to work exactly as before, while the hydraulic system benefits from the improved viscometric behaviour of the fluid alone.

Equipment type	Skid Steer Loader	Compact Excavator	Small-size Excavator	Mid-size Excavator	Modern Mid-size Excavator	
Type of work	Loading, carrying & dropping	Trench digging	Loading, carrying & dropping	Digging, grading, travelling & idling	Digging	
Ref. fluid	ISO 46 VI 140*	ISO 46 VI 140*	10W genuine oil	ISO 46 HM	ISO 46 HM	
Tested fluid	ISO 68 VI 200*	ISO 68 VI 200*	ISO 46 VI 195*	ISO 46 VI 170**	ISO 46 VI 157*	ISO 32 VI 190*
Efficiency Improvement	**up to 14%**	**up to 15%**	**up to 30%**	**up to 35%**	**up to 6%**	**up to 12%**
Productivity Improvement	**up to 16%**	**up to 14%**	**up to 25%**	**up to 35%**	**up to 6%**	**up to 14%**

* shear stable fluid
** very shear stable fluid

Improvement in Efficiency and Productivity obtained through high VI shear stable fluid and appropriate ISO grade in all trials

Figure 7.1. Generalized view on mobile equipment hydraulics field trial results at Evonik across the applications.

The combined evidence of two decades of field trials provides strong confirmation that the improved efficiency demonstrated in the excavator case study is representative of a broader principle: high-VI, shear-stable multigrade fluids enable both economically and environmentally superior operation of mobile hydraulic equipment, enhancing productivity while simultaneously reducing fuel consumption and emissions. This makes fluid-based efficiency improvement an immediately deployable, low-risk, and cost-effective lever for sustainable performance enhancement across the mobile hydraulics sector.

8. REFERENCES

[1] Ahluwalia, R. K., Wang, X., Star, A. G., & Papadias, D. D. (2022). Performance and cost of fuel cells for off-road heavy-duty vehicles. International Journal of Hydrogen Energy, 47(15), 8880–8898. https://doi.org/10.1016/j.ijhydene.2022.01.144

[2] GEHL TCO Practical Guide, Online brochure accessed at 30.01.2026, https://www.gehl.com/sites/default/files/2022-10/guide_tco_interactive_en_us.pdf

[3] BDEW, Strom:Verbrauch nach Verbrauchergruppen. Retrieved from https://www.bdew.de/service/daten-und-grafiken/strom-verbrauch-nach-verbrauchergruppen/, accessed on 30. May 2025

[4] Statistisches Bundesamt. Preise für Nicht-Haushalte. Jahresverbrauch 2000-20000 MWh. Retrieved from https://www-genesis.destatis.de/datenbank/online, accessed on 25. January 2025.[4] Fleiter, T., Hirzel, S., Worrell, E., 2012. *The characteristics of energy-efficiency measures – a neglected dimension*. Energy Policy 51, 502–513.

[5] Gasoline-Germany.com, ADAC/Statista, destatis.de accessed at 30.01.2026

[6] Alibert, M. J., Görlitzer, H., Herzog, S. N., Niveau, C.D. "*Efficiency Advantages in Vane, Piston and Gear Pumps – What High VI Hydraulic Fluids Can Do for You*!", 5th International Fluid Power Conference, Aachen, 2006

[7] Alibert, M. and Schimmel, T., "*The importance of viscosity for hydraulic fluid efficiency - What can we learn from a decade of fluid development*?", 8th International Fluid Power Conference, Dresden, 2012

[8] Palm, J., & Bryngelson, E. (2023). *Energy efficiency at building sites: barriers and drivers*. Energy Efficiency, 16, Article 7.

Digitalization & Standardization: as enablers for the future of industrial hydraulics

Dr.-Ing. Mark C. Krieg & Uwe Haas

Authors affiliation, Bosch Rexroth AG Industrial Hydraulics

Abstract

The industrial hydraulics sector is undergoing a profound transformation driven by evolving customer expectations, commoditization, digital mindsets, automation pyramid restructuring, and demands for rapid time-to-benefit. This paper highlights five key trends and outlines strategic opportunities to address them, emphasizing a necessary shift from component-centric to system-oriented thinking. Analysis of recent conference topics reveals a progression from initial component digitalization to connectivity, data-based services, and finally to system integration, all converging to meet demands for sustainability, resilience, and increased automation. Three practical use cases – Digital Machine Engineering, Hydraulic Hub, and CytroConnect – exemplify proactive responses to these trends, showcasing how digital twins, accessible knowledge platforms, and predictive maintenance are being implemented. A core argument is advanced that effective cross-manufacturer collaboration, essential for navigating these complexities, critically depends on the adoption and active development of common standards. The Asset Administration Shell (AAS) is introduced as a pivotal enabler for seamless, vendor-independent interoperability. The paper concludes by stressing the imperative for industry-wide engagement in standardization initiatives (e.g., VDMA, IDTA, ZVEI) and proposes future advancements like the Asset Orchestration Platform (AOP), all built upon the foundational role of the AAS to ensure the continued competitiveness and innovation leadership of fluid power technology.

Keywords. Digitalization, Industry 4.0, Digital Twin, Asset Administration Shell (AAS), Standardization, Cross-Manufacturer Collaboration, System Thinking, Interoperability, Fluid Power Technology

1. INTRODUCTION AND OBJECTIVES

The industrial hydraulics sector is currently undergoing significant transformation, driven by several interconnected trends that present both challenges and substantial opportunities.

Firstly, we observe a shrinking pool of traditional hydraulic knowledge alongside the **unjustifiably declining reputation of hydraulics, which is perceived as a mature technology.** Furthermore, **rapidly evolving customer expectations** driven by B2C user

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

experiences set the benchmark for B2B customer requirements. This trend opens an opportunity to innovate by developing user-friendly software solutions, we can simplify hydraulics. Furthermore, leveraging deep application knowledge and the expertise of internal system integrators, combined with insights from data-driven use cases in the consumer market, will be vital for future success.

Secondly, the industry faces increasing **commoditization, Hardware-as-a-Product (HaaP) models, and regulatory protectionism.** To counter the diminishing differentiation on component level, companies have the chance to provide solutions that offer added value across the entire product life cycle. By leveraging a leading market position, businesses can foster deep customer integration and commitment and actively shape the regulatory landscape through strong market presence and networks.

Thirdly, there's a clear emergence of an **increasingly digital mindset among customers.** A new generation of clients is entering the market with less traditional application knowledge but with distinct digital expectations. This shift creates an opportunity to learn from the success of data-driven business models prevalent in the consumer market. Moreover, monetizing hydraulic know-how and software (SW) solutions will be a key strategy for growth.

Fourthly, we are witnessing the **opening of horizontal layers within the automation pyramid.** By leveraging hydraulic know-how, it is possible to expand customer relationships within the ecosystem, independent of hardware and ecosystem. Proactively co-creating ecosystems with strategic partners and integrating vertically will be essential to capitalize on this trend.

Finally, there is an overarching **demand for fast time-to-benefit.** Shorter life cycles are becoming the norm, with the "China speed" benchmark dictating rapid development and deployment. The opportunity here lies in accelerating the digitization of internal processes to boost efficiency. Additionally, developing adaptable hardware platforms and shifting flexibility into software will be critical to meet the urgent need for quick results and innovation.

2. STATE OF WORK

The main topics of presentations at conferences in recent years reveal a clear trend. Around 2020, the primary focus was on component digitization – how can sensor data be acquired and initial digital functions implemented in valves or systems. In 2022, the discussion shifted towards connectivity and initial data-based services. The aim was to network components and generate initial added value from the acquired data, such as simple condition monitoring. Conferences around 2024 increasingly addressed system integration and efficiency improvements. The focus was on how networked components interact within a system to, for example, increase energy efficiency or improve overall equipment effectiveness (OEE). Today, we must converge these lines of development to address major industry trends: the call for greater sustainability (e.g., through decarbonization), the need for resilient systems, and the increased demand for automation driven by the shortage of skilled workers. This article uses three concrete use cases to demonstrate how these trends can be addressed and

derives from this the compelling need for industry-wide, standards-based cooperation. [1][2][3]

3. METHODOLOGY AND RESULTS

3.1. View of components

We provide three practical examples in this chapter, showcasing how we have proactively responded to and integrated the prevailing trends.

Use Case 1: Digital Machine Engineering – Efficiency from the Start
This approach drastically cuts development and commissioning times by enabling continuous digital development and simulation of entire machines, including hydraulics. A digital twin of the machine is assembled from component digital twins, allowing validation of system behavior and control logic before physical assembly. This significantly reduces errors, costs, and time-to-market, building on advances in component digitalization and system integration.

Use Case 2: Hydraulic Hub – Digitalizing and Accessing Knowledge
The Hydraulic Hub is an online platform offering 24/7 access to knowledge, data, and processes for hydraulic components for service technicians and maintenance personnel. It bridges the gap between the physical and digital world, allowing users to access data sheets, maintenance instructions, and spare parts lists via unique identifiers (e.g., QR codes). AI-powered "Smart Diagnostics" aid in rapid fault identification, addressing the skilled labor shortage and trend towards data-driven services by digitally preserving and applying expert knowledge.

Use Case 3: CytroConnect – From Reaction to Anticipation
CytroConnect is an IoT solution enabling condition monitoring and predictive analytics for hydraulic systems. By analyzing real-time operational data (pressure, temperature, oil condition), it prevents unplanned downtime and allows for proactive maintenance planning. Learning algorithms assess component health and predict remaining lifespan. CytroConnect exemplifies predictive maintenance and increased plant availability, responding to the demand for greater resilience and productivity

3.2. Necessity of system orientation

The necessity to shift from component-centric to system thinking arises primarily from the increasing complexity of modern engineering projects and organizational structures. While component-based approaches optimize individual parts, they often fail to account for how these components interact and influence the behavior of the larger system. This can lead to unforeseen consequences, suboptimal performance, and the creation of new problems while attempting to solve others.[4]

System thinking, in contrast, offers a holistic understanding by considering the interdependencies and relationships among various components, stakeholders, and the environment. [4] This broader perspective enables better decision-making, as engineers can identify underlying causes of issues rather than just addressing symptoms, anticipate unintended consequences, and align technical choices with overall strategic objectives. [4]

It fosters greater resilience, adaptability, and the ability to innovate more effectively by revealing critical connections and the "bigger picture," ultimately leading to more robust and sustainable solutions.[5]

3.3. Standards: Enabling Seamless Cross-Vendor Systems

In today's complex industrial landscape, cross-manufacturer collaboration has become indispensable, and its effectiveness is intrinsically tied to the adoption of standards. Without agreed-upon standards, collaborative efforts between different manufacturers would be severely hampered by issues of incompatibility and miscommunication.

Firstly, standards ensure interoperability. When multiple manufacturers contribute components or systems to a larger solution, standards provide a common language and set of rules for how these diverse elements should interact. This allows for seamless integration, preventing the need for costly and time-consuming custom adaptations between every new combination of products.

Secondly, standards enhance efficiency and consistency. By establishing common practices, protocols, and quality metrics, standards streamline workflows across different organizations. This reduces ambiguity, errors, and redundant efforts, leading to faster development cycles and higher quality outcomes for all collaborators.

Thirdly, standards foster trust and transparency. A shared understanding of expectations and requirements, facilitated by standards, builds confidence among collaborating parties. It clarifies responsibilities, simplifies quality control, and allows for greater accountability, strengthening the partnership. [6]

Finally, standards enable scalability and innovation. With a standardized framework, new partners can be integrated more easily, and new technologies can be developed and adopted without disrupting existing systems. This accelerates innovation by allowing manufacturers to build upon a stable, common foundation rather than reinventing core functionalities with each project. In essence, standards act as the crucial connective tissue, transforming disparate components from various manufacturers into a cohesive and functional system.

3.4. The Asset Administration Shell

The Asset Administration Shell (AAS) is a standardized digital representation, or digital twin, of a physical or virtual asset, playing a crucial role in Industry 4.0. It acts as a digital identity card, structuring and providing all relevant information and functionalities of an asset throughout its entire lifecycle.

The primary benefit of the AAS is to enable seamless integration and communication between machines, equipment, and IT systems, regardless of the manufacturer. This fosters true vendor-independent interoperability, breaking down data silos and facilitating efficient and secure data exchange across different systems and company boundaries. By providing a common, machine-readable database and standardized interfaces, the AAS supports real-time data transmission, advanced analytics, and the development of new business models,

ultimately leading to optimized processes, increased efficiency, and greater flexibility in industrial production. [7] [8]

4. BRIEF INTERIM SUMMARY AND APPEAL

The preceding chapters underscore the critical necessity of collaboration to effectively navigate and address emerging industry trends. This collaborative imperative inherently dictates the joint development of common standards. Organizations must actively engage in standardization efforts within relevant industry associations. For instance, contributing to bodies like the VDMA, which considers standardization a core task for corporate management and promotes the competitiveness of the mechanical engineering industry , or the IDTA, which focuses on maintaining data protection standards and facilitating secure international data transfers Similarly, participation in the ZVEI, which sees standardization as an essential integrative task for bringing together stakeholders and accelerating the digital and green transformation , is crucial. Furthermore, formalizing commitment through signing a Letter of Intent (LoI) with organizations like the ZVEI further solidifies the dedication to such shared standardization goals. Active involvement in these and similar initiatives is not merely beneficial but essential for ensuring interoperability, fostering trust, enhancing efficiency, and ultimately driving innovation across the industry. [9] [10]

5. OUTLOOK - ORCHESTRATION OF ASSETS WITH THE AAS

For the Asset Administration Shell to reach its full potential, tools must make its use simple and efficient for a multitude of use-cases. One such tool is the Asset Orchestration Platform (AOP). This platform leverages the Asset Administration Shells (AAS) provided by component suppliers to radically simplify the integration (or "onboarding") of components into systems and the coupling of systems with each other, e.g., in a production process. Instead of manually sifting through data sheets and individually integrating a proprietary protocol, the AOP reads a component's data and capabilities directly from its AAS and orchestrates its interaction within the system – a "plug-and-produce" approach for industrial applications.

6. SUMMARY

Fluid power technology demonstrates its innovative strength through the consistent development of digital solutions. Systematic monitoring of trends and their driving factors (like in chapter 1 described) is essential to serve as a basis for deriving necessary strategic measures. The path to a successful future is clearly defined and based on three inseparable pillars:

Defining the standard together: With AAS, driven by IDTA and Fluid 4.0, the standard for the digital twin has been defined.

Consistently using the standard: Use cases such as Digital Machine Engineering and solutions like the Hydraulic Hub and CytroConnect, supported by tools like AOP, translate the standard into practice.

Collaboration: Industry-wide cooperation to establish these standards ensures the competitiveness and innovation leadership of the entire sector.

By embarking on this path together, fluid power technology is ideally equipped for a successful and profitable future.

7. REFERENCES

[1] S. Haack, M. Krieg “Industrial Hydraulics: Now – Next – Beyond” in IFK2020 Dresden 2020

[2] M. Krieg “Digitalization in industrial hydraulics – added values for machine builders and their customers” in IFK2022 Aachen 2022

[3] M. Krieg “Software-defined Industrial Hydraulics” in IFK2024 Dresden 2024

[4] Tim Hardwick is a Strategy & Transformation Consultant specialising in Technology Strategy & Enterprise Architecture (2023/07) in Harnessing Complexity - The Power of Systems Thinking in Engineering at https://quantumfields.co.uk/architecture/harnessing-complexity-the-power-of-systems-thinking-in-engineering (30.01.2026)

[5] “7 Reasons Why Systems Thinking is Important for Innovation in Organizations” (2024/03) at https://www.6sigma.us/systems-thinking/why-is-systems-thinking-important/ (30.01.2026)

[6] “Benefits Of Standardization In Collaborative Processes” at https://fastercapital.com/topics/benefits-of-standardization-in-collaborative-processes.html (30.01.2026)

[7] IDTA “Digital Twin – the heart of Industry 4.0” at https://industrialdigitaltwin.org/en/technology (30.01.2026)

[8] ZVEI “Examples of the Asset Administration Shell for Industrie 4.0 Components” 04/2017 https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2017/April/Asset_Administration_Shell/ZVEI_WP_Verwaltungschale_Englisch_Download_03.04.17.pdf (White Paper) (30.01.2026)

[9] ZVEI “Policy paper on standardisation” (11/2021) at https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2021/November/Grundsatzpapier_Normung/Policy-paper-on-standardisation-2021-11-10.pdf (30.01.2026)

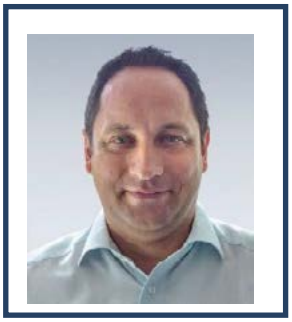
[10] Helga Jost “Standardization Key tool for the mechanical engineering industry” at https://www.vdma.eu/documents/34570/4877456/Flyer+Standardization+EN.pdf/c92e66a2-47f1-eeae-3759-6fab407b49f9?t=1648036762192 (30.01.2026)

Biographies



Dr.-Ing. Mark C. Krieg VP Development Business Unit Industrial Hydraulics Bosch Rexroth AG

Dr. Mark Krieg has been Vice President Development of the Business Unit Industrial Hydraulics since February 2018. Main R&D focus areas are industrial valves and controls, cylinders and servo-hydraulic actuators, power units and overall providing additional customer values by digitalization.
Prior to Bosch Rexroth he was responsible for the development of hydraulic pumps and motors within Parker Hannifin. He earned his Ph.D. degree in engineering at the Fraunhofer-Gesellschaft.



Uwe Haas Business Digital Officer Business Unit Industrial Hydraulics Bosch Rexroth AG. He has held his current position since March 2025. He joined Bosch Rexroth in 2012, taking on various roles within the company's digitalization departments. He began his career as a Consultant within a Management and IT Consulting company, after earning a Diploma in Computer Science (Dipl.-Inform. (FH)).

Machine Learning- Based Health Monitoring of High-Power Electro-Mechanical Actuator for Non-Road Mobile Machinery Using Experimental Data

Vinay Partap Singh[1,2,*], Tatiana Minav[1]

1 – IHA lab, Faculty of Engineering and Natural Sciences, Tampere University, Finland.

2 – Danfoss Power Solutions, Nordborg, Denmark.

** - vinaypartapsingh@tuni.fi , vinay.singh@danfoss.com*

ORCID - https://orcid.org/0000-0002-0699-2712

Abstract.

This paper presents a deployable and interpretable framework for the health monitoring of high - power electro-mechanical actuators (EMAs) used in non-road mobile machinery (NRMM), turning standard sensor data into compact predictors, a health index, and finally a fused ensemble that tracks degradation for real-time maintenance.

Using experimental run-to-life data, four correlation-based coefficients identified in a previous study are computed to capture cross-sensor dynamics. An early-life baseline is established, and a bounded, monotone health index is derived using Mahalanobis distance-based deviation, persistence gating, and outlier-robust accumulation. Several machine learning regression models are trained and compared; three complementary models—Gaussian process regression, a bagged-tree ensemble, and a compact multilayer perceptron are selected for further process.

The models are then deployed in Simulink where raw sensor signals are transformed into correlation-based coefficients, which feed three trained regressors. The outputs of these individual regressors are fused via a performance-weighted scheme and smoothed with a rolling-window average to produce a bounded health index. In a separate 200-cycle dataset withheld from all training, the Simulink implementation tracks degradation accurately and suppresses transients. After an initial overshoot due to a shorter adaptive window, the HI remains close to 0.4 on a scale of 0 to 1, which is within the expected range for the withheld data from the HI map of the training data. The fused HI remains interpretable, monotonic in expectation, robust to brief operational disturbances, and consistent with training behaviour, and providing a practical basis for real-time maintenance decisions for high-power EMAs used in NRMM.

Keywords. Electro-mechanical actuators, non-road mobile machinery, machine learning regression, health monitoring.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

ACRONYMS

- **EMA** – Electro-Mechanical Actuator
- **NRMM** – Non-Road Mobile Machinery
- **HI** – Health Index
- **MD** – Mahalanobis Distance
- **ML** – Machine Learning
- **MLP** – Multi-Layer Perceptron
- **GPR** – Gaussian Process Regression
- **RMSE** – Root Mean Square Error
- **FDI** – Fault Detection and Isolation
- **PHM** – Prognostics and Health Management
- **PRV** – Pressure Relief Valve
- **UI** – User Interface
- **EDA** – Exploratory Data Analysis
- **STD** – Standard Deviation
- **LOF** – Local Outlier Factor
- **MCSA** – Motor Current Signature Analysis

1. INTRODUCTION

Electro-mechanical actuators (EMAs) are rapidly being proposed for adoption in non-road mobile machinery (NRMM) due to their combination of high energy efficiency, precise motion control, and compatibility with the broader trend toward the electrification of heavy machinery. Unlike hydraulic systems, EMAs convert electrical energy directly into mechanical motion without the need for pressurised fluid, reducing energy losses and eliminating the complexity of pumps, valves, and hoses. This not only improves system level efficiency but also simplifies integration with modern, software-based control architectures that are increasingly prevalent in NRMM [1].

Despite these advantages, EMAs in NRMM operate under harsh and variable conditions that can accelerate component degradation. High mechanical loads, impact forces, and repetitive duty cycles impose wear on mechanical components, such as screw drives, bearings, and gears. Moreover, exposure to dust, moisture, and temperature extremes can degrade lubrication, increase friction, and cause corrosion. The electric motor and power electronics are also subject to thermal stress, insulation breakdown, and electrical component aging. Importantly, the majority of these degradation processes are not immediately apparent in the output of the actuator until a critical threshold is crossed, leading to sudden and unexpected failures. In contrast to hydraulic systems where issues, such as leaks or pressure drops, often provide an early warning, EMA faults can progress silently, making them more difficult to detect without continuous monitoring.

Research into health monitoring and prognostics for EMAs has grown steadily over the last few years for NRMM, with interest driven largely by aerospace applications and the broader adoption of more-electric systems. Early work in this area focused primarily on fault detection and isolation (FDI), aiming to identify actuator malfunctions as they occur rather than tracking gradual degradation. For example, a comprehensive prognostics and health management (PHM) framework for EMAs is developed by [2] by integrating model-based

fault detection with a data-driven prognosis. Their system was validated on aerospace representative actuators, demonstrating accurate isolation of injected faults, such as motor winding failures and gear jams. However, as with much of the early PHM research, this work relied heavily on simulated fault scenarios and laboratory testbeds operating under simplified duty cycles, limiting its applicability to the variable, high-load operating environments found in NRMM. Subsequent studies extended EMA health monitoring by incorporating mechanical degradation mechanisms into the analysis. For example, the work on screw backlash and fatigue estimation examined the evolution of mechanical wear in transmission components, particularly in ball-screw assemblies [3]. The aforementioned study modelled wear progression and fatigue using analytical relationships and parameter estimation techniques, sometimes coupled with laboratory experiments under controlled loading. While they provided important insight into mechanical degradation pathways, they typically did not operate under the diverse and harsh conditions typical of NRMM service. A recurring theme across the literature is the use of synthetic or fault-injection data in place of genuine run – to - failure datasets.

A test rig has been developed for the experimental evaluation of a flight control EMA, with the ultimate goal of developing health monitoring algorithms in [4]. However, the actuator used in this study has a nominal load capacity of 300N. Furthermore, to generate the test data, the test rig has been subjected to axial loads up to 70% higher than nominal in order to accelerate degradation. The same research group proposes a general health monitoring approach for EMAs based on statistical process control methods using electric motor currents in [5]. In that study, the actuator does not reach failure, but testing was stopped at an event frequency of 60%. Another paper proposes a change detection algorithm for health monitoring of the same small EMA [6]. This algorithm compares the statistical features of the current experiment with past ones in batches, rather than providing an online estimation. Hence, the algorithm assumes that the external conditions on the EMA remain the same each time it is tested, which limits its application for real-time health monitoring. Two points stand out in the last three studies mentioned: first, the size of the EMA, with a nominal load capacity of 300 N, which is significantly low for any work function of NRMM; second, the assumption that external conditions (i.e., load, disturbances) remain the same during all test sessions, which is highly unlikely for any NRMM. The recent study by [7] notes that many approaches to actuator health assessment rely on artificially induced faults, such as deliberate increases in backlash or step changes in motor parameters, to generate degradation data for algorithm testing. While useful for algorithm development, these data often represent abrupt failures rather than the gradual degradation experienced in the real-life operating conditions. Furthermore, artificially induced faults can bias algorithms toward detecting the specific fault types injected, reducing generalizability.

In terms of sensing and feature extraction, multiple studies have explored multivariate feature spaces derived from actuator signals. These include time-domain statistics (i.e., mean, RMS, skewness, and kurtosis), frequency-domain parameters from vibration or current spectra, and residuals from physics-based models. For instance, [8] investigates condition monitoring through multiple sensor channels, applying statistical analysis to extract discriminative features. Similarly, the work in [9] discusses combining motor current signature analysis (MCSA) with vibration and temperature data to detect early anomalies. These approaches underscore the potential of feature fusion for more reliable health estimation. Feature extraction and multi-model fusion have been used for condition

monitoring of an EMA in a train transmission in [10]. The study employs correlation-based feature selection and, through a Multimodal Fusion Improved Transformer Network, achieves better accuracy compared to traditional models. However, the study uses a 1:2 scaled-down experimental setup, and faults are induced manually rather than through natural degradation. That said, in most cases, the outcome is an anomaly score or classification label, not a continuous, interpretable health score.

The use of advanced machine learning (ML) methods for EMA health monitoring has also been investigated. Supervised learning techniques are applied in [11] to predict the actuator condition from sensor features. While these methods demonstrate strong offline performance, they are often computationally intensive, involving high-dimensional feature vectors and complex nonlinear models. Such complexity poses challenges for embedded real-time deployment, especially in NRMM applications where on-board computational resources may be limited. Moreover, many ML-based approaches in the literature are trained and validated on datasets that do not represent the full life cycle of the actuator, making them less robust to long-term drift and variability in operating conditions.

Another limitation is the focus on small to medium-sized EMAs, especially in aerospace and robotics contexts. In contrast, high-power EMA of the type used in heavy construction, mining, and agricultural machinery exhibit different thermal, mechanical, and electrical behaviours due to their larger components, higher load levels, and more severe environmental exposure. This gap is particularly significant given that high-power EMAs are now being deployed more frequently in NRMM applications, where predictive maintenance is essential.

This highlights a clear research gap where work on the health monitoring of high-power EMA working on duty cycles of NRMM is lacking so far.

This work directly addresses this gap by developing a health index for a high-power EMA using exclusively experimental degradation data from a long-term run-to-failure test. The method fuses multiple statistical features into a single normalised scale, enabling continuous tracking of actuator health. This work proposes a framework for constructing and predicting a health index for high-power EMAs operating under realistic duty cycles using experimental data. The starting point for this study is our previous research, in which a systematic feature selection procedure identified four statistical parameters from multi-sensor EMA data that exhibit consistent and monotonic degradation trends across the life cycle of the actuator [12]. Derived from load force, actuator velocity, and electric motor torque, these features form the basis for the HI construction.

A key challenge in this study is the absence of a clearly observed zero-health failure point. During the experiments, the actuator had an unexpected mechanical failure which is not directly related to the gradual degradation [12], which prevents the direct mapping of sensor data to a complete health trajectory from 1 (perfect condition) to 0 (end of life). To address this, the current study employs the Mahalanobis distance method [13, 14], which has been widely adopted in condition monitoring for combining multiple correlated features into a single metric that quantifies deviation from the healthy baseline. MD is advantageous in this context because it is scale-invariant, accounts for feature correlations, and produces a monotonic measure of degradation when applied to well-chosen features. The HI is then

obtained by mapping MD values to a normalised scale where '1' corresponds to the healthy state and '0' to the end of life.

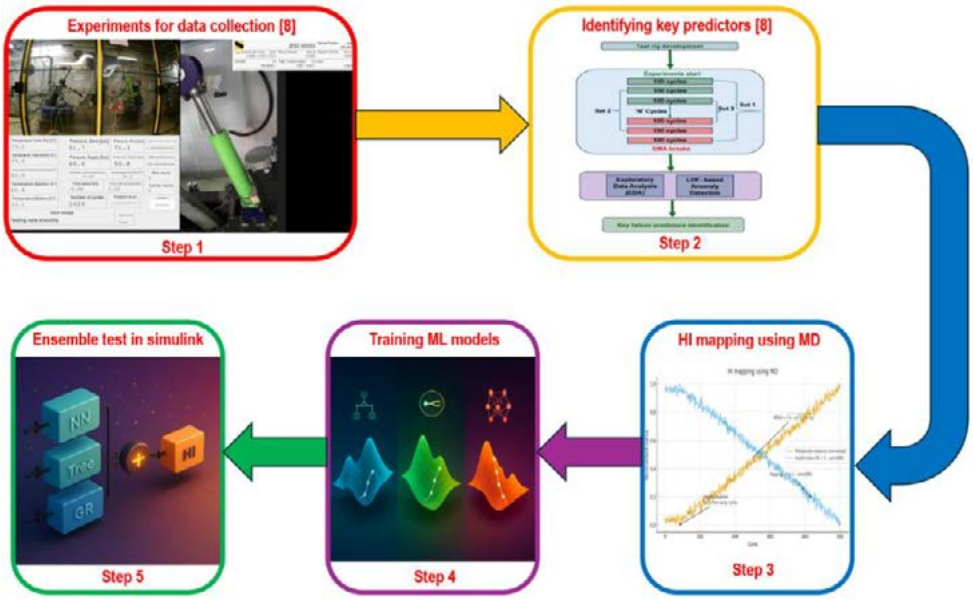


Figure 1. Overview of the study methodology

Figure 1 shows the overall process used for the study. In the above figure, Step 1 and Step 2 are covered in detail in our earlier study [12], which outlines the test rig, experiments, and the identification of key predictors of degradation. The current study carries forward the results from Step 2 and converts the key predictors from Step 2 into a score for HI using the MD. In Step 4, using MATLAB, several ML regression models are evaluated, and finally, the three best performing models in terms of the accuracy and size of the model are selected. These are a tri-layer feedforward neural network, a Gaussian process regression, and a bagged ensemble of decision trees. Each model is trained on the majority of the dataset, with a small, randomly selected segment of 200 cycles withheld for independent validation. Finally, in Step 5, the selected models are integrated into a Simulink-based ensemble, where their predictions are combined through a weighted moving average. The moving window length adapts over time (2, 10, 50, and 100 cycles), enabling early predictions with limited data and improving the stability as more observations become available. This adaptive windowing ensures that actionable health estimates can be generated immediately after startup, without waiting for extended operating histories. Finally, the ensembled model is tested in Simulink using the withheld 200-cycle experimental dataset, representing data that the models have never encountered during training.

The remainder of the paper is structured as follows. The next section of the paper introduces the test rig, the experimentation, and the key predictors of degradation identified in the previous study. Section 3 provides a detailed description of the data labelling process for the health index using MD. Section 4 then presents the training of machine learning regression models. Section 5 elaborates on the formation of the ensemble and its implementation in

Simulink as well as the testing of the framework on randomly excluded data. Finally, Section 6 concludes the paper with remarks and outlines potential future research directions.

2. TEST RIG, EXPERIMENTS, AND KEY PREDICTORS OF DEGRADATION

The long run-to-failure test was conducted on a purpose-built, seesaw-type test bench at Tampere University. The EMA under test is a ball-screw linear actuator driven by a servomotor and loaded by a hydraulic cylinder whose chamber pressures are regulated by electronically controlled pressure-relief valves (PRVs). In combination with anti-cavitation valves, the PRVs let the hydraulic load act as a passive brake in both motion directions: as the EMA moves, pressure builds in the opposing chamber until it exceeds the PRV cracking pressure, producing a controllable, bi-directional resistive force. The test rig with control and monitoring UI is shown in Figure 2 [12].

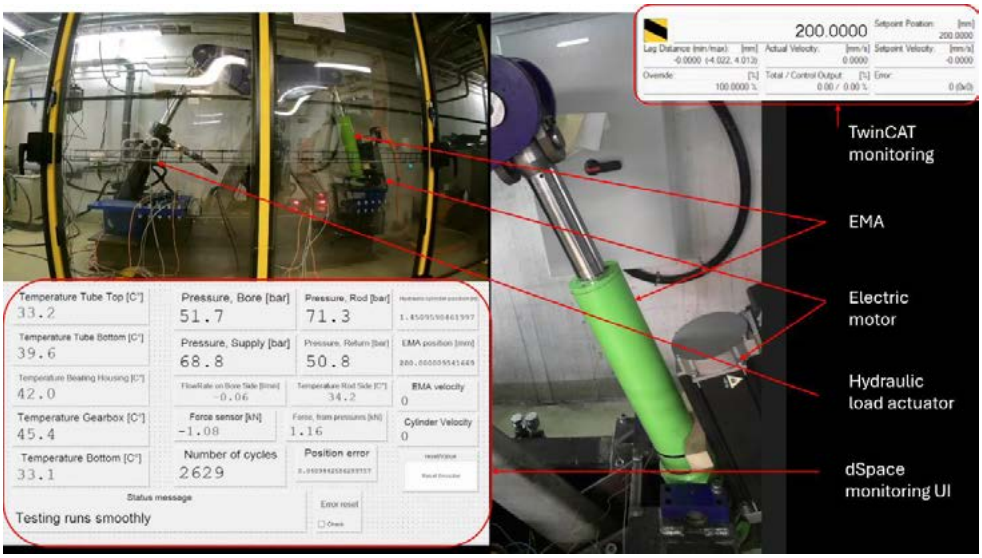


Figure 2. Test rig with monitoring UI [12]

A dedicated 20-bar PRV in the tank line maintains sufficient backpressure to overcome pressure drops across the anti-cavitation valves, and additional passive PRVs protect the circuit against fault-induced over-pressurisation which is important for safe unattended operation. The tested EMA has a 500mm stroke and is powered by an electric motor of 11.62 kW, and a nominal torque rating of 37 Nm at 3000 rpm. Whereas the hydraulic cylinder for loading is 150/80-550 mm and is controlled by an electronic PRV with a maximum operating pressure of 350 bar and maximum flow of 120 L/min.

The test rig has been fitted with numerous sensors on both sides, the hydraulic loading side and tested EMA side. On the tested EMA side, five temperature sensors are placed on different locations on the EMA, and the electric drive is used to measure the position and speed of the EMA as well as the torque and speed of the electric motor. On the hydraulic side, both the chamber pressures are measured, and a 250kN bi-directional load cell records the external force; the calculated cylinder force from these pressures is used to cross-validate

the load-cell readings. The hydraulic cylinder position is monitored using a draw-wire position sensor and used as a safety interlock against drifting to stroke limits by comparing it with EMA position readings. Data collection and the control are split between a dSPACE MicroLab Box (hydraulic load control and centralised logging) and Beckhoff TwinCAT (electric motor control), with dSPACE recording all signals. The test rig was built to operate unsupervised and work 24/7 and hence fitted with numerous safety features against any accidents and failures.

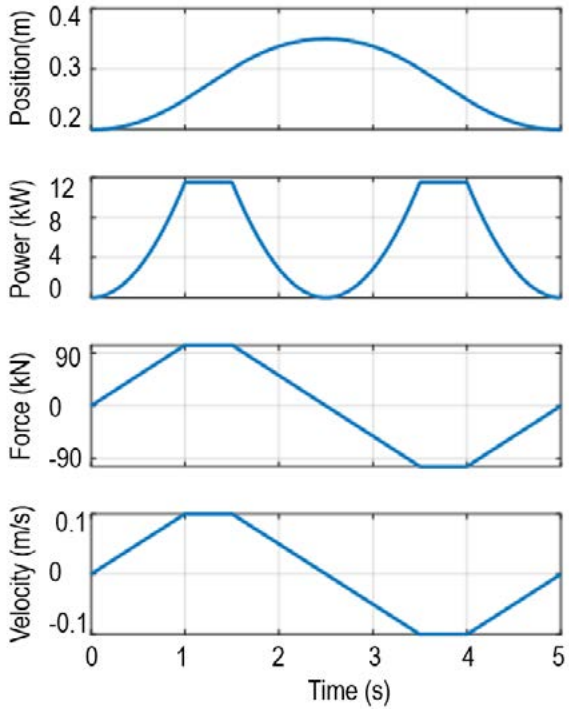


Figure 3. Utilized duty cycle for testing

One cycle for the tested duty cycles is shown in Figure 3. The duty cycle, designed with industrial partners, is an aggressive push-pull sequence: the actuator is driven at its maximum 100 mm/s within a mid-stroke window under a targeted load of 95kN. The bench is capable of regenerative energy via a proportional valve, but regeneration was not used in the current tests. As there is no active cooling, a 20-second pause follows each cycle to allow heat to conduct from internal interfaces (e.g., the screw nut) to the monitored surfaces; automated interlocks pause operation above safe temperature limits and resume once conditions normalise.

2.1. Key predictors of degradation

To quantify how degradation emerges in statistics rather than in any single raw channel, our previous study [12] compared early-life versus late-life operation using three matched pairs of datasets: (i) first 100 cycles versus last 100 cycles; (ii) second 100 cycles versus second-to-last 100 cycles; and (iii) third 100 cycles versus third-to-last 100 cycles. A sliding window was applied within each dataset to compute the summary features-mean, standard deviation

(STD), and skewness on the principal signals of load force, motor torque, and EMA velocity. Correlations among these windowed features were then evaluated, emphasising changes from the beginning to the end of life.

The exploratory data analysis (EDA) revealed consistent weakening of several cross-signal relationships as the actuator wore. Specifically, correlations that were strong and positive early on diminished markedly by the end of life. Complementing EDA, we performed local outlier factor (LOF) anomaly detection directly on the raw signal pairs (load - torque) and (load- velocity) which showed that late-life data occupies regions farther from the "healthy" cluster, with heavier distribution tails toward high anomaly scores [12].

From nine candidate correlation metrics (i.e., combinations of feature types across signals), four emerged as key predictors of degradation because they changed monotonically and consistently across all three dataset pairs:

1. **Correlation of the STD of load force with the STD of motor torque (C1)**
2. **Correlation of the STD of load force with the STD of actuator velocity (C2)**
3. **Correlation of the skewness of load force with the skewness of actuator velocity (C3)**
4. **Correlation of the mean of load force with the mean of actuator velocity (C4)**

In the present study, these four correlations are employed as key predictor signals to map the actuator's health index and to train the ML regression models.

A more detailed description of the test rig, experiments, and the procedure used to obtain the four signals lies outside the scope of the present study, as the primary objective here is to develop a health monitoring framework based on the results of our previous work. For comprehensive details regarding the test rig, experimental procedures, and signal acquisition process, the reader is kindly referred to our earlier study [12].

3. HEALTH INDEX AND DATA LABELLING

This section describes the construction of a scalar health index utilised to label data for supervised ML regression. The procedure operates on cycle-segmented signals and the four correlation-based predictors defined in Section 2, converting their windowed behaviour into (a) a multivariate deviation from a healthy baseline, (b) a smooth exceedance measure with persistence control, (c) a cumulative damage variable with outlier-robust scaling, and (d) a bounded HI suitable for training and evaluation.

The three signals of load force, actuator velocity, and electric motor torque are segmented into cycles using motion and reversal logic; cycles with insufficient samples are discarded to maintain statistical reliability (minimum sample count specified in Table 1). Low-activity cycles are filtered by an activity threshold defined as a fraction of a healthy reference quantile, ensuring that correlation estimates are not biased toward zero by quiescent operation. Within a rolling window of fixed length (see Table 1), it computes Spearman rank correlations for the predictor pairs defined in Section 2.

For distinct ranks, the coefficient ρ is

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \quad (1)$$

Here d_i is the rank difference and n is the number of observations in the window.

To quantify degradation, the Mahalanobis distance was employed to measure deviations from baseline correlations:

$$MD = \sqrt{(X - \mu)^T \sum^{-1} (X - \mu)}, \quad (2)$$

Here X is the current correlation vector, μ is the baseline mean, and $\sum$ is the covariance matrix. MD was transformed into an exceedance score using a sigmoid function:

$$e = \frac{1}{1 + e^{-\frac{MD - z0}{k}}}, \quad (3)$$

Here *Z0* sets the sensitivity threshold, and k controls the sigmoid steepness. Persistence gating was applied requiring exceedances in at least 4 of the last 7 cycles to reduce false alarms.

Damage D accumulation used an exponential relation:

$$D = \sum e^{\beta(MD - z0)}, \quad (4)$$

Here β regulates the exponential rate, ensuring smooth, monotonic accumulation. Damage was capped at the 95th percentile to avoid extreme outliers, Finally Health Index is:

$$HI = 1 - \frac{D}{D_{95\%}} (1 - h_{min}), \quad (5)$$

Here, h_{min} denotes the minimum allowed HI.

Table 1 summarises the notation and hyperparameters used in this study. While the equation coefficients *(e.g., ρ and MD*) are computed dynamically from the data, the hyperparameters (e.g., *β, z0 and k)* were optimised through a grid search in MATLAB and cross-validation within literature-based ranges. The final selected values are reported in Table 1.

Table 1. Hyperparameters used to calculate HI

Parameter	Description	Value
MinSample	Minimum data samples per cycles to eliminate bad cycle data	500
Z0	Exceedance sensitivity threshold	3
k	Sigmoid Scaling	0.25
β	Damage exponential factor	1.2
h_{min}	Minimum HI	0.20

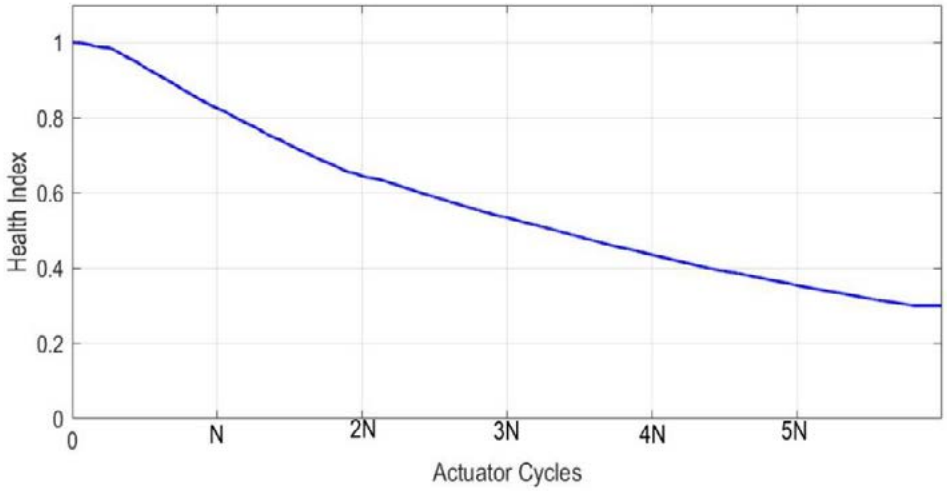


Figure 4. Health index mapping with actuator cycles

Applying the above pipeline to our full run-to-failure dataset gives a strictly non-increasing HI trajectory. The HI values remain near unity during early operation, decline gradually with time without collapsing to zero. This trajectory is presented in Figure 4, where the final HI is mapped with the number of cycles of the actuator. To respect the confidentiality constraints of the manufacturer of the tested EMA, the abscissa is de-identified: the numerical cycle count is replaced with a non-numeric value. The data of 200 cycles was withheld from the entire process to test the framework by deploying in Simulink at a later stage. The withheld data was from the dataset between the 4N and 5N points in Figure 4.

The monotonic decline of the HI demonstrates that the proposed methodology reliably captures progressive degradation of the actuator. Importantly, the absence of abrupt fluctuations indicates robustness against noise and transient effects in the sensor data. The gradual convergence towards a lower bound further confirms the stability introduced by the chosen hyperparameters and strategy. These characteristics collectively ensure that the HI provides an interpretable and reproducible representation of actuator health over its lifecycle.

It shall be noted that the proposed method, the selected hyperparameters, and the resulting HI trajectory are only for the present dataset and experimental conditions. We do not claim this representation to be a universal standard, as the actual health evolution of other actuators may differ. The results shown here are derived from a single run-to-failure experiment and therefore should be interpreted as an illustrative case study rather than a generalised benchmark. Neither this should be interpreted as the actual HI of the tested actuator, since the actuator failure was not directly linked to the gradual degradation, as explained earlier in Section 1.

4. TRAINING OF MACHINE LEARNING REGRESSION MODELS

The data and the HI prepared in Section 3 is used for the training of the ML regression models to estimate the HI directly from the four correlation-based predictors (Section 2).

The training data is divided into 90% for training in a 5-fold cross-validation and 10% for testing. Here, it is worth noting that this 10% testing data is different from the 200 cycles kept exclusively and randomly outside from the dataset for final testing in Simulink at later stages. Several model families were explored, and the three best performers were selected for complementary strengths are selected finally. These models are a Gaussian process regression, bagged decision tree ensemble, and feed-forward neural network.

Gaussian Process Regression (GPR): This provides flexible, non-parametric modelling with built-in uncertainty quantification which means it predicts behaviour while also estimating confidence in the prediction. The trained model has a zero-mean basis and an isotropic exponential kernel. Kernel scale and signal variance were determined by marginal-likelihood optimisation in the MATLAB regression learner app. The optimised sigma value is 1.7054e-4 which is obtained by the Bayesian optimiser of the MALTAB. The model achieved a test **RMSE** of **0.0254**, the lowest error among all candidates. The parity plot shown in Figure 5(A) shows tight adherence to the identity across the full health range with a minimal bias at the extremes, indicating an accurate and smoothly varying fit.

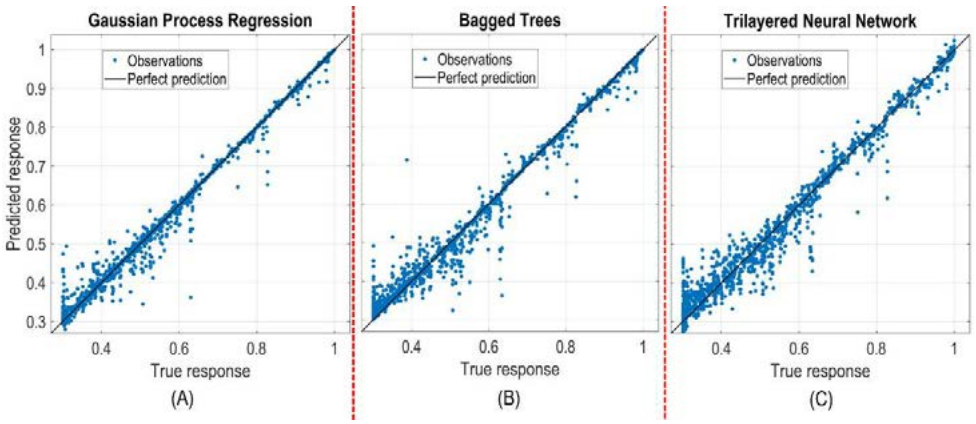


Figure 5. Regression models and predicted versus true response

Bagged Decision Tree Ensemble: An ensemble of 50 bootstrap-aggregated regression trees was fitted, with all predictors available at each split and a minimum leaf size of 1. No input standardisation was required. Bagging reduces the estimator variance by averaging high-variance trees, yielding a robust predictor that captures local nonlinearities without strong parametric assumptions. On the test set the ensemble delivered an **RMSE** of **0.0342**, the second best among the three used models. The parity plot shown in Figure 5(B) exhibits near-linear alignment with a modest dispersion in the mid-range.

Feed-Forward Neural Network (MLP): The last selected and trained model was a compact multilayer perceptron with three fully connected layers (sizes 60–300–50) and ReLU activations. Inputs for this were standardised, and a regularisation strength (Lemda) of 8e-5 was applied. Training used a maximum of 2000 iterations with cross-validation guiding early stopping and hyperparameter choice. The network provides a sufficient capacity to model smooth nonlinearities while remaining small enough to generalise from a modest feature set. On the test set, it achieved an **RMSE** of **0.0378**. The parity plot in Figure 5 (C) shows accurate tracking the near high and low ends of the HI with slightly conservative

predictions in the lower-mid range. Owing to its compact size, the MLP offers very high inference throughput and a small model size, which are attractive for embedded deployment and closed-loop evaluation.

5. ENSEMBLE AND IMPLEMENTATION

The three regressors developed in Section 4 were deployed in Simulink as a real-time estimation chain that converts raw measurements into an operational HI. The architecture shown in Figure 6 implements three independently trained regression models as parallel branches in a Simulink environment. The same cycle segmentation, activity gating, and rolling-window settings used during labelling are embedded in this feature block to avoid a train–test mismatch. An upstream feature-extraction block processes the raw sensor signals of load force, actuator velocity and motor torque to compute the four correlation coefficients (C1–C4) described in Section 3. Once computed, the four features are fed concurrently into the three trained regression models explained in the previous section. Each branch is encapsulated as a standalone block with the exact hyperparameters learned during offline training; the blocks run synchronously on the same time steps, ensuring that all model outputs correspond to the same cycle. The modular design also simplifies diagnostic logging and makes it straightforward to replace or update individual regressors without affecting the rest of the system. Collectively, the three models provide redundant health estimates derived from complementary inductive biases, nonparametric kernel regression, variance-reduced ensemble partitioning, and parametric deep learning, thus meeting reliability requirements for critical machinery.

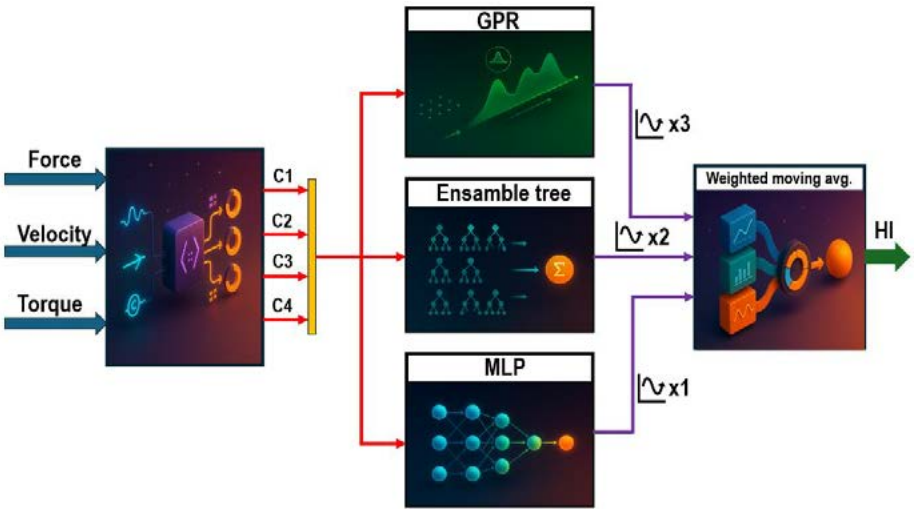


Figure 6. Simulink-implemented ensemble architecture

All three regressors are retained to exploit complementary inductive biases, i.e., kernel smoothing (GPR), partition-based nonlinearity (trees), and universal function approximation (MLP), so that the fused estimator is more robust to regime shifts and single-model failure modes than any constituent alone. Each model returns an instantaneous HI estimate at every cycle as shown in Figure 7. The GPR is the most accurate and stable across the range; the tree ensemble is competitive but slightly more variable in the mid-range; and the MLP reacts

fastest but exhibits significant transients during the first few cycles, including a brief overshoot above the nominal upper bound of 1. As the operation proceeds, all three models converge toward consistent estimates in the mid-range. Although the MLP (neural network) exhibits transient overshoot at the beginning, its high sensitivity to rapid nonlinear changes increases ensemble diversity and early-change detectability, while the performance- and interpretability-weighted fusion prevents it from dominating, yielding a more resilient and stable estimate overall, as explained next in the section.

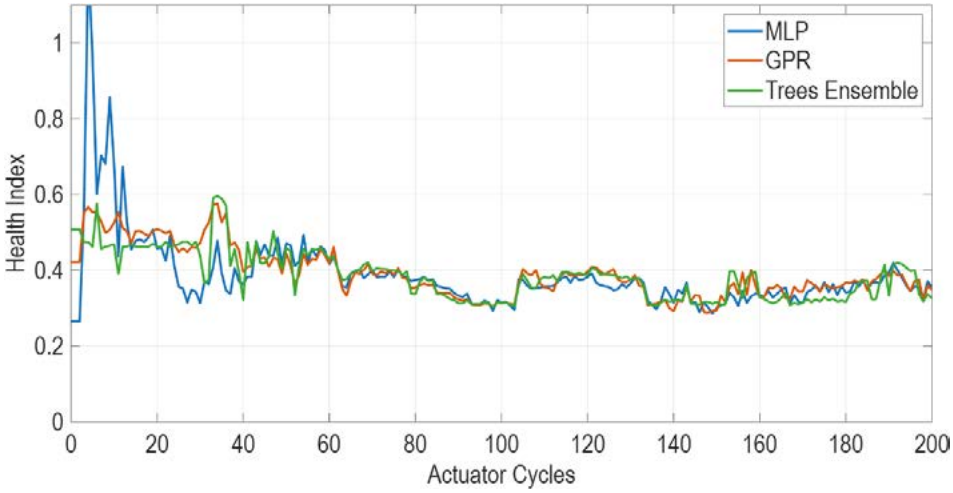


Figure 7. Individual output of instantaneous HI of regression models

The next step assigns fixed model weights to combine the three instantaneous estimates into a single operational HI on a rolling weighted average window. Weights were chosen using two principles. The first is predictive performance in the training stage in Section 4, where the GPR achieved the lowest RMSE, followed by the bagged trees, and finally the MLP. The second is model transparency at the global level, which is operationally valuable during deployment and auditing: GPR provides a principled predictive variance that can be surfaced as a confidence indicator; the tree ensembles support aggregated importance diagnostics and split-wise summaries; and the MLP, while fast, is less transparent as commonly called a black-box model [15]. Balancing these considerations, the fusion uses weights of 3:2:1 for GPR, trees, and MLP, respectively. This choice privileges the most accurate and interpretable models while retaining the responsiveness of the MLP as a minority voice that can nudge the fused estimate when rapid, genuine changes occur.

Since health evolves slowly while measurements and model outputs exhibit short-term variability, the fused weighted signal is smoothed by an adaptive moving-average window whose length increases as operational data accumulates. Immediately after start-up, only a handful of cycles are available; a 2-cycle window is therefore used to produce actionable estimates with minimal latency. This window maximises responsiveness but carries the largest variance, so it is used only long enough to cover the very early regime. Once sufficient data have accumulated, the window expands to 10 cycles, which reduces variance while still allowing the estimate to follow any real operating-regime transitions. After further operation, the window grows to 50 cycles, where the estimator begins to align with the slow time scale of mechanical wear rather than with short-term fluctuations. Finally, it settles at

100 cycles for steady operation, providing strong suppression of residual noise and model idiosyncrasies and enforcing temporal coherence consistent with gradual degradation. Transitions occur only when the next window can be fully populated with sufficient cycle count. Importantly, this adaptive smoothing acts after the instantaneous, weight-based fusion: the correlation windowing used for label construction in Section 3 is kept fixed and separate, ensuring methodological consistency while giving the online estimator the flexibility to manage its own bias–variance trade-off as evidence accumulates.

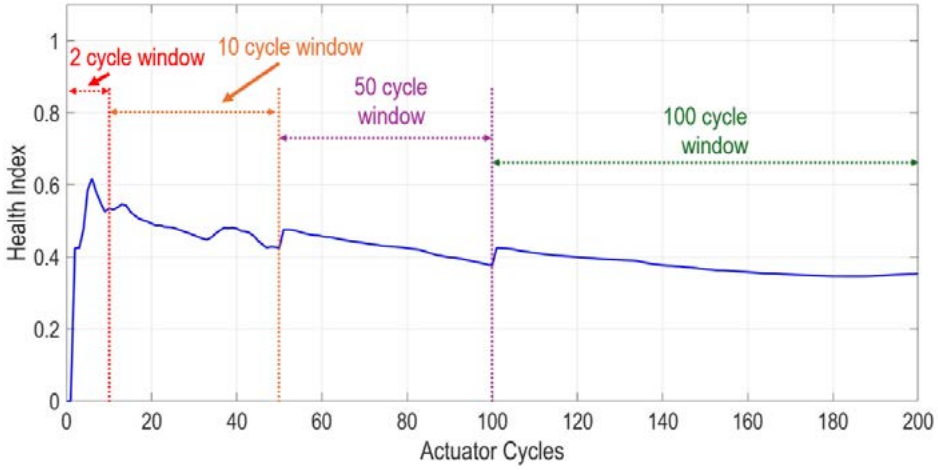


Figure 8. Final weighted average HI with varying window

Figure 8 presents the final HI after the fusion of three independent model estimations over the rolling window average. Vertical markers in the figure indicate the scheduled transitions of the adaptive window (i.e., 2→10, 10→50, and 50→100 cycles). Small, short-lived impulses appear at these markers due to a reweighting toward longer-horizon averages; they are not indicative of mechanical self-healing. Apart from these controlled adjustments, the fused curve is smooth, monotone in expectation, and bounded within the operational range. Here it should be noted that all results use a contiguous 200-cycle segment from the middle of the dataset that was randomly withheld from training and cross-validation; none of the models had prior exposure to these data. The response in Figure 8 remains within the interval expected from the held-out data shown in Figure 4. To protect the industry partner's confidentiality, the exact numerical range is intentionally omitted in Figure 4; however, the 200-cycle segment analysed here lies between the 4N and 5N markers in that figure, which corroborates the HI produced by the Simulink-deployed ensemble of Figure 8. At start-up, when the rolling window is short the fused HI is slightly elevated, driven by the initial peaks of the MLP visible in Figure 7. As the window length increases and more cycles are incorporated, the estimate stabilises and settles squarely within the 4N–5N interval, i.e., the region to which the test segment belongs.

Taken together, the Simulink ensemble provides a timely and stable HI estimate that generalises to previously unseen operating segments, reconciling early responsiveness with long-horizon coherence. These results validate the end-to-end pipeline—from feature extraction through weighted fusion with adaptive averaging, as a practical basis for real-time monitoring and maintenance decision support for high power EMAs.

6. CONCLUSION

This paper presented an end-to-end framework for the health monitoring of a high-power electro-mechanical actuator, from experiment design to real-time deployment. Building on correlation-based predictors established in our previous study, we defined a scalar health index that maps windowed feature relationships to a bounded health scale via an activity-gated, Mahalanobis-distance pipeline with persistence and damage accumulation.

Using the HI label and predictors, the ML regression models are trained, and the three best performing models are selected for their complimentary advantages. The selected regression models are a Gaussian process regression, a bagged-tree ensemble, and a tri-layered neural network. The three models were then imported into Simulink and fused by a weighted moving average that emphasises accuracy and transparency. An adaptive temporal window (i.e., 2→10→50→100 cycles) provided immediate estimates at start-up and increasing stability as more data accumulated. On a contiguous 200-cycle segment drawn randomly from the dataset and withheld from all training and validation, the fused output tracked the expected gradual degradation while suppressing early overshoot; brief impulses coincided with scheduled window transitions rather than physical recovery. Overall, the study provides a reproducible and deployable framework from raw measurements to robust health estimation.

7. ACKNOWLEDGEMENTS AND DECLARATION

This research is part of the Business Finland project EMMA-2 (2471/31/2021). The authors gratefully acknowledge this funding and support by Business Finland.

The authors gratefully acknowledge the contribution of Ville Raunio, Zakhar Afansev, Matias Alenius, and Toni Haapakoski from the Faculty of Engineering and Natural Sciences of Tampere University, for their contribution in building the test rig and carrying out experiments.

At the time of submission of this study, the first author, Dr. Vinay Partap Singh, was employed at Danfoss Power Solutions, Denmark; however, this study and all the experiments were solely carried out at the IHA lab of the Faculty of Engineering and Natural Sciences of Tampere University, Finland.

OpenAI's ChatGPT was used to generate some graphics used in Figure 1 and Figure 6; this does not affect or reflect the methodology or results of the study. The authors take full responsibility for these graphics and for the content of the paper.

8. REFERENCES

[1] Huang, X., Yan, W., Cao, H., Chen, S., Tao, G., Zhang, J. (2024). Prospects for purely electric construction machinery: Mechanical components, control strategies and typical machines. Automation in Construction, 164, 105477.

[2] Balaban, E., Saxena, A., Narasimhan, S., Roychoudhury, I., Koopmans, M., Ott, C., Goebel, K. (2015). Prognostic health-management system development for electromechanical actuators. Journal of Aerospace Information Systems, 12(3), 329-344.

[3] Isturiz, A., Vinals, J., Abete, J. M., Iturrospe, A. (2012). Health monitoring strategy for electromechanical actuator systems and components, screw backlash and fatigue estimation. Recent Advances in Aerospace Actuation Systems and Components.

[4] Previdi, F., Maccarana, Y., Mazzoleni, M., Scandella, M., Pispola, G., Porzi, N. (2018, June). Development and experimental testing of a health monitoring system of electro-mechanical actuators for small airplanes. In 2018 26th mediterranean conference on control and automation (MED) (pp. 673-678). IEEE.

[5] Mazzoleni, M., Previdi, F., Scandella, M., Pispola, G. (2019). Experimental development of a health monitoring method for electro-mechanical actuators of flight control primary surfaces in more electric aircrafts. IEEE Access, 7, 153618-153634.

[6] Mazzoleni, M., Scandella, M., Maccarana, Y., Previdi, F., Pispola, G.,Porzi, N. (2018, August). Condition monitoring of electro-mechanical actuators for aerospace using batch change detection algorithms. In 2018 IEEE conference on control technology and applications (CCTA) (pp. 1747-1752). IEEE.

[7] Gómez de Mariscal, B., Villar Domínguez, A., Gallego, A., Fernández, J. A., Novillo, E., López de Calle, K., Fernández, A., Egaña, I. (2025, May 21–23). Health monitoring system for electromechanical actuation system: Pre-flight status assessment. In Recent Advances in Aerospace Actuation Systems and Components (pp. 119–126). Toulouse, France.

[8] Bajarunas, K., Baptista, M. L., Goebel, K., Chao, M. A. (2024). Health index estimation through integration of general knowledge with unsupervised learning. Reliability Engineering & System Safety, 251, 110352.

[9] Kim, S., Kim, N. H., Choi, J. H. (2020). Prediction of remaining useful life by data augmentation technique based on dynamic time warping. Mechanical Systems and Signal Processing, 136, 106486.

[10] Shi, C., Zhao, S., Chen, X., Wang, S., Liu, D. (2025). Condition monitoring of train transmission systems based on multimodal fusion improved transformer network. Sound & Vibration, 59(2), 2904-2904.

[11] Atamuradov, V., Medjaher, K., Camci, F., Zerhouni, N., Dersin, P., Lamoureux, B. (2020). Machine health indicator construction framework for failure diagnostics and prognostics. Journal of signal processing systems, 92(6), 591-609

[12] Raunio, V., Alenius, M., Minav, T. (2025, May). Identifying key predictors of failure in electro-mechanical actuator. In *International Conference in Aerospace Actuation Systems and Components* (pp. 127-134). INSA de Toulouse.

[13] Niu, G., Singh, S., Holland, S. W., Pecht, M. (2011). Health monitoring of electronic products based on Mahalanobis distance and Weibull decision metrics. Microelectronics Reliability, 51(2), 279-284.

[14] Jin, X., Wang, Y., Chow, T. W., Sun, Y. (2017). MD-based approaches for system health monitoring: a review. IET Science, Measurement & Technology, 11(4), 371-379.

[15] Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115

Biographies



Vinay Partap Singh received his master's degree in mining machinery engineering from IIT(ISM) Dhanbad, India in 2019, and Doctor of Science (Technology) from Tampere University, Finland in 2025. He is currently working as a Design Engineer in Danfoss Power Solutions in Denmark.



Tatiana Minav is an Associate Professor at Innovative Hydraulics and Automation lab of Tampere University. She has over 18 years of expertise on system electrification of non-road machinery and electro-hydraulic actuators solutions.

Digitalization in Sustainability for Fluid Power Components and Systems

Moritz Neltner, Andreas Schumacher, Robert Rahmfeld

Danfoss Power Solutions GmbH & Co OHG, Krokamp 35, 24539 Neumünster, mneltner@danfoss.com

Abstract.

Tracking energy losses and global warming potential (GWP) in mobile hydraulic machines and individual components is of growing importance due to an increase in regulations limiting the total GWP of operations and products. Based on data from a wheel loader performing a Y-cycle load operation, this paper computes the energy flow within the hydraulic system using a combination of measurements and system simulation. The energy flow can be used for system optimization. The calculated energy losses and GWP are shared with actors along the value chain using the Asset Administration Shell.

Keywords. Digitalization, simulation, asset administration shell, global warming potential, mobile hydraulics.

1. INTRODUCTION

The EU directive on the energy performance of buildings requires member states to set limits on the total global warming potential (GWP) of new buildings from 2030. The total life cycle GWP include emissions emitted during the construction phase of a building [1]. Already since July 2025, the Danish building regulations (BR18) include mandatory GWP limits for the combined emissions during transportation to and from a construction site (phase A4) and during construction and assembly (phase A5) [2]. BR18 requires the measurement and reporting of consumed energy for vehicles heavier than 1 ton, motivating the live monitoring of energy usage and energy optimization of construction equipment.

Using a wheel loader as an example vehicle, this paper shows how a static simulation model can be used to augment standard measurements to enable component level energy consumption calculations. The aggregated data can be used to fulfill regulations like BR18 while the individual data can be used for system efficiency optimization. This individual energy data can be shared using standardized interfaces like the Asset Administration Shell (AAS) [3] as tested in the research project Fluid 4.0 [4].

In the next section the wheel loader and the load cycle are introduced. A wheel loader enables the analysis of two common hydraulic functions: propel and cylinder-based work functions. Next the simulation model is introduced. It is based on quasistatic analysis of the individual components. Section 4 discusses the energy losses of individual components and highlights potential savings. Section 5 presents how the results of section 4 can be

postprocessed to fulfill regulations and outlines how these values can be shared effectively using the AAS. Lastly, the results are evaluated and future steps discussed.

2. WHEEL LOADER DEMONSTRATOR

The test vehicle is a mid-sized wheel loader shown in **Figure 2.1**. It is equipped with sensors for pressures, rotational speeds, positions, control currents, temperature, and flows. However, not all measurements are required for the energy flow calculations. Notably the flow measurements which are not provided in standard applications are not required. During a Y-cycle, four hydraulic functions were investigated: the hydrostatic drivetrain, articulated steering, lifting of the boom through two parallel cylinders and tilting of the bucket.



Figure 2.1: Wheel loader used for testing.

2.1. Wheel Loader

A modified Paus RL 8.7 wheel loader with a weight of 6.5 tons is the test vehicle. It has three hydraulic circuits of which two are discussed in this paper. First, the hydrostatic drivetrain consists of a single motor and a single pump. Second, the work functions (WF) are driven by a single open circuit pump. The fan drive system is not discussed in this paper but also contributes to the load on the main engine.

The drivetrain consists of the variable displacement axial piston pump Danfoss H1P-078 and the variable bent-axis motor Danfoss H1B-110 connected to a Dana Spicer 367 gearbox. All measurements were collected in first gear. The investigation of this paper focuses on the hydraulic components. This means that losses in the gearbox and the engine are not considered. A simplified overview of the propel and work function hydraulic components can be seen in **Figure 2.2**. Load sensing, load holding, pilot, cooling and emergency functions are not shown and not considered in the simulation.

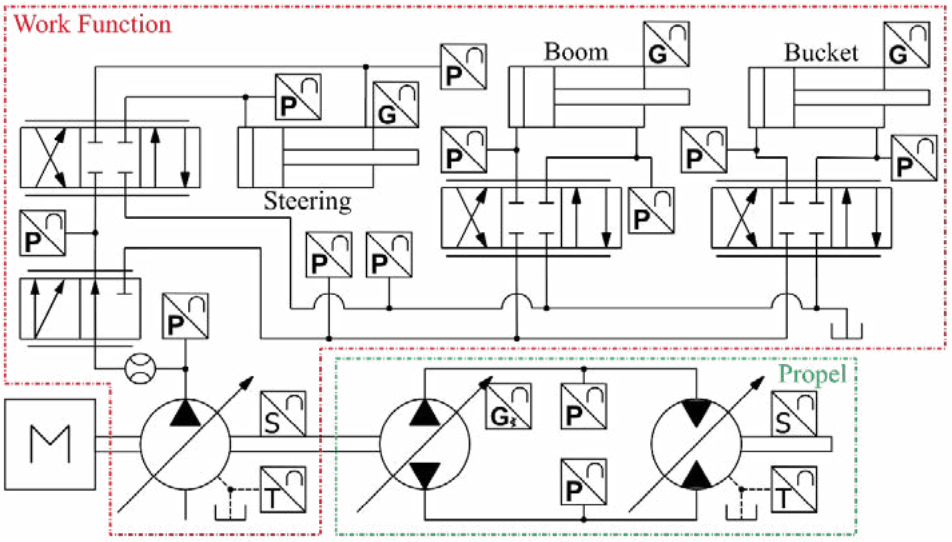


Figure 2.2: Overview of hydraulic circuit and measurements taken.

In the propel circuit the following quantities were measured: Pump and motor shaft speed, pump displacement angle, line pressures at motor and pump side, and motor drain temperature.

The open circuit hydraulics drives four work functions: articulated steering, boom lifting, bucket tilting, and auxiliary hydraulics. During testing, no auxiliary hydraulics were connected, leaving three hydraulic functions. All functions are driven by a single load sensing open circuit Danfoss S45-75 pump. The pump outflow is passed through a priority valve, giving priority to the steering function. A Danfoss OSPC250 steering unit controls the flow to the steering cylinder. Flow to the boom and bucket cylinders is controlled by two proportional Danfoss PVG valves. All returning flow from the steering unit and the proportional valves is filtered and cooled. The boom cylinders are equipped with a ride control unit. While the ride control unit was not actively used in any of the tests, the automatic system still precharged the accumulator for potential use.

Measurements were recorded with a sampling frequency of 500 Hz. The position of the cylinders was captured using draw-wire position sensors. Pressure was measured at the pump side of the steering unit and proportional valves, at both sides of all cylinders, and the pump outflow. The temperature was recorded in the pump case flow and after the cooler. Furthermore, the outflow of the open circuit pump, flow into the ride control unit and into the steering unit were measured.

2.2. Load Cycles

The Y-cycle was performed as shown in **Figure 2.3** on loose gravel. To start the cycle, the wheel loader is positioned with an empty bucket and raised boom at the drop-off pile on the right (1). Splitting the cycles at the drop-off ensures that the start and end states of the cycles are as similar as possible, because the boom is at a defined position. First, the wheel loader reverses and starts lowering the boom. The vehicle switches directions at point 2 and drives

to the left pile (3). When reaching point 2, the boom might not be fully lowered yet. At point 1 the bucket is filled and the boom raised while driving in reverse to the initial turning point (4). The cycle ends by dropping the gravel at point 1. In total 8 cycles were recorded.

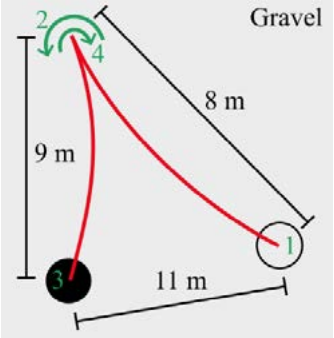


Figure 2.3: Wheel loader test cycle.

Figure 2.4 shows the vehicle speed and boom cylinder displacement with corresponding pressures during the test cycle. During the reverse driving from point 1 to 2, the propel delta pressure is negative during acceleration and positive while braking. After passing point 2 and driving forward, this relationship switches. Pressure during acceleration is now positive. While driving from 1 to 3, the boom is lowered. The boom pressure spikes when the wheel loader hits the gravel pile just before point 3. With a full shovel the boom pressure is higher and exhibits oscillation.

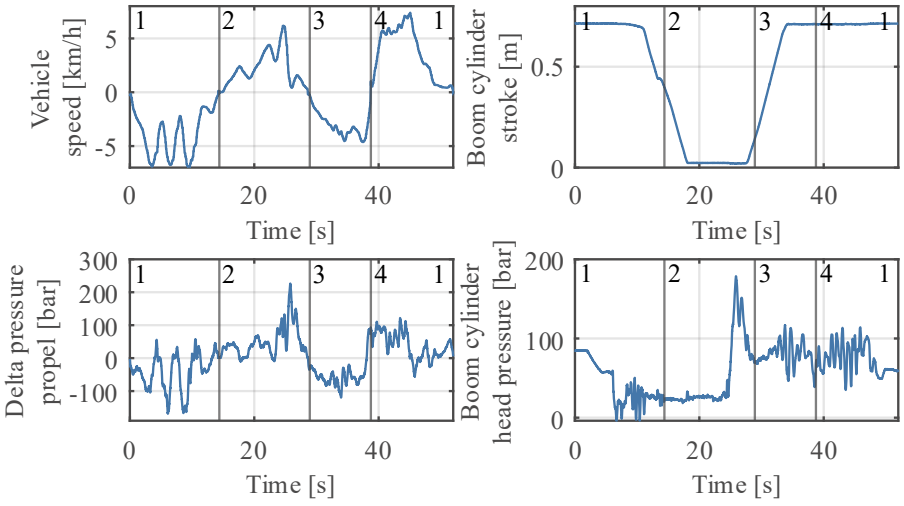


Figure 2.4: Measurement examples for test cycle.

3. SIMULATION MODEL

A quasistatic simulation is used to augment the measurement data and calculate the power flow through each component in every timestep. The simulation model is modular, allowing to swap measurements for simulation components and vice versa. Compared to an analysis only based on measurements, this approach allows replacing hard-to-measure hydraulic quantities like volumetric flow rates by simulation models. On the other hand, hard-to-simulate or unknown hydraulic components can be replaced with measurements. Compared to a full dynamic simulation, no complex environment needs to be simulated. In the following the overall simulation methodology as well as the component models and required external inputs are presented.

The main components of the hydrostatic system are modeled by a component model defining its hydraulic, mechanical, and electrical connections. Every component model defines linear and non-linear relationships between its connections. For example, a pressure relief valve model might define two equations:

1. The inflow is equal to the outflow as a linear equation.
2. The relationship between throughflow and pressure-drop as a non-linear equation through a look-up table.

If the relationship between throughflow and pressure drop is unknown, only the first equation can be defined. Instead of defining the second equation, pressure can be measured on both sides of the valve such that the system can still be solved.

To solve the complete system, the linear and non-linear equations of the individual component models are assembled into a system of linear and a system of non-linear equations using the hydraulic, mechanical, and electrical connections of the components. The system of linear equations is solved first and the solution substituted into the non-linear system. Next, the measurement values are substituted into the non-linear system, and the system is solved using a numerical solver. In postprocessing steps, the system solution is transferred to the individual component models and the power flow through each component is calculated. The modeling framework is implemented in MATLAB.

The level of detail of the component models depends on the available product knowledge and availability of measurement data. While a lot of information is known about the hydrostatic pumps and motor, less information is available for the valves and cylinders. In the following, first the models of the work function circuit and then of the propel circuit are presented.

The performance characteristics of the open circuit pump are known allowing a detailed model. The Polymod efficiency map based on [5] is used to define the non-linear relationship between pressure, shaft speed, displacement angle, temperature and flow. Due to the detailed model, no measurement of the displacement angle is necessary.

In contrast to the open circuit pump, the 4/3-valve performance characteristics are not known. The valves connections are modelled as generic pressure drops. Based on the valve control signal, flow continuity between the connected ports is ensured. Furthermore, the model ensures fluid flows from high to low pressure. The amount of fluid flow is determined by the connected cylinders.

Pressure and position measurements are available for all cylinders. The cylinder model is lossless. Speed, inflow, and outflow are calculated using the time derivative of the position. The cylinder force is computed using the differential cylinder area and the measured pressures.

The propel pump and motor are modelled analogous to the work function pump. Motor and pump shaft speed as well as both line pressures are measured. One more measurement is needed to specify the power flow through the drivetrain. As the control of the pump is equipped with an angle sensor, the displacement of the pump can be determined. Through the efficiency maps of pump and motor, the system flow and motor displacement are calculated by the model. As the flow through the system is based on the pump equations, the pump efficiency model and measurements specifying the evaluation point need to be accurate.

In summary, the simulation model calculates flow, torque and force based on pressure, control signals, pump angle, linear speed, and rotational speed measurements. In a postprocessing step power flow across the connections and power loss in each component are calculated. The presented model is not solvable for all measurement times. This is likely due to the violation of the quasistatic assumption of the simulation or errors in measurement as discussed in section 4.

To validate the simulation for the solvable states, the propel motor displacement and work function pump flow are used. Unlike in a real-world application, the flow rate of the work function pump was measured and is plotted along the simulation result on the right-hand side of **Figure 3.1**. The overall fit between measurement and simulation is good, but for certain time periods, for example around the 30 s mark, the simulation has an offset to the measured value.

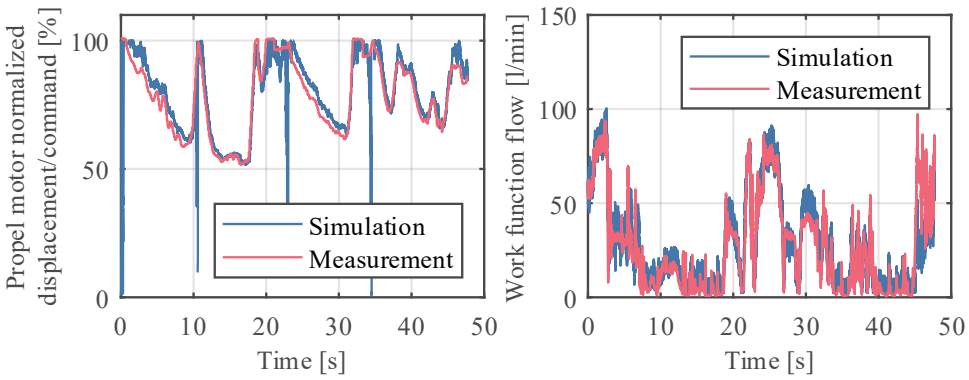


Figure 3.1: Comparison between simulation and measurement data.

No flow rate was measured in the propel hydraulic circuit. However, due to the mechanical feedback link in the motor control, the motor displacement correlates well to the motor control signal. Hence, the simulated normalized motor displacement is compared to the normalized motor displacement command on the left-hand side of Figure 3.1. Like in the work function circuit, the simulation matches the commanded value well most timesteps. However, around the times when the simulation is not solvable, the simulated displacement

drops sharply. As the simulation failures are of short duration, the impact on results concerning a complete Y-cycle is small.

4. RESULTS

This section presents the results derived from augmenting the measurement data with the simulation model and the subsequent energy calculations. The Y-cycle described in section 2.2 was performed 8 times by a non-professional driver. The power flow between components, during different operating phases and differences between cycles are discussed. The analysis focuses on the main power flow, neglecting control flows, cooling flows, and dynamic effects. As the sample size is small and the data was gathered in a test setup and not during regular operations, the results are only illustrative.

First, the power flow through the hydraulic system is discussed. The Sankey diagram shown in **Figure 4.1** shows the power flow between the components during one particular Y-cycle. Red arrows mark the useful work performed by the boom, bucket, steering unit, and wheel relative to the absolute engine output power. Roughly half of the energy provided by the engine is used for the work functions, the other half is used for the propel application. About a third of the energy provided by the engine is turned into useful work. In the median cycle 28 % are used at the wheel, 4 % at the boom, and about 1 % for steering and bucket tilting.

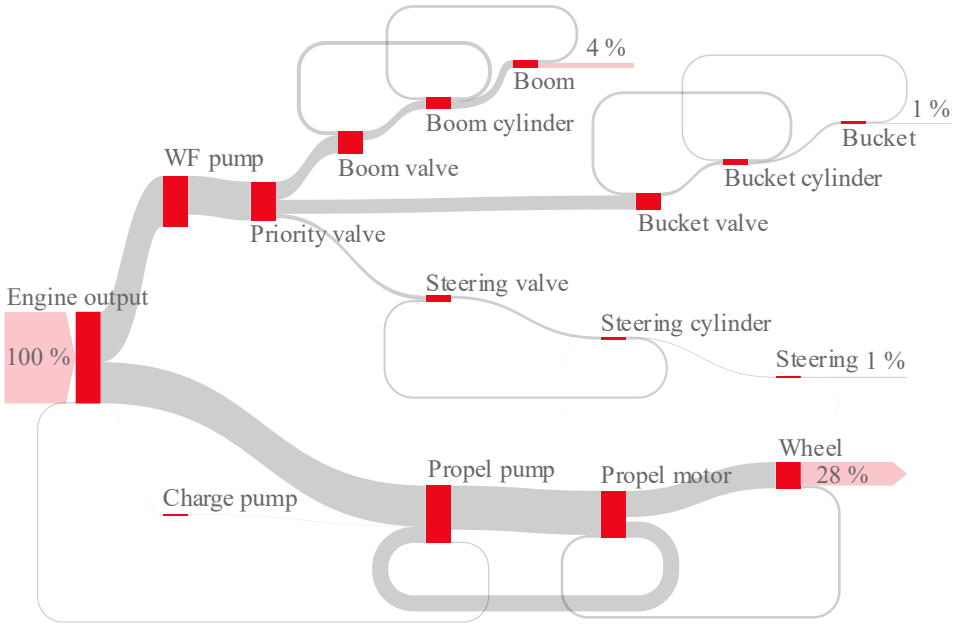


Figure 4.1: Energy flow during 1 Y-cycle. Red arrows mark hydraulic system energy inflow and outflow.

In the work function system, the priority valve splits the work function flow between the cylinders and the steering function. The losses in the priority valve are minor. The highest relative losses are in the throttling valves to the cylinders. As Figure 4.1 shows the energy

flow across a complete cycle, it includes raising and lowering of the boom. During raising unnecessary losses occur in the valve as the pressure level between the cylinders differ. While lowering, all potential energy from boom and bucket is converted to heat. A hydraulic circuit which allows for energy recovery, could recover energy during lowering. Roughly 42 to 76 % of the energy required to move the bucket and 26 to 36 % for the boom are provided back at the cylinders during cylinder retraction. The large spread in recoverable energy for the bucket highlights the influence of vehicle operation on energy saving potential.

The propel system has a median efficiency of roughly 64 % engine to wheel. Only about 8 % of the energy provided to the wheel is returned to the motor during braking. The rest of the energy is lost to friction, especially due to the loose gravel ground. A comparatively large energy flow from the propel motor to the propel pump can be observed due to the low pressure return flow at around 22 bar.

All results are based on measurements with a sampling rate of 500 Hz. All calculations have been repeated with lower sampling rates. With the 500 Hz computations as a baseline, the relative error stays below 1 % for sampling rates up to 20 Hz. For the same sampling rate, the relative error in the work function components is larger than for the propel components. This means, sensors capturing the work function hydraulics need to have higher sampling rates to achieve the same accuracy level.

For a few timesteps, no solution within the physical limits of the components could be found. In these cases adjacent solutions were linearly interpolated. The most common cause for unsolvable system states is the propel motor on maximum displacement not being able to consume all flow produced by the pump. Likely reasons for the surplus of flow on motor side are the neglect of control losses for the propel motor, violation of the quasistatic assumption and measurement errors. The issues might be alleviated by smoothing the measurements before processing in the simulation model.

Figure 4.2 gives an overview of the energy outflow of the system while the same Y-cycle was performed 8 times. The top 4 energy sinks (red borders in Figure 4.2) are the useful work performed for the steering, bucket, boom and propel function. All other items are losses in the hydraulic circuit. While the energy provided by the engine to the work function and propel circuit is similar (see. Figure 4.1), more work is being done by the propel circuit. This is caused by the efficiency of the propel circuit being larger than that of the work function circuit. While the efficiency of the rotating units is larger than the efficiency of the valves, the absolute energy loss is comparable, because the rotating units handle a larger energy flow.

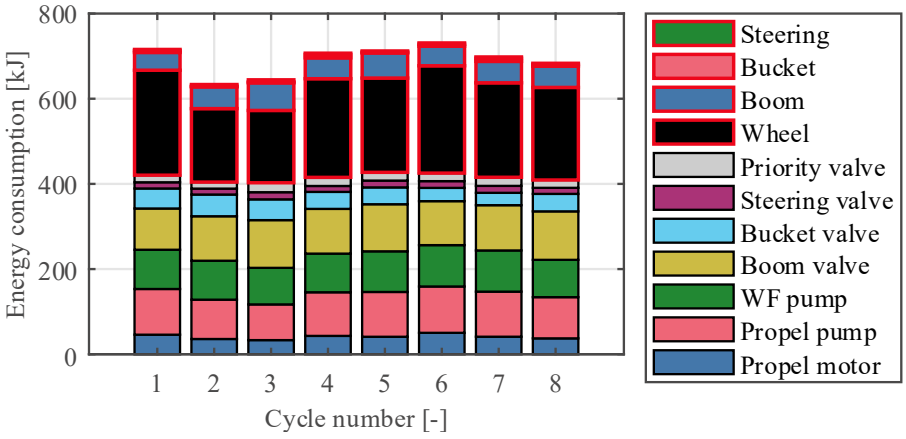


Figure 4.2: Work performed (red border) and energy loss of individual components.

The median energy used in a single Y-cycle is 700 kJ. In all cycles the same work was performed by the same operator. The large spread of total energy consumed, 14 % of the median, highlights the importance of repeated field measurements. To acquire exact energy consumption and emission values, measurements as presented in this paper are required. Accurate knowledge of energy consumption is the basis for GWP calculations in the next section. Furthermore, the energy data can be shared directly via the AAS as demonstrated in the Fluid 4.0 research project and explained in the next section.

5. EMISSION CALCULATION AND SHARING

The energy flow data presented in section 4 is the basis for computing key climate emission indicators. An overall GWP value for the wheel loader is calculated based on the total energy output of the engine and then attributed to the components of the hydraulic system. The total GWP potential can be used to fulfill building regulations like BR18. Individual GWP of the components can be used to calculate the lifetime GWP of individual products. To share the GWP information along the value chain, the standardized interface of the AAS is used. This facilitates the further automatic processing of the emission information into Life Cycle Assessments (LCAs) and other documents required to meet regulations.

Building regulations like BR18 not only stipulate overall emission limits for materials and the use phase of a building, but also specific limits for transportation to and from a construction site (phase A4) and during construction and assembly (phase A5). While the calculations for phase A4 can be performed based on the distance driven by trucks, the energy consumed during construction and assembly needs to be measured directly. For diesel fueled machines, this can be achieved by measuring the fuel consumption or using the engine's specific fuel density and engine energy output calculated in section 4. By using time stamps, the latter approach can be used to assign emissions to the corresponding projects, when a construction machine is performing work for multiple projects on one tank filling. By multiplying the engines' specific fuel consumption of 221 g/kWh and the GWP (including life-cycle emissions) of diesel of 4.13 $kgCO_2e/kg$ [2] with the engine's power output, the median GWP for one Y-cycle is calculated to 178 gCO_2e. Note that the GWP of

diesel varies with the calculation scope determined by the regulation. The value provided by BR18 not only includes the direct effects of diesel combustion, but also emissions generated during sourcing and transportation.

Following the same calculations, the GWP for individual components may be calculated. The GWP for individual components may be used as a data point in the components LCAs. Due to the versatility in use of hydraulic components and the strong impact on energy losses, only real-world measurements can provide an accurate value for the GWP of a component in a specific system. If, for example, the same valve is used in a work function system with accumulator and energy recovery, less energy loss in the valves during lowering is expected. While the proposed method provides an accurate GWP value for a component in a specific system and use case, it is difficult to compare it with the GWP of other products without testing them under the same conditions. On the other hand, a GWP value based on testing defined operating points does not represent the real GWP value of a component, but it ensures that GWP values between similar products are easily comparable.

The calculated emission data is needed by a variety of parties along the value chain. In the example of a rented wheel loader working on a building, the raw emission data might be used by the construction site management, the rental company, vehicle manufacturer, and hydraulic component manufacturers. While the construction site management and rental company are primarily interested in aggregated vehicle level data, the vehicle and component manufacturers have an interest in emissions of particular components. Field data for components ensures that emission reporting is based on real life and thus actual emissions are tracked and can be reduced over time.

To facilitate the sharing of emission data with the different parties in varying granularity, a standardized exchange format is needed. A digital machine-readable format is preferred, because the emission data needs to be aggregated and processed at the different parties before being ready for reporting. One exchange format which is gaining traction in the fluid power industry is the AAS [6]. The AAS framework specifies the technical aspects of data exchange and the data content on a semantic and structural level. The content level structure is described in submodules. For example, a submodule “Energy Monitoring”, facilitating sharing of component level energy data, is currently developed in the research project Fluid 4.0 [4]. The prototype of the submodule Energy Monitoring was used to provide the energy data of this paper in a standardized format to the Fluid 4.0 project partners.

Multiple submodules holding carbon footprint data are already released or under review. The submodel Carbon Footprint can be used to share carbon footprints which were calculated according to ISO 14067 or similar standards [7]. However, two further submodules related to product emission data are in review by the IDTA: Digital Battery Passport – Part 3: Product Carbon Footprint and Product Related Environmental Data [8]. Multiple submodules holding the same or similar data hinders efficient automatic processing of the information within. Ideally every piece of information would have a well-defined unique place inside the AAS. In the future, it is important to consolidate overlapping submodules to create submodules which are general enough to apply to a wide range of assets while providing a unique place for specific information.

6. CONCLUSION AND OUTLOOK

Using a wheel loader as a test machine, this paper demonstrates how standard measurement data can be augmented using a static simulation model to analyze energy flow within a hydraulic system and compute global warming potential (GWP) of vehicle operation to fulfill emission regulations of construction sites. The information about energy flow between components is helpful for system optimization and emission reporting of individual components. Furthermore, the energy and emission information can be shared with parties along the value chain using the Asset Administration Shell (AAS) as a standardized format.

The energy loss of hydraulic components and thus the overall system efficiency is strongly dependent on the load cycle for each component. The individual load cycles depend not only on the machine load cycle, but also on the operator. This makes energy calculation based purely on system simulation difficult, as defining realistic load conditions is challenging. On the other hand, measuring all flow and pressure quantities in a real hydraulic system to determine energy losses, is costly and time consuming. The methodology outlined in this paper reduces the need for a complex sensor setup by augmenting the test data using static simulation models for the individual components. The resulting energy flows show large variations of 14 % in energy consumption between cycles of the same task, highlighting the need for real world measurements to achieve accurate results. Accurate live data is required to fulfill emission limits targeting the use phase of a machine like the Danish building regulation BR18. For planning purposes, it needs to be investigated if the GWP results can be generalized to different drivers, cycles or even machines of different sizes using a larger data set.

To facilitate accurate emission reporting for all stakeholders, emission data needs to be communicated along the value chain. An AAS allows standardized machine-readable communication of the emission data to various parties. Next to emission data, per component energy data can be shared to facilitate system optimization and lifetime load calculations.

7. ACKNOWLEDGEMENT

This work has been funded by the European Union (EU) and supported by the German Federal Ministry for Economic Affairs and Energy (BMWE) in the project Fluid4.0 – Revolutionizing Fluid Engineering: Implementation of digitalization for fluid power 4.0 in the cross-industry and cross-manufacturer data space using asset administration shells, submodels and demonstrators (project number 13IK039N). The views and opinions expressed are solely those of the authors and do not necessarily reflect the views of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.



8. REFERENCES

[1] European Parliament and Council of the European Union, *Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the energy performance of buildings (recast) (Text with EEA relevance)*. 2024. [Online]. Available: http://data.europa.eu/eli/dir/2024/1275/oj

[2] Social- og Boligministeriet, *BEK nr 271: Bekendtgørelse om ændring af bekendtgørelse om bygningsreglement 2018 (BR18)*. 2025.

[3] Working Group Specifications, *Specification of the Asset Administration Shell Part 1: Metamodel (Version 3.1.2)*. doi: 10.62628/IDTA.01001-3-1-2.

[4] "Fluid 4.0 Research Project." Dec. 01, 2025. [Online]. Available: https://fluid40.de

[5] T. Kohmäscher, R. Rahmfeld, H. Murrenhoff, and E. Skirde, "Improved Loss Modeling of Hydrostatic Units: Requirement for Precise Simulation of Mobile Working Machine Drivelines," in *Volume 4: Design, Analysis, Control and Diagnosis of Fluid Power Systems*, Seattle, Washington, USA: ASMEDC, Jan. 2007, pp. 195–206. doi: 10.1115/IMECE2007-41803.

[6] M. Becker, S. Karaoglu, F. Makansi, J. Radtke, and K. Schmitz, "Industry 4.0: Review of the State of the Art in Fluid Power Research and Industry," *Int. J. Fluid Power*, Dec. 2025, doi: 10.13052/ijfp1439-9776.2643.

[7] IDTA, *Submodel Carbon Footprint*, 02023, Mar. 2025. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2025/03/IDTA-02023_Submodel_CarbonFootprint.pdf

[8] IDTA, "Registered AAS Submodel Templates." Jan. 12, 2026. [Online]. Available: https://industrialdigitaltwin.org/en/content-hub/submodels

Biographies



Moritz Neltner received a bachelor's degree in mechanical engineering in 2020 and a master's degree in theoretical mechanical engineering in 2022 from Hamburg University of Technology. He received a master's degree in technology management from the Northern Institute of Technology Management in 2023. Since 2023 he works at Danfoss Power Solutions as a simulation engineer focusing on system simulation, data automation, control, and condition monitoring for mobile hydraulics.



Dr. Robert Rahmfeld is the Senior Director Engineering for Hydrostatics at Danfoss Power Solutions, being technically responsible for hydrostatic axial piston units. He is with Danfoss over 20 years, holds several patents, and is author or co-author of more than 60 scientific journal and conference papers. Dr. Rahmfeld completed his PhD on new displacement controlled hydraulic linear actuators at the Technical University of Hamburg-Harburg in 2002 under supervision of Professor Monika Ivantysynova, and his MSc in mechanical engineering at Duisburg University in 1996. He was also holding a postdoc position in fluid power for 2 years at the Technical University of Hamburg-Harburg and Purdue University, Indiana, USA. Dr. Rahmfeld represents Danfoss within the VDMA fluid power board and is also the chairman of the corresponding fluid power research fund with more than 60 member companies. He was awarded in 2024 by Purdue University with the *Monika Ivantysynova Medal* for technology advancements of hydrostatic units and their controls, and for leadership in fostering industry - academic initiatives to promote fluid power.



Senior Director **Dr. Andreas Schumacher** is responsible for Sustainability and Product Compliance for the complete Danfoss Power Solutions segment. He joined Danfoss in 2007 as a simulation engineer after his PhD on mobile machinery tire and driving dynamics simulations at the Technical University of Braunschweig under the supervision of Professor H.-H. Harms. The career development over the nearly 20 years in Danfoss increased his responsibilities and also changed from Engineering to the current role. Dr. Schumacher is also author and co-author of over 30 journal and conference papers as well as presentations. As part of his current role, he is representing Danfoss Power Solutions externally in the field of mainly Sustainability in regulatory, scientific/research and business contexts and further publications like publicly available podcasts (e.g. on Spotify).

Predicting leakage and operating conditions in axial piston pumps using vibration-based machine learning

Simon Knoll, Svenja Horn and Jürgen Weber

Institute of Mechatronic Engineering, Technische Universität Dresden,

E-mail: simon.knoll@mailbox.tu-dresden.de, fluidtronik@mailbox.tu-dresden.de

Abstract.

Condition monitoring for axial piston pumps is often hindered by high sensor and infrastructure costs. This paper presents the groundwork for a low-cost, edge-computing "virtual sensor" that uses vibration-based machine learning (ML) to predict pump operating conditions (OCs) and leakage. Using data from a test rig with multiple artificially generated fault types, ML models (XGBoost and Multilayer Perceptron) were evaluated.

Results show that direct leakage prediction from vibration data is accurate for known faults ($R^2 \geq 0.97$) but fails to generalize to unknown faults ($R^2 > 0.50$). This prediction for unknown faults improves significantly (e.g., $R^2 \approx 0.80$) when OCs are provided as features. It is shown that these OCs (pressure, speed, and stroke) can be accurately predicted ($R^2 > 0.96$) from the same vibration signal. This high accuracy, however, is critically dependent on a robust training dataset that includes various fault types.

Finally, edge-computing feasibility is confirmed by deploying a Multilayer Perceptron model to an ESP32-S3 microcontroller starting at less than 10 €. This work validates the key data requirements for developing a cascaded, low-cost smart sensor.

Keywords. Economic fault detection, condition monitoring, machine learning, Edge-Computing.

1. INTRODUCTION

Condition Monitoring (CM) for fluid power systems, particularly axial piston pumps, is a subject of growing importance. These pumps are critical components in many industrial settings, such as steel mills, where they can be difficult to access for maintenance. Traditional maintenance relies on fixed replacement intervals, which is often costly and does not prevent spontaneous, expensive failures. Predictive Maintenance (PM) offers a needs-based alternative, but its implementation in existing systems faces significant hurdles.

Integrating new sensors into established industrial plants is often complicated and cost intensive. Furthermore, many applications involve sensitive process data, making cloud-based monitoring solutions undesirable due to data security concerns. These practical and economic constraints point toward the need for low-cost, stand-alone retrofit solutions that perform calculations locally. This "Edge-Computing" approach, using low-power microcontrollers, has become increasingly viable. While much recent research focuses on high-performance Deep Learning algorithms, this paper targets a robust, low-hardware solution by focusing on predicting leakage flow, a highly interpretable health index for axial piston pumps.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

To achieve this without an expensive flow meter, a "virtual sensor" concept is proposed. This paper demonstrates the ground work for a cascaded regression method where vibration data from a single, low-cost accelerometer—identified in previous work as a highly promising sensor for ML-based monitoring [1]—is first used to predict the pump's operating conditions (pressure, speed, and stroke). This predicted operating condition information is then, in a second step, used to improve prediction of the pump's leakage flow.

This paper is structured as follows: Chapter 2 provides an overview of the existing work in fault detection for Axial Piston Units (APU). Chapter 3 introduces the artificially generated faults, the test rig, the installed sensors, and the machine learning (ML) methods. Chapter 4 presents the results of the ML algorithms including the proof-of-concept implementation and performance of this approach on a low-cost microcontroller for edge deployment.

2. STATE OF THE ART

Condition Monitoring (CM) and Predictive Maintenance (PM) are established fields for hydraulic systems, motivated by the high economic cost of unexpected failures and the growth of the Industrial Internet of Things (IIoT). A significant amount of research focuses specifically on axial piston units (APUs), developing CM systems for both mobile and stationary applications. While model-based fault detection exists [2, 3], data-driven methodologies using sensor data (e.g., pressure, acceleration) are gaining interest.

Most research employs Machine Learning (ML) for fault classification, utilizing algorithms like K-Nearest Neighbors [4], SVMs [5], or Deep Learning (DL) [6, 7] to identify the type of fault or a severity measure defined by multiple classes. Notably, Duensing [22] predicted wear progress in APU using data from actual degradation rather than artificial faults, identifying Multilayer Perceptrons and Random Forests as the most effective models. Keller et al. [4] identified specific acceleration and operating parameters as essential for classifying valve plate damage. Aligning with this, previous work [1] validated the triaxial acceleration sensor for ML tasks and demonstrated that higher sampling rates improve prediction quality, motivating the 16.4 kHz rate used here. Some approaches omit manual feature extraction using DL pipelines. For instance, Tang et al. [6] utilize wavelet transforms to convert signals into images for DL classification. While effective, this lacks an interpretable, continuous degradation measure necessary for maintenance planning. Additionally, data-driven models often struggle with small datasets, requiring transfer learning or simulated data [8, 9].

Addressing this, regression models enable continuous condition prediction. Schimke et al. [10] applied Random Forests and CNNs to predict damage severity. To simulate cavitation erosion holes are drilled in specific areas of the valve plate. The area of the drilled-out portion then functions as the regression target. Their use of 96 operating conditions highlights the importance of dataset diversity. They emphasize that inclusion or exclusion of damage types in the training dataset significantly influences prediction accuracy.

Other studies quantify degradation via abstract "Health Indices" (HIs) [11, 12]. However, unlike physical metrics such as leakage flow, HIs often lack the direct meaning required to set concrete maintenance thresholds. While Liu et al. [9] proposed a "virtual flow sensor" to predict leakage directly, their reliance on simulated data limits real-world relevance. Furthermore, high-performance architectures like PINNs [13], CNNs [10], and LSTMs [14, 15] demand computational resources often unavailable in low-cost, retrofit scenarios.

Walani et al. [16] argue for lightweight models on edge platforms to ensure real-time processing and low latency, though hardware limitations restrict model complexity.

Consequently, this review identifies a gap for an APU monitoring solution that is simultaneously: 1) Low-Cost & Edge-Deployable, avoiding heavy DL models; 2) Interpretable, by predicting physical metrics like leakage; and 3) Robust against variable conditions and unknown faults, a challenge for static classification models.

3. EXPERIMENTAL AND METHODICAL SET-UP

This section provides an overview of the artificially worn pump parts, the test rig and sensors, followed by a description of the measurement procedure and a brief introduction of the methodical approach for the implementation of ML algorithms. More details as well as the pump efficiencies for the different faults are presented in [17].

3.1. Artificially worn pump parts

Figure 3.1 illustrates three representative types of axial piston pump damage, each reproduced as realistically as possible. To identify the location and severity of a typical valve plate cavitation damage, simulations were conducted at elevated speeds, complemented by measurements at very low suction pressures. Real cavitation erosion at three different intensities (I, II, III) was induced locally in the transition zone between low and high pressure using a sonotrode, compare Figure 3.1 a).

Axial clearance between slipper and piston, a failure that can result in total pump breakdown, was implemented with a tensile testing machine. The required axial force was estimated via finite element analysis (FEA), as shown in Figure 3.1 b). All nine slipper-piston assemblies were then elongated until reaching a 179 µm clearance using a tensile testing machine, with a custom fixture to clamp both parts without damaging the sliding interfaces.

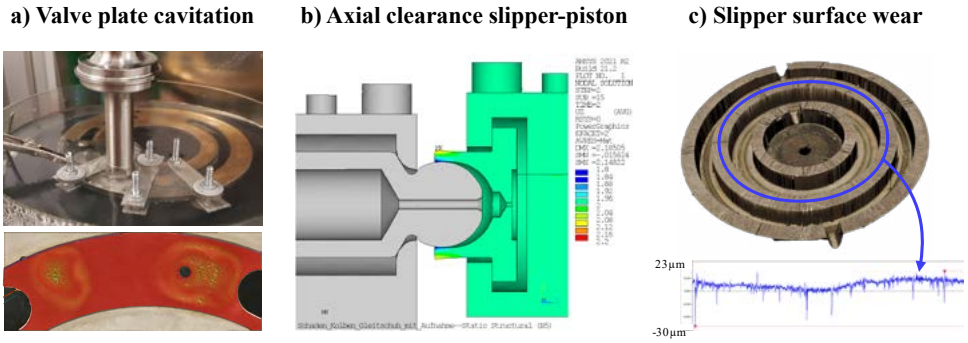


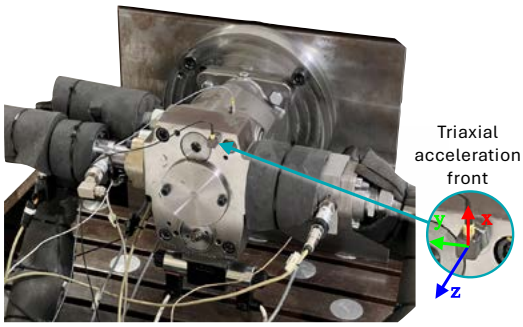
Figure 3.1 Artificially damaged pump parts, manufactured as realistically as possible (more details in [17])

Finally, slipper sealing lands were intentionally damaged using a file, and the achieved damage was quantified by profilometer measurements, as depicted in Figure 3.1 c).

3.2. Pump Test-Rig and Measurement Procedure

A standard test rig setup, designed for pump efficiency measurements, was equipped with high-frequency pressure and acceleration sensors. These were sampled at 16.4 kHz and

located at the pump's front plate and housing, compare Figure 3.2. The triaxial acceleration sensor at the front plate provided the most informative signals for machine learning applications. In this study, all three acceleration components were included in the predictive pipeline. Suction pressure was maintained at 1.8 bar to prevent cavitation throughout all operating conditions. Low-frequency pressure and temperature data were acquired to monitor and control the test environment. Measurements were conducted across the full operating range of the variable displacement APU, covering six pressure levels, four stroke values and four rotational speeds, resulting in 96 distinct operating conditions. Most faults resulted in minor efficiency losses, with the exception of severe valve plate cavitation III, as clearly shown by the leakage results (see Figure 4.1).



Property	Value
Sensor	MMF KS903.10
Sensor type	triaxial piezoelectric accelerometer
Mounting method	threaded stud
Upper frequency limit (3 dB)	11 kHz (X), 22 kHz (Y/Z)
Sampling Frequency	16.4 kHz

Figure 3.2 Mounting location and technical information for the triaxial acceleration sensor used in this work.

The study's findings are based on a representative subset of 28 operating conditions that are presented in Table 1. Each measurement lasted 10 seconds, sufficient to capture multiple pump rotations at each speed. Since temperature significantly affects leakage and vibration damping, the inlet temperature was controlled at 40°C ± 3°C. In practical applications, temperature variations can be measured cost-effectively and incorporated into the evaluation process.

Table 1: Overview of representative set of 28 operating conditions (OCs) grouped by stroke value relative to the maximum

100 % stroke		**75 % stroke**		**50 % stroke**		**25 % stroke**	
n/rpm	Δp/bar	n/rpm	Δp/bar	n/rpm	Δp/bar	n/rpm	Δp/bar
500	200	500	150	500	50	500	100
1000	100	500	250	1000	150	500	250
1000	250	1000	50	1000	300	500	300
1500	150	1000	200	1500	50	1000	150
1500	200	1500	100	1500	250	1500	50
2000	50	1500	300	2000	100	2000	200
2000	300	2000	150	2000	200	2000	250

Initially, a reference measurement was conducted under healthy conditions without any damaged parts (“new” condition). This data serves as a baseline for comparison with subsequent measurements involving damaged parts. Each defect was measured separately, without combining multiple defects.

3.3. Methodical approach for implementing machine learning

This section details the standardized methodology applied to all supervised machine learning tasks performed in this paper using Python (v3.12.3). As illustrated in Figure 3.3 the approach encompasses data acquisition, data preprocessing, feature engineering, model selection, hyperparameter optimization and final model evaluation. Additionally, a proof-of-concept edge deployment on an ESP32 micro controller is presented for one selected model.

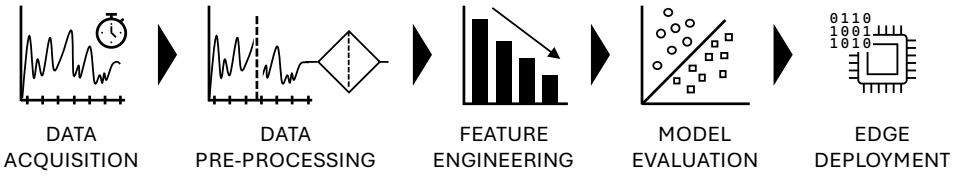


Figure 3.3 Methodical approach to implement machine learning algorithms.

3.3.1. Data segmentation and feature engineering

The raw acceleration signals exhibit cyclostationary behaviour, with patterns repeating for each pump revolution. To create a high volume of training samples, the data is split into smaller segments, each containing two full pump revolutions. Each segment underwent normalization (zero mean, unit variance) to emphasize signal shape over absolute amplitude across operating conditions. Feature extraction was performed using TSFresh (v0.20.2) [18], a Python library validated in prior studies [1, 17].

Table 2: Selected feature types ordered by count in the final top 700 (names matching the TSFresh implementation)

Feature Type	Amount in final selection	Feature Type	Amount in final selection
fft_coefficient	384	linear_trend	9
change_quantiles	126	number_cwt_peaks	6
agg_linear_trend	62	mean_abs_change	3
ar_coefficient	26	longest_strike_below_mean	3
autocorrelation	22	longest_strike_above_mean	3
partial_autocorrelation	21	count_above_mean	3
number_peaks	15	count_below_mean	3
fft_aggregated	11	abs_energy	3

From an initial candidate set of approximately 2,400 features, a preliminary reduction using univariate p-value analysis selected the top 16 feature types based on median ranking. This yielded 1,600 spectral and time-domain features, which essentially describe the shape and patterns of the signal. A second p-value analysis further reduced this selection to a final set of 700 features, detailed in Table 2. The most significant predictors included time-series autocorrelation, longest strikes above and below mean as well as spectral attributes (FFT coefficients, skewness, variance, kurtosis) up to 1.7 kHz equivalent frequency. Finally, the feature matrix was standard-scaled to ensure equal contribution during training. The regression label for each sample is represented by the unscaled mean leakage flow for the corresponding time segment.

3.3.2. Model selection

To balance performance with deployment feasibility, two regression models are employed. First, **XGBoost** (v2.0.3) [19], a decision-tree-based ensemble, serves as a robust benchmark due to its proven success in prior studies. Second, a **Multilayer Perceptron** (MLP) using Scikit-Learn (v1.5.0) [20] is selected specifically for its computational efficiency, making it ideal for resource-constrained edge deployment.

3.3.3. Hyperparameter optimization and training

Before the models are optimized and trained, 20 % of the entire dataset is separated as a final, untouched hold-out set for the calculation of performance metrics. The remaining 80 % of the data is used for hyperparameter optimization. The optimization for both models is conducted using the Optuna framework (v3.6.1), an automated, trial-based hyperparameter optimization tool [21]. This process employs a GroupKFold cross-validation strategy with 5 folds per trial. For this strategy, the data is grouped by the specific "damage kind" (e.g. valve plate cavitation II, slipper wear). This ensures that for each fold of the optimization process, the validation set contains data from at least one damage type that the model has not been exposed to during training in that fold. The objective of this approach is to develop final deployment models that are as robust as possible and generalize well to novel or unseen fault conditions. The hyperparameter search spaces for both models are represented in Table 3 in the appendix.

3.3.4. Final model evaluation

Performance results in the following sections are generated using the 20 % hold-out set that was segregated prior to hyperparameter optimization. The evaluation on this hold-out set is conducted using a KFold cross-validation strategy with 10 folds and random sampling aiming to represent the real-world deployment of pretrained models. In each iteration, a model is instantiated with the optimized hyperparameters, trained on 90% of the development set, and evaluated against the fixed hold-out set. This process is repeated 10 times, ensuring that varying samples in the original set are used for training. The aggregated results from these 10 folds provide a robust and comprehensive assessment of the model's final performance in a deployment setting.

4. PREDICTING PUMP LEAKAGE

4.1. Critical faults rise leakage flow

Leakage flow measurement is a widely used indicator of hydraulic pump condition, as increased leakage often signals wear and the need for replacement. However, clamp-on flow meters are expensive (beginning from 1000 €) and typically require a straight section of pipe or specific housing to ensure accurate readings, making them less practical for many installations. In-line flow meters, on the other hand, are difficult to implement in existing systems and can be prone to failures. For these reasons, the vibration signal should be used to predict the pump's leakage. In comparison to traditional fault analysis as for example the classification in "good" and "bad" or the use of a "health index" [12], operators that have a long-standing experience with their pumps and systems know the critical leakage values to decide whether the pump needs to be exchanged. As a further advantage, leakage has the potential to also work for different pump sizes if normalized. Not all pump faults - such as those involving the shaft or bearings - directly affect leakage, so relying solely on leakage measurements may miss some mechanical issues. Therefore, vibration analysis provides a more comprehensive assessment, allowing for better detection of both hydraulic and mechanical faults. In addition to predicting leakage, other models can monitor the vibration signal to detect fault types that do not measurably increase leakage, a capability demonstrated in previous fault classification work [1, 17].

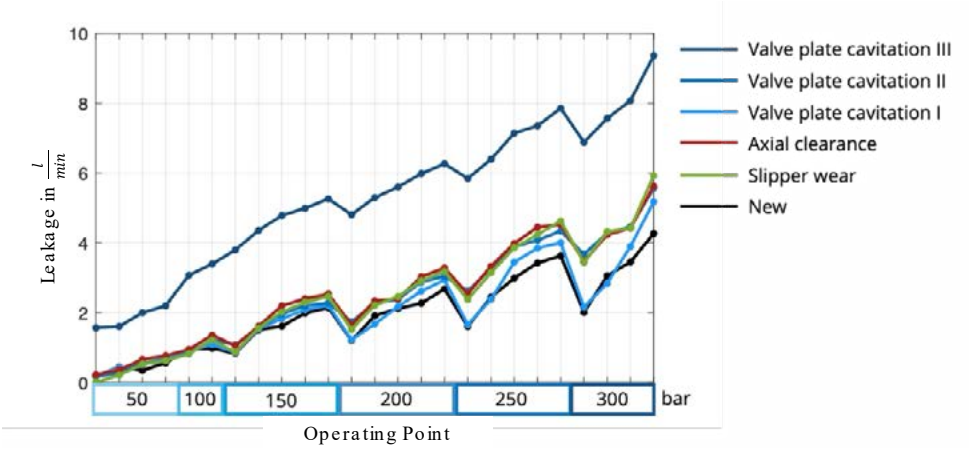


Figure 4.1 Leakage of the different faults over many operating conditions at maximum stroke (100 %) with rising speed for each pressure level

Figure 4.1 illustrates the measured leakage for the maximum stroke (100 %) for the different faults across a wide range of operating conditions, clustered by pressure levels and rising speeds for each pressure. Measurements were conducted using a VSE VS1 gear type flow rate sensor recording the total external leakage flow of the APU at a sampling frequency of 330 Hz. All faults are situated within the minor wear range, causing only slight increases in pump leakage, except for valve plate cavitation III, which shows a more pronounced effect. The figure demonstrates that higher pressures and speeds particularly accentuate leakage differences between fault types. Additionally, leakage is shown to be significantly dependent on operating conditions and temperature, supporting the hypothesis that machine learning

models could improve fault detection by incorporating additional variables such as speed and pressure.

4.2. Correlation between pump vibration and fault types

While leakage indicates severe faults (e.g., valve plate cavitation III), it shows only minor increases for defects like slipper wear (Figure 4.1). To capture a holistic view of pump health, a vibration-based ML approach is employed. The correlation between vibration and condition is analysed using heatmaps of the y-acceleration amplitude against the "pump order". Due to cyclostationary behaviour, spectral peaks occur at harmonics of the piston frequency f_p, where Z is the piston count and the n rotational speed. Unlike raw frequency, the pump order normalizes the signal frequency f against f_P, enabling comparisons across varying speeds:

$$pump\ order = \frac{f}{f_p} = \frac{f}{Z * n}$$

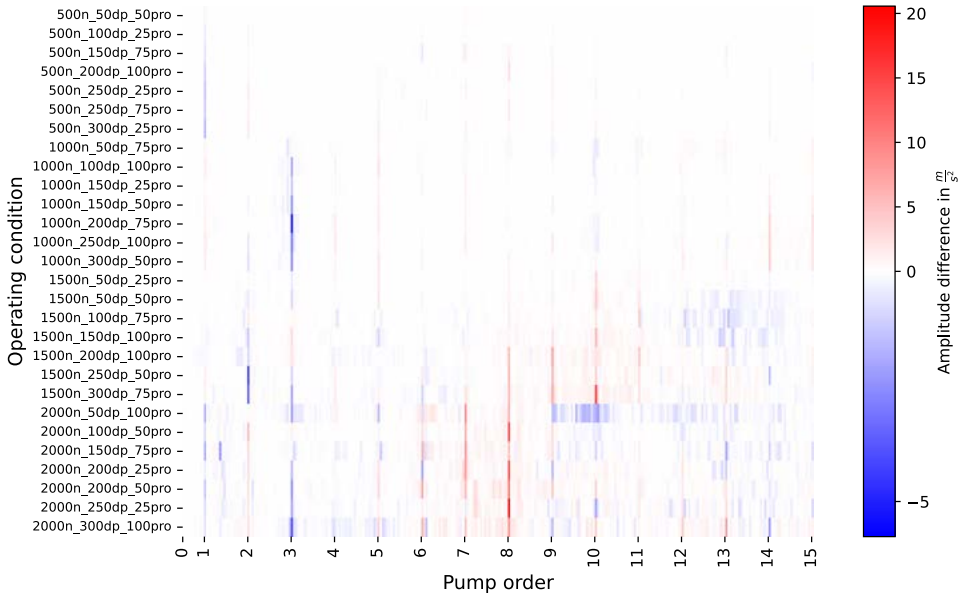


Figure 4.2 Differential amplitude spectrum of the y-acceleration signal highlighting changes from new pump condition to valve plate cavitation damage (stage III) across varying operating conditions.

Figure 4.2 illustrates the spectral difference between a healthy pump (new condition) and a pump exhibiting valve plate cavitation (stage III) as a function of pump order and operating condition. The difference spectrum directly reveals how the fault alters the acceleration amplitude at each harmonic order: blue regions indicate amplitude reductions relative to the healthy baseline, while red regions denote amplitude increases induced by the cavitation damage. Notably, certain orders, such as the 3rd harmonic, exhibit particularly pronounced differences, and the magnitude of these deviations varies significantly across operating conditions. This order-dependent and operating-point-sensitive behaviour confirms that

spectral analysis of the acceleration signal provides a robust means of detecting and characterizing condition changes in the pump.

4.3. Leakage flow prediction with supervised learning

The dataset of the 28 OCs (see Table 1) is used, including the data from the new condition, axial clearance, slipper wear, as well as valve plate cavitation I and III. The fault of valve plate cavitation II is not included, as it will be used to judge the model's ability to perform on unknown damages. When training deployment models, data may only be available for low and high fault intensity (e.g. valve plate cavitation I and III). During the deployment of the final model, actual faults are expected to lie in between these extremes as represented by valve plate cavitation II. The triaxial sensor is evaluated, whereby for this investigation all three axes x, y, and z are used at the full sampling rate of 16.4 kHz.

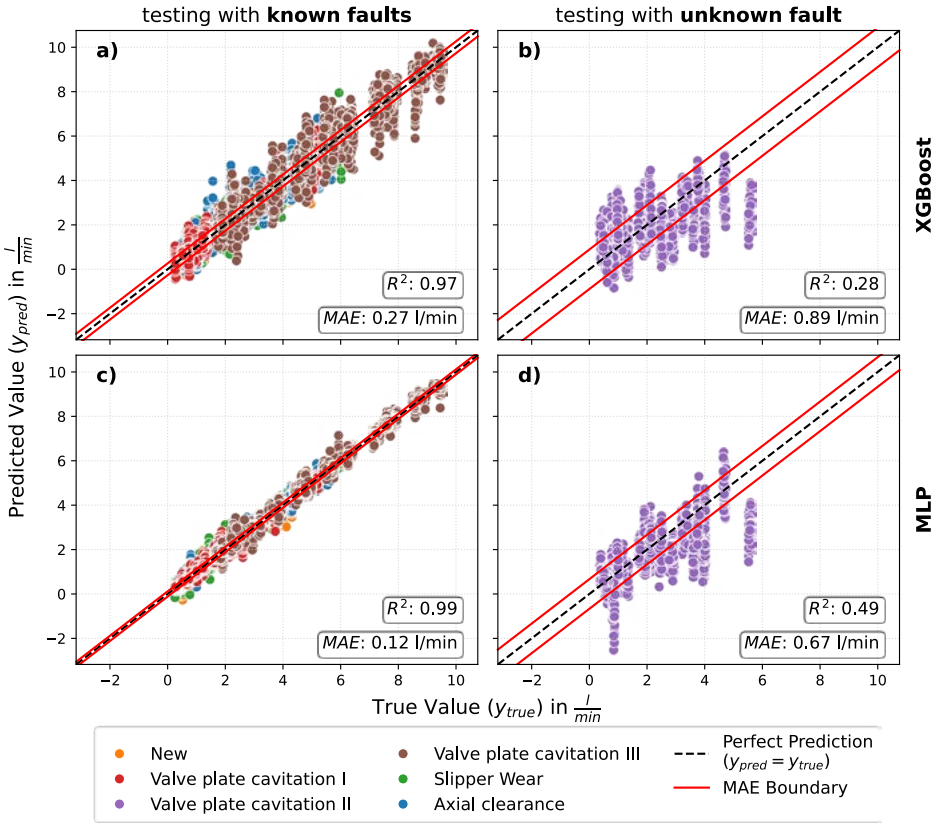


Figure 4.3 Comparison of leakage prediction with known pump conditions and unknown valve plate cavitation II damage not seen during training using XGBoost and MLP regression models.

Figure 4.3 a) shows the prediction result of the XGBoost regression model on known fault types. When predicting the leakage, it achieves a mean absolute error (MAE) of 0.27 l/min and a coefficient of determination (R^2) of 0.97. This high regression quality can be attributed to the fact that the same operating points and fault types were included in both the training

and test data. As a next step, the dataset valve plate cavitation II is used for prediction, as it is still completely unknown for the model. The prediction results are depicted in Figure 4.3 b) and show a larger variation and a MAE of 0.89 l/min with R^2 of 0.28. This lower performance is expected when predicting samples with an unknown fault type.

Figure 4.3 c) shows the same kind of scatter plot for predictions made using the Multilayer Perceptron model. Prediction quality improves across the board when compared to the previous XGBoost predictions reaching a MAE of 0.12 l/min when predicting for known fault types. Performance on the unknown valve plate cavitation II damage also improves to 0.67 l/min of mean absolute error in Figure 4.3 d). Judging by this first analysis the Multilayer Perceptron model shows a higher adaptability to the unknown fault type combined with better performance metrics on known damages than the XGBoost model.

For the valve plate cavitation II predictions (Figure 4.3 b) and d)), both models clearly reveal distinct operating conditions through the dispersion of the vertical lines. Since leakage flow strongly depends on these parameters (speed, pressure and stroke/displacement), the objective is to enhance the prediction of leakages under unknown faults by incorporating operating condition information as model features.

4.4. Leakage flow prediction with known operating condition

Up to this point all features used by the models are derived from acceleration time series data. For the following analysis, the operating parameters speed (revolutions per minute), pressure (bar) and stroke (% of maximum displacement) are added to the feature space. Using these features, models can now deduce that, for example high speed and stroke values lead to higher overall leakage across all fault types. As can be seen in Figure 4.4 the addition of operating parameters improves prediction quality for the unknown valve plate cavitation II fault and both models. The XGBoost Regression model show the biggest improvement for predictions on this fault type, increasing R^2 values from below 0.4 to above 0.8 and almost halving MAE to 0.47 l/min. The Multi-Layer-Perceptron shows smaller improvements for the unknown fault type with a higher variance of R^2 values between folds.

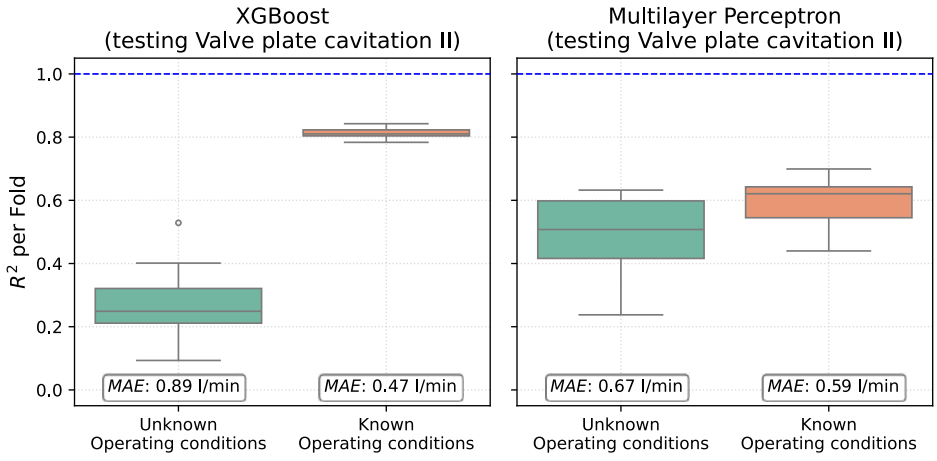


Figure 4.4 Improvement of leakage flow prediction when testing on unknown valve plate cavitation II fault and adding operating conditions to the feature space

4.5. Operating condition determination using vibration data

As demonstrated in the previous section, incorporating operating parameters (speed, pressure, and stroke) as features enhances the model's ability to predict leakage, especially for unknown fault types (Figure 4.4). However, this crucial information is often unavailable in retrofit scenarios, as dedicated sensors for pressure or stroke (i.e. swashplate angle sensors) may not exist or are unavailable to the CM system. The following analysis, therefore, investigates the feasibility of using the same high-frequency acceleration data to predict the operating conditions themselves.

Determining the operating condition from the vibration signal enables three key functions for a 'smart sensor'. First, it allows the recording of load history. By logging these conditions, load patterns become visible, which are important for assessing the system state. Second, it enables indirect monitoring. In applications without dedicated sensors, parameters like flow rate can be derived directly from the vibration data. Finally, it improves predictive maintenance. Analysing the connection between wear and load allows for a better prediction of potential pump failures. To achieve this, the following sections will test the ability of machine learning models to determine the operating condition (speed, pressure, and stroke value) using *only* the acceleration data, with the goal of realizing this functionality within a single sensor.

While the rotational speed can be analytically calculated from the frequency spectrum using spectral and order analysis, a similar analytical approach for pressure (e.g., correlating the amplitude of the first harmonic) was found to be unreliable. While a clear trend exists for the healthy pump, this correlation breaks down completely in the presence of faults, as the fault signatures interfere with the pressure-based amplitudes. This high complexity and variance, especially under fault conditions, lends itself to the application of machine learning models to predict speed and the remaining stroke parameter.

4.6. Operating condition prediction with ML-models

To predict the operating point from vibration data, both XGBoost and MLP regression models are employed. The feature space and the representative set of 28 operating conditions (Table 1) are identical to those used in the leakage prediction tasks.

A key challenge is that a deployed model will initially only be trained on data from a healthy pump. To simulate this, models are first trained exclusively on data from the undamaged pump (new) and then tested on data from the faulty components. As shown in Figure 4.5, this approach fails.

While the models achieve near-perfect accuracy on undamaged test data and XGBoost can still predict speed on faulty samples with $R^2 > 0.8$, the R^2 scores for pressure and stroke drop to negative values. This is attributed to the substantial impact of damage on the vibration spectra (see Section 4.2), which indicates the models have overfitted to the specific patterns of the healthy pump and cannot generalize to unknown fault conditions.

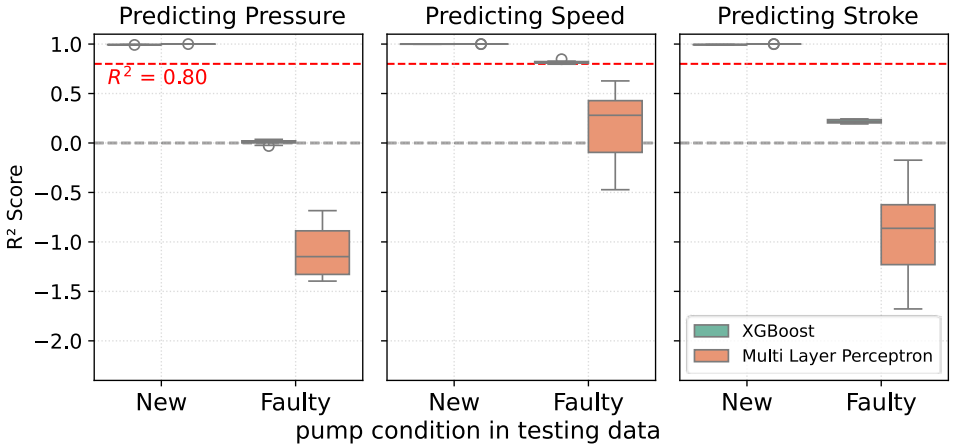


Figure 4.5 Prediction of pressure, speed and stroke with XGBoost and MLP on new and faulty condition training samples, with only new condition samples used for training

While training solely on healthy pump data yields limited generalization capability, extending the training set to include fault conditions significantly improves model performance. Figure 4.6 demonstrates this effect: when trained on a comprehensive dataset comprising all fault types (new condition, valve plate cavitation stages, axial clearance, and slipper wear), both XGBoost and MLP models achieve R^2 values exceeding 0.98 for all operating parameters.

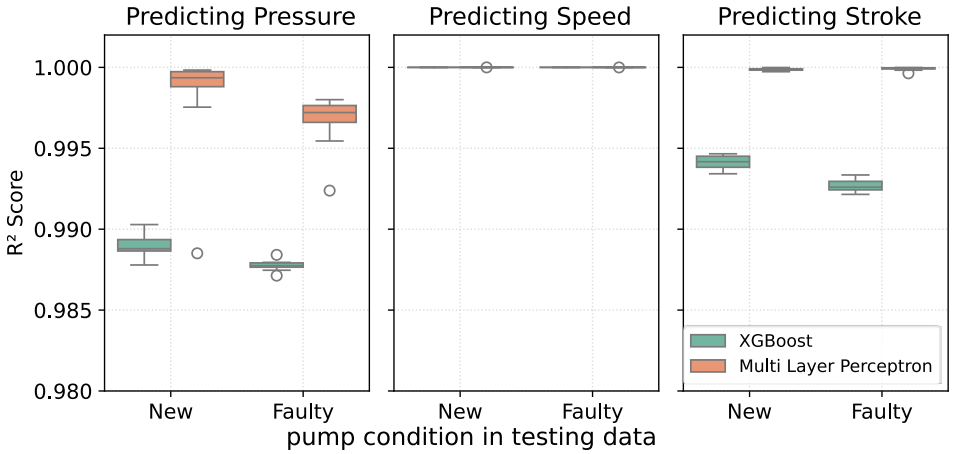


Figure 4.6 Prediction of pressure, speed and stroke with XGBoost and MLP on new and faulty condition training samples, with samples from all fault types used for training

This analysis confirms that accurate operating condition prediction solely from acceleration data is feasible, provided the training dataset captures the full range of expected pump conditions. Since industrial applications initially provide only healthy pump data, an adaptive training method is proposed. As the pump ages, the model is continuously retrained using its own predictions as 'quasi-true' labels, as long as internal performance metrics

remain stable. This process could allow the model to adapt emerging degradation patterns and maintain high OC prediction quality over the pump's service life.

4.7. Proof of concept for Edge Deployment

Implementing the model directly on an embedded system enables a secure, self-contained "smart sensor". This study validates the deployment of the MLP model on a low-cost ESP32-S3 microcontroller (< 10 €). Although XGBoost showed slightly higher accuracy in certain analyses, the MLP was selected as it lends itself more readily to efficient conversion into C++ code. The final model architecture accepts 703 input features (700 base features and three operating parameters). It consists of 3 hidden layers with 238, 41, and 187 neurons respectively, and produces a single output using the tanh activation function.

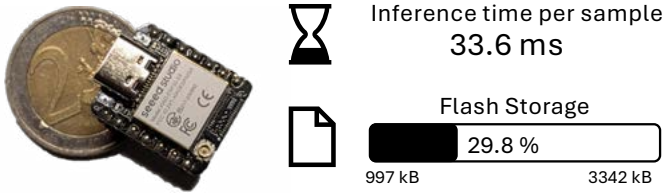


Figure 4.7 Validation of the 703-feature MLP model on a 10 €-ESP32 microcontroller: despite its complexity, the model occupies only 29.8 % of the storage and enables real-time predictions

For deployment, the optimized MLP model used in Section 4.4 (with knowledge of operating conditions) was converted to a C++ implementation via the ONNX format. Validation was performed by sending test samples to the microcontroller to measure real-world inference performance. As shown in Figure 4.7, the compiled code requires 997 kB, occupying only 29.8 % of the available flash storage. The average inference time is minimal at 33.6 ms per sample. This confirms that complex leakage prediction is viable on low-cost hardware.

5. SUMMARY AND OUTLOOK

This paper successfully demonstrated the foundational work for a low-cost, vibration-based "virtual sensor" for axial piston pumps and highly diverse operating conditions. The key findings are: 1) Predicting leakage flow for unknown fault types is only possible when the pump's operating conditions (OCs) are included as features. 2) These OCs (pressure, speed, and stroke) can be accurately predicted ($R^2 > 0.98$) using ML models from the same vibration signal. 3) This high accuracy is critically dependent on a robust training dataset that includes examples of various fault conditions; models trained only on healthy data fail to generalize. 4) The entire leakage flow prediction can be deployed on a low-cost (<10€) ESP32-S3 microcontroller with a minimal footprint (997 kB) and fast inference time (33.6 ms per sample), confirming its viability for edge computing.

The outlook for this work is focused on developing a fully autonomous, cascaded prediction system. This involves first predicting the OCs from the vibration signal (e.g., analytically calculating speed and using ML for pressure and stroke), and then feeding these predicted OCs into a second model to accurately estimate leakage. Future work will also explore more sophisticated feature engineering, such as using the heatmap analysis to define custom

frequency bands and combining the leakage prediction with anomaly detection for event recognition. Also feature types fit for leakage flow prediction may not represent the best features for OC prediction further motivating more in-depth feature engineering approaches. For a fully standalone system, calculating features directly on the microcontroller is essential. Future work will explore the ESP32-S3's dual-core design to run feature extraction and model inference in parallel. On a systemic level, the connectivity of modern microcontrollers opens the possibility of using Federated Learning (FL) a method where only model parameters, not raw data, are shared—to train robust models across a fleet of pumps without compromising data security.

6. ACKNOWLEDGEMENT

The presented research activities are part of the project “Causal Condition Monitoring von Verdrängereinheiten durch Verschleiß- und Schadensdetektion in Korrelation zur Betriebs- und Lastsituation” (Ref. No. 22012 BR/1). The authors would like to thank the Fluid Power Research Fund of the VDMA for the funding and support.

7. APPENDIX

Table 3: Hyperparameter optimization search space (unspecified parameters use defaults). Parameter names match variable names in both models.

MLP Parameter	Distribution	XGBoost Parameter	Distribution
learning_rate	[1e-4, 1e-1]	learning_rate	[0.01, 1]
alpha (L2 Regulization)	[1e-5, 1e-1]	n_estimators	[1, 1000]
max_iter	[200, 1000]	max_depth	[1, 32]
batch_size	[16, 256]	min_child_weight	[1, 10]
activation	ReLU, Tanh	subsample	[0.5, 1.0]
Number of hidden layers	[1, 3]	colsample_bytree	[0.5, 1.0]
Layer Dimensionality	[32, 256]		

AUTHOR CONTRIBUTIONS

As the main author, S.K. was responsible for designing and refining the condition monitoring procedure, handling the edge implementation, and creating the written content and figures for the ML and CM sections. S.H. was responsible for the figures and written content regarding artificial fault creation, the experimental setup, and the leakage flow analysis. J.W. acted as the overall scientific supervisor.

REFERENCES

[1] S. Horn, S. Knoll, A. Shorbagy, M. Lenz, and J. Weber, “Economic Fault Detection for Axial Piston Pumps Under Variable Operating Conditions,” International Journal of Fluid Power, vol. 26, no. 2, pp. 319–342, 2025, doi: 10.13052/ijfp1439-9776.2628.

[2] C. Dong, J. Tao, H. Sun, Q. Chao, and C. Liu, “Inverse transient analysis based calibration of surrogate pipeline model for fault simulation of axial piston pumps,” Mechanical Systems and Signal Processing, vol. 205, Dec. 2023, doi: 10.1016/j.ymssp.2023.110829.

[3] N. Kumar, R. Kumar, B. K. Sarkar, and S. Maity, “Condition monitoring of hydraulic transmission system with variable displacement axial piston pump and fixed displacement motor,” in Materials Today: Proceedings, Elsevier Ltd, 2019, pp. 9758–9765. doi: 10.1016/j.matpr.2020.09.327.

[4] N. Keller, A. Sciancalepore, and A. Vacca, “Demonstrating a Condition Monitoring Process for Axial Piston Pumps with Damaged Valve Plates,” International Journal of Fluid Power, vol. 23, no. 2, pp. 205–236, 2022, doi: 10.13052/ijfp1439-9776.2324.

[5] Y. Zhao, A. Jiang, W. Jiang, Y. Zhang, and S. Zhang, “Hydraulic pump fault diagnosis by modified slime mold algorithm optimized support vector machine,” Measurement, vol. 254, p. 117878, Oct. 2025, doi: 10.1016/J.MEASUREMENT.2025.117878.

[6] S. Tang, Y. Jiang, T. Zhou, Y. Zhu, and K. M. Lim, “Bayesian algorithm optimized deep model for multi-signal fault identification of hydraulic pumps,” Alexandria Engineering Journal, vol. 121, pp. 465–483, May 2025, doi: 10.1016/j.aej.2025.02.084.

[7] S. Tang, Z. Yan, Y. Xue, K. M. Lim, and Y. Zhu, “Deep transferable model with multi-head attention for small sample fault diagnosis of hydraulic pumps,” Engineering Applications of Artificial Intelligence, vol. 160, Nov. 2025, doi: 10.1016/j.engappai.2025.111808.

[8] S. Liu, Y. Zhang, C. Ai, and P. Liang, “Semi-supervised multi-adversarial domain-adaptive fault diagnosis for hydraulic pumps from pressure simulation data to experimental data,” Advanced Engineering Informatics, vol. 68, Nov. 2025, doi: 10.1016/j.aei.2025.103658.

[9] M. Liu, G. Kim, K. Bauckhage, and M. Geimer, “Data-Driven Virtual Flow Rate Sensor Development for Leakage Monitoring at the Cradle Bearing in an Axial Piston Pump,” Energies, vol. 15, no. 17, Sep. 2022, doi: 10.3390/en15176115.

[10] A. Schimke, O. Gnepper, and E. Boos, “Data-Driven Damage Assessment of Axial Piston Pumps,” in Proceedings of the 30th ICE IEEE/ITMC Conference on Engineering, Technology, and Innovation: Digital Transformation on Engineering, Technology and Innovation, ICE 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICE/ITMC61926.2024.10794220.

[11] T. Wang, S. Liu, Y. Fu, and J. Ma, “Health state evaluation index system for aviation piston pumps based on asynchronous multi-source information fusion,” Results in Engineering, vol. 27, Sep. 2025, doi: 10.1016/j.rineng.2025.107076.

[12] Q. Chao, Y. Shao, C. Liu, and X. Yang, “Health evaluation of axial piston pumps based on density weighted support vector data description,” Reliability Engineering and System Safety, vol. 237, Sep. 2023, doi: 10.1016/j.ress.2023.109354.

[13] Q. Chao, Y. Hu, and C. Liu, “Physics informed neural networks for detecting the wear of friction pairs in axial piston pumps,” Reliability Engineering and System Safety, vol. 261, Sep. 2025, doi: 10.1016/j.ress.2025.111144.

[14] Ö. Özmen, C. Sinanoğlu, A. Caliskan, and H. Badem, “Prediction of Leakage from an Axial Piston Pump Slipper with Circular Dimples Using Deep Neural Networks,” Chinese Journal of Mechanical Engineering (English Edition), vol. 33, no. 1, Dec. 2020, doi: 10.1186/s10033-020-00443-5.

[15] Y. Zhang, S. Wang, C. Zhang, H. Dui, and R. Chen, “Application of physics-informed machine learning in performance degradation and RUL prediction of hydraulic piston pumps,” Reliability Engineering and System Safety, vol. 261, Sep. 2025, doi: 10.1016/j.ress.2025.111108.

[16] C. C. Walani and W. Doorsamy, “Edge vs. Cloud: Empirical Insights into Data-Driven Condition Monitoring,” Big Data and Cognitive Computing, vol. 9, no. 5, May 2025, doi: 10.3390/bdcc9050121.

[17] S. Horn, A. Shorbagy, M. Lenz, S. Knoll, M. Winter, and Prof. J. Weber, “CAUSAL CONDITION MONITORING FOR AXIAL PISTON PUMPS: PREDICTION OF DIFFERENT FAULTS UNDER VARIABLE OPERATING CONDITIONS,” pp. 181–204, Oct. 2024, doi: 10.13052/RP-9788770047456.014.

[18] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, “Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package),” Neurocomputing, vol. 307, pp. 72–77, Sep. 2018, doi: 10.1016/J.NEUCOM.2018.03.067.

[19] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 13-17-August-2016, pp. 785–794, Mar. 2016, doi: 10.1145/2939672.2939785.

[20] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[21] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2623–2631, Jul. 2019, doi: 10.1145/3292500.3330701.

[22] Y. Duensing, “Zustandsüberwachung hochdynamischer bidirektionaler Axialkolbenpumpen mittels maschinellen Lernens,” 2025. doi: 10.18154/RWTH-2025-10412.

Biographies



Simon Knoll

Simon Knoll graduated from the Technical University of Dresden in 2025 with a diploma degree in Mechanical Engineering, specializing in fluid power. During his studies, he focused on research related to the slipper–swashplate interface in axial piston pumps and using machine learning with sensor data to monitor pump conditions. He is currently working as a research assistant at the Chair of Fluid-Mechatronic Systems.



Svenja Horn

Svenja Horn received her diploma degree in Aircraft Construction from the Technical University of Dresden in 2018. Since then, she has been working as a research assistant at the Chair of Fluid-Mechatronic Systems, focusing on improving the efficiency and service life of axial piston pumps. Her research concentrates on tribological phenomena at the slipper–swashplate interface, which she investigates using simulations, a hydrostatic tribometer and pump measurements. Furthermore, she applies machine learning algorithms for pump condition monitoring and lifetime prediction.

Improved System Efficiency through Drained Case Operation of Axial Piston Units

Dipl.-Ing. (FH) Thorsten Stiller, Dr.-Ing. Torsten Kohmäscher

CLAAS Selbstfahrende Erntemaschinen GmbH
Thorsten.Stiller@claas.com, Torsten.Kohmaescher@claas.com

Abstract

This article presents a drained case approach for axial piston units in mobile hydraulics to minimize splashing (churning) losses and enhance system efficiency. Utilizing a Venturi valve and existing flushing flows, the system evacuates an oil-air mixture from the motor housing without additional energy input. Experimental validation demonstrated evacuation times of 30–55 seconds and significant reduction in power loss. The solution is applicable with slightly modified standard hydraulic components and supports significant energy savings and lower operating costs in agricultural machinery.

Keywords: Hydraulic efficiency, axial piston units, splashing losses, mobile hydraulics, energy savings.

1. INTRODUCTION

The demand for increased efficiency and reduced CO_2 emissions in mobile machinery, such as self-propelled agricultural machines, has enabled innovation in hydraulic drive technology. Hydraulic drive technology is applied in open-circuit and closed-circuit arrangement to distribute the engine power throughout the machine and towards the point of power consumption. Some modern agricultural machinery – like the JAGUAR 1000 series in Figure 1 – makes significant use of this hydraulic power distribution.



Figure 1. JAGUAR 1000 - Hydraulic system and power distribution

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

The cylinder-based work functions are supplied by the constant pressure hydraulics (CPH system) which is an open-circuit system which is supplied by a variable work pump at constant 205 bar pressure level. High-power motor-based work functions are implemented as closed-circuit hydrostatic transmission. In total, 4 closed-circuit hydrostatic transmissions are utilized on the JAGUAR 1000 – the ground drive and the feeder drive as well as the variable header drives 1 and 2. Axial piston hydrostatic pumps and motors – in swash plate and bent axis design – are integral to the implementation of closed-circuit transmission due to their reliability, robustness, and power density. Improvements in these critical components can lead to significant cost savings and environmental benefits which will help hydrostatic drives to compete in the market with electric alternatives.

However, splashing losses caused by the rotation of the internal components (rotational kit) in the oil-filled housings remain a significant source of inefficiency [1, 2]. Reducing power losses will prevent substantial heat dissipation into the oil and simplify the thermal management on the mobile machine. Draining the oil from the case of the hydrostatic components significantly reduces the power loss and is referred to as drained case operation at CLAAS. This paper examines the innovative system solution for drained case operation and minimized hydro-mechanical power losses in mobile machines as introduced to the market at Agritechnica 2025.

2. FUNDAMENTALS OF HYDROSTATIC TRANSMISSIONS

Hydrostatic pumps and motors are components which are design to convert mechanical power into hydrostatic power (pump) and back to mechanical power (motor). The hydrostatic transmission (HST) in Figure 2 combines pump (1) and motor (2) into a closed-circuit system that provides an infinitely variable ratio from the pump input shaft towards the motor output shaft. This closed circuit must be pre-charged via the charge pump (3), filter (5), relief (4) and charge valves (6) to reduce air content in the oil and thereby increase the stiffness of the drive. Heated oil is flushed from the system into the cooler (9) via flushing (7) and purge valve (8). High pressure relief valves (10) protect the system against overload. Due to the dual high-pressure lines, the HST can be operated in four quadrants – forwards, reverse, uphill, and downhill.

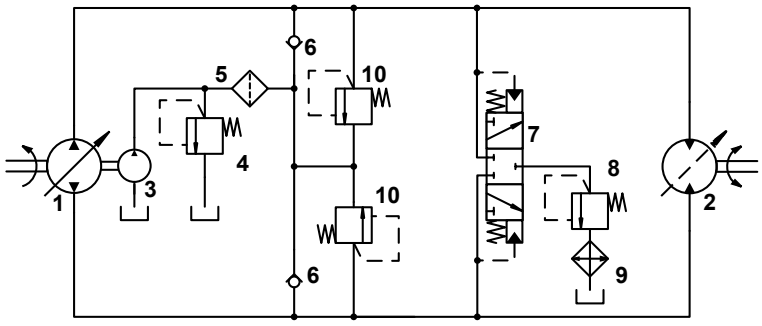


Figure 2. Hydrostatic transmission in closed-circuit system [3, 4]

Most of the hydraulic flow constantly circulates inside the closed-circuit system to transmit the hydraulic power from the pump to the motor. Consequently, the required exchange volume in the hydraulic tank is reduced and it can be designed smaller.

Axial piston units (pumps & motors) are complex components that are developed for efficient conversion of mechanical into hydraulic energy and vice versa. In mobile machinery the power of the Diesel engine is converted into hydraulic flow and pressure by the hydraulic pump. The hydraulic power is converted back to mechanical power and motion by the hydraulic motor. Axial piston units operate on the principle of axial movement, in which pistons slide dynamically within cylindrical bores, driven by a swash plate (pump) or by a bent axis (motor).

Losses in hydrostatic units are differentiated in volumetric and hydro-mechanical losses. While volumetric losses impact the flow of the hydrostatic unit, hydro-mechanical losses impact the torque. Splashing losses are caused by the rotation of the internal kit of the hydrostatic unit in the oil-filled housing (see Figure 3, left) and are considered part of the hydro-mechanical losses. These losses are strongly dependent on the rotational speed of the hydrostatic unit und reach significant values. Due to higher speed levels of hydrostatic motors, the splashing losses in hydrostatic motors are generally higher. Draining the oil from the housing (Figure 3, right) prevents the splashing and thereby eliminates the splashing losses in the hydrostatic unit.

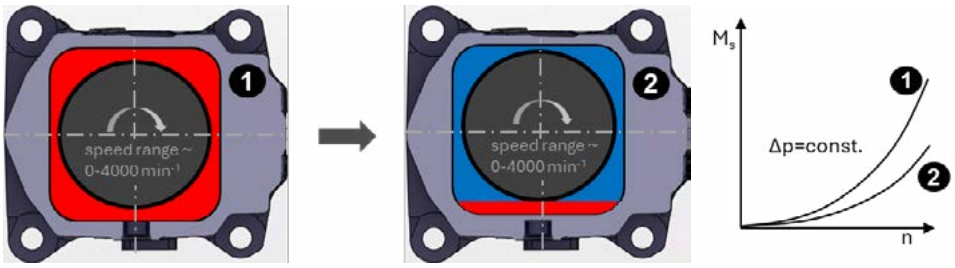


Figure 3. Sectional view of hydrostatic unit with full (1) and drained (2) housing

The introduced benefits of drained case operation are well known to the industry since many years [5]. The component suppliers can develop the hydrostatic units to allow drained case operation, but the OEM must develop a system solution around this hydrostatic unit to get drained case operation started. With the introduced system solution, the benefits of drained case operation outweigh the additional system cost.

In 2025, Grimme presented the outcome of a real-life study of drained case operation on a self-propelled beet harvester. The four motors of the hydrostatic ground drive are drained from oil by air from the compressed air system of the machine. There was no negative effect identified on the components or the hydrostatic system and the expected savings from test stand measurements were reproduced in real life on a real machine [6].

3. DEVELOPMENT OBJECTIVES AND TECHNICAL CHALLENGES

During the development of the JAGUAR 1000 series, new hydraulic functions had to be considered in the thermal management of the machine. It became obvious that drained case operation had to be implemented to control the losses in the hydraulic system. Further development of the serial solution for drained case operation on the CLAAS SCORPION telehandler allowed for reduction on the power losses.

Summary of the key requirements during the development:

- Integration into compact, cost-sensitive mobile hydraulic system
- Minimize splashing losses in standard hydrostatic components during field and road operation (close collaboration with supplier required)
- Leverage existing flushing flows in the Venturi nozzle to avoid additional power consumption
- Rapid and complete evacuation of the motor housing during operation
- Effective de-aeration of the evacuated oil to prevent cavitation and maintain lubrication in the hydrostatic units
- Compatibility with supplier specifications (e.g. temperature, pressure limits)

The following paragraphs will describe the implementation of the drained case solution in the CLAAS JAGUAR series.

4. DRAINED CASE OPERATION

Drained case operation minimizes splashing losses and leads to a more efficient machine operation. By implementing a suction and ventilation system, the hydrostatic unit can continuously drain fluid from the housing. Reduced thermal load on the system reduces the power needs for cooling of the hydraulic fluid.

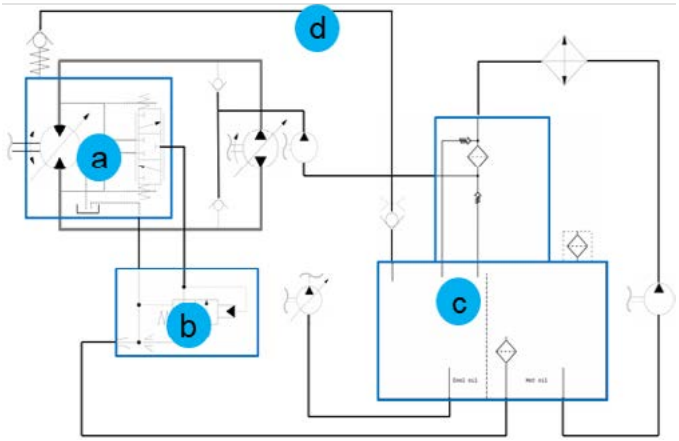


Figure 4. Conceptual schematic of the system solution for drained case operation

4.1. System Layout and Key Components

The system solution for drained case operation is introduced as a conceptual schematic in Figure 4. It becomes obvious that drained case operation cannot be supplied by a single supply, however it must be introduced by the OEM as system integrator.

The key components of the drained case system solution (see Figure 4):

- Hydrostatic motor (a) – Transforms hydrostatic to mechanical power in the closed-circuit transmission and is the source of splashing losses in standard design.
- Venturi nozzle (b) – Utilizes the loop flushing flow from the closed-circuit hydrostatic transmission as driving force for the suction from the hydrostatic motor.
- Hydraulic tank and filter system (c) – Separates the air from the hydraulic oil and allows for additional volume from drained hydrostatic motor.
- Ventilation system (d) – Essential for continuous suction of oil-air mixture from the hydrostatic motor.

A robust system solution requires an optimized alignment of components and performance parameters. The interaction between the different machine parts must be analyzed carefully and the required adjustments to the system must be carried out with care. The decision was taken that the Venturi principle shall be used to implement the drained case feature.

4.2. Hydrostatic Motor

The hydrostatic motor is an essential part of the closed-loop hydrostatic transmission and transforms the hydrostatic power into mechanical power which can be utilized in the ground drive or the header drive of the JAGUAR 1000. This main functionality remains intact, but the motor requires some adjustments to allow drained case operation.

The Venturi principle which was selected as the driving force for the evacuation of the housing requires a driving hydraulic flow as the power input to the nozzle. With the loop flushing in the hydrostatic motor, this driving force is already existing in the hardware of the hydrostatic motor. However, this flow must be re-routed from the motor housing to an external port to be connected to the Venturi nozzle.

Adjustments to the hardware and the operation conditions of the hydrostatic motor can only be realized in close cooperation with its supplier. For the initial implementation of the drained case operation, the HMV swash-plate variable motors from LHY were modified to fit the requirements – as single and dual variant. In particular, the motor also had to be prepared to be robust in drained case operation.

The following modifications to the HMV were necessary:

- Re-routing of loop flushing flow to an external port as input to the Venturi nozzle
- Evacuation port at the bottom of the unit (mounting position of Venturi nozzle)
- External flushing supply for the main bearings of the hydrostatic unit

Besides these modifications, test runs for the hydrostatic units were necessary to confirm temperature levels in drained case operation that lead to sufficient viscosity and lubrication in the key tribological interfaces. Figure 5 introduces a cut-away view of the hydrostatic motor which provides an impression of the rotating parts of the hydrostatic kit in the hydrostatic motor.

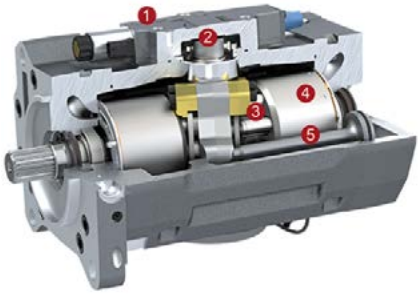


Figure 5. HMV swash-plate variable motor from LHY [7, 8]

4.3. Venturi Valve

The Venturi valve utilizes the Bernoulli effect to generate a suction effect from the loop-flushing flow which is leveraged to evacuate the hydrostatic units. Thereby enabling the evacuation without any additional energy input.

The Bernoulli effect states that the local pressure p_2 of a fluid decreases from the static pressure p_1 as the speed of this fluid increases from v_1 to the accelerated speed v_2. Purposefully increasing the speed of a fluid – e.g. by means of a nozzle – creates suction which is the working principle behind jet pumps. This effect is also well known in pump or valve design and can lead to cavitation as fast jets of fluid occur.

Accordingly, the pressure drop in the suction pressure p_2 versus the pressure in the driving flow p_1 is dependent on the speed of the fluid before (v_1) an after the nozzle (v_2) according to equation 4.1.

$$p_2 = p_1 - \frac{\rho}{2} \cdot (v_2^2 - v_1^2) \qquad (4.1)$$

With the continuity assumption for the hydraulic flow

$$v_1 \cdot A_1 = v_2 \cdot A_2 \qquad (4.2)$$

and the calculation of a round area

$$A = \frac{\pi}{4} \cdot d^2 \qquad (4.3)$$

the dependency on the diameter of the nozzle (d2) versus the diameter of the inlet of the Venturi valve (d1) becomes visible in equation 4.4.

$$p_2 = p_1 - \frac{\rho}{2} \cdot v_1^2 \cdot \left(\frac{d_1^4}{d_2^4} - 1\right) \qquad (4.4)$$

The effect of suction starts as soon as the suction pressure drops below the pressure in the housing of the hydrostatic unit. These assumptions are only valid for ideal, unobstructed, and frictionless flow which cannot be guaranteed in real life applications. Nevertheless, this is the basis for the development of the Venturi nozzle in the CLAAS application.

Figure 6 presents a schematic with the key characteristics of the Venturi valve which consists of the jet nozzle, the intake nozzle, the mixing chamber, and the diffusor. The main design parameters that were developed and confirmed by real experiments are labelled in the schematic. Another critical parameter is the flow rate and pressure in the driving flow from the loop-flushing valve of the hydrostatic motor.

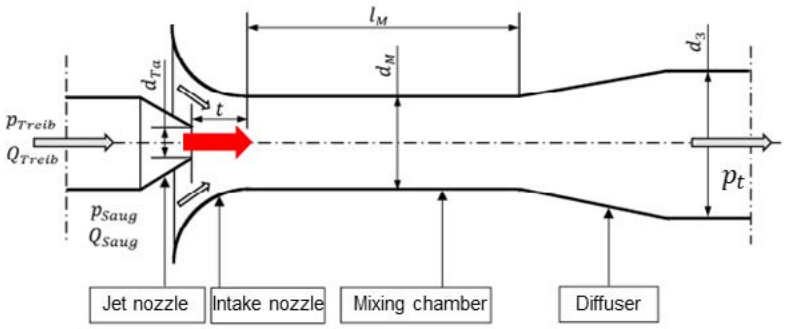


Figure 6. Schematic of the Venturi nozzle applied in the CLAAS Venturi valve

The key advancement in the presented system solution is the utilization of the existing loop flushing flow as the driving force for the suction effect – an efficiency improvement versus existing solutions in the market.

4.4. Hydraulic Tank and Filter System

The hydraulic tank serves as the reservoir for the pendulum volume of the hydraulic system which includes the volume resulting from draining the housings of the hydraulic motors. In the hydraulic system of the JAGUAR 1000, the tank is essential for filtration, thermal separation, as well as the de-aeration of the hydraulic fluid. Figure 7 shows a cut-away view of the tank (left) and the results of the CFD simulation that was performed to optimize the tank design.

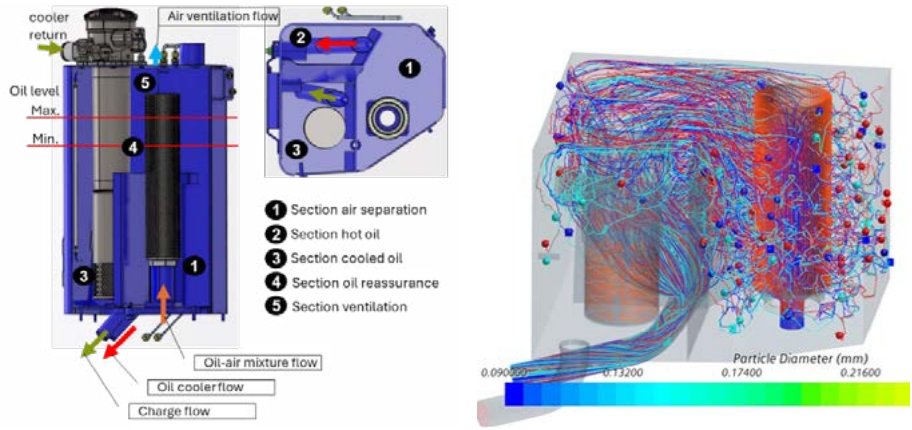


Figure 7. Desing of the hydraulic tank and visualization of CFD simulation

The drained oil-air mixture from the motor housings enters the hydraulic tank through the air separation element which the oil flow passes from bottom to the top in section 1 for the

air separation. After transition to the “hot oil” section 2 the oil it pumped through the oil cooler and returns into the tank via the filter element into the “cooled and cleaned oil” section 3. From this section, the flows for charging of the closed-circuit systems and the work hydraulics are taken. The connection port to the ventilation system (5) is in the cover plate on the top of the hydraulic tank. This port is kept free of oil to guarantee circulation of air in all operation conditions.

The design of the hydraulic tank was optimized for thermal management and de-aeration by means of CLAAS internal CFD simulations. The target for the de-aeration of the oil in the application is volumetric air content of less than 1%.

4.5. Ventilation System

When thinking about the drained-case solution, the need for the ventilation system of the motor housings is not obvious. For the continuous and fast suction of an oil-air mixture from the motor housings, the system must feed air back to the housings. In figure 8 this Venturi valve is mounted to the bottom drain port of the housing while the ventilation port is in the top of the housing.

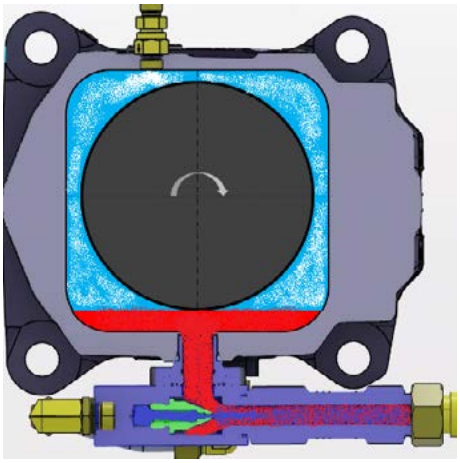


Figure 8. Cross section of drain motor with ventilation (top) and Venturi valve (bottom)

The illustration in figure 9 visualizes the ventilation system, the essential air circuit for the successful drained-case operation. The flow of air through the top port in the cover plate of the hydraulic tank is restricted by a check valve and a restricting orifice. With this restriction, the pressure level in the motor housings can be kept lower.

There is another pre-loaded check valve (0,2 bar) in the top of the motor housing to prevent any oil from entering the ventilation system. When oil enters the ventilation system, draining the motor housings is significantly delayed.

The oil-air mixture is drained from the motor housings by means of the loop-flushing flow for the closed-circuit hydrostatic transmission. The purge valve, included in the Venturi valve, secures the minimum pressure level on the low-pressure side of the closed-circuit system. The air is fed back to the hydraulic tank as oil-air mixture and is separated again in the separation element.

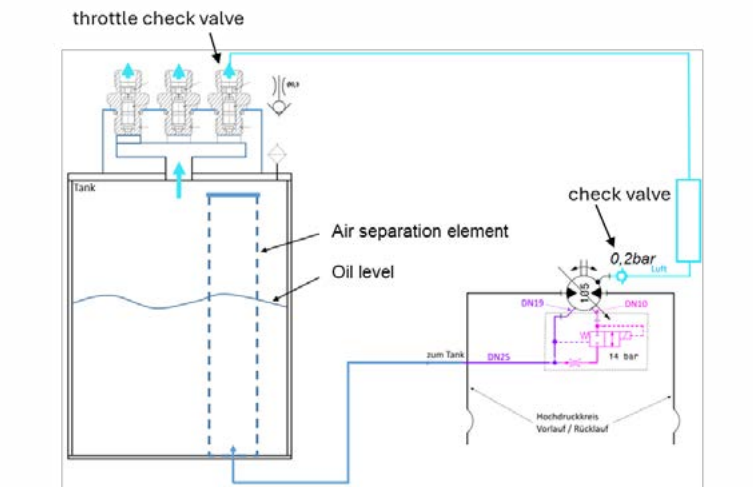


Figure 9. Illustration of the ventilation system for a single motor

The ventilation system is closed towards the environment to keep water, dust and other particles away from the hydraulic system.

5. System Integration in Mobile Machinery

The integration of the system for drained-case operation into a self-propelled forage harvester – or similar mobile machinery – requires careful design modification across various mechanical and hydraulic components. The interaction between these different machine parts – including pumps, motors, filters, and tank – must be analysed thoroughly before actions can be taken.

The introduced solution for drained-case operation is purely hydro-mechanical and does not require any electric actuation. Evacuation of the housings is initiated solely by the loop-flushing flow which starts as soon as sufficient pressure difference is build-up in the hydrostatic transmission. Temperature sensors and software is introduced only as monitoring system to prevent the operation outside of specified limits.

The various modifications to components require close collaboration with the respective suppliers of hydrostatic components, tanks, valves, and filtration system. The additional hydraulic lines of the ventilation system are the only modification which is visible for the outside. This modification requires additional design space on the machine and causes some additional system cost.

The assessment of the operating conditions of the machine is essential to determine the main sources for the heat load on the hydraulic system in critical situations. The data from telematics systems which is available today helps to identify the peaks loads on the cooling system. This way, the required modification and the additional system cost can be balances against other counter measures. Thoughtful introduction of the drained-case operation can lead to a competitive advantage in terms of efficiency and machine performance.

6. VALIDATION AND TESTING

Validation and extensive testing are critical to confirm the operational integrity and functionality of the system solution for drained-case operation. On the one hand, the system must fulfil the performance requirements. On the other hand, the system must be safe and reliable. Precise data acquisition for monitoring of pressures, speeds, temperatures, and fluid flows provided valuable insights into system performance.

For many years, the optimization of the Venturi nozzle was researched and implemented on development machines in the CLAAS hydraulics workshop. The main objective was the utilization of the loop-flushing flow as the driving force. Only this way, the benefits of drained case operation could lead to the required loss reductions on system level.

As mentioned before, the collaboration with our suppliers was essential for the success of the market introduction. The hydraulic motors had to be reliable in drained-case operation at high rotational speeds and pressure loads. Design modifications were confirmed in extensive test series which lead to the approval of the new CLAAS operation conditions. Some temperature thresholds in the monitoring system had to be adjusted based on the created experiences.

Machine level integration tests of the drained-case solution were necessary to confirm that the modifications had no negative impact on machine operation. For the ground drive, which is very relevant for the customer experience, a standard test series was executed with accelerations, decelerations, load changes, reversals, forwards and reverse driving. Finally, the system got approval for serial introduction.

During the testing, it was confirmed that an evacuation time of 30 to 55 seconds was reached by the system. Based on the dynamics in the thermal behaviour of a mobile machine, this duration is considered as “rapid”. Especially when considering the absence of any additional power consumption for the evacuation.

Additionally, the targeted savings in power loss of 5 to 10 kW per hydraulic motor were confirmed during the testing on machine level. This confirmed two things at once. The evacuation of the motor housing was complete and the expected savings from test stands could be reproduced in real machine operation.

Measurements of the volumetric air content in the oil behind the tank element for air separation confirmed a consistent value below 1%. The experience with this value and the required threshold for reliable machine operation was very limited at CLAAS and our suppliers. However, based on long-term validation runs on development machines it was concluded that the existing system design is well suited for the CLAAS lifetime requirement.

This final validation and testing also confirmed that the original development targets were achieved and fulfilled.

7. CONCLUSION AND FUTURE STEPS

This paper has described the transformation approach for reducing the power loss of a system with hydrostatic transmissions by means of a system solution for drained-case operation. The operation of the drained-case system is based on the Venturi effect. High velocity loop-flushing flow through the Venturi nozzle creates a low-pressure zone which is

sucking an oil-air mixture from the motor housing into the hydraulic tank. The air is fully separated from the oil in the tank and a volumetric air content below 1% is maintained. The results which were obtained in the testing and validation phase confirm the effectiveness of the drained-case operation in significantly reducing the hydraulic losses.

The drained case system demonstrated:

- Significant reduction in splashing losses in hydrostatic components
- Utilization of the existing loop-flushing flow which additional power consumption
- Reliable operation across all tested speed and load conditions
- Compatibility with standard hydraulic components and CLAAS machine system

As a conclusion it can be stated that drained-case operation of axial piston units offers a practical and effective method for improving the efficiency of hydraulic transmissions. It is particularly applicable for self-propelled agricultural machinery where space, cooling capacity, and cost constraints are critical.

The achieved savings in power loss of 5 to 10 kW per hydrostatic component (dependent on operation conditions) are sufficient to compensate for the additional system cost. Additional system cost is mainly created by the Venturi valve, the ventilation system, and the air separation element in the hydraulic tank. Reduced power loss translates directly into higher productivity of the agricultural machine and lower operational cost for the machine operator.

As the markets continue to focus on efficiency and fuel consumption, the assessment of power losses will be very relevant for future improvements on system level. The drained-case operation can be one contribution to the benefit of hydrostatic transmissions in the competition with alternative drive technologies.

Future work will focus on expanding the compatibility with various hydraulic components and exploring application in additional CLAAS machines. Efficiency modelling and evaluation of historic data for machine operation will support the introduction of drained-case operation to other machines and functions.

8. REFERENCES

[1] Ivantysyn, J., Ivantysynova, M.: Hydrostatic Pumps and Motors: Principles, Design, Performance, Modelling, Analysis, Control and Testing. Tech Books International, 2001.

[2] Schmitz, K., Murrenhoff H.: Fundamentals of Fluid Power – Part 1: Hydraulics. Shaker Verlag, 2018.

[3] Geimer, M.: Mobile Working Machines. SAE International, 2020

[4] Kohmäscher, T.: Modellbildung, Analyse und Auslegung hydrostatischer Antriebsstrangkonzepte. Dissertation RWTH Aachen, Shaker Verlag, 2008

[5] Rahmfeld, R., Marsch, S., Göllner W., et al.: Efficiency Potential of Dry Case Operation for Bent-Axis Motors 8th International Fluid Power Conference (IFK), Dresden, 2012

[6] Biemann, C.: The CO2 Reduction Potential Exemplified by a Hydraulic Drive System for Self-Propelled Harvesting Machines. VDI-Berichte Nr. 2465, 2025

[7] Lasaar, R., Schneider, H., Dückinghaus, H.: Development of new hydrostatic motor in swashplate design, 73rd International Conference on Agricultural Engineering LAND.TECHNIK AgEng (VDI-MEG). Hanover, 2015

[8] LHY.: Dry Case for HMV 02 D, Fact Sheet Dry Case Technology, https://lhy.com/product/hmv-02-d-dry-case/, retrieved on 05.01.2026

Biographies



Dipl.-Ing. (FH) Thorsten Stiller received the diploma degree in mechanical engineering in 1997. He is currently working as development engineer in the System Technology Hydraulics for CLAAS Selbstfahrende Erntemaschinen in Harsewinkel, Germany. His work experience is mainly focused on hydrostatic ground drive systems for harvesting machinery.



Dr.-Ing. Torsten Kohmäscher received the diploma degree in mechanical engineering from RWTH Aachen University in 2003 and the doctorate degree from Institute for Fluid Power Drives and Controls at RWTH Aachen University in 2008. He is currently working as the Head of System Technology Hydraulics for CLAAS Selbstfahrende Erntemaschinen in Harsewinkel, Germany. His research areas include simulation, layout, design, and mechatronic control of hydrostatic ground drives – throughout his university and professional life.

Real time fault detection in hydraulic system for wheel loader application

Fabio Scolari, Federica Grossi, Francesco Pintore, Ciro Mariniello

All the authors are with CNH, fabio.scolari@cnh.com, federica.grossi@cnh.com, francesco.pintore@cnh.com, ciro.mariniello@cnh.com

Abstract.

This paper presents the development and implementation of a real-time simulation model for the hydraulic implements of a wheel loader, with a focus on fault detection applications. The model replicates the behavior of the hydraulic system controlling the boom and bucket, incorporating components like load sensing pumps, control valves, and double-acting actuators. It was engineered for real-time performance without sacrificing accuracy, making it suitable for embedded systems. After the validation phase - performed using experimental data from a sensor-equipped wheel loader - the model was then deployed on an embedded platform, where it processes live sensor inputs and compares predicted actuator positions with actual measurements. Significant deviations can indicate faults such as leaks, wear, or sensor drift, allowing early detection and operator alerts. This approach enhances safety and reduces downtime. Ultimately, the model functions as both a virtual sensor and diagnostic tool, and its methodology is applicable to other electro-hydraulic systems promoting wider adoption of virtual simulation in off-highway equipment.

Keywords. Fault detection, real-time, digital twin, virtual model, off-highway equipment, hydraulics.

1. INTRODUCTION

Construction and off-highway machinery are becoming increasingly automated and sensorized, so simulation technology offers a powerful approach for enhancing operational reliability and predictive maintenance [1].

Fault detection in hydraulic systems, particularly for heavy-duty mobile machinery such as wheel loaders, is essential for ensuring operational reliability, safety, and productivity. The main fault detection methods [2] for mobile equipment and hydraulic systems can be classified as follows: Model-Based methods [3][4], which utilize mathematical or physical models to predict expected system behavior and detect faults by comparing measured signals with model outputs; Data-Driven methods [5][6][7], which employ statistical or machine-learning techniques to identify anomalous patterns in sensor data without relying on explicit physical models; Signal-Based methods [8][9], analysing time-domain or frequency-domain

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

characteristics of pressure, flow or vibration signals; and Hybrid approaches [10], such as Digital Twin [11][12]frameworks that combine model-based and data-driven techniques by integrating physical models with real-time data.

In the frame of model-based fault detection methods, the main approaches are the following: Parity Equation-Based Detection [13] describes the use of parity equations (analytical redundancy) in mobile hydraulic systems to generate residuals and identify faults through threshold-based deviations; Nonlinear Observer with Adaptive Robustness [4] introduces a nonlinear adaptive robust observer addressing uncertainties and disturbances for detecting sensor faults, fluid contamination, and supply pressure deviations; Interactive Multiple Model (IMM) Filters [14] Applies IMM filtering with smooth variable-structure filters to detect and isolate friction and leakage faults in hydraulic actuators; Multi-Model Observer-Based Diagnosis [15] utilizes a bank of parallel models (observers) representing normal and fault conditions where the model with minimal output error indicates the current system state; Comprehensive Parity-Space and Observer Approaches [16] presents a unified treatment of observer-based residual generation, parity-space techniques, Kalman filters, and isolation strategies in a state-space framework.

Traditional maintenance approach on off-highway vehicles is time-based, that means preventive actions are scheduled regardless of the actual condition of the components [17][18]. Time-based approach relies on routine inspections and replacements, which can lead to unnecessary system downtimes and inefficiencies.

The target of this project is failure and predictive maintenance analysis and observation using a hard real time model integrated on physical vehicle electronic control unit (ECU) during standard manoeuvres. In this paper the focus is on the high-pressure circuit of a wheel loader, a heavy-duty construction machine equipped with a front implement used for digging, lifting and transporting bulk materials like soil, gravel, or debris.

In fluid power systems, the application of a failure detection systems can significantly enhance operational reliability, safety and productivity. These techniques can be implemented across various components such as pumps, valves, and motors, enabling early detection of wear or malfunction, preventing failures that could compromise safety.

The condition monitoring of single components inside a hydraulic circuit is described in many papers using different techniques. Raduenz et Al. in [19] describe a method for condition monitoring and fault detection on proportional valves based on spool position and supply current monitoring to detect deviations from healthy behavior, while Kumar et Al. in [20] investigated the condition monitoring of hydrostatic transmission by developing a detailed mathematical model of the system and the analysis of the effects of the loss gap that leads to state the pressure/flow deviation as a key indicator for the fault detection. Applications to hydraulic pumps can be found in [9] and [21]. In particular [9] investigates the use of wavelet analysis for real-time health diagnosis of hydraulic pumps, focusing on pulsation discharge pressure signals rather than vibration signals. Traditional methods (limit checking, spectrum analysis, logic reasoning) struggle due to high noise levels and low signal-to-noise ratios in pressure signals. The proposed method uses wavelet decomposition that provides sub-band signals with distinguishable features; patterns and amplitudes of wavelet coefficients correlate with defect types and this method works on-line in real time without affecting pump operation thus providing robust fault detection in hydraulic systems.

Paper [21] introduces a deep learning approach for anomaly detection in hydraulic piston pumps with demonstrated improvements in diagnostic accuracy. Additionally, in the context of wheel loader applications, [7] proposed a fault diagnosis of a Wheel Loader, but instead of using model based fault detection it exploits a hierarchical approach based on artificial neural networks (ANN) and fuzzy logic in order to identify the subsystem of the vehicle affected by the fault and the type of fault with inputs from 24 sensors.

Building upon this background, the present paper integrates model-based fault detection with a real-time embedded Digital Twin deployed on the vehicle's ECU. This hybrid strategy leverages both physics-based modeling and data-driven insights, thereby enhancing robustness and minimizing false alarms due to noise or disturbances. A time-counter mechanism further filters transient anomalies, ensuring that only persistent deviations above threshold levels trigger fault alerts.

In this activity, a dynamic model of the hydraulic circuit controlling the boom and bucket actuation was developed to simulate the subsystem's behavior under real operating conditions. The simulation model includes key elements such as Load Sensing (LS) pump, control flow valves, and hydraulic double acting actuators. The model was designed with real-time computational efficiency as a core requirement, without compromising model fidelity. This enables its deployment on embedded hardware ensuring responsiveness suitable for continuous monitoring in live operation.

To validate the model, experimental data were collected from instrumented wheel loader operating under various load conditions and duty cycles. Sensor data, including cylinder pressures, actuator displacements, and control inputs, were used to calibrate and correlate the simulation model. The simulation results demonstrate a good degree of correlation with the measured system behavior, confirming its suitability for use as a real time replica.

The correlated model was implemented on a real-time embedded computing platform. The model receives live input signals from the machine's onboard sensors and computes the expected positions of the boom and bucket in real time. This virtual output is continuously compared with actual position measurements of the actuators. Deviations between the predicted and measured positions provide a basis for fault detection. Differences beyond predefined thresholds may indicate potential faults such as hydraulic leakage, actuator wear, sensor drift, or electronic control system anomalies. By analysing these deviations, the system can notify the operator of emerging issues. This early fault detection capability enhances safety, reduces unplanned downtime, and contributes to the overall robustness of the machine.

The proposed solution thus acts as both a virtual sensor and a diagnostic tool in heavy machinery. The work demonstrates how physics-based modeling, when combined with real-time computation and sensor integration, can lead to significant advances in machine reliability. The methodology and implementation described in this paper are generalizable to other systems, not only electro-hydraulic, promoting wider adoption of virtual simulation in off-highway equipment.

2. DESCRIPTION OF THE SYSTEM

2.1. Wheel Loader

The Wheel Loader considered in this work is a Case 821G, see Figure 2.1, which main characteristics are listed in Table 2.1.



Figure 2.1. Wheel loader Case 821G

Table 2.1. Main vehicle features

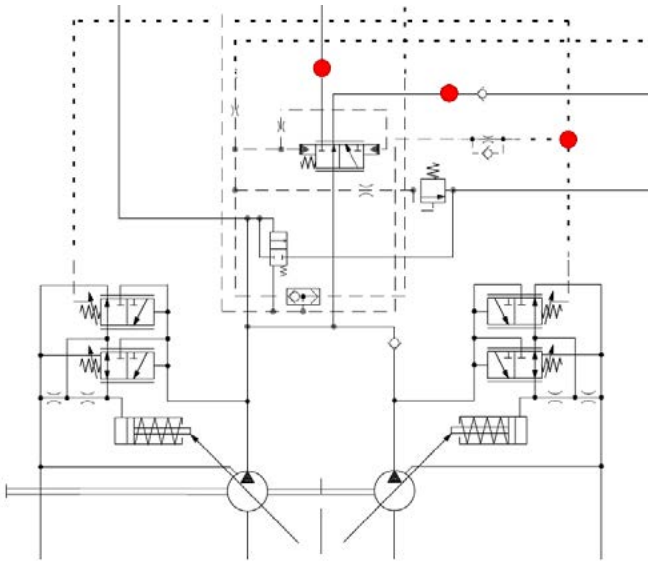
Maximum Engine Power	172 kW
Hydraulic Pump	Tandem, variable displacement
Maximum Oil Flow	236 l/min

2.2. Documentation

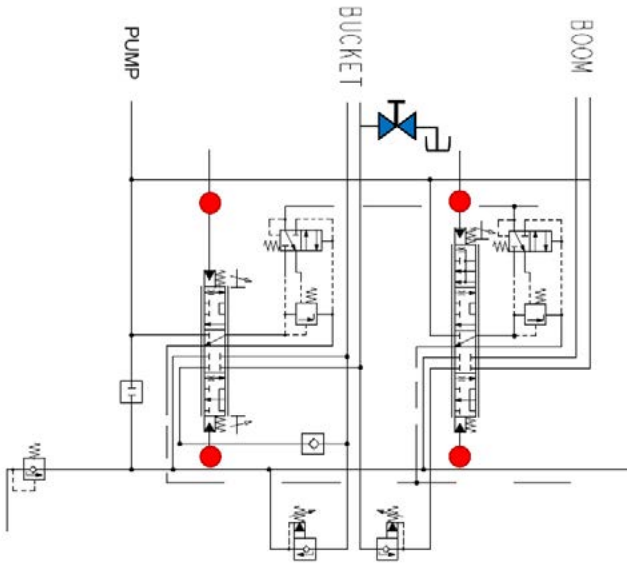
The following Documentation is the starting point for the work:

- Hydraulic Schematic
- Mechanical Design 3D drawings
- CAN Configuration
- Electric/Electronic Interface

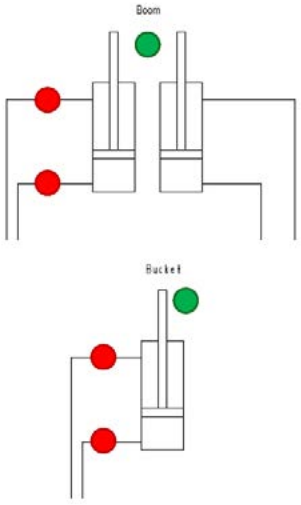
The main parts of the high-pressure hydraulic circuit of the vehicle are shown in Figure 2.2.



(a)



(b)



(c)

Figure 2.2. ISO drawings of main components: (a) CCLS hydraulic twin pumps; (b) Directional valves; (c) Boom and bucket cylinders. Red dots indicate the position of pressure sensors; green ones the displacement sensors.

It is a Load Sensing system fed by two twin variable displacement pumps, that work with two slightly different values of pump margin to ensure the pumps works in sequence, Figure 2.2a. A priority valve is present to provide oil first to steering, the most critical function; the

excess flow goes to the implement, boom and bucket. In the case study vehicle, the boom has two cylinders, while the bucket just one, Figure 2.2c. Directional valves are placed to control the flow to boom and bucket, Figure 2.2b: they are hydraulic piloted valves, actuated by a pilot pressure signal provided by electronically controlled valves. These are actuated by the electric signal of the joystick and generate the correct pilot signal to move the spools of the directional valves. In this way, the movement of the driver can control the movement of the implement.

2.3. Procurement and Physical Installation

In order to acquire the needed measurements, the vehicle has been set up with sensor installation. Then a preliminary verification of possible installation issues is made, to ensure that the measurement equipment is correctly in place.

Several sensors were installed on the wheel loader; their position is indicated with a spot in Figure 2.2:

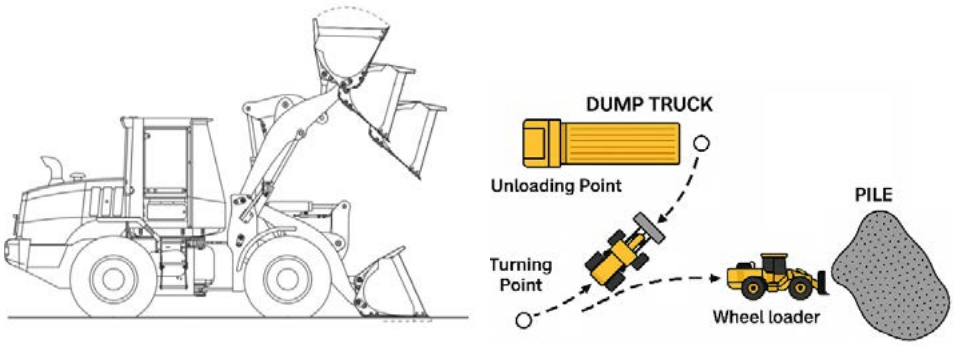
- Pressure sensors (red dots) to measure: boom and bucket pressure in both side of the cylinders; pilot pressure on directional valves, load sensing pressure and pump delivery pressure.
- Electric signals provided by joystick.
- Linear transducers (green dots) to measure the cylinders displacement of boom and bucket. Cable transducers were chosen since the displacement of the hydraulic cylinders is on the order of hundreds of millimeters, see Figure 2.3.

All vehicle CAN signals have been acquired. The sampling rate is 1000 Hz to capture even the fastest system dynamics.



Figure 2.3. Detail of bucket cylinder with displacement sensor installed.

The tests performed are aligned with the standard test procedure of Roading/Transport and Y Cycle. The Wheel Loader implement movements are schematized in Figure 2.4, see the Service Manual of CASE 821G and [22].



(a) (b)

Figure 2.4. (a) Wheel Loader implement movements. (b) Scheme of an Y loading cycle.

2.4. Physical Test

The next step is the physical test on the vehicle equipped with all the needed sensors. The test is led following a Test Procedure, which defines how to perform typical maneuvers such as Roading, Transport and Y Cycle, [22].

3. SIMULATION MODEL

3.1. Model Description

A model of the hydraulic circuit of vehicle has been developed in Matlab/Simulink environment, focusing on the hydraulic high-pressure circuit of front implement (Boom and Bucket). The model consists of the hydraulic power supply unit (Hydraulic Pumps), the pilot pressure valves, the directional valves and the implements actuators (Boom and Bucket cylinders), see Figure 2.2.

All components have been carefully parametrized by combining information obtained from component catalogs, suppliers, or by drawings from internal company documentation. Then the model has also been tuned to match the experimental data.

In more details, the pump model includes volumetric and hydromechanical efficiencies and the absolute and differential compensators are modeled to have a correct prediction of the swashplate angle and its related dynamics. The pump speed is an input to the model and it is calculated from the engine speed (available from CAN) as:

$$\omega_{Pump} = \omega_{Engine} \cdot R \tag{3.1}$$

where R is constant ratio of the pump drive.

The flow rate provided by the pump is calculated by the model as follows:

$$Q_{Pump} = \alpha \cdot V_d \cdot \omega_{Pump} \cdot \eta_v(\omega_{Pump}, \Delta p_{Pump}) \tag{3.2}$$

Where α is the fractional displacement of the pump, V_d the nominal displacement and η_v the volumetric efficiency expressed as a function of speed and pump differential pressure Δp_{Pump}.

The electronic valves that generate pilot pressure signals are pressure reducer valves: they reduce the upstream pressure of a low pressure line in order to maintain the desired pressure value to correctly pilot the directional valves. The model of such electronic valves contains the real software characteristics used in production to achieve the actual relation between joystick actuation and pilot pressure. During the simulation, the actual current signals – wired from the joystick - are used as input of this model subsystem.

The directional valves (including the post compensator) are modelled using the orifice equation:

$$Q_{Valve} = C_d \cdot A \cdot \sqrt{\frac{2 \cdot |\Delta p_{Valve}|}{\rho}} \cdot sign(\Delta p_{Valve}) \tag{3.3}$$

Where C_d is the discharge coefficient (assumed 0.7), A is the metering area, Δp_{Valve} the differential pressure and ρ the oil density. The metering areas of both the main spool and the compensator are provided by supplier to accurately calculate the flow rate across the valves. The differential pressure Δp is given by the difference between the pump pressure and the pressure signals of sensors on boom and bucket actuators.

The displacement of the boom/bucket cylinders is computed from the valve flow by:

$$x = \frac{1}{A_{Cylinder}} \int Q_{Valve}(t) dt \tag{3.4}$$

where $A_{Cylinder}$ is the area of the cylinder.

A picture of a portion of the Matlab/Simulink model is shown in Figure 3.1 with a focus on variable displacement twin pumps, directional valve, compensator and bucket cylinder.

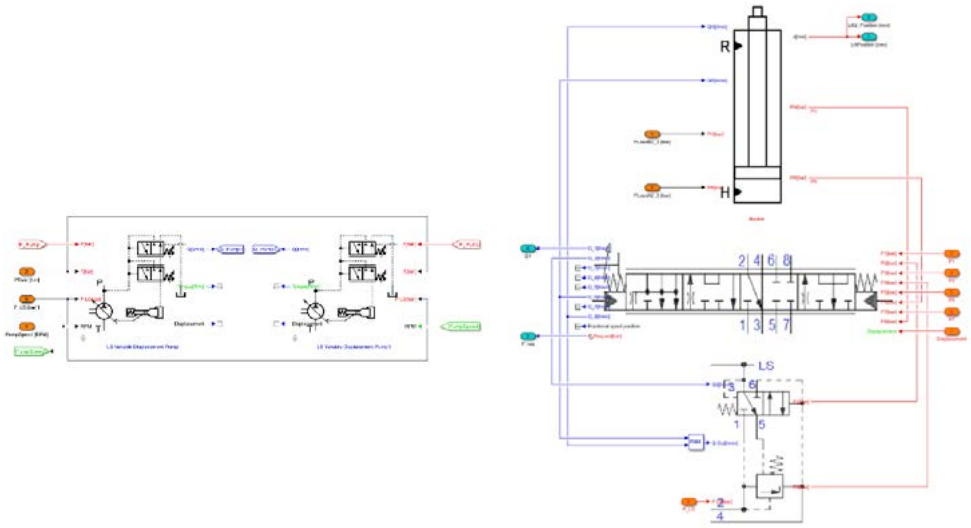


Figure 3.1. Sketch of Simulink model with twin LS pumps, directional valve, compensator and bucket cylinder.

3.2. Model Validation

The model was validated through experimental data, replicating a Y cycle. The graphs in Figure 3.2, show the comparison between experimental data and simulation results of the boom and bucket displacements over time. The y-axes are normalized using a reference value for confidentiality reason.

The correlation between the experimental and simulated curves is high enough for the fault detection purpose. Considering only the bucket and its rising phase, which is the movement studied in this work (falling phase is neglected for fault detection), the maximum deviation between the two curves does not exceed 5%. This deviation is small enough for the purposes of fault detection activity, confirming the robustness of the model and its ability to reproduce real-world behavior.

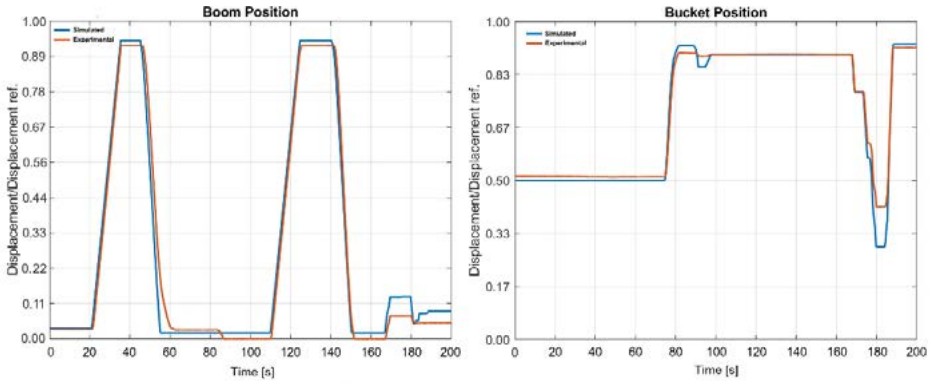


Figure 3.2. Comparison between simulation results and experimental data of Boom and Bucket position.

3.3. Hard Real Time Model on ECU

The validated Simulink model was first discretized then converted and deployed on the vehicle ECU by integrating it within the ECU software. The model in the ECU is able to run in real time and some checks have been done to verify that the results are the same as the Simulink model. The Figure 3.3 shows the bucket position over time, simulation in Simulink in dashed line and real time model with solid line.

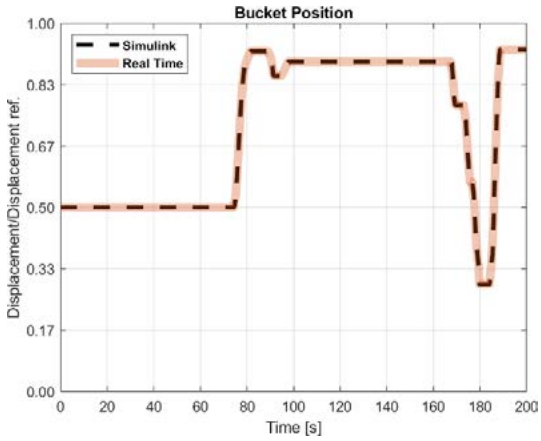


Figure 3.3. Comparison between Simulink simulation and real time model of bucket position.

4. FAILURE DETECTION

Most of the claims on Wheel Loader about the implement regard a slowdown of the cylinders and one of the root causes can be the external drain on Boom or Bucket.

4.1. Physical Test with induced Failure

The study focused on leakage failures because it represents the wear of various components inside the machine, such as directional valves, fittings and hose connections, as well as internal leakage of hydraulic cylinders due to seal wear. During the life of an off-highway vehicle, harsh working conditions and severe environments can lead to damage to some of these components.

In order to test if the mathematical model is able to recognize a possible failure an additional experimental activity was performed, inducing a slowdown only on the bucket. To achieve this, a 3-way valve was inserted in the circuit between the directional valve and the cylinder, Figure 2.2b, to drain oil to the tank so inducing a leakage in the circuit: in this way, a lower amount of oil is sent to the actuator and so the bucket movement is slower. Figure 4.1 shows a 3D image of the 3-way valve highlighted by a circle placed on the bucket hose.

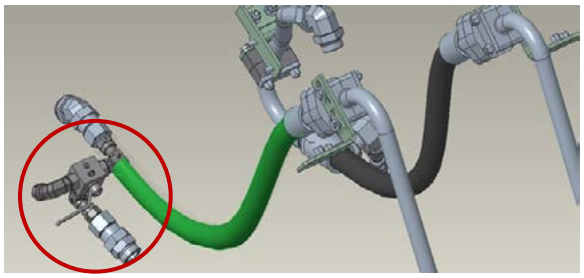


Figure 4.1. CAD representation of 3-way valve positioned on the bucket hose.

The effect of the induced external drain leads to a slower movement of the bucket. To test the model accuracy, several valve openings were set to achieve different bucket speed variations, representative of the cases that can occur during the useful life of the machine. Therefore, the valve opening was set by monitoring the average bucket cycle time on a Y-cycle. In this section, the results are shown for two limit cases according to internal criteria. Taking as reference the simulation output of the bucket position in healthy condition, when a high drain is set then the cycle time increases up to 80% of the lifting time; when a small drain is set then the cycle slowdown is lower and the cycle time increase is about 10% of the lifting time, see Figure 4.2 where axes are normalized using a reference value for confidentiality reason.

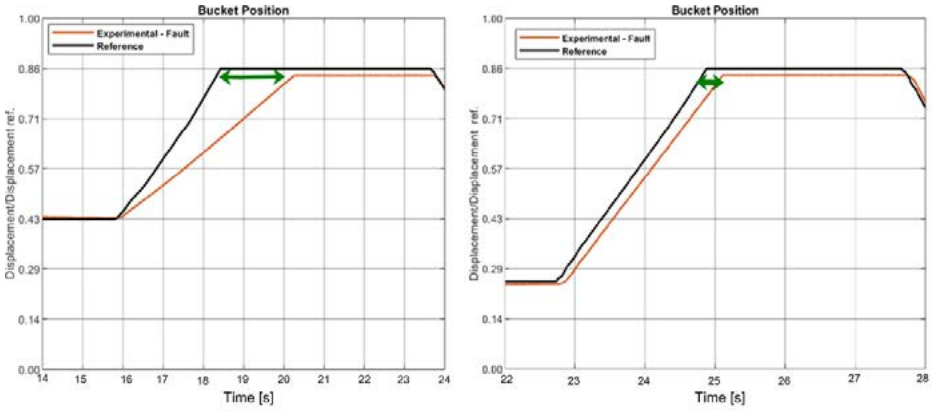


Figure 4.2. Bucket position in case of high and low drain with respect to healthy condition.

The fault detection algorithm checks the kinematic of the bucket between experimental measurement and the simulation results of the Hard Real Time Model. A flowchart of the algorithm is shown in Figure 4.3. When a deviation in the speed of the bucket movement is detected, the algorithm is able to assess whether this is happening during the bucket actuation by monitoring the rising pilot current. The procedure also checks the time interval of the deviation with respect to the manoeuvre time (obtained by the pilot current): this prevents that numerical speed deviations trigger a fault detection. Monitoring the speed instead of the position of the actuator, leads to a higher accuracy since any position offset due to numerical deviations of the model are neglected therefore avoiding false positives in fault detection. Threshold values were estimated accordingly internal criteria and are confidential.

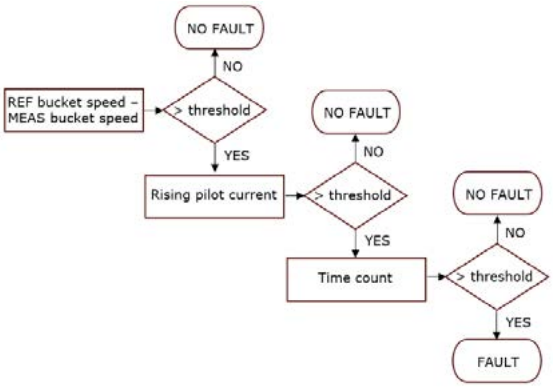
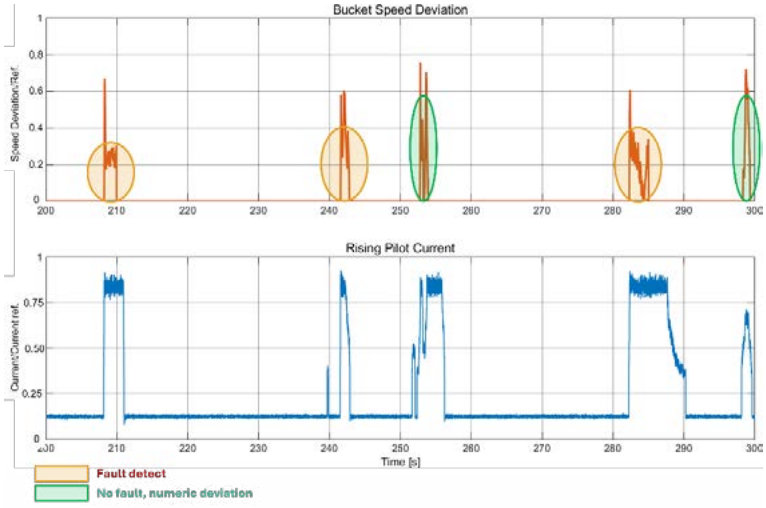


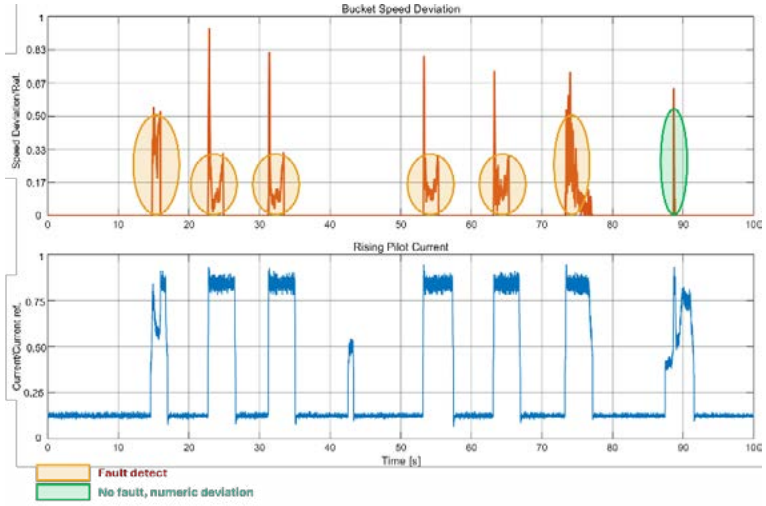
Figure 4.3. Flowchart of fault detection algorithm.

The tests are presented in two conditions: high and low drain. Figure 4.4 (a) shows the plot of the bucket speed deviation together with the rising pilot current (linked to the manoeuvre time) in the high drain condition while (b) is related to the low drain condition. The external drain valve has a constant opening during the entire manoeuvre. The described algorithm detected the faults in correspondence of the parts of the signal highlighted within the yellow

circles, since both the speed deviation and the rising pilot current time thresholds are reached. The speed deviations highlighted with green circles are due to numerical deviations and are not detected by the algorithm because the rising current time threshold is not reached.



(a) High drain condition



(b) Low drain condition

Figure 4.4. Bucket speed deviation together with the rising pilot current in the high drain condition (a) and in low drain condition (b).

The study was limited to the bucket cylinder: once verified the effectiveness of the proposed approach in detecting faults in a hydraulic circuit, the methodology can be extended to the boom cylinder as well. Considering the study of a single actuator allowed to reduce the experimental tests, which are the most expensive part from an industrial perspective, without losing the generality of the methodology. Since the methodology proved promising, the

aspiration is to extend the analysis to a greater number of actuations and a wider range of fault types.

5. CONCLUSION

This paper presented the exploitation of a real time model deployed on real ECU of a wheel loader for fault detection purposes. The development and implementation of a real-time simulation model for the hydraulic implements of a wheel loader is the basis to replicate in real time the behaviour of the healthy system.

The proposed model is deployed on a real-time embedded computing platform, acquiring live data from onboard sensors to estimate boom and bucket positions. These estimates are continuously compared against measured actuator positions. Deviations exceeding predefined thresholds serve as indicators of potential faults, including hydraulic leakage, actuator degradation, sensor drift, or control system anomalies. This model-based fault detection framework enables early anomaly identification, enhancing operational safety, reducing unscheduled downtime, and improving system reliability.

References

[1] N. Hivarekar, S. Jadav, V. Kuppusamy, P. Singh and C. Gupta, "Preventive and Predictive Maintenance Modeling", *2020 Annual Reliability and Maintainability Symposium (RAMS)*, Palm Springs, CA, USA, 2020, pp. 1-6, doi: 10.1109/RAMS48030.2020.9153636.

[2] R. Isermann, "Fault-Diagnosis Systems - An Introduction from Fault Detection to Fault Tolerance" Springer Berlin, Heidelberg 2006, doi: 10.1007/3-540-30368-5

[3] E. Fischer and J. Weber, "Data analysis for the evaluation and design of a model-based fault detection in independent metering mobile hydraulics," *Proc. 11th Int. Fluid Power Conf.*, Aachen, Germany, 2018.

[4] P. Garimella and B. Yao, "Model-based fault detection of an electro-hydraulic cylinder," in *Proc. American Control Conference*, Portland, OR, USA, 2005, pp. 3141–3146.

[5] H. Tao, P. Jia, X. Wang, and L. Wang, "Real-Time Fault Diagnosis for Hydraulic System Based on Multi-Sensor Convolutional Neural Network," *Sensors*, vol. 24, no. 2, p. 353, Jan. 2024.

[6] P. Rune, "Machine learning methods for vehicle predictive maintenance using off-board and on-board data", 2014, Halmstad University, School of Information Technology, Halmstad Embedded and Intelligent Systems Research (EIS), CAISR - Center for Applied Intelligent Systems Research. Volvo Group Trucks Technology, Malmö, Sweden.

[7] Z. Zhang, X. Shao and D. Yu, "Fault Diagnosis of a wheel Loader by Artificial Neural Networks and Fuzzy Logic", 2006 *IEEE Conference on Robotics, Automation and Mechatronics*, Bangkok, Thailand, 2006, pp. 1-5, doi: 10.1109/RAMECH.2006.252704.

[8] J. Dai, J. Tang, S. Huang, and Y. Wang, "Signal-Based Intelligent Hydraulic Fault Diagnosis Methods: Review and Prospects," *Chinese Journal of Mechanical Engineering*, vol. 32, no. 1, pp. 1–15, 2019.

[9] Y. Gao, Q. Zhang, X. Kong, "Wavelet-based pressure analysis for hydraulic pump health diagnosis". *Trans. ASAE 2003*, doi: 10.13031/2013.13952.

[10] L. Wang, Y. Liu, H. Yin, and W. Sun, “Fault diagnosis and predictive maintenance for hydraulic systems based on digital twin,” *AIP Advances*, vol. 12, no. 6, pp. 065213-1–065213-10, Jun. 2022.

[11] L. Rosa, C. Branco, “Digital Twin-Enabled Fault Detection for Suspension Systems in Autonomous Mining Haulage Vehicles” *SAE Technical Paper* 2024, doi: 10.4271/2024-36-0170.

[12] J. Li, J. Wang, “Advanced digital twin-enabled fault diagnosis framework for unmanned vehicle systems,” 2024 *Meas. Sci. Technol.*, doi: 10.1088/1361-6501/ad3a8e.

[13] E. Fischer and J. Weber, "Data analysis for the evaluation and design of a model-based fault detection in independent metering mobile hydraulics," *Proc. 11th Int. Fluid Power Conf.*, Aachen, Germany, 2018.

[14] A. Saeedzadeh, S. Habibi, M. Alavi, and P. Setoodeh, “A robust model-based strategy for real-time fault detection and diagnosis in an electro-hydraulic actuator using updated interactive multiple model smooth variable structure filter,” *ASME J. Dyn. Syst., Meas., Control.*, vol. 145, no. 10, pp. 101006-1–101006-12,Oct. 2023.

[15] M. Muenchhof, “Multi-model based fault detection and diagnosis of a hydraulic servo axis,” *IFAC 17th World Congress*, Seoul, Korea, July 2008.

[16] Z. Wang and Y. Shen, Model-Based Fault Diagnosis: Methods for State-Space Systems, Springer, 2023.

[17] S. Nagy and I. Lakatos, “Maintenance Strategies in the Automotive Industry,” *Periodica Polytechnica Transportation Engineering*, vol. 46, no. 2, pp. 73–78, 2018, doi: 10.3311/PPtr.11341.

[18] IBM Corporation, “Preventive Maintenance,” IBM Knowledge Center, 2020. Available: https://www.ibm.com/docs/en/maximo-manage/8.0.0?topic=overview-preventive-maintenance

[19] H. Raduenz, Y. Ernesto A. Mendoza, D. Ferronatto, F. J. Souza, P. Bastos, J. M. C. Soares, V. Juliano De Negri, “Online fault detection system for proportional hydraulic valves”, *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 2018, doi: 10.1007/s40430-018-1240-3.

[20] N. Kumar, R. Kumar, B. K. Sarkar, S. Maity, “Condition monitoring of hydraulic transmission system with variable displacement axial piston pump and fixed displacement motor”, *Materials today proceedings*, 2021, doi: 10.1016/j.matpr.2020.09.327.

[21] M. Choi, C. Lee, S. Park, M. Hwang, H. Oh, “Frequency-enhanced network with self-supervised learning for anomaly detection of hydraulic piston pumps”, *Expert Systems with Applications*, 2025, doi: 10.1016/j.eswa.2025.127662.

[22] R. Filla, “Optimizing the trajectory of a wheel loader working in short loading cycles”, *13th Scandinavian International Conference on Fluid Power*, 2013, Linköping; Sweden, doi: 10.3384/ecp1392a30.

Biographies



Fabio Scolari received the bachelor's and master's degree in mechanical engineering in 2016 and 2018 respectively, and the PhD in Industrial Engineering in 2022 from University of Parma, Italy, with a dissertation on tribological improvement of hydraulic pumps with texture surfaces. He is currently working as system modelling engineer at CNH. His research areas include simulation, hydraulic system modelling, fuel economy analysis, high performance systems in off-highway vehicle applications.



Federica Grossi received the bachelor's and master's degree in electronic engineering in 2003 and 2006 respectively, and the PhD in Electronics and Telecommunications in 2010 from University of Modena and Reggio Emilia, Italy, with a dissertation on dynamic modelling and control of hybrid vehicles. She is currently working as senior system modelling engineer at CNH. Her research areas include modelling, simulation, software development, fuel economy analysis, high performance systems in automotive and off-highway vehicle applications.



Francesco Pintore received the bachelor's and master's degree in mechanical engineering and the Ph.D. in High Mechanics and Automotive Design & Technology from University of Modena & Reggio Emilia, Italy. He is Manager of High-Performance System Simulation at CNH. He is responsible for High Fidelity 0D/1D model development, looking on entire vehicle performance evaluation, interfacing and providing modelling support to multiple Teams across CNH. He is also responsible for interfacing Modelling Team with electrification, Hydraulics, Driveline and Powertrain Teams across the Company's entire agriculture and construction product portfolio.



Ciro Mariniello received the master's degree in Automation Engineering from University of Sannio in 2011. He started to work in 2011 as Control system engineer with Segula as a resident engineer in Magneti Marelli, in 2012 he started to work in CNH as system modelling engineer, in 2013 he became System modelling European manager. Currently working as System modelling & Dynamic simulator Manager within DA&S Department in CNH.

Enhancing Energy Efficiency in Drilling: Innovative Pipe Handling with Hydraulic Energy Recovery

Sai Kishore Ayalasomayajula[1,*], Jürgen Binder[2]

[1]Herrenknecht AG, Schwanau, Germany,
[2]Herrenknecht Vertical GmbH, Schwanau, Germany,
E-mail: ayalasos@herrenknecht.com
**Corresponding author*

Abstract.

Geothermal drilling operations involve frequent lifting and lowering of heavy tubulars, making pipe handling one of the most energy-intensive rig subsystems. In conventional designs, the gravitational potential energy released during lowering is dissipated as heat through throttling valves. This paper presents a novel hydraulic energy-recovery concept that integrates a load-compensation cylinder and a gas-charged accumulator to capture and reuse this otherwise wasted energy. The system architecture was modelled in Siemens Amesim, and parametric studies were conducted to optimize accumulator pre-charge and volume. Simulations predict energy-recovery efficiencies up to approximately 70% per lifting cycle, with corresponding reductions in hydraulic-power demand. A prototype was implemented on a Herrenknecht vertical rig and verified during factory acceptance tests, which demonstrated substantial reductions in operating pressure from about 176 bar to 99 bar when compensation was active confirming the predicted load-sharing effect. The concept provides a compact, retrofit-capable solution for improving the energy efficiency of hydraulic pipe-handling systems and offers a practical pathway toward reduced rig-level power consumption and enhanced sustainability in geothermal drilling operations.

Keywords. Hydraulic energy recovery, pipe handling, accumulator, load compensation, tripping, drilling energy efficiency.

NOMENCLATURE

E_{dead}	Mechanical energy required to move the handler's dead load [kJ]
E_{comp}	Mechanical energy contributed by the compensation system [kJ]
E_{uncomp}	Uncompensated residual energy supplied by the HPU [kJ]
$E_{acc,usable}$	Usable accumulator energy (isothermal approximation) [kJ]
$E_{rec,available}$	Available recovered energy after hydraulic losses [kJ]
η_{Comp}	Compensation efficiency = $E_{Compensated}$ / $E_{Required}$

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

η_{hyd}	Hydraulic round-trip efficiency
p_0	Accumulator pre-charge pressure [bar]
p_1	Accumulator minimum working pressure [bar]
p_{max}, p_{min}	Maximum and minimum accumulator pressures [bar]
V_0	Accumulator effective gas volume [L]
F(h)	Cylinder force as a function of displacement [N]
h	Cylinder stroke displacement [m]
P_{hyd}	Hydraulic power [kW]
Q	Volumetric flow rate [L/min]
η	Overall efficiency factor
N	Number of pipe joints handled per bit change

1. INTRODUCTION

Geothermal and other deep-well applications are gaining renewed importance as global energy policy and security concerns accelerate investment in dispatchable, low-carbon energy sources. Among all drilling activities, the tripping process 'repetitive lifting and lowering of heavy tubulars' remains one of the most energy-intensive operations on a drilling rig. In conventional hydraulic systems, the gravitational potential energy released during lowering is dissipated as heat across throttling valves, resulting in substantial energy loss and increased cooling demand.

To address this inefficiency, the present work investigates a hydraulic energy-recovery subsystem integrated into a horizontal pipe handler. The subsystem captures the energy released during lowering operations and reuses it to assist subsequent lifting cycles. A nitrogen-charged hydraulic accumulator, connected via a dedicated compensation cylinder and control block, temporarily stores hydraulic energy that would otherwise be wasted. This stored energy is then discharged to support the boom cylinder during lifting, thereby reducing the load on the hydraulic power unit (HPU).

The overarching goals of this research are threefold:

1. Quantify the energy requirements of typical pipe-handling cycles and determine the recoverable portion through simulation and testing.
2. Evaluate the effectiveness of the accumulator-based compensation system in reducing mechanical and hydraulic power demand.
3. Validate the simulation results through factory acceptance testing (FAT) and establish a framework for future on-rig performance measurements.

The proposed concept supports broader industry efforts to improve rig-level energy efficiency, reduce emissions, and enable hybridization of hydraulic subsystems with electrical energy-storage technologies. By focusing on a modular retrofit-capable component, the pipe handler this approach provides a practical step toward measurable sustainability gains in geothermal drilling operations.

2. LITERATURE REVIEW

Energy recovery in hydraulic systems is well established in mobile and construction machinery, particularly for applications involving cyclic lifting and lowering motions. In such systems, gravitational potential energy released during lowering is often dissipated as throttling losses. Accumulator-based energy regeneration concepts have been shown to capture this energy and reuse it during subsequent lifting operations, thereby reducing peak actuator loads and instantaneous hydraulic power demand [1]. While high levels of energy recovery can be achieved locally at actuator level through accumulator-assisted regeneration, net system-level energy savings are strongly influenced by hydraulic losses, control strategy, and the proportion of non-compensated operating phases. Experimental and simulation studies show that integration effects limit the direct translation of local recovery ratios into overall reductions in hydraulic power-unit energy demand [2]. In industrial practice, accumulator-based systems are commonly applied for peak shaving and short-term energy buffering, particularly in excavators and cranes, where appropriate control strategies have demonstrated measurable reductions in fuel and electrical energy consumption [3].

For drilling applications, several studies have adapted energy recovery concepts to hoisting and tripping operations. Quan et al. [4] and Zhang [5] demonstrated potential-energy recovery in drilling rigs to reduce motor power demand during lifting; however, these approaches primarily addressed large-scale vertical hoisting systems and did not consider compact, high-cycle pipe-handler mechanisms typical of modern geothermal drilling rigs. More recent research has expanded toward system-level energy management and hybrid drilling-rig architectures. Chupin [6] reported diesel-consumption reductions of 20–30% through integrated energy storage and optimized generator scheduling, highlighting the importance of system-level coordination rather than isolated subsystem optimization. Beyond individual subsystems, improving energy efficiency has become a strategic objective in geothermal drilling operations, driven by operational cost and sustainability considerations [7], [8]. Operational studies further indicate that tripping efficiency and fuel consumption significantly affect overall rig performance and cost structure [9], [10].

3. SYSTEM DESIGN

This section describes the design and functional architecture of the pipe handler equipped with an integrated hydraulic energy-recovery subsystem. The design objective is to minimize the net energy demand on the HPU by recovering gravitational potential energy during lowering and reusing it to assist lifting operations.

3.1. Design Objectives

The main design goals were defined as follows:

- Capture and store gravitational potential energy during the lowering phase of trip-out operations.
- Reuse stored energy to assist lifting during trip in operations, reducing instantaneous HPU demand.

- Maintain full operational safety, controllability, and compatibility with existing pipe-handler geometries.
- Enable compact integration or retrofit within the envelope of standard geothermal pipe-handling systems.

3.2. Functional Description

The horizontal pipe handler is an automated handling system used on geothermal drilling rigs to lift, rotate, and position tubulars such as drill pipes, casing, and tubing between the catwalk and the well centre. A typical handling cycle involves lifting a horizontally positioned drill pipe (up to 13.72 m length and 4500 kg mass), rotating it to a vertical orientation, and handing it over to the top drive within a single operating sequence.

For such operations, Figure 3.1. shows a representative force diagram of compensation relative to the force required to support the pipe handler dead load, with and without an additional pipe weight. The area under the curve corresponds to the potential energy that can be recovered during lowering.

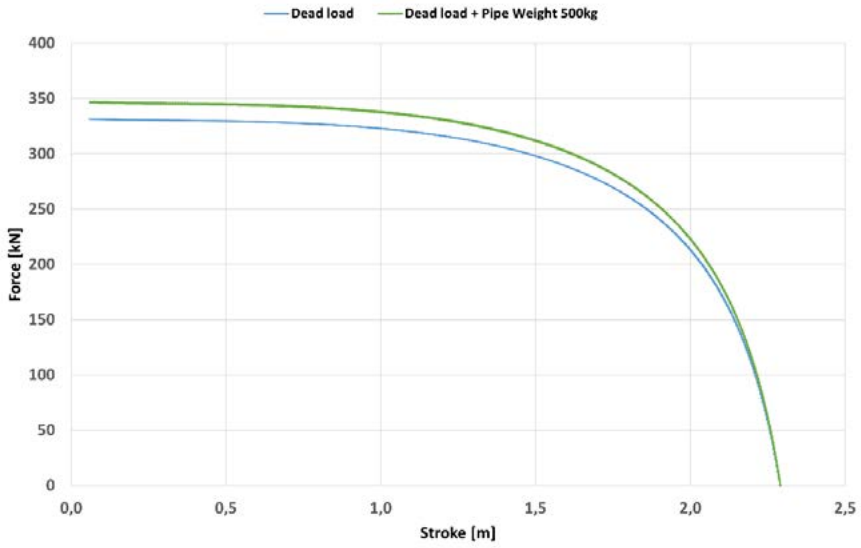


Figure 3.1.: Required Force Diagram of Compensation Cylinder

3.3. Functional Architecture

The pipe-handling system comprises a mechanically coupled boom assembly with hydraulically independent lifting and compensation circuits, as illustrated in Figures 3.2 and 3.3. The primary lifting and lowering motion are provided by the boom cylinder (1), which is driven directly by the hydraulic power unit (HPU, 6) and equipped with load-holding valves to ensure secure positioning in the event of pressure loss. In parallel, two load-compensation cylinders (2) are mechanically coupled to the boom frame but hydraulically decoupled from the HPU and instead connected to a nitrogen-charged accumulator (3) via a dedicated control manifold (4).

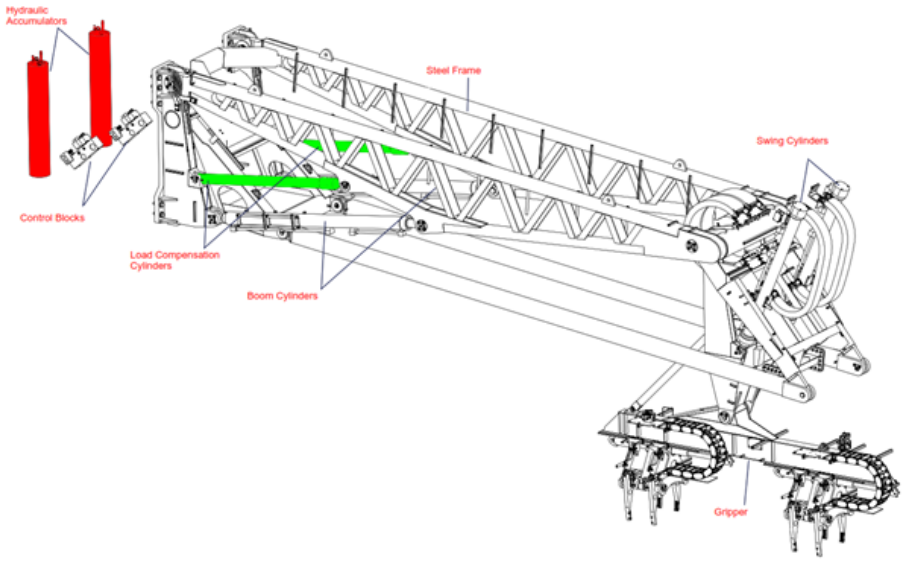


Figure 3.1.: Novel Pipe Handler with load Compensation Cylinder

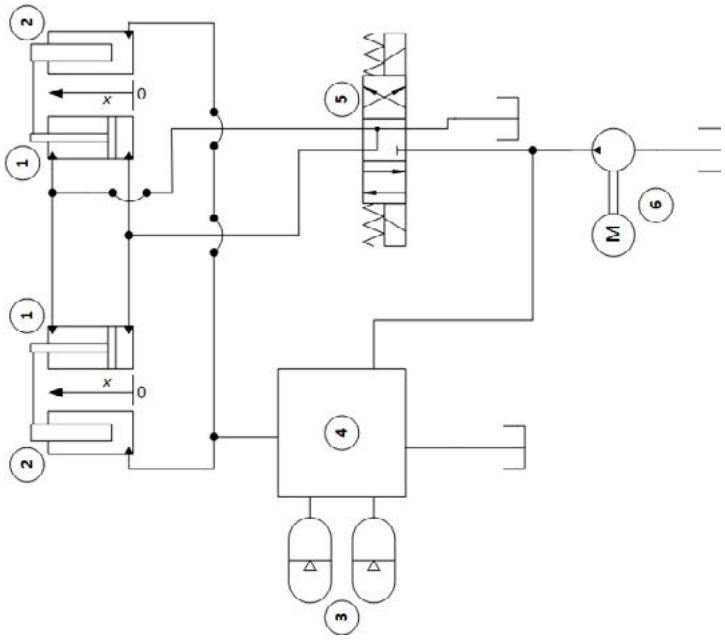


Figure 3.3.: Hydraulic Schematic Novel Pipe Handler

During lowering, hydraulic energy is transferred to the accumulator through compression of the nitrogen gas, while during subsequent lifting this stored energy is released to support the boom cylinder and reduce peak HPU power demand. The control block (5) regulates

hydraulic flow between the compensation cylinders and the accumulator and incorporates directional, check, and relief valves to manage energy transfer, provide over-pressure protection, and isolate the compensation circuit when inactive.

3.4. System Coordination and Monitoring

System operation is monitored using standard pressure and position sensors, providing feedback on cylinder and accumulator pressures as well as boom position relative to the well centre. These signals are processed by a programmable logic controller (PLC), which sequences valve operations, coordinates accumulator charging and discharging, and establishes interlocks with external handling systems. This coordinated control enables repeatable motion and controlled energy exchange during cyclic tripping operations, resulting in reduced instantaneous HPU power demand over successive cycles.

4. ENERGY RECOVERY MECHANISM

The proposed hydraulic energy-recovery mechanism operates cyclically through two complementary phases - lowering capture and lift assist - across the two primary operational modes of the pipe handler: Trip In (running pipe into the well) and Trip Out (removing pipe from the well). Each Trip consists of two sub-cycles, one loaded and one unloaded, forming a complete energy-exchange loop between the boom cylinder, the load-compensation cylinder, and the hydraulic accumulator.

4.1. Trip In: Lifting Loaded, Lowering Unloaded

During Trip In, the handler performs two sequential motions:

Cycle part 1 – Lifting loaded: The handler lifts the dead load plus a pipe segment from the catwalk to the well centre. In this motion, the accumulator discharges previously stored hydraulic energy into the compensation cylinder, producing an assisting force that reduces the effective pressure required in the boom cylinder. This lowers the hydraulic power demand from the HPU and results in smoother lifting dynamics.

Cycle part 2 – Lowering unloaded: After transferring the pipe to the well centre, the handler returns to the catwalk carrying only its dead load. During this lowering phase, the compensation cylinder retracts, forcing hydraulic fluid into the accumulator and re-charging it. Gravitational potential energy from the handler's structure is thus converted into stored hydraulic energy, preparing the system for the next lifting sequence.

4.2. Trip Out: Lowering Loaded, Lifting Unloaded

The Trip Out process follows the inverse energy flow:

Cycle part 1 – Lowering loaded: The handler lowers a pipe segment from the well centre to the catwalk, supporting the combined dead load and pipe mass. The boom cylinder retracts under load while the compensation cylinder directs fluid into the accumulator. The

gravitational energy of the pipe and structure is stored as compressed gas energy rather than dissipated as throttling losses.

Cycle part 2 – Lifting unloaded: Once the pipe is placed on the catwalk, the handler returns to the well centre with only the dead load. The accumulator assists this motion by releasing part of its stored energy into the compensation cylinder, thereby reducing the required boom-cylinder pressure and the hydraulic energy drawn from the HPU. This balanced exchange closes the energy loop for the Trip Out sequence. Summary of Trip cycles and energy flow is presented in Table 4.1.

4.3. System-Level Control and Energy Flow

Over consecutive trips, this alternating sequence of accumulator charging (during lowering) and discharging (during lifting) maintains a nearly steady-state energy balance.

The control block governs this exchange by routing hydraulic flow between the compensation cylinder and the accumulator while preventing over-pressurization. Check and relief valves ensure that accumulator charging is smooth, and that stored energy is released only during commanded lift phases. The result is a cyclic reduction of peak HPU power and smoother hydraulic operation with less throttling loss.

Operation	**Cycle**	**Motion**	**Load Condition**	**Accumulator Energy Flow**
Trip In	Cycle part 1	Lift (Catwalk → Well Centre)	Dead load + pipe	Discharge – assist lift
	Cycle part 2	Lower (Well Centre → Catwalk)	Dead load only	Charge – recover energy
Trip Out	Cycle part 1	Lower (Well Centre → Catwalk)	Dead load + pipe	Charge – recover energy
	Cycle part 2	Lift (Catwalk → Well Centre)	Dead load only	Discharge – assist lift

Table 4.1.: Summary of operational cycles and corresponding energy flow in the pipe handler

4.4. Interpretation

In this sequence, each complete Trip In or Trip Out operation alternates between an energy-storage and an energy-release phase. The oil level in accumulator is maintained within the designed range (p_0 to p_1) by the control block & PLC, ensuring consistent assistance during loaded lifts and safe capture during lowering phases. Over multiple trips, this results in lower HPU energy demand, smoother motion, and measurable reductions in overall hydraulic consumption compared with conventional systems.

5. MODELING, SIMULATION, AND VALIDATION

5.1. Simulation Model

A detailed component-level model was implemented in Siemens Amesim to analyze the dynamic behaviour of the hydraulic load-compensation subsystem. The model incorporated nonlinearities of the accumulator gas law, valve pressure drops, and cylinder dynamics. Parametric sweeps were carried out to optimize accumulator pre-charge, volume, and valve sizing.

Representative input parameters used in the simulations are:

Parameter	**Value**
Boom-cylinder effective area	0.012 [m²]
Boom-cylinder stroke	3.2 [m]
Compensation-cylinder effective area	0.012 [m²]
Compensation-cylinder stroke	2.2 [m]
Typical handled joint mass	500 [kg]
Structural dead load (self-weight)	18,000 [kg]
Accumulator nominal volume	120 [L]
Ambient temperatures tested	–5 [°C], +20 [°C] , +40 [°C]

Table 5.1.: Simulation Parameters

The performance indicator was the compensated-energy fraction, defined as the ratio between energy supplied by the compensation subsystem and the total mechanical energy required by the handler.

The mechanical work during one pipe-handling cycle is obtained by integrating the measured or simulated force–displacement curve:

$$E_{work} = \int_{h_0}^{h_1} F(h)dh \qquad (5.1)$$

For discrete data, the trapezoidal approximation was applied:

$$E_{work} \approx \Sigma_{i=1}^{n-1} \left(\frac{h_{i+1} - h_i}{2}\right)(F_i + F_{i+1}) \qquad (5.2)$$

The accumulator's usable isothermal energy between minimum and maximum pressures is expressed as:

$$E_{acc,usable} = p_0 V_0 \, ln\left(\frac{p_{max}}{p_{min}}\right) \qquad (5.3)$$

Accounting for hydraulic round-trip efficiency η_{hyd}:

$$E_{rec,available} = \eta_{hyd} \cdot E_{acc,usable} \qquad (5.4)$$

The compensation efficiency is then:

$$\eta_{Comp} = \frac{E_{Compensated}}{E_{Required}} \qquad (5.5)$$

5.2. Simulation Results

The total energy required for pipe handling was computed by numerical integration of simulated force-displacement curves (Equation 5.1). All results are expressed in SI units (N, m, J, kJ).

Dead-Load Work

From the measured boom-cylinder dead-load curve:

$$E_{dead} = \int_{h_0}^{h_1} F_{dead}(h)dh = 625.2\ kJ$$

This value represents the total mechanical work per cycle that must be supplied by the HPU when the compensation system is inactive.

Compensation Simulation and Temperature Dependence

Parametric simulations with optimized accumulator pre-charge p_0 generated compensation-cylinder force profiles for several ambient temperatures (Figure 5.1.).

The variation in compensation force with ambient temperature arises from the thermodynamic behaviour of the nitrogen-charged accumulator. For a given pre-charge pressure and volume, higher temperatures lead to higher gas pressure levels, resulting in increased compensation forces, while lower temperatures reduce the available force accordingly. The observed hysteresis between the lifting and lowering curves is primarily caused by heat transfer between the gas and the ambient air. The small discrete force steps near the end positions of the stroke originate from valve cracking pressures, numerical switching of control elements in the simulation model, and damping effects at the mechanical stroke limits. These effects are inherent to realistic hydraulic systems and are intentionally retained in the model to reflect non-ideal operating behaviour.

For the selected operating point at +20 °C, the integrated work areas under these curves (Figure 5.2., Figure 5.3.) yield the remaining uncompensated energy E_{uncomp} that the HPU must supply when compensation is active.

In these simulations, the integrated energy segments E_A+E_B+E_C correspond to the mechanical work that remains to be supplied by the HPU after the compensation system has acted.

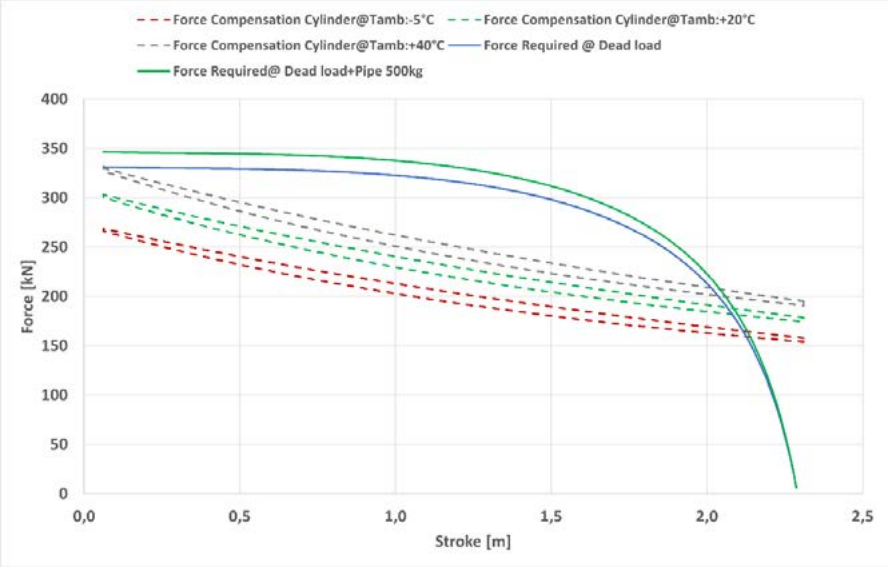


Figure 5.1.: Simulated compensation cylinder Force vs Stroke for varying ambient temperature

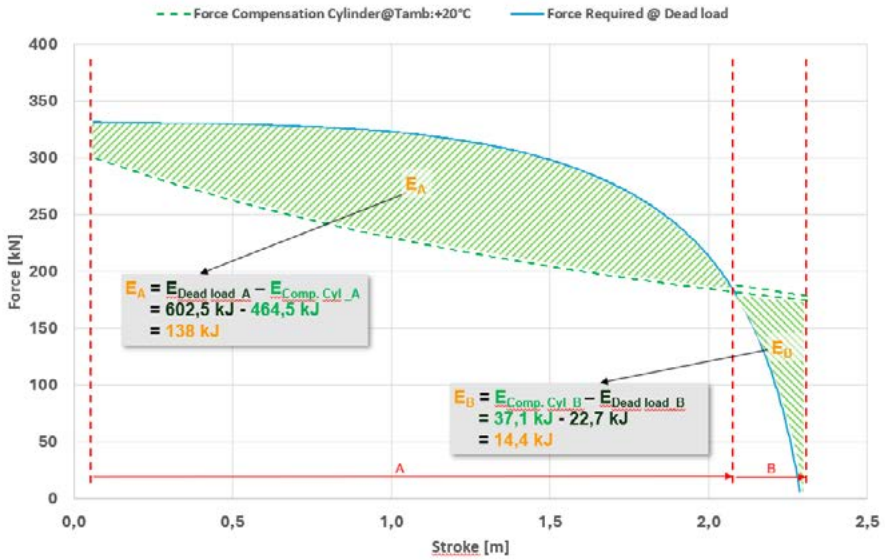


Figure 5.2.: Lift-phase forces and residual (uncompensated) energy segments

Accordingly, this represents the uncompensated energy:

$$E_{uncomp} = E_A + E_B + E_C \qquad (5.6)$$

The effective mechanical energy provided by the compensation system is therefore given by

$$E_{comp} = E_{dead} - E_{uncomp} \qquad (5.7)$$

Hence, when the compensation subsystem is active, the HPU needs only to provide E_{uncomp} while the balance E_{comp} is contributed by the accumulator-assisted compensation circuit. The corresponding fractional reduction in HPU energy demand is E_{comp}/E_{dead}.

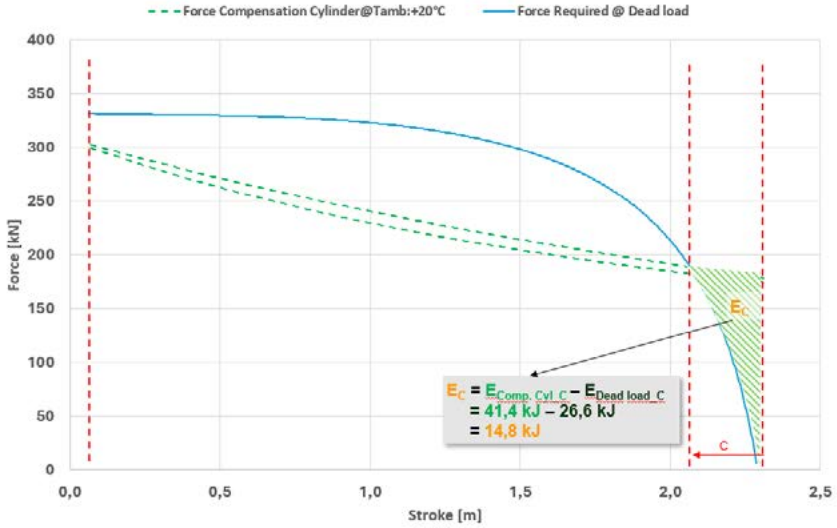


Figure 5.3.: Lowering-phase forces and residual (uncompensated) energy segments

Case 1 – Baseline Dead Load

$$E_A = 138.0\ kJ, E_B = 14.4\ kJ, E_C = 14.8\ kJ$$

Hence Uncompensated work:

$$E_{uncomp} = 167.2\ kJ$$

Compensation contribution:

$$E_{comp} = 625.2 - 167.2 = 458.0\ kJ$$

Energy reduction relative to dead-load energy:

$$\frac{E_{comp}}{E_{dead}} = 0.7326 = 73.26\%$$

Case 2 – Dead Load + 500 kg Pipe

$$E_A = 165.7\ kJ, E_B = 14.4\ kJ, E_C = 14.8\ kJ$$

Hence Uncompensated and compensated work are:

$$E_{uncomp} = 194.9\ kJ$$

$$E_{comp} = 653.9 - 194.9 = 459.0\ kJ$$

And therefore, Energy reduction is :

$$E_{comp} \,/\, E_{dead,new} = 0.7019 = 70.19\%$$

5.3. Results Summary

For the two investigated load cases, the mechanical compensation contribution was 458 kJ and 459 kJ, corresponding to actuator-level energy reductions of 73.3% (dead load) and 70.2% (dead load + 500 kg pipe). It is important to note that the compensation ratios of 70-73% reported here refer to the mechanical load compensation at the actuator level, i.e., the fraction of lifting work provided by the accumulator-assisted compensation cylinder rather than by the boom cylinder. These values therefore describe local energy sharing within a single lifting or lowering cycle and do not yet represent net system-level energy savings at the HPU level.

6. RIG / HPU SCALING FOR FULL DRILL BIT CHANGE

Building on the actuator-level results presented in the previous section, the following analysis scales the pipe-handler energy demand from individual lifting and lowering cycles to a complete drill bit-change operation. In this context, it is essential to distinguish between the local mechanical load compensation achieved within a single cycle and the resulting net hydraulic energy demand at the power-unit level, which is additionally influenced by hydraulic losses, Trip Out operations performed without pipe load, and non-compensated operating phases.

The scaling analysis estimates the HPU energy demand for a full bit-change operation - that is, completely removing a 2,500 m drill string and reinserting it with a new drill bit. Each trip of the pipe handler (PH) corresponds to the movement of one drill pipe joint. A full bit-change therefore consists of successive **Trip Out** and **Trip In** sequences.

Let:

- Dead-load mechanical energy per cycle $E_{dead} = 625.2\ kJ$
- Dead-load + pipe (500 kg) cycle $E_{dead,new} = 653.9\ kJ$
- Compensation energy (dead-load cycle) $E_{comp,1} = 167.2\ kJ$
- Compensation energy (+500 kg cycle) $E_{comp,2} = 194.9\ kJ$
- Pipe joint Length $L_j = 13.72\ m$
- Borehole Depth $d = 2500\ m$
- Number of joints $N = {}^{d}/_{L_j} = 183$

6.1. Per-Cycle and Per-Trip HPU Energy

For hydraulic round-trip efficiency η :

$$E_{C1}(\eta) = E_{dead,new} - \eta E_{comp,2} \qquad (6.1)$$

$$E_{C2}(\eta) = E_{dead,new} - \eta E_{comp,1} \qquad (6.2)$$

Trip energy (valid for Trip In or Trip Out):

$$E_{Trip}(\eta) = E_{C1}(\eta) + E_{C2}(\eta) \qquad (6.3)$$

Here η denotes hydraulic round-trip efficiency (fraction of simulated mechanical compensation that is available hydraulically).

Reference (no-compensation) Trip: $E_{no_comp,Trip} = E_{dead,new} + E_{dead} = 1279.1\ kJ$

Full Drill bit change total:

$$E_{total}(\eta) = 2 \cdot N \cdot E_{Trip}(\eta) \qquad (6.4)$$

Estimated savings:

$$Savings(\eta) = \left(1 - {E_{Trip}(\eta)}/{E_{no\ comp,Trip}}\right) \cdot 100 \qquad (6.5)$$

6.2. Numerical Results

Hydraulic Efficiency η	E_{C1} (kJ)	E_{C2} (kJ)	E_{Trip} (kJ)	E_{Trip} (kWh)	E_{Total} (kWh)	Energy per m (kJ/m)	Est. Savings (Mech)
1.0	459.0	458.0	917.0	0.255	93.23	134.3	**28.3 %**
0.9	478.5	474.7	953.2	0.265	96.91	139.6	**25.5 %**
0.8	498.0	491.4	989.4	0.275	100.59	144.9	**22.7 %**
0.7	517.5	508.2	1025.6	0.285	104.27	150.2	**19.8 %**

Table 6.1.: Simulated mechanical energy per Trip and for full bit-change (2500 m, 183 joints)

The results summarized in Table 6.1. represent the net HPU energy demand for a complete drill bit-change operation when the compensation system is active and Energy per meter refers to total HPU energy normalized by drill-string length. Although actuator-level analysis indicates mechanical load compensation ratios of up to approximately 70-73%, the corresponding system-level HPU energy savings are reduced to approximately 20-30% once hydraulic losses, Trip out operations performed without pipe load, and non-compensated operating phases are considered. These values therefore provide a realistic estimate of the achievable energy reduction under practical operating conditions.

7. PROTOTYPE VALIDATION AND PRELIMINARY RIG TESTS

The hydraulic compensation subsystem and its associated control components were installed on the vertical rig and evaluated during the factory acceptance tests (FAT). These initial tests were performed to confirm basic functionality, safety, and control logic prior to field commissioning. Although limited in duration, they provided valuable insights into system performance under real load conditions.

7.1. Factory and Preliminary Observations

During FAT, a clear difference in boom-cylinder pressure was observed between compensated and uncompensated configurations. The reported pressures represent characteristic average lifting pressures evaluated over the loaded lifting phase used here as a representative measure for estimating instantaneous hydraulic power. The actual pressure varies continuously along the cylinder stroke due to geometry, friction, and dynamic effects.

- **Without compensation (cylinder inactive):** required pressure 176 bar
- **With compensation active:** required pressure: 99 bar

The relative reduction is therefore:

$$Pressure\ reduction = 1 - \left({}^{p_{active}}/_{p_{inactive}}\right) \tag{7.1}$$

$$Pressure\ reduction = 1 - \left({}^{99}/_{176}\right) = 0.4375 = 43.75\%$$

This reduction corresponds directly to the fraction of lifting force now provided by the compensation circuit rather than by the boom cylinder and hence indicates that the HPU only needed to deliver the residual, or uncompensated, energy E_{uncomp}.

Interpreting pressure reduction as hydraulic power reduction Assuming actuator volumetric flow is approximately unchanged between compensated and uncompensated cases, hydraulic power scales with pressure according to $P_{hyd}(t) = p(t)Q(t)$.

Using the observed pressure change from 176 bar to 99 bar:

$$\frac{p_{active}}{p_{inactive}} = \frac{99}{176} = 0.5625 = 56.25\%$$

i.e. a reduction of 1-0.5625 = 0.4375 or 43.75% in instantaneous hydraulic power. Applied to the analytically estimated, uncompensated reference energy demand for a complete drill-bit change operation energy range varies $P_{inactive} \approx 93$ to104 kWh, this implies:

$$P_{active} = 0.5625 \cdot P_{inactive} = 52\ to\ 58.5\ kWh$$

This simplified estimation is directionally consistent with the mechanical load compensation observed at actuator level, while the lower magnitude reflects the net system-level reduction expected once hydraulic losses and non-compensated operating phases are included.

7.2. Comparison with Simulation and FAT Results

Simulation results predict actuator-level mechanical compensation efficiencies of approximately 70-73% under idealized, loss-free hydraulic conditions. In contrast, the 56.25% pressure and corresponding power reduction observed during factory acceptance testing (FAT) represents realistic, loss-inclusive performance, accounting for hydraulic friction, throttling, line pressure losses, and non-ideal accumulator thermodynamics.

The FAT results confirm the fundamental energy-saving principle of the hydraulic compensation system and validate the simulation predictions at actuator level under realistic

operating conditions. The observed pressure and power reductions demonstrate effective load sharing by the accumulator-assisted compensation cylinder. It is emphasized that these reductions are derived from characteristic phase-averaged lifting pressures and the resulting instantaneous hydraulic power during specific lifting phases. Net energy savings over a complete drill bit-change operation, however, result from time integration of hydraulic power across all compensated and non-compensated motions and should therefore not be inferred directly from phase-specific reductions.

8. CONCLUSION

This paper investigated the application of accumulator-assisted hydraulic energy recovery to drilling pipe-handling systems, with the objective of reducing hydraulic power-unit (HPU) energy demand during cyclic tripping operations. A combined simulation and experimental approach were employed to quantify both actuator-level mechanical load compensation and resulting system-level energy effects under realistic operating conditions. The study focused on a retrofit-oriented solution compatible with existing geothermal drilling rigs and standard hydraulic architectures

At actuator level, the accumulator-assisted compensation cylinder provided a substantial share of the lifting work during cyclic operation. Simulation results demonstrated mechanical load compensation ratios of approximately 70-73%, indicating that a significant fraction of the required lifting energy can be supplied locally by stored hydraulic energy rather than by the boom cylinder. Experimental force–stroke measurements confirmed the predicted trends and revealed characteristic nonlinearities and hysteresis effects, which were attributed to nitrogen gas compressibility, temperature-dependent pressure variation, and frictional losses within the hydraulic system.

Factory acceptance tests (FAT) further showed pronounced reductions in operating pressure and instantaneous hydraulic power during compensated lifting phases. A pressure decrease of approximately 56% relative to uncompensated reference operation was observed, corresponding to a 43.75% reduction in instantaneous hydraulic power during these phases. It is emphasized that these reductions represent phase-specific, instantaneous effects and should not be directly interpreted as total energy savings without considering the full operating cycle.

To assess rig-level impact, the actuator-level results were scaled to a complete drill bit-change operation, accounting for both compensated and non-compensated motions. The resulting analysis showed that, despite high local compensation ratios, the net HPU energy reduction for a full bit-change operation is limited to approximately 20-30%. This difference arises from hydraulic losses, unloaded return strokes, and operating phases in which compensation is inactive. These findings highlight the importance of system-level integration and time-integrated power analysis when evaluating hydraulic energy recovery concepts.

From a practical perspective, the proposed solution offers a favourable balance between achievable energy savings and implementation effort. The accumulator-based compensation concept can be integrated into existing pipe-handling systems with minimal structural modification and moderate additional component complexity. Under typical geothermal

drilling conditions, the resulting reduction in electrical energy demand suggests a payback period on the order of 1-3 years, depending on duty cycle and energy costs.

Overall, the results confirm that accumulator-assisted hydraulic energy recovery represents a technically robust and economically attractive option for improving the energy efficiency of drilling pipe-handling operations. While actuator-level compensation ratios are high, realistic system-level assessments remain essential to avoid overestimation of achievable savings. Future work will focus on extended field trials to validate long-term performance, quantify temperature and aging effects, and refine control strategies for further optimization of rig-level energy efficiency.

9. REFERENCES

[1] J. Zhu, C. Shen, Q. He, S. Li, P. Dai, and X. Li, 'Boom potential energy regeneration method for hybrid hydraulic excavators', IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3386741.

[2] T. C. Do, D. G. Nguyen, T. D. Dang, and K. K. Ahn, 'A boom energy regeneration system of hybrid hydraulic excavator using energy conversion components', Actuators, vol. 10, no. 1, 2021, doi: 10.3390/act10010001.

[3] C. DigitalHydraulics, 'Digital hydraulics for energy recovery in mobile machinery', Control and Automation Journal, vol. 15, no. 1, pp. 55–62, 2023.

[4] L. Quan, X. Zhou, and Y. Wang, 'An energy-saving oil drilling rig for recovering potential energy and decreasing motor power', Energy Conversion and Management, vol. 52, no. 1, pp. 359–365, 2011, doi: 10.1016/j.enconman.2010.06.053.

[5] L. Zhang, 'Design of energy-saving winch system for drilling rig based on energy recovery', Applied Mechanics and Materials, vol. 86, pp. 352–356, 2011, doi: 10.4028/www.scientific.net/AMM.86.352.

[6] E. Chupin, 'Energy storage systems for drilling rigs', Journal of Petroleum Exploration and Production Technology, vol. 12, no. 3, pp. 341–350, 2022, doi: 10.1007/s13202-021-01363-6.

[7] International Energy Agency (IEA), The Future of Geothermal Energy, IEA Report, Paris, France, 2023.

[8] IPIECA, Energy Efficiency Compendium: Drilling Rigs, Technical Publication, London, UK, 2023.

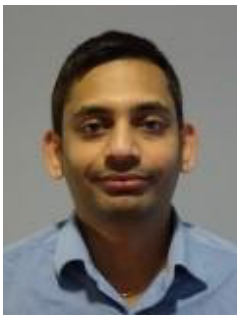
[9] Tripping Performance Analysis, Quantifying Invisible Lost Time in Tripping Operations, Technical Report, 2022.

[10] M. İ. Yildiz and A. Yilmaz, 'Fuel consumption analysis and cost overview of mechanical vehicles in petroleum drilling rigs', European Mechanical Science, vol. 8, no. 4, pp. 296–300, 2024, doi: 10.26701/ems.1249536.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the collaboration and technical support from HYDAC Technology GmbH in the development and tuning of the hydraulic accumulator system. Special thanks are extended to the Herrenknecht Vertical testing and commissioning teams for their assistance during the factory acceptance tests and prototype setup. Their contributions were essential to realizing the successful implementation and initial validation of the hydraulic compensation concept.

Biographies



Sai Kishore Ayalasomayajula received his Bachelor's degree in Mechanical Engineering from JNTU Kakinada, India in 2009 and his Master's degree in Mechanical/Mechatronics Engineering from Fach Hochschule Kaiserslautern, Germany in 2018. He has over 13 years of professional experience in hydraulic system development and engineering applications. Since 2020, he has been working at Herrenknecht AG, where he currently serves as Group Leader Hydraulics. Prior to this, he worked in various hydraulic development roles at several companies in Germany and India. His professional interests include hydraulic system design, energy recuperation, and rig automation, with a focus on sustainable and efficient drilling technologies.



Jürgen Binder began his career with a three-year apprenticeship as a mechanic before joining Herrenknecht AG in 1982. Over the course of four decades, he has worked across machining, assembly, field service, and engineering functions. Following a degree in Mechanical Engineering, he led engineering of small tunneling systems and later advanced to Project Manager for global tunneling projects. With the establishment of Herrenknecht Vertical GmbH, he became General Manager, responsible for the technical development of advanced deep-drilling systems and currently serves as Technical Director. His key interests include engineering innovation, renewable energy systems, and cost-optimized equipment design for sustainable drilling operations.

Dextreme MAX: A Digital Displacement® Excavator System for Energy Recovery and Reduced Losses

Matteo Pellegri[1], Paul Marshall[2], Casper Olesen[3], Erik Westergaard[3]

[1]*Danfoss Power Solutions GmbH & Co. OHG, matteo.pellegri@danfoss.com*
[2]*Danfoss Scotland Ltd., paul.marshall@danfoss.com*
[3]*Danfoss Power Solutions ApS, colesen@danfoss.com, ewestergaard@danfoss.com*

Abstract.

We present Dextreme MAX, a hydraulic architecture applied to a 30-tonne electric excavator based on a Digital Displacement pump/motor with multiple independently controlled outlets. The system provides independent actuator supply, eliminates flow-sharing losses, and enables energy recovery / hydraulic transformation via a boom H-bridge valve. On standardized JCMAS cycles, the converted machine reduced DC energy use by 49% in grading and 31% in dig-and-dump, with negligible impact on cycle time; across a representative duty mix this corresponds to ≈35% lower battery power and >53% longer uptime. These gains primarily arise from reduced throttling; additional benefits come from boom-lowering energy recovery. Results indicate that MAX can reduce required battery capacity and cost while maintaining productivity, accelerating electrification of heavy machinery.

Keywords. Digital Displacement, Excavator, Energy recovery, Electrification, Hydraulic transformation.

1. INTRODUCTION

As climate targets tighten, the inefficiency of hydraulic off-highway machinery has become a pressing concern. Excavators, one of the largest and most energy-intensive off-highway segments, typically achieve only 20–30% hydraulic efficiency [1]. Whilst for ICEs this drives high fuel consumption and CO_2 emissions, for electric excavators, this inefficiency is especially costly: batteries must be sized for a full shift, and as the most expensive component, this makes electrification economically challenging. Conventional systems (Negative and Positive Flow Control, Load Sensing) suffer large throttling losses and lack energy recovery [2]. Numerous advanced concepts such as displacement control [3,4], independent metering valves (IMV) [5,6], multi-chamber cylinders [7], two-pressure-rail hybrids [8], and secondary controlled actuators [1] have shown 15–50% efficiency gains in research prototypes. Yet commercial adoption remains minimal due to cost, scalability, complexity, a radical overhaul of the system, and uncertain ROI.

Digital Displacement® (DD) technology offers a more evolutionary path: DD pumps combine high efficiency across the load range with multi-outlet capability [9]. Prior work on a 16-tonne excavator showed that a single “multi-outlet” DD pump could decouple boom/bucket from arm/swing functions, reducing throttling losses while retaining the original hydraulic layout. Reported results included productivity improvements of 10–15%

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

or fuel savings of 20–30% [10,11] in a diesel machine compared to the traditional NFC system with a tandem-swashplate unit. However, flow-sharing valve blocks as well as a lack of energy recovery still limited the further efficiency improvements.

This paper introduces Dextreme MAX, a novel architecture developed during a two-year project funded by the UK Department for Energy Security and Net Zero (DESNZ).

2. DEXTREME MAX SYSTEM DEVELOPMENT

The Dextreme MAX system aims at eliminating the flow-sharing central valve block by using a DD pump/motor with four independently controlled outlets to supply each actuator of the excavator, removing flow-sharing losses and enabling independent actuator control as well as energy recovery and hydraulic transformation. This new architecture builds on top of the SWAP and FLEX systems presented in [9,10].

2.1. Technology Background: The DD1x0D Digital Displacement Pump/Motor

The DD pump/motor is a radial piston unit based on electronically controlled poppet valves, enabling each cylinder to be switched on or off in real time according to crankshaft position [12] via a dedicated real-time controller, the DPC30. Flow is therefore modulated digitally at the cylinder level, with cylinders producing a full/partial pumping or motoring stroke or idling with negligible loss. This yields high efficiency across the full displacement range, combined with fast dynamics and precise flow control. The DD1x0 family is the first generation of large-format DD machines for prime movers in the 120–250 kW class, targeting 20–40 tonne excavators [13]. A key innovation is the combination of pump/motor capability with ten independently controlled outlets, each operating at its own pressure level. This allows independent multi-actuator supply and new modes of hydraulic transformation. In fact, the DD1x0D can operate as a transformer: some outlets motor while others pump or idle, enabling direct energy transfer between circuits at different pressures (Figure 2.1).

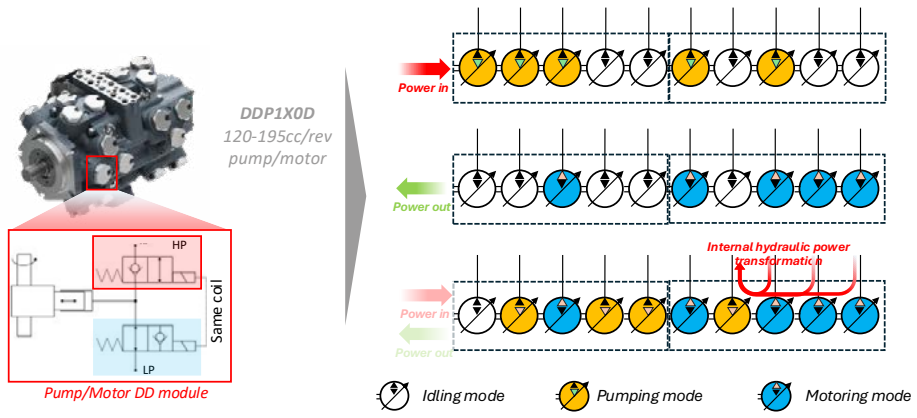


Figure 2.1. The DD1x0D Pump/Motor combines pump/motor/transformer functionality with ten independent outlets

Overrunning energy, such as from boom lowering, can be returned to the drivetrain or transformed to support other actuators. Component tests of the DD1x0 have shown pumping

and motoring efficiencies above 94.5% in the main excavator operating range (150–250 bar, 1400–1800 rpm), with peak motoring efficiency of 96.7%. These results confirm the DD1x0's potential for very high efficiency together with system-level functions like energy recovery and transformation unavailable with conventional pumps [13].

2.2. The DX300 Dextreme MAX System: concept and implementation

The MAX system was demonstrated for the first time on a 30-tonne battery-electric excavator, selected to meet the objectives of the UK Red Diesel Replacement Phase 2 (RDR2) programme [14], which aims to accelerate decarbonisation of off-highway machinery. The project sought to prove that higher hydraulic efficiency can reduce required battery capacity, cutting cost and charging demand while maintaining productivity. Although validated on an electric platform, the same principles directly apply to diesel machines, where lower hydraulic losses reduce fuel use and emissions.

A DEVELON DX300LC-7 crawler excavator was chosen as the test platform. Originally diesel-powered with a two-service NFC hydraulic system, it was obtained in an electric configuration from STAAD B.V., which replaces the engine with an electric drivetrain and battery system (Table 2.1). The model offered representative scale, short delivery time, and strong technical support, enabling integration of the MAX concept to a Technology Readiness Level 7 prototype (TRL7 [15]).

Table 2.1. Electric Drivetrain Specifications

Parameter	Value	Notes / Components
System HV voltage	670-790 VDC	Battery supply voltage
Nominal motor speed	1900 rpm	
Maximum motor speed	2200 rpm	
Continuous motor power	145 kW	Limited at machine level (motor rated 174 kW)
Motor	Danfoss PMI375-T800-1800	Permanent magnet synchronous motor
Inverter	Danfoss EC-1200-450-L240A	DC–AC phase inverter
Motor controller (VCU)	Danfoss MC050-010	Vehicle control unit
Installed Battery Capacity	3x140kWh	1 fixed + 2 swappable

The Dextreme MAX hydraulic architecture comprises four primary services – boom, arm, bucket, and swing – each supplied independently by the DD1x0D. The unit features ten controllable outlets that can be dynamically grouped through a digital distributor reallocating capacity to the actuator requesting it [9]. A schematic comparison between the baseline and MAX configurations is shown in Figure 2.2.

For this prototype, full decoupling of all four services was not implemented due to project time limits and the need to retain the original main control valve (MCV) without altering its core functions. Energy-flow analysis showed that keeping the MCV for the arm and bucket circuits introduced acceptable losses: during high arm-in demands (> 50 % operator input), the "arm 2" and bucket spools continue to share flow through the coupled MCV arrangement.

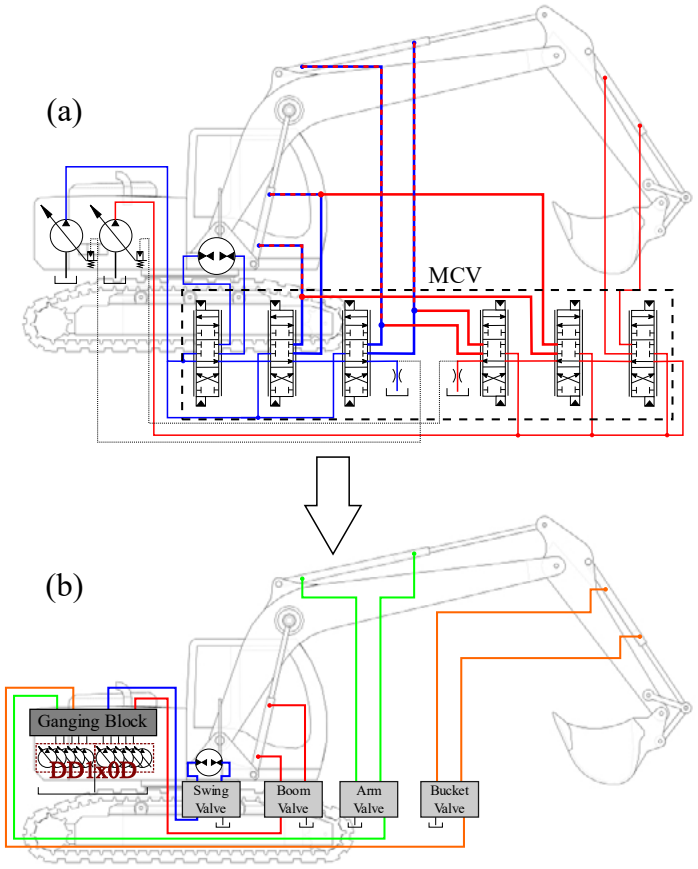


Figure 2.2. Layouts of the original NFC system (a) and the Dextreme MAX system (b).

Based on previous studies [7,8], the boom and swing were identified as the main candidates for energy recovery, as they dominate potential and kinetic energy exchange during excavator cycles. The boom typically offers several times higher recovery potential than other actuators and about two to four times that of the swing. Consequently, both functions received dedicated valve services in the MAX demonstrator. Because the DD1x0D operates in an open circuit, practical swing-energy recovery could not be implemented within the project timeframe. Instead, the swing used an independent metering valve to remove flow-sharing and metering losses, while only the boom employed a full H-bridge for regeneration and pressure amplification. Table 2.2 summarizes the configuration of the baseline NFC and MAX systems. To maximize the possibility of energy recovery and simultaneous actuators operation, a DD180D (2x180cc/rev) has been selected, replacing the original 140cc/rev swashplate tandem unit. In addition to the DD180D, two supporting subsystems were added:

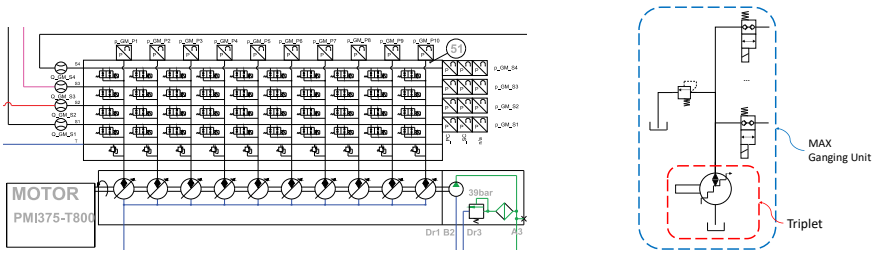
- Ganging block – a digital distributor enabling dynamic grouping of pump outlets
- Boom and swing valve – a dedicated arrangement providing advanced control of these actuators

Table 2.2. Service/actuators configurations comparing the original NFC and Dextreme MAX systems

Service ID	**Original**	**MAX configuration**
Service 1	-	Swing
Service 2 (2)	Bucket / Boom 1 / Arm 2	Bucket / Arm 2
Service 3 (1)	Swing / Boom 2 / Arm 1	Arm1
Service 4	-	Boom

2.2.1. The Ganging Block

This valve block, referred to here as the ganging block, connects each service to each DD outlet ("triplet" – three pistons each), enabling the allocation of large capacity to individual services when required. A set of digitally controlled valves is introduced to distribute the flow supplied or absorbed by the DD1x0D among the working actuators. The hydraulic schematic of the ganging block is shown in Figure 2.3.



(a) (b)

Figure 2.3. (a) Schematic of the ganging manifold; (b) MAX ganging unit consisting of a DD1x0 Triplet and four 2/2 valves, and a PRV connected in parallel.

The valves selected for the ganging block are 2-way 2-position, high pressure bi-directional, pilot operated, poppet type screw-in cartridge valve (Danfoss SBV11-12-C). Each valve is normally closed (no free-flow direction) and allows bidirectional flow when energized. In addition, the manifold integrates ten safety relief valves, one per pump outlet.

2.2.2. Boom valve - the H-bridge

Because the DD1x0D is an open-circuit pump/motor, additional valve functionality was required for full bidirectional control of the linear actuators. A dedicated H-bridge valve was therefore developed to manage independent meter-in and meter-out flows, providing anti-cavitation, pressure amplification, and energy recovery during overrunning motions. Building on previous work on independent metering and secondary control [1,16], the design aimed to minimize throttling losses while enabling controlled energy exchange. In the Dextreme MAX system, the H-bridge supports the following key functions:

1. Four-quadrant operation – provide full control of actuator motion across all load and velocity directions, ensuring symmetric functionality in extension and

retraction. For the boom, quadrant I corresponds to resistive extension (lifting) and quadrant II to overrunning lowering.

2. Anti-cavitation (quadrants II and IV) – supply make-up flow to prevent cavitation when the actuator is driven by overrunning loads
3. Flow regeneration (quadrant II) – transfer flow from the head side to the rod side during overrunning lowering.
4. Pressure amplification / differential mode (quadrant II) – apply controlled pressure on the rod-side to raise effective load pressure while reducing flow demand. For the DD1x0D, this maintains constant power transfer with less processed flow, reducing conveyance losses and freeing capacity for other actuators.
5. Normal metering-out (quadrants II and IV) – allow conventional throttling when energy recovery or transformation is not possible.

Together, these modes provide the flexibility, safety, and fallback behaviour required for practical excavator operation.

Several H-bridge topologies were evaluated. A four-cartridge concept ([16]) and Valvistor-type elements were excluded due to flow-capacity limitations and development time constraints. To ensure availability and reliability, the final design employed Danfoss PVG components, as shown schematically in Figure 2.4.

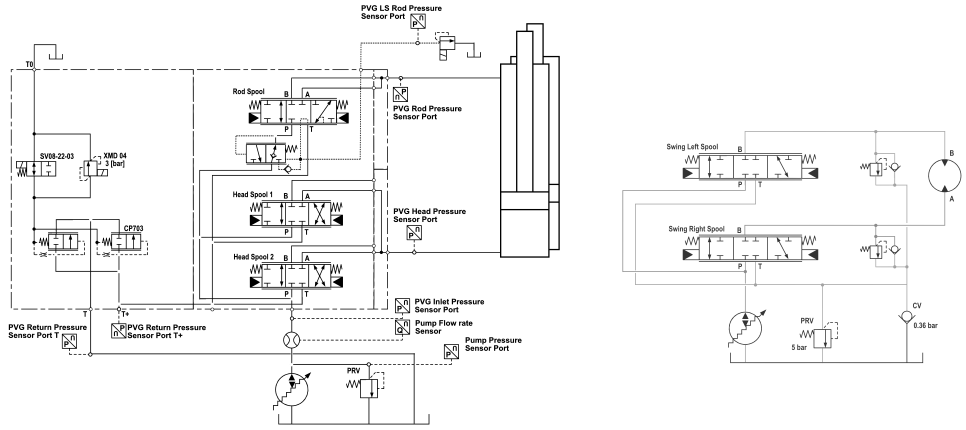


Figure 2.4. Simplified schematic of the Boom H-Bridge and Swing valve

The implemented boom H-bridge employs three PVB valves (2 × PVB128 and 1 × PVB256) to meet the requirements for four-quadrant operation, anti-cavitation, regeneration, pressure amplification, and safe metering-out while handling the excavator's peak flows. The circuit connects pump to head (P–A), pump to rod (P–B), head to tank (A–T), and rod to tank (B–T), enabling full control of both chambers. For stable rod-side pressure regulation, the PVB256 compensator was modified so that the main spool remains fully open while an actively piloted compensator in the rod branch maintains chamber pressure. A pilot relief valve (PRV) defines the pressure reference dynamically, providing faster and more stable control than direct control of the PVG spool. This approach is consistent with [17], which showed the efficiency and safety benefits of active pressure regulation. The resulting PVG-

based layout delivers the required four-quadrant performance using standard, commercially available hardware, ensuring timely and practical system integration.

2.2.3. Swing valve

The swing service valve was implemented using the same independent-metering PVG concept adopted for the boom, but with a simplified layout consisting of two PVB128 as shown in Figure 2.4. All PVB blocks for boom and swing were physically integrated into a single large PVG assembly, interconnected through a custom hydraulic integrated circuit (HIC) block. This approach allowed the services to be grouped in a compact package while providing the necessary functionality for the MAX system.

2.3. Control Logic

The introduction of four independent services in the Dextreme MAX system required a complete redesign of the control strategy compared with previous SWAP and FLEX implementations. The controller must manage flow allocation, displacement distribution, pressure regulation, torque sharing, and boom-specific recovery logic, while ensuring seamless operator control and compatibility with the machine's electric prime mover. Figure 2.5 summarizes the control architecture, highlighting the interaction between displacement control, pressure controllers, torque management, and H-bridge logic.

2.3.1. Flow Allocation and Elastic Pump Distribution

Similar to positive flow control (PFC) systems the joystick pilot pressures are mapped through look-up tables to generate service flow demands. The DD pump/motor provides multiple discrete outlets ("triplets"), which must be assigned dynamically to the active services. In general, the system tries to satisfy the capacity request of each actuator. When the demand exceeds the total pump capacity, proportional displacement allocation is enforced. The 2/2 valves in the ganging block are switched on and off accordingly. As each service is handled separately by the pump controller, the flow is adjusted by controlling the displacement fraction of the instantaneous allocated capacity of the service.

2.3.2. Pressure Limiting

The MAX controller allows for pressure control for all services. A PI controller regulates pump displacement accordingly to limit losses over relief valve, which is common control practice when accelerating the swing function. The pressure setpoint can be changed dynamically based on the operating conditions: for example, because the MAX architecture decouples the boom and swing into separate services, the flow-sharing characteristic of the baseline system during lift-and-rotate operation is lost. To re-establish the expected interaction between these functions, a flow-sharing emulating algorithm is introduced. In this case, the pressure of the swing is limited to the boom service pressure, thereby reproducing the coordination observed in conventional architectures while preserving the independence of the MAX layout.

2.3.3. Torque Control

The torque controller determines the aggregate pump torque request and interfaces with the prime mover. Only the positive torque contributions of each service are considered when

calculating total demand; negative torques from motoring services are excluded to avoid disturbing the expected flow distribution. An anti-lift-off strategy is applied to reduce motoring demand when the shaft speed exceeds a defined threshold, preventing overspeed. In addition, motoring is disabled entirely when calculated negative power falls below the minimum limit specified by the electric motor control unit. Under these circumstances the H-Bridge metering functionality maintains productivity by ensuring requested boom down speed. These safeguards ensure compatibility between hydraulic energy recovery and the electric drivetrain.

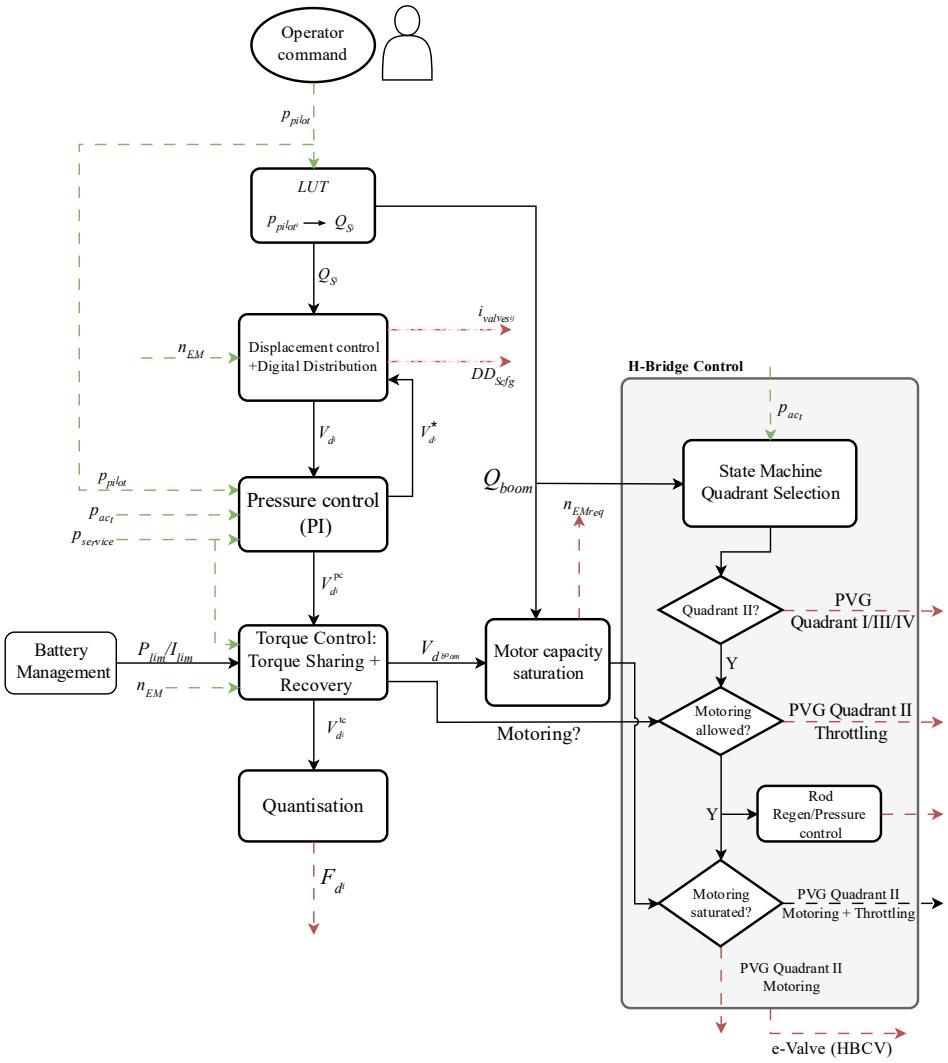


Figure 2.5. Control structure for the Dextreme MAX system

2.3.4. Boom Control Logic

Boom control is critical to the overall efficiency of the MAX system. The boom controller is structured around four quadrants of actuator operation:

- Quadrant I – Resistive extension (lifting).
- Quadrant II – Overrunning retraction (lowering).
- Quadrant III – Resistive retraction.
- Quadrant IV – Overrunning extension.

During resistive operation (Quadrants I and III), the pump supplies the head or rod chamber in a conventional manner. The head spool remains fully open, and motion is regulated entirely through displacement control of the DD180D, with smooth behaviour enabled by its fast dynamic response. Overrunning retraction (Quadrant II, i.e. boom lowering) presents the greatest control challenge. The key issue is managing cavitation risk in the rod chamber while handling surplus flow from the head side. This is addressed through differential actuation, in which part of the head flow is routed to the rod side. In this mode, the rod chamber pressure is actively regulated: either by partial opening of the rod spool or, preferably, by a piloted pressure compensator in the rod branch. The compensator, controlled via an external PRV, provides faster and more stable response. Rod pressure set-points vary with load: a higher-pressure setting is used for light bucket loads to maximize energy recovery, while lower pressure is set for heavy loads to avoid excessive head pressurization or when motoring is not possible. During motoring, because the machine has finite capacity, any excess head flow beyond the motoring limit is throttled to tank. While this introduces some loss, it ensures safe boom lowering and maintains productivity.

2.4. Control System Implementation

2.4.1. Data acquisition

To evaluate the performance of the prototype system, a comprehensive data acquisition (DAQ) alongside the control setup was installed on the excavator. The objective of this setup was to capture all relevant parameters from the working hydraulics – such as arm, boom, bucket, swing, and track systems – as well as from the DD pump outlets. The measurement system consisted of five DEWE-43 data acquisition units from Dewesoft, enabling synchronized high-speed recording across all channels covering hydraulic, mechanical, and electrical domains, allowing a detailed analysis of machine performance, energy flow, and control behaviour. The instrumentation covered the following main sensor groups:

- Hydraulic pressure sensors – installed at each DD pump outlet (service ports), actuator chambers, main control valve, and pilot lines to monitor system pressures and load conditions.
- Flow meters – flow meters on each DD outlet to measure service flow rates and calculate hydraulic power. A screw flowmeter was installed on the Boom Service for better accuracy at high flows during motoring.
- Position sensors – string potentiometers on the boom, arm, and bucket cylinders, and an IMU for the swing for displacement and motion tracking.
- Electrical power sensor – a high current and voltage transducer monitors DC supply from the battery to the inverter.

2.4.2. System controllers

The control system architecture (Figure 2.6) was designed to provide real-time control of the DD system, electric powertrain, and auxiliary subsystems. The setup combines a SpeedGoat Mobile real-time target machine (M3) with two Danfoss XL104-0000.

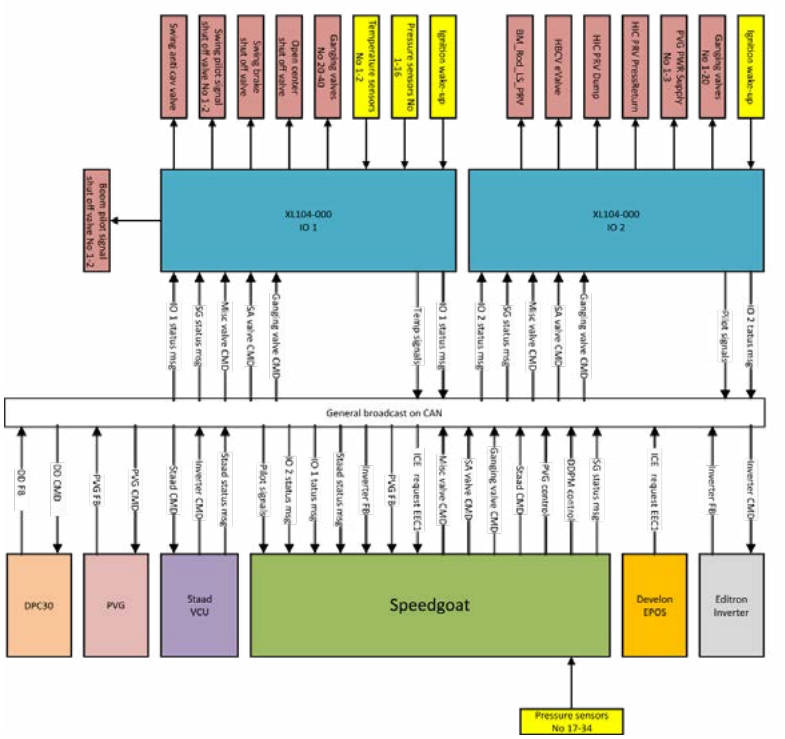


Figure 2.6. Control system architecture showing the SpeedGoat real-time target as the main controller and two Danfoss XL104-0000 units as distributed I/O.

The SpeedGoat unit is equipped with two IO133 modules for analog and digital I/O and one IO602 CAN interface module. It serves as the main real-time controller, executing the system's supervisory control, coordination of subsystems, advanced algorithms such as motoring and digital distribution, power limiting, and fault handling. The control logic was developed entirely in MATLAB / Simulink, and SpeedGoat was chosen for its capability for rapid prototyping. A corresponding Simscape model of the excavator was used to develop and verify the algorithms in real-time before deployment.

Two Danfoss XL104-0000 controllers act as distributed I/O nodes that interface directly with the hydraulic valves, pilot signals, and pressure sensors. These units provide the required output power to drive the proportional and on/off valves that exceed the current-handling capacity of the SpeedGoat I/O modules. They also serve as functional safety controllers, implementing local safety logic and redundancy for critical valve operations. All components communicate via CAN HS and J1939 protocol including the Danfoss Editron inverter, Develon EPOS, and STAAD VCU, which exchange commands and feedback signals over different CAN networks.

3. TESTING

The RDR2 program required a field trial to demonstrate TRL 7 and controlled testing to compare system performance before and after conversion. To quantify performance, a comprehensive characterisation of the original configuration was undertaken. The tests were designed to capture single-function behaviour, function-prioritisation performance, and simulated duty-cycle operation. Three types of simulated duty-cycle tests were performed:

1. JCMAS air grading (Equivalent of ISO11152-2) (Figure 3.1a)
2. JCMAS air dig and dump (Equivalent of ISO11152-2) (Figure 3.1b)
3. Real dig and dump with gravel (Figure 3.1c)

All tests were designed for repeatability and could be executed both before and after conversion. Operators were instructed to work as quickly as possible, as this typically yields the most consistent results.



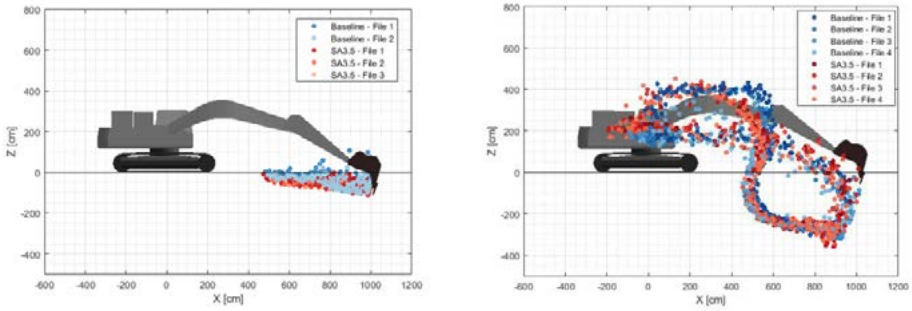
(a) (b) (c)

Figure 3.1. (a) JCMAS Air-Grading, (b) Air-Dig and Dump, (c) Real Dig and Dump

Baseline testing of the conventional system was conducted at the Danfoss Application Development Centre (ADC) in Nordborg. To meet project timelines, the post-conversion tests were repeated at a dedicated site near Edinburgh, providing a consistent basis for quantitative comparison and enabling safe operation in MAX mode. Because calibrated gravel was unavailable, real dig-and-dump tests could not be repeated; however, the JCMAS air cycles offered a controlled and repeatable means of assessing Dextreme MAX system efficiency relative to the Nordborg baseline.

Ensuring consistency across test cycles was challenging, as the baseline tests were conducted by a different operator at a separate site nearly a year earlier. The bucket-tracking tool was used to verify motion repeatability, and outlier cycles were excluded from the energy analysis using plotting tools. A sample comparison of bucket paths for the JCMAS tests is shown in Figure 3.2.

Cycle time influenced repeatability since the DDP180D pump's 30 % larger displacement enabled faster operation than the baseline unit. To align cycle durations, motor speed setpoints were adjusted for the converted machine tests, with active control of the EM speed during boom down to maximize energy recovery.



(a) (b)

	BOOM [M]	ARM [M]	BUCKET [M]	SWING [°]	
BASELINE	0.68	0.86	0.74	88	DnD
DEX. MAX	0.72	0.86	0.7	95	
ERROR %	**+5.8%**	**0%**	**-5.4%**	**+8%**	
BASELINE	0.33	1.23	-	-	Grading
DEX. MAX	0.30	1.23	-	-	
ERROR %	**-9.1%**	**0%**	-	-	

Figure 3.2. Comparison of bucket traces for JCMAS test – (a) Grading, (b) DnD; table with error percentage between average actuators extensions across tests.

3.1. Test results

Several tests were conducted for both the baseline and the converted machine using sets of 10 cycles per test for both grading and dig and dump. To ensure a fair comparison, the actuators' extensions have been evaluated, and clear outliers have been discarded. The errors in the actuators extensions (rotation) are between 9% and 0%. In case of the grading, the MAX system boom extension is 9% lower than the baseline, but the absolute value is 0.03m which is difficult to control in a fast and dynamic cycle. Average of the selected cycles results are presented in Table 3.1 and Table 3.2 for air grading and dig-and-dump operations, respectively.

Table 3.1. Air Grading results

Configuration	Cycle Time (s)	DC Power (kW)	DC Energy (kJ)
Baseline	6.78	120.5	817.0
MAX	7.29	56.9	414.8
% difference	+7.5%	-52.8%	**-49.2%**

Table 3.2. Air Dig and Dump (DnD) results

Configuration	Cycle Time (s)	DC Power (kW)	DC Energy (kJ)
Baseline	14.44	99.9	1442.6
MAX	14.59	68.2	995.0
% difference	+1%	-31.7%	**-31%**

General field operation of excavators implies a mix of grading, dig and dump and idling. Each OEM uses a different mix of those with dig and dump being the largest portion of the mix. For electric machines, we can exclude the idle portion as the electric motor can easily be stopped and started when needed. Assuming a duty cycle of 30% air grading and 70% air digging, the MAX system would reduce battery power consumption by 35% without significantly compromising work rate. This would result in 53% longer operating hours.

3.2. Data analysis

Table 3.3 summarizes the energy balance in a series of JCMAS DnD cycles between the baseline and MAX configurations, highlighting where the efficiency improvements originate. Losses associated with the inverter/motor, DD pump, baseline pump, and the AUX systems are estimated using numerical loss models; actuator mechanical losses are embedded in the reported actuator net energy. The different contributions are calculated using a mix of measured data and simulation models hence the small difference with the reported DC energy in Table 3.2.

Table 3.3. Breakdown of the system losses and contribution of energy savings

Energy [kJ]	Baseline	MAX	Δ	% of savings contribution
DC	1449.7	1021.6	**428.0**	
Pump Losses	167.3	85.2	82.1	19.2%
AUX	61.9	33.6	28.3	6.6%
EM Losses	71.5	50.6	20.9	4.9%
Actuator Net	481.8	457.8	24.0	5.6%
Meter In Loss	307.7	142.5	165.2	38.6%
Meter Out Loss	207.4	272.3	-64.9	-15.2%
Open Center Loss	93.9	38.6	55.3	12.9%
Relief Valve	58.1	40.7	17.4	4.1%
Recovered Energy	0.0	-99.7	99.7	23.3%

The total energy savings are primarily attributed to reductions in pump losses, metering losses, and the introduction of energy recovery. Pump losses decrease by 82.1 kJ (19.2 % of total savings), reflecting the improved part-load efficiency of the DD pump. A significant 220.5 kJ reduction in combined open-center/meter-in losses (51.5%) confirms the

effectiveness of independent flow-matching control in minimizing metering losses during multi-actuator operation. The recovery system contributes an additional 99.7kJ (23.3 %), corresponding to energy regenerated during boom lowering. Meter-out losses are not improved and are slightly increased. This is primarily due to the higher flow levels through the arm and bucket circuits, while the same conventional valves were retained.

Overall, approximately half of the total energy savings result directly from reduced hydraulic throttling, with the remaining half achieved through energy recovery and pump efficiency improvements. Figure 3.3 shows a detail breakdown of energy for an Air Dig and Dump cycle of the Dextreme MAX system. It is clear that a significant portion of the energy is still lost in meter out for bucket and arm. This is due to the usage of the original MCV for those actuators. These losses can be addressed with a different valve design in a system developed from scratch. Additionally, a significant part of the boom potential energy is lost through metering: this was due to an error in the control logic parameters preventing a complete opening of the boom load holding valve during boom down operation, wasting more than 30kW of potential power during peak flows. Unfortunately, the issue was discovered after the tests were completed and it will be fixed in the future.

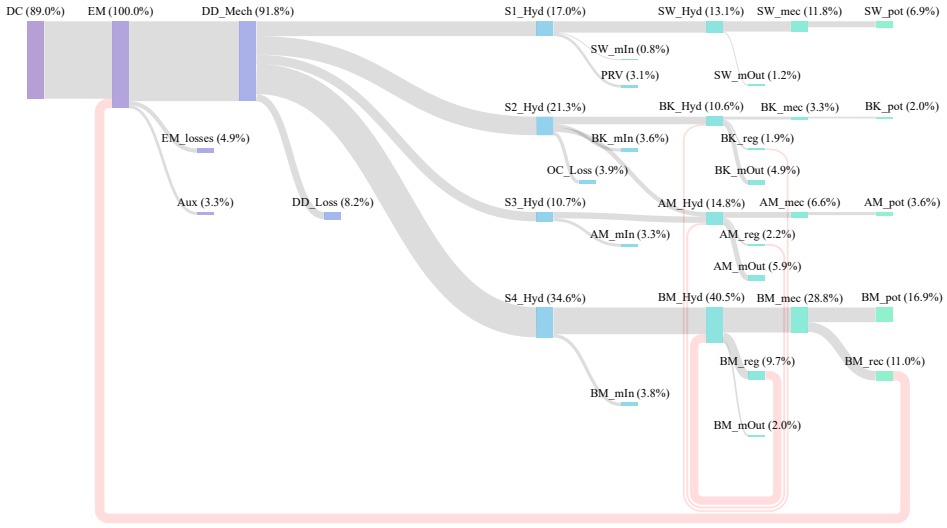


Figure 3.3. Sankey chart of the energy flow for the MAX system in a JCMAS DnD cycle.

4. CONCLUSION AND FUTURE WORK

The Dextreme MAX prototype successfully demonstrated that a fully digital, multi-service electro-hydraulic architecture can significantly improve the efficiency of a 30-tonne electric excavator without compromising performance. Tests conducted on standardized duty cycles showed energy consumption reductions of 31–49%, corresponding to a 35% decrease in electric energy demand for typical operations. A large part of the energy savings stemmed from reduced throttling losses due to independent actuator control, while approximately one-quarter resulted from energy recovery during boom lowering. While the results confirm the potential of DD technology for large excavators, time constraints, some control and hardware limitations hindered full system optimization. Specifically, incomplete opening of

the boom Hose Burst Control Valve prevented full energy recovery, and the current control logic can be refined to improve transient performance and motoring efficiency.

Future work will therefore focus on:

- Enhancing control performance and addressing the identified valve control issue to enable complete boom energy recovery.
- Conducting real dig-and-dump testing once the machine returns to the Danfoss Application Development Centre (ADC) for validation under representative working conditions.
- Investigate and extend the concept to include an electrified swing drive (e-swing), which represents an obvious next step given the electric platform; this addition is expected to further reduce losses and enable additional energy recuperation.

ACKNOWLEDGEMENT

Part of this work was carried out under the Red Diesel Replacement Phase 2 (DEPOWER2) project, *Digital Electro-Hydraulic Power Module with Energy Recovery – Phase 2*, supported by the UK Department for Energy Security & Net Zero (DESNZ) through the Net Zero Innovation Portfolio (NZIP). Project Lead Organisation: Danfoss Scotland Ltd. Project Number: RDR2-01-DFS (con_4308). Project Duration: 04 July 2023 – 30 June 2025.

NOMENCLATURE

F_d	Fraction of Displacement [0-1]
n	Shaft speed
p	Pressure
P	Power
pc, tc	Pressure control/torque control
Q	Flow
S_i	Service [i=1-4]
V_{di}	Service capacity cc/rev [i=1-4]

5. REFERENCES

[1] Heybroek, K., “On Energy Efficient Mobile Hydraulic Systems: With Focus on Linear Actuation,” PhD Thesis. Linköping University, 2017.

[2] Williamson, C., Zimmerman, J. D., and Ivantysynova, M., “Efficiency Study of an Excavator Hydraulic System Based on Displacement-Controlled Actuators,” presented at the FPMC, University of Bath, UK, 2008.

[3] Busquets, E., and Ivantysynova, M., “The World’s First Displacement-Controlled Excavator Prototype with Pump Switching—A Study of the Architecture and Control,” presented at the 9th JFPS International Symposium on Fluid Power, Matsue, Japan, 2014.

[4] Zimmerman, J. D., “Toward Optimal Multi-Actuator Displacement-Controlled Mobile Hydraulic Systems,” PhD Thesis. Purdue University, 2012.
[5] Ketonen, M. S. M., Minav, M., and Pietola, M., “Simulation Study of a Digital Hydraulic Independent-Metering System for an Excavator Application,” presented at the The 15th Scandinavian International Conference on Fluid Power, Linköping, Sweden, 2017.
[6] Ge, L., Quan, L., Li, Y., Zhang, X., and Yang, J., “A Novel Hydraulic Excavator Boom Driving System with High Efficiency and Potential Energy Regeneration Capability,” *Energy Conversion and Management*, Vol. 166, 2018, pp. 308–317. https://doi.org/10.1016/j.enconman.2018.04.046
[7] Heybroek, K., Vael, Ir. G., and Palmberg, P. J.-O., “Towards Resistance-Free Hydraulics in Construction Machinery,” Linköping University, 2012.
[8] Vukovic, M., “Hydraulic Hybrid Systems for Excavators,” Shaker Verlag, Aachen, 2017.
[9] Pellegri, M., Green, M., Macpherson, J., McKay, C., and Caldwell, N., “Applying a Multi-Service Digital Displacement® Pump to an Excavator to Reduce Valve Losses,” presented at the 12th International Fluid Power Conference, Dresden, Germany, 2020.
[10] Green, M., MacPherson, J., Caldwell, N., and Rampen, W., “DEXTER: The Application of a Digital Displacement® Pump to a 16 Tonne Excavator,” 2018.
[11] Macpherson, J., Williamson, C., Green, M., and Caldwell, N., “Energy Efficient Excavator Hydraulic Systems with Digital Displacement® Pump-Motors and Digital Flow Distribution,” presented at the FPMC2020, 2020.
[12] Rampen, W. H. S., “Digital Displacement Hydraulic Piston Pump,” The University of Edinburgh, 1992.
[13] Pavlis, V., McMillan, C., Nørgård, C., and Caldwell, N. J., “Digital Displacement Motoring Characteristics of Dynamic Energy Recovery and Hydraulic Transformation,” presented at the FPMC2023, Sarasota, Florida, USA, 2023. https://doi.org/10.1115/FPMC2023-111913
[14] Department for Energy Security & Net Zero and Department for Business, Energy & Industrial Strategy, “Red Diesel Replacement Competition: Phase\,2 Successful Projects,” Government Report, UK Government, September 2023.
[15] Manning, C., “NASA - Technology Readiness Levels,” SCaN (Space Communications and Navigation) Program. Retrieved 29 August 2025. https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/
[16] Hutcheson, J., Abrahams, D., Green, M., and Rampen, W., “Motion Control of a Hydraulic Cylinder with a Digital Displacement Pump- Motor,” presented at the The 13th International Fluid Power Conference, 13. IFK, Aachen, Germany, 2022.
[17] Heybroek, K., Larsson, J., and Palmberg, J.-O., “MODE SWITCHING AND ENERGY RECUPERATION IN OPEN-CIRCUIT PUMP CONTROL,” presented at the The 10th Scandinavian International Conference on Fluid Power, Tampere, Finland, 2007.

Biographies



Matteo Pellegri received the BSc (2012) and MSc (2014) degree in mechanical engineering from *Università degli studi di Parma* , and the PhD in Engineering from *Purdue University* in 2018, respectively. He is currently working as a Senior Engineering Manager for Danfoss focussing on systems using Digital Displacement technology. He authored publications in different areas of fluid power and he has been serving as a reviewer for many highly respected journals.



Paul Marshall graduated from the University of Strathclyde in 2016 with a master's degree in mechanical engineering. Since then, he has worked on developing and testing prototype off-highway vehicles using Digital Displacement hydraulic systems for Artemis Intelligent Power and Danfoss Scotland. As a Senior Systems Engineer, he is currently working alongside OEM partners, providing technical leadership on projects aiming to improve the system efficiency of excavators through the application of Digital Displacement pumps and novel hydraulic systems.



Casper Olesen received his M.Sc. degree in Electro-Mechanical System Design from Aalborg University, Denmark, in 2011. He is currently a Senior Systems Engineer at Danfoss Power Solutions, Denmark. His work focuses on the development and analysis of system architectures integrating hydraulic and electrical components, with an emphasis on improving overall performance, comfort and efficiency.



Erik Westergaard is a Staff Engineer with Solutions R&D department of Danfoss Power Solutions. Solutions R&D is responsible for developing system solutions for off-highway machinery through the value-added combination of hydraulic and electrical components with control software. Based out of Danfoss' Nordborg, Denmark, location of the Application Development Center, Erik brings 14 years of off-highway vehicle system-centric experience, blending his practical approach from a farming background with a strong technical education. His contributions focus on steady advancements in product and system development, varying from operator comfort and productivity to system efficiency.

Environmental Sustainable Pneumatics

Dr. Wolfgang Gauchel, Victoria Pohl, Daniel Ditterich

Festo SE & Co. KG
wolfgang.gauchel@festo.com, victoria.pohl@festo.com, daniel.ditterich@festo.com

Abstract.

Environmental sustainability has become a key priority in production and manufacturing engineering, driven by increasing regulatory requirements and customer expectations for reduced environmental impact. Of course, this is also due to the realization that the current economic model cannot be continued for ecological reasons. The transition to climate-neutral and resource-efficient production creates high pressure to adapt to pneumatics. After all the problem lies in the high energy demand (compared to electric drives) and the associated high CO_2 emissions. The assumption that customers switching to green electricity will solve the problem could be too short-sighted.

The focus of this paper is on drive technology, other areas such as vacuum technology are only dealt with in passing. After an introduction including a summary of the European Ecodesign for Sustainable Products Regulation (ESPR) [1] in chapter 1, chapter 2 deals with the possibilities for improving the energy efficiency of pneumatic actuators, including compressed air generation and distribution, as the greatest challenge with regard to CO_2 neutrality. Chapter 3 describes the Product Carbon Footprint (PCF) of pneumatic components, i.e. the CO_2 emissions caused during their production. An initial assessment with regard to the circular economy is provided in chapter 4. Finally, the topics mentioned are summarized in chapter 5 in the form of a comparison between pneumatic and electric drive technology.

1. INTRODUCTION

The German Federal Environment Agency [2] defines six principles that can be used to determine eco-design – and which can therefore be used as a benchmark for environmental sustainable pneumatics. These are

- **Durability**
  An area in which pneumatics has significant strengths due to its reliability and robustness.
- **Repairability**
  Repair kits for pneumatic actuators have been offered for decades. Due to the modularity of their design, maintenance units and valve terminals also have the option of replacing

a defective module, e.g. a valve. In this way, the entire system, in our example the valve terminal, is quickly put back into operation.

- **Resource efficiency**
  This aims to reduce the consumption of energy, materials and water in the production process and during the product life cycle. Behind the water consumption during the use phase but also in the production of pneumatic components, a tick can be set very quickly. The material consumption of pneumatics is also comparatively low, as described in chapter 3 on the Product Carbon Footprint cradle to gate (PCF ctg). Energy consumption, on the other hand, is the central challenge, which is dealt with in detail in chapter 2.
- **Lack of problematic substances**
  In pneumatics, this is usually the case, apart from lead in piezo valves or using of per- and polyfluoroalkyl substances (PFAS) [3]. The last ones are depending on the application used in seals and lubricants [4, 5], partly also in other plastic components.
- **Renewable raw material**
  are hardly used in pneumatics. The majority of pneumatic components consist of aluminum and steel, both of which can be easily and economically recycled (see chapter 4.4). There is potential for further development with regard to renewable raw materials in the field of seals and lubricating greases. Of course, tubes, valve housings, sensor housings or various cover plates made of plastic based on renewable raw materials are also a possibility.
- **Circular Economy (CE)**
  Due to their simple design and robustness, pneumatic components are well suited for the CE-strategies RePair and ReUse as well as ReCycling (see chapter 4). The ReFurbish and ReManufacture strategies, on the other hand, are hardly used at the moment, so they are a future option.

In the following subchapters, important aspects of 'Environmental Sustainable Pneumatics' are discussed. These are energy efficiency (chapter 2), Product Carbon Footprint (chapter 3) and Circular Economy (chapter 4). Chapter 5 summarizes the results.

2. ENERGY EFFICIENCY

Energy efficiency has been a topic in pneumatics for many years. And numerous publications have been published on this subject, including some with an overview character [6, 7, 8].

The energy transition from fossil fuels to CO_2-neutral energy sources requires significant improvements in the energy consumption of pneumatic systems. This is because the demand for electrical energy is increasing significantly as a result of the energy transition – whether for the operation of cars (electromobility), for heating buildings (heat pumps) or for hydrogen generation (electrolyser). In these and many other cases, electricity is needed. Accordingly, (green) electrical energy is and remains a rare resource that should only be used efficiently in the future. For pneumatic applications, this also results in the need for efficiency, as electrical energy is used to generate compressed air (see chapter 2.1). To illustrate the need for energy-efficient pneumatics with a concrete example, reference can be made from the sustainability report of Festo from 2024 [9].

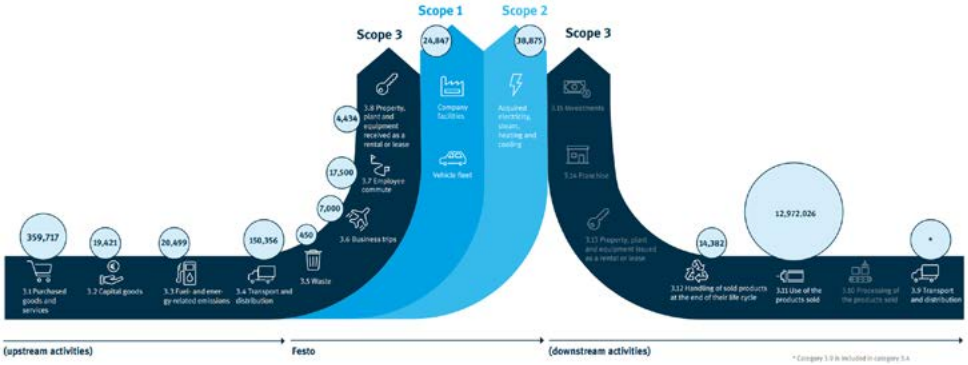


Figure 1. Company Carbon Footprint (CCF) 2024 of Festo [9]

The report shows the Company Carbon Footprint (CCF) of Festo, see Figure 1. 95% of the CCF of Festo is caused by Scope 3.11, the 'use of product sold'. Festo also sells components such as electric actuators, sensors, valves, valve terminals, ... that have a direct electrical energy consumption. But the 'indirect' electrical energy consumption for compressed air generation is by far the largest part of Scope 3.11. About half of the remaining 5% of the CCF is Scope 3.1, the CO_2equivalent (CO_2e) of the 'purchased goods and services'. The next largest share of just under 1% of the CCF is caused by the delivery of supplier or (own) preliminary products to our factories (Scope 3.4, 'transport and distribution'). Both factors are well described by the aggregation of the PCF ctg of the products sold (see chapter 3).

Festo has set itself the goal of reducing greenhouse gas emissions in Scope 3 by 90 per cent by 2050 compared to the base year 2022 [10]. This ambitious climate target was officially validated by the Science Based Targets initiative (SBTi) in February 2025. In the Target Dashboard of the SBTi [11] not only Festo can be found. It reads a bit like the 'Who's Who' of the pneumatics industry and its suppliers: SMC Corporation, Emerson Electric Co. (Aventics), IMI plc. (Norgren), Parker Hannifin Corporation, Schunk SE & Co. KG, Trelleborg AB, ... This demonstrates the importance of CO_2-reduction for the pneumatic industry and their customers, which are also widely represented there.

In the following subchapters, the entire compressed air chain is described from an energy perspective, from compressed air generation (chapter 2.1) to its distribution (chapter 2.2) to the operation of machines and systems (chapter 2.3). Chapter 2.4 concludes the topic of energy efficiency with the control of pneumatic actuators.

2.1. Compressed air generation

Pneumatic energy efficiency starts with the compressor. Since thermodynamic effects inevitably generate a lot of heat during the compression and generation of compressed air [12], and this cannot be used pneumatically, compressor manufacturers are very concerned with the topic of heat recovery [13, 14, 15]. The recovered heat can be used, for example, for heating purposes or to generate required process heat. There are also solutions to heat rooms with the compressors in winter and to cool them in summer by means of absorption chillers [16].

In summary, future improvements in compressor technology to recover heat and reduce specific power consumption for air compression [6] will help pneumatics reduce their CO_2 emissions in Scope 3.11.

In the future, it may become a niche application that compressed air is no longer generated centrally via large compressor stations, but decentral on site. Initial solutions for a decentral pressure supply, e.g. for pneumatic grippers, are already available on the market [17, 18]. Decentralized compressors for vacuum generation for suction cups are also offered on the market for several years [19, 20].

2.2. Compressed air distribution

After generation, the compressed air is distributed in smaller or larger compressed air networks. These must be appropriately dimensioned and designed with a view to the connected loads. To avoid losses due to leaks, regular inspection and monitoring of the compressed air network is helpful. Accordingly, there are various providers – from start-ups to pneumatics suppliers – that are active in this area [21, 22]. Comparable offers are also available from compressor manufacturers [23]. The monitoring of the compressed air network is also partly linked to a management system that determines the most efficient combination of compressors supplying the compressed air network for the current compressed air consumption [24]. Acoustic cameras from various manufacturers have proven to be a practical tool for easy detection of the exact location of the leak [25, 26].

2.3. Design and operation of machines and systems

The basics for efficient design and efficient operation of pneumatic systems in machines and plants are known for a long time. As an example, a white paper by Festo from 2013 describes basic approaches to reducing compressed air consumption [27], cf. Figure 2.

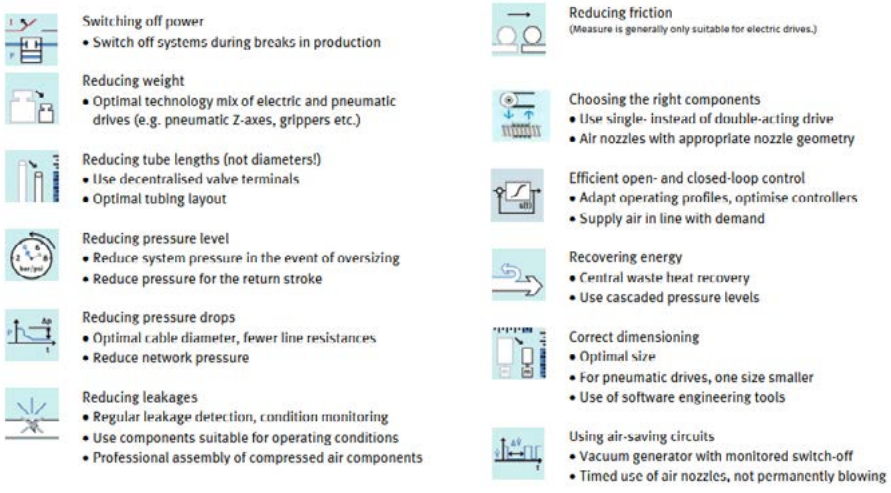


Figure 2. Measures to reduce compressed air consumption [27]

As well as for the compressors and the compressed air network, there are now also solutions for monitoring individual machines and systems. One example is the SFAM flow sensor with integrated pressure sensor from Festo, which can be used to detect compressed air fluctuations, leaks and other anomalies as part of a maintenance unit [28]. In addition, there are also monitoring systems that enable compressed air shutdown or pressure reduction in the event of detected downtimes of the machine/system [29, 30]. E²PME, a start-up from Marly (CH), goes one step further. They offer a data logger that can be used to record and evaluate the pneumatic and electrical power consumption of a machine. In the event of a standstill, electrical circuits can be switched off in addition to the compressed air, which, according to E²PME, allows savings of between 25 and 50 % in stand-by mode [31].

2.4. Pneumatic drive technology and its control

For the individual drive, the correct design and dimensioning is first of all necessary, as made possible e.g. by the web tool pneumatic simulation from Festo [32]. This is based on the Pneumatic Frequency Ratio (PFR), which describes the relationship between the pneumatic natural frequency and the application-specific frequency [33].

-17% ← Change of extension time
+6% ← Change of retraction time
-67% ← Change of air consumption per double stroke

-67% ← Measurement result
-67% ← Simulation result

not appl. ← Non-functional due to technical/safety constraints (s. explanation)

Energy saving measure / Load case	Design measures				Energy saving circuits						Energy saving systems			
	Pressure reduction		Single-act. cylinder		Supply air cut-off		Short circuit		3-phase circuit		PCC Blue Mader		Festo Motion Terminal	
Cylinder size	Ø32	Ø50	Ø32	Ø50	Ø32	Ø50	Ø32	Ø50	Ø32	Ø50	Ø32	Ø50	Ø32	Ø50
1-1 $\Omega_e \approx 1.3$ $\Omega_r \approx 1.6$	-3% -3% -67%	0% 0% -67%	+100% +680% -45%	+158% +355% -45%	-2% -8% -9%	-4% -10% -14%	+78% +67% -45%	+80% +80% -45%	-4% -8% -66%	-2% -15% -69%	not applicable		-15% -10% -27%	-11% -25% -7%	+1% -9% -31%
1-2 $\Omega_e \approx 2.6$ $\Omega_r \approx 3.2$	+7% +6% -50%	+4% +11% -51%	0% +430% -45%	+8% +245% -45%	-3% -4% -6%	-8% -15% -16%	+4% +2% -45%	-5% +41% -45%	-2% -5% -64%	0% -14% -69%			-33% -24% -22%	-13% -4% -75%	+4% -18% -37%
1-3	-3% -1% -67%	+3% 0% -3%	not appl.	not appl.	-17% -9% -9%	-6% -9% -16%	not appl.	not appl.	-1% -11% -55%	-4% -17% -67%			-16% -9% -10%	-6% -12% -50%	+2% -1% -33%
2-1	-22% +46% -18%	+8% +9% -6%	+67% +317% -45%	+100% +283% -45%	-13% -7% -8%	-8% -7% -10%	+6% +92% -45%	+4% +96% -45%	-39% -27% -58%	-37% -33% -58%			not appl.	+2% +7% 0%	+7% +15% 0%
2-2	5% +18% -6%	+5% +8% -6%	+10% -84% -46%	+111% +287% -45%	-4% -4% -1%	-2% -3% -4%	not appl.	not appl.	not appl.	not appl.			not appl.	-6% +7% 0%	+4% -9% 0%
3-1	-6% +65% 23%	+15% +43% 20%	+71% -108% -16%	+89% +317% -45%	-2% -4% -9%	-3% -10% -12%	+153% +64% -45%	-8% -108% -45%	-14% -27% -39%	-25% -12% -10%			-4% -11% 6%	+62% +66% -5%	+19% +14% -20%
3-2	-22% +61% 23%	+33% +6% -4%	not appl.	not appl.	-4% -10% -9%	-5% -18% -20%	-34% -108% -45%	-117% +97% -45%	-23% 0% -62%	-21% -23% -60%			-12% -15% -25%	-17% +6% -67%	-4% +4% -60%
3-3	-5% -30% -10%	-1% -31% -5%	not appl.	not appl.	-4% -8% -9%	0% +5% -22%	not appl.	not appl.	-20% +1% -65%	-24% -31% -67%			-28% +2% -16%	-17% -18% -67%	-7% -11% -57%

Figure 3. Change in performance and air consumption of energy-saving circuits in comparison to the standard meter-out circuit [34]

The PFR is also discussed in a recent publication in the Journal for Fluid Power [34] by comparing control strategies for pneumatic drives with regard to energy consumption and cycle time for different, typical applications (see Figure 3). As this publication shows, the supply air cut-off offers a way to save up to 20% of energy and time compared to pneumatic meter-out circuit in a wide range of applications. Instead of a 5/2-way valve, this control technology uses a 5/3-way valve with a closed center position, which is switched to the middle position when the end position is reached. As a result, the drive chamber is no longer filled with compressed air. The final pressure in the end position remains low, which saves energy. This has the additional effect that less pressure needs to be exhausted to start the countermovement. Thus, cycle time is shortened. If energy efficiency becomes more important, such a solution may get a stronger footprint in the market.

As can be seen in the right part of Figure 3, more complex energy saving systems are required to achieve greater savings. One example for such a solution is the Motion Terminal VTEM of Festo with the ECO drive app, which is operated at the minimum required pressure depending on the load [35]. At the end of the movement, there is therefore no further increase in pressure in the drive chamber – comparable to the supply air cut-off.

Another example of an energy-saving system is Smart Switching Lite, which is available on the new valve terminal VTUX from Festo [36]. This uses a 5/4-way valve that has two center positions (see Figure 4). In these, the air supply to the ventilation chamber is interrupted, while the ventilation path remains open. This means that no further compressed air is consumed after switching in one of these two middle positions. In principle, as with the supply air cut-off described above, the pressure build-up is stopped – only earlier, during movement and not only when the end position is reached. Accordingly, the pressure in the drive chamber is even lower in the end position – and even less pressure has to be released after the countermovement has been initiated. This further reduces the cycle time.

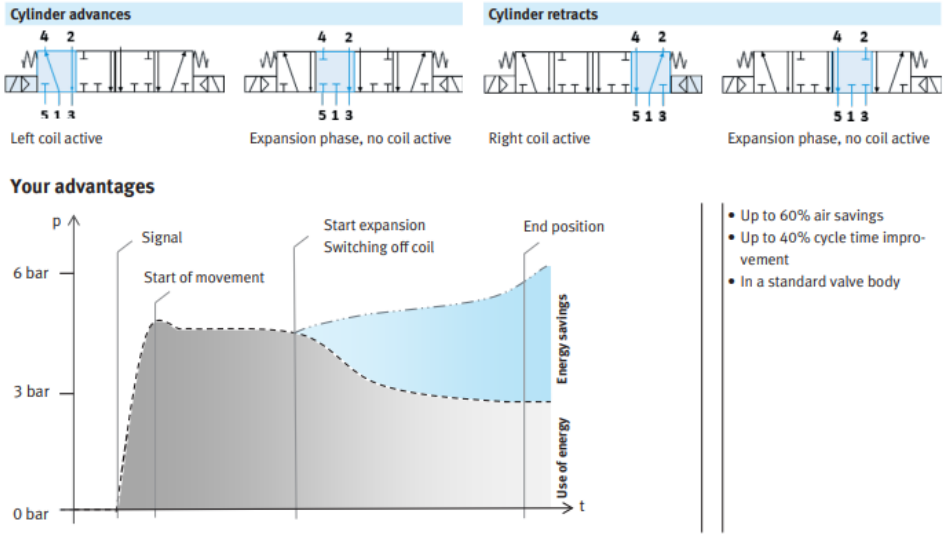


Figure 4. How Smart Switching Lite works [36]

The ECO drive with the Motion Terminal VTEM and Smart Switching Lite on valve terminal VTUX are two examples of what Festo calls Controlled Pneumatics [37]. This describes solutions in which, in contrast to today's common switching pneumatics, control of one or more state variables are possible. This also includes, but is not limited to, servo pneumatics, i.e. the position control of cylinders. Pressure or flow control are also part of Controlled Pneumatics.

In the field of pneumatic vacuum generation, energy savings are made possible by solutions that integrate the ejector itself as well as valve technology to control the supply air and sensors. Such vacuum generator systems reduce the amount of air required to achieve a target vacuum to a minimum. In this way, air-saving functions and functions for energy and process control are possible [38, 39].

3. PRODUCT CARBON FOOTPRINT (PCF)

For a complete CO_2 comparison of different drive solutions, not only the energy consumption in operation and the associated CO_2 consumption is important, but also a 'starting value' is required for the comparison. This is because without the drive having carried out a single stroke, energy has already been consumed to generate the components used, thus CO_2 emissions were generated.

According to the Greenhouse Gas Protocol (GHG) [40], a distinction is made between a Product Carbon Footprint cradle-to-gate (PCF ctg) and cradle-to-grave. PCF ctg is 'a partial life cycle inventory, including all emissions and removals from material acquisition through to when the intermediate product leaves the reporting company's gate'. Cradle-to-grave also includes the use phase of the product (see chapter 2) and the emissions at the end of the life cycle. In the following, PCF is always understood to be PCF ctg. This is because it offers the required 'starting value' for a CO_2-comparison of different drive solutions, comparable to the historical costs [41] in a total cost of ownership (TCO) analysis [42].

The methods for determining this PCF ctg are described theoretically, for example in EN ISO 14067 [43] or the VDMA Guideline 'Calculation of the Product Carbon Footprint in Mechanical and Plant Engineering' [44]. In practice, however, there are some challenges, following is an example:

A pneumatic and an electric drive solution are to be compared with each other for a specific application. On the pneumatic side, all components must be taken into account, from the air preparation unit, covering the fittings and tubes, the valve terminal up to the throttle and cylinder. And these components must be included into the comparison in the right configuration, for example the cylinder with the correct diameter and stroke, the selected end position damping, the desired piston rod shape, ...
In the case of air preparation units and valve terminals, only the proportion required for the individual drive should be considered. The ratio to be taken into account must be calculated even smaller if several drives are supplied with compressed air via one valve – as is usually, for example, in the control of clamps in the body shop of the automotive industry.

It is therefore necessary to provide customers with good support in the design tools in order to be able to compare systems appropriately with each other. In any case, the quality of the data taken into account to determine the PCF ctg must be declared. This is because one

pneumatic supplier only calculates the PCF ctg approximately, the other puts a lot of effort into it. The difference between the two approaches can only be made clear to the customer by information on data quality. Concerning the question how to calculate the data quality of subsystems or even entire drive solutions by using the data quality of individual components or parts, it may be necessary to refine the regulation.

Further refinement of the regulatory framework is required, how PCF ctg calculations can be compared with each other in a simple and clear manner – from individual components to subassemblies to complex systems. This is because a major advantage of pneumatics lies in its low investment costs. Can it really be assumed that a customer will place the extensive documentation of different suppliers next to each other for each of the large number of cost-effective components? And compare them in detail? The 'Example presentation PCF reporting' in the VDMA Guideline [44] comprises 5 pages – unfilled.

The low investment costs of pneumatics correlate with a low PCF ctg. This is because pneumatic components for a drive solution are much more compact and have a lower mass than comparable electric drive solutions. This reduces both the material required in production and the energy required during production, for example when using machinery. The lower PCF ctg of pneumatic compared to electric drive technology can also have a monetary effect. This is because the aggregation of the material share of the PCF of all products sold should result in Scope 3.1 ('purchased goods and services', see chapter 2) of the CCF. As SBTi allows to compensate emissions from the CCF under certain conditions – i.e. certificates from climate protection projects for CO_2-sinks [cf. 45] – a low PCF and CCF also become relevant from an economic perspective. It should be noted, however, that CO_2-compensation is only permitted as a supplementary measure for unavoidable residual emissions that are subject to strict quality requirements [46, 47].

The Carbon Border Adjustment Mechanism (CBAM) [48] is the EU's tool to set a fair price for CO_2-emissions from the production of carbon-intensive goods imported into the EU and to promote cleaner industrial production in non-EU countries. In the medium term, this could have the effect that low CO_2-equivalents of materials and the products produced from them also become economically cheaper.

4. CIRCULAR ECONOMY (CE)

Circular economy (CE) is a 'is a model of resource production and consumption that involves sharing, leasing, reusing, repairing, refurbishing, and recycling materials and products, to extend product life cycle for as long as possible' [49]. In the meantime, with ISO 59000 [50] there exists a family of standards concerning CE. And CE has found its way into legislation. In line with the EU's 2050 climate neutrality goal under the Green Deal, the European Commission proposed in March 2022 the first package of measures to speed up the transition towards a CE, as announced in the Circular Economy Action Plan [51, 52]. CE is also part of the Ecodesign for Sustainable Products Regulation (ESPR), which entered into force in 2024 [53, 54]. As Figure 5 shows, ecodesign requirements such as durability and reliability are described there, in which pneumatics has strengths.

(a) durability;
(b) reliability;
(c) reusability;
(d) upgradability;
(e) repairability;
(f) the possibility of maintenance and refurbishment;
(g) the presence of substances of concern;
(h) energy use and energy efficiency;
(i) water use and water efficiency;
(j) resource use and resource efficiency;
(k) recycled content;
(l) the possibility of remanufacturing;
(m) recyclability;
(n) the possibility of the recovery of materials;
(o) environmental impacts, including carbon footprint and environmental footprint;
(p) expected generation of waste.

Figure 5. Ecodesign Requirements, ESPR Article 5.1 [53]

Other requirements, such as energy and resource efficiency, have already been described in chapters 2 and 3. Requirements such as repairability and recyclability are directly linked to the CE.

In our view, the most important CE-strategies for pneumatics as presented in Figure 6 are ReUse (chapter 4.1), RePair (chapter 4.2), ReFurbish and ReManufacture (chapter 4.4) as well as ReCycle (chapter 4.4) as final step. They will be discussed in the mentioned chapters under perspective of pneumatics. To define the mentioned Re-Strategies, reference is made to ISO 59004 [55], which describes the 'vocabulary' of CE (see Table 1).

One challenge to realize CE in the field of pneumatics can be described using the example of seals. DIN 7716 describes the requirements for the storage of rubber and rubber products [56]. According to this, care must be taken 'to ensure that the products are stored stress-free, i.e. without tension, pressure or other deformations'. However, this is not the case in the installed state of seals. If, for example, for a cylinder or a valve the Re-Strategy ReUse should be used, there may be a significant time interval between the first and the second use. But how long can this be? A week, a month, a year? Or even several? With ReFurbish, too, the question arises, whether seals must be replaced in principle. Or is the sealing still in such a good condition that it can be used in the refurbished component without any significant risk? For lubricating greases, bearings and guides or electronic components [57], there are comparable challenges.

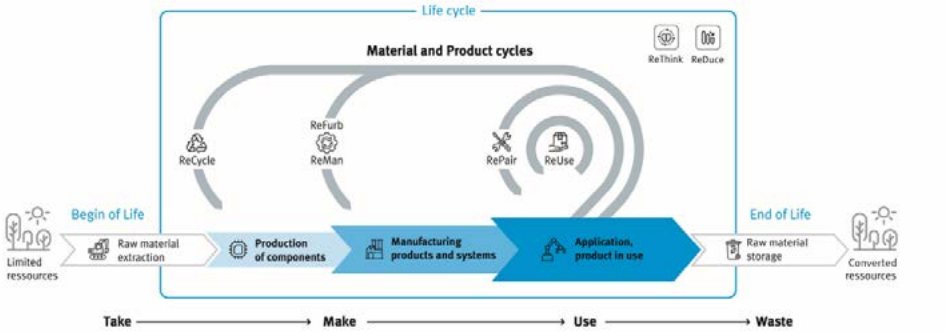


Figure 6. Re-Strategies of the CE

Table 1 – Definition of Re-Strategies in ISO 59004:2024 [55]

Action	Description
RePair	action to restore a product to a condition needed for the product to function according to its original purpose.
ReUse	use of a product or its component parts after their initial use, for the same purpose for which they were originally designed
ReFurbish	restore an item, during its expected service life, to a useful condition for the same purpose with at least similar quality and performance characteristics
ReManufacture	return an item to original condition from both a quality and performance perspective using an industrial process
ReCycle	activities to obtain recovered resources or use in a process or a product, excluding energy recovery

In order to decide on the selection of a suitable Re-Strategy at the end of the first service life, a large amount of information about the component is therefore necessary. In the Factory-X project [58], a proposal for structured data storage (cf. Figure 7) was created for this procedure, called R-Grading. This makes clear that the trends of CE and digitization must be considered together. Moreover, a Digital Twin offers the opportunity to securely and consistent store the large amount of data required over the long service life of the many components, so that a robust decision for a Re-Strategy can be made on this basis.

General Information	Operational Data	History	Physical State
Operator Location	**Environment** • Temperature • Humidity • UV radiation	**Crash Incidence** • Emergency shutdowns • Crash time	**Component operational/functional**
Date of Installation		**Maintenance Operations**	**Technical State**
Expected Wear	**Operating time** • Processing Time • Down Time • Operation Time	**RePair Operations**	**Optical State**
Installation Location		**ReFurb/ReMan Operations**	
	Sensor Data and time series • Pressure Level • Operating Voltage	**Number of power on/off cycle**	
	Compressed Air Quality	**Predictive Maintenance, Condition Monitoring** **Singular Events** • Pressure too low • Pressure spikes • Voltage to low • Voltage spikes **Gradual Changes** • Leakage • Friction	
	Load and stress		

Figure 7. Structure to implement CE-relevant data in an Asset Administration Shell

4.1. RePair

The repair of pneumatic components has a long tradition. The high robustness and reliability of pneumatic components as well as the simple design of the components are certainly reasons for this. Because under these conditions, a repair is worth it, i.e. the replacement of seals or the re-greasing of the components. Spare parts for this are available from Festo, for example, via a spare-part catalog [59], as well as videos to help with the repair [60]. Even tools and repair equipment are available [61].

All the examples mentioned aiming to ensure that the end customer's maintenance staff can repair the pneumatic components on its own. To send back components to the manufacturer for repair is currently not very common in pneumatic drive technology – in contrast to electric drive technology.

As shown in Table 1, there is a definition for RePair according to ISO 59004:2004 [55] ‘restore a product to a condition needed for the product to function according to its original purpose’ But what does this mean in practice? There is a distinction between regenerative repair, i.e. the restore and further use of defective parts or components, and replacement repair, i.e. the replacement of worn or defective components with new components [62]. Replacement repair is probably most common for pneumatic components. This is because a machine/system is repaired by, for example, replacing a defective valve terminal. This is quick and makes it possible for the production plant not to come to a standstill or to restart quickly. The removed valve terminal is repaired, e.g. by replacing a defective valve or another defective module. The valve is repaired by changing the pilot valve, for example. The pilot valve ...
The way in which the repair of pneumatic components will become established and at which of the levels mentioned must and will be tested on the market.

4.2. ReUse

There is probably a long history in the field of ReUse of pneumatic components, even if this is usually not transparent for the manufacturers of pneumatic components. What stops a user from utilizing a valve terminal, a maintenance unit, a pneumatic actuator or even more complex systems such as hall and robot installation platforms (see Figure 8) [63] in the next machine/system when the original machine/system is demounted?

In some cases, the end user is even actively supported in ReUse. For valve terminals and maintenance units, for example, Festo offers a modification aid that supports reconfiguration and thus, reduces the effort required for ReUse [64].

What happens if a machine/system is to be taken out of service and then will be demounted? At latest, the contracted waste disposal company will take a close look at the entire structure and demount valuable components for which it sees a market and ReUse them separately from the remaining parts. It seems, that this is already happening today, because you can find used pneumatic components on various internet portals [65, 66].

In order to increase the ratio of reused components, it would be helpful in future if the component itself – for example via a Digital Twin – provides information about its current 'state of health' and how much of its statistical lifespan it already has consumed. The technical terms on this in the course of the CE are 'Health Management' and 'Remaining Useful Life' (RUL) [67]. This requires access to the data mentioned in Figure 7.

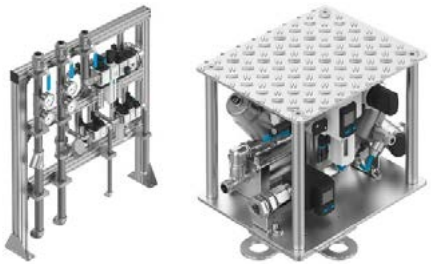


Figure 8. Examples for Hall and Robot Plates (HIP, RIP) [63]

4.3. ReFurbish and ReManufacture

According to the definition of ISO 59004:2004 [55, see Table 1], the take-back/repurchase of components and systems, their reconditioning in different quality levels and the ultimately resale is referred to as ReFurbish and ReManufacture.

While there is a wide range of options for reconditioning motors and motor controllers for electric drives and drive systems [68, 69, 70], this is still largely unknown in pneumatics. The customers of pneumatics, i.e. the machine builders, are already one step ahead. For example, there are offers with integrated pneumatics for rebuild machines [71], pre owned machines [72] or used machines [73]. The general labels for these different terms are ReFurbish and ReManufacture.

Just as with ReUse (see chapter 4.2), Digital Twins are helpful for realizing ReFurbish and ReManufacture for components. This is because the manufacturer of the component needs

information about the usage history of the specific product in order to set a buyback price. How intensive was the use? Under what environmental conditions was it used? Has maintenance and servicing been carried out? If so, how? This and other information (see Figure 7) must be provided by the owner of the component so that the manufacturer of the component is willing to pay a higher price than the scrap price. Only by sharing data an economic benefit for all partners involved can be derived from ReFurbish or ReManufacture.

A specific challenge of pneumatics is that only limited data is available. With this little data, you can certainly achieve important insights in the sense of predictive maintenance [cf. 74]. But since far more data is available for electric drives, more or more accurate information can be derived from it. This fact could result in a disadvantage for pneumatics in the future with regard to ReFurbish and ReManufacture. In the end, it leads to the question already mentioned in the introduction to chapter 4, with which data should a Digital Twin be filled? Fortunately, there are currently some research projects related to fluid power technology such as Factory-X [58], Fluid 4.0 [75] or CliCE-DiPP [76], which deals with the development of Digital Twins based on the Asset Administration Shell (AAS).

For pneumatic components and systems, this raises the question of which of the large number of products a ReFurbish and/or ReManufacture makes sense from an economic as well as an ecological perspective? This is because the deployment, operation and maintenance of a Digital Twin are associated with costs and CO_2e. Even the return transport from the owner/user to the manufacturer cannot be achieved without economic and ecological costs. This significantly limits the number of pneumatic components for which a corresponding offer makes sense.

To put it another way: For typical C-parts such as screws, nuts or washers [77] a ReFurbish or ReManufacture probably doesn't make sense, so not for a Digital Twin. But there is obviously a market for used machines, otherwise there would be no corresponding offers. And somewhere between these two extremes, there is a borderline at which a Digital Twin and the associated CE offerings become ecologically and economically useful.

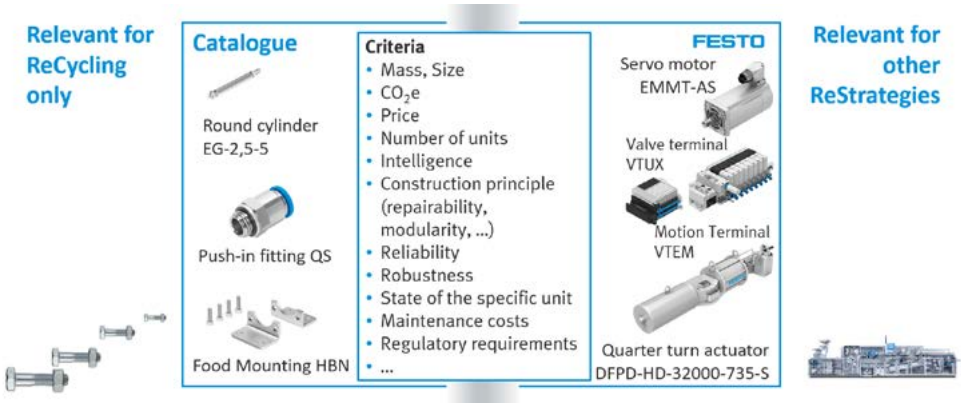


Figure 9. Possible criteria for deciding whether a Digital Twin and the services based on it make sense for CE

This borderline probably runs directly through the catalog of pneumatics suppliers as Festo. To make it concrete with the example of a pneumatic cylinder. Festo offers cylinders with piston diameters from 2.5 mm to 735 mm [78, 79], the product weight for these pneumatic cylinders varies from 2 g to 1.5 kg. It is therefore obvious that the same Re-Strategy does not make sense for all pneumatic cylinders. Figure 9 shows some possibilities for criteria that can be used to draw the mentioned borderline.

4.4. ReCycle

In terms of their mass and very roughly considered, pneumatic components consist of a lot of aluminium, some plastic (in a wide variety of variants) and steel (see the distribution of production materials of Festo [9] shown in Figure 10). In addition, there are some other materials in smaller quantities, e.g. copper, zinc and brass.

For the aluminium production bauxite is needed, which has been declared in the European Union since 2020 as a Critical Raw Material [80]. According to the fifth list of EU critical materials [81], which was published in 2023, aluminium itself is also classified as a Critical Raw material [82, cf. also 83].

Aluminium can be easily recycled as a high-quality metal. The End-of-Life Recycling Rate (EoL RR, [84]) was 69% in Europe in 2021 [85]. Iron and thus steel has an even higher EoL RR, which was already 75% in 2015 [85]. The recycling of plastics is much more complex. For further details, please refer to [86]. With regard to pneumatic components, it should also be mentioned that there are initial considerations to ReCycle sealings [87].

Due to the simplicity of the design of pneumatic components, only a limited number of raw materials are required. On the electronic circuit of a motor controller as example, many more components and raw materials (including critical ones) are used. The recycling processes for electric components are correspondingly more complex to recover the raw materials [88]. Of course, this applies in the same way to the electronic components in pneumatic valve terminals or sensors.

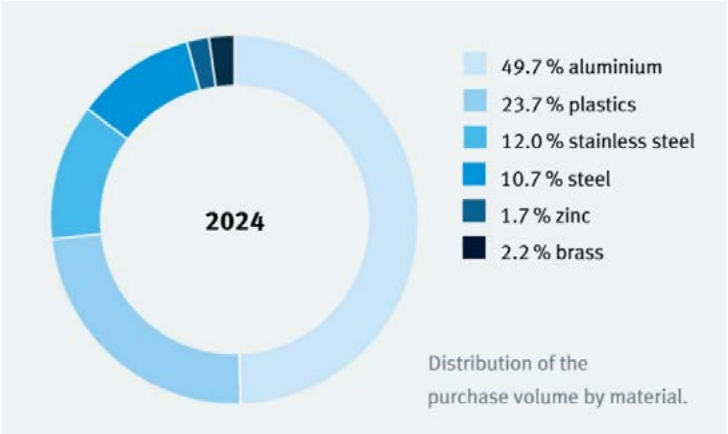


Figure 10. Distribution of production materials of Festo by weight [9]

To illustrate it with another example. Only a few magnets are used in pneumatic drive technology. Although magnet-free electric motors are now available [89], rare-earth magnets [90] are commonly used for powerful motors. And rare-earth elements are critical raw materials [81, 83].

5. SUMMARY AND OUTLOOK

Significant changes are required for pneumatic drive technology to successfully complete the energetic transformation towards CO_2 neutrality on a broad scale. This applies in particular to the high energy and thus CO_2 consumption during the use phase. Controlled pneumatics with solutions such as ECO drive with the VTEM Motion Terminal or Smart Switching Lite as an option for the new VTUX valve terminal offer potential for this. This is associated with an increasing integration of sensor technology, ideally without any significant impact on reliability and robustness of pneumatic systems. This is important in view of a second megatrend in addition to environmental sustainability, which is currently transforming fluid power technology: digitalization. There are definitely opportunities for pneumatics here – even and especially in competition with electric drives.
However, this requires pneumatic research and development to focus more on the integration of sensor technology and the additional customer benefits that can be derived from it (Soft sensors, Functional Mock-Up Units (FMU), predictive maintenance, health index, RUL, ...).

On the other hand, pneumatics offers a number of fundamental advantages that may be relevant in terms of ecodesign. Pneumatic products, for example, are robust and reliable – as significantly fewer components are required for a pneumatic movement compared to an electric movement. System reliability is calculated by multiplying the reliability of all individual installed elements. Another example of the advantages of pneumatics is the simple and cost-effective commissioning. With a 24V power supply (if there is a power supply at all), pneumatic components are suitable for the extra-low voltage range [91]. They are therefore of course subject to the requirements for general product safety [92] - but not the Low Voltage Directive [93]. Accordingly, no qualified electrician in accordance with DGUV V3 [94] is usually required.

Figure 11 concludes this paper with a general comparison of electric and pneumatic drive technology (based on [95 and 96]) as well as a specific comparison in terms of environmental sustainability. In times of change to climate-neutral and resource-efficient production combined with the digital revolution, it is important to reduce the weaknesses of pneumatics – without reducing its strengths.

Criterion	Standard Pneumatics	Smart Switching Lite	Electric Spindel Drive	Electric Tooth Belt Drive
System Properties				
Number of Positions	2-3	2	any desired	any desired
Quality of movement	★★	★★	★★★★★	★★★★
Dynamics	★★★★	★★★★	★★★	★★★★★
Force	★★★★	★★★★	★★★★★	★★
Accuracy	★★★	★★★	★★★★★	★★★
Flexibility	★	★★	★★★★★	★★★★★
Power density	★★★★★	★★★★★	★★★	★★★
Simplicity, Installation effort	★★★★★	★★★★	★★★	★★★
EMC/Ex-proof	★★★★★	★★★★★	★★	★★
Acquisition costs	★★★★★	★★★★★	★★	★★★
Operation costs	★★	★★★	★★★★★	★★★★★
Reliability and Robustness				
Reliability	★★★★★	★★★★	★★★	★★
Load stiffness	★★	★	★★★★★	★★★★
Overload capacity	★★★★★	★★★★★	★★	★★
Robustness against changes of environmental conditions	★★★★★	★★★★	★★★	★★★
Avoidance of decentral heat generation	★★★★★	★★★★★	★★	★★
Noise level, Low vibration	★★	★★	★★★★★	★★★★★
Environmental Sustainability				
Energy Efficiency	★★	★★★	★★★★★	★★★★★
Load-dependent energy consumption	★★	★★★★	★★★★★	★★★★★
PCF c2g	★★★★★	★★★★★	★★	★★★
Circular Economy (CE)				
CE - RePair	★★★★★	★★★★★	★★	★★★
CE - ReUse	★★★★	★★★★	★★★	★★★
CE - ReMan, FeFurb	★★	★★	★★★★	★★★★
CE - ReCycle	★★★★★	★★★★★	★★★	★★★

Figure 11. Comparison of pneumatic and electric drive technology

6. ACKNOWLEDGMENT

We would like to thank the German Federal Ministry for Economic Affairs and Energy (BMWE former BMWK) for financially supporting the work as part of the Factory-X project (funding code: FKZ: 13MX001S). We also would like to thank the German Federal Ministry for Economic Affairs and Energy and the DLR Project Management Agency for financially supporting the work within the ‘Climate-neutral Circular Economy enabled by Digital Product Carbon Pass’ (CliCE-DiPP) project (funding code: 01MN23023G).

7. REFERENCES

[1] EU Commission, Ecodesign for Sustainable Products Regulation, https://commission.europa.eu/energy-climate-change-environment/standards-tools-and-labels/products-labelling-rules-and-requirements/ecodesign-sustainable-products-regulation_en, retrieved on, 16.01.2026

[2] Deutsches Umweltbundesamt, Eco-Design Kit, Ökodesign-Prinzipien, https://ecodesignkit.de/oekodesign-prinzipien, abgerufen am, 16.01.2026

[3] Wikipedia, PFAS, https://en.wikipedia.org/wiki/PFAS, abgerufen am, 16.01.2026

[4] Solveig Adler, PFAS ban - drastic effects for fluid power, VDMA e. V., https://www.vdma.eu/en/viewer/-/v2article/render/158165218, 28.11.2025, abgerufen am 16.01.2026

[5] Nora Menzel, PFAS in Schmierstoffen, CHEMIETECHNIK, Hüthig Medien GmbH, https://www.chemietechnik.de/energie-umwelt/pfas-in-schmierstoffen/1940063, 09.08.2025, abgerufen am 16.01.2026

[6] Energy efficiency in production in the drive and handling technology field (EnEffAH) - Basic principles and measures; 2012 https://www.eneffah.de/EnEffAH_Broschuere_engl.pdf, abgerufen am 16.01.2026

[7] Jan Hepke, 'Energetische Untersuchung und Verbesserung der Antriebstechnik pneumatischer Handhabungssysteme', PhD dissertation, TU Dresden, Germany, 2016

[8] Abschlussbericht Potenzialstudie Energie- / Kosteneinsparung in der Fluidtechnik, CLIMATE CHANGE 19/2021, von Tobias Radermacher, Marcel Merx, André Sitte, Vladimir Boyko, Institut für Mechatronischen Maschinenbau, Technische Universität Dresden, im Auftrag des Umweltbundesamtes, Oktober 2020

[9] Sustainability Report Compact 2024, Festo, https://www.festo.com/PDF_Flip/corp/Festo_CS-Report_2024/en/index.html, abgerufen am 16.01.2026

[10] Corporate sustainability, Our contribution to sustainable development, Festo SE & Co. KG, https://www.festo.com/de/en/e/about-festo/sustainability-and-responsibility/corporate-sustainability-id_3698/, abgerufen am 16.01.2026

[11] Target Dashboard SBTi, https://sciencebasedtargets.org/target-dashboard, abgerufen am 16.01.2026

[12] W. Gauchel, S. Hülsmann, L. Müller, 'On the thermodynamics of pneumatic compression and expansion', The 13th International Fluid Power Conference, 13. IFK, March 21-23, 2022, Aachen, Germany

[13] An introduction to air compressor heat recovery, Atlas Copco AB, https://www.atlascopco.com/en-uk/compressors/greenproduction/energy-recovery-systems, abgerufen am 16.01.2026

[14] Heat Recovery, Otto Boge GmbH & Co. KG, https://www.boge.com/en/heat-recovery/, abgerufen am 16.01.2026

[15] Heat Recovery, Kaeser Kompressoren SE, https://www.kaeser.com/int-en/products/rotary-screw-compressors/heat-recovery, abgerufen am 16.01.2026

[16] Wäsche, Timo; Kilpper, Renate, Fleißige Hummel kühlt mit Kompressorabwärme, Chemie Produktion Anlagen Verfahren (cav), Heft 04/2025, https://www.baelz.de/baelz-aktiv/baelz-publikationen/absorptionskaelteanlagen, abgerufen am 16.01.2026

[17] Electro-pneumatic unit, Series AEPN, Zimmer Group, https://www.zimmer-group.com/en/products/components/clamping-and-braking-technology/aggregates/aepn#/, abgerufen am 16.01.2026

[18] Compact Compressor CRP Serie, SMC Corporation, https://www.smcworld.com/newproducts/en-jp/23/crp/, abgerufen am 16.01.2026
[19] Vacuum Generators End-of-Arm, J. Schmalz GmbH, https://www.schmalz.com/en/products/vacuum-technology-for-automation-301607/vacuum-components-301608/vacuum-generators-307617/vacuum-generators-end-of-arm-332585, abgerufen am 16.01.2026
[20] VGC10, compact, customizable electrical vacuum gripper, OnRobot A/S, https://onrobot.com/en/products/vgc10-vacuum-gripper, abgerufen am 16.01.2026
[21] Compressed air and leakage management software, WRS Energie + Druckluft GmbH, https://wrs-energie.de/en/, abgerufen am 16.01.2026
[22] Reduce compressed air cost: Energy Saving Services, Festo SE & Co. KG, https://www.festo.com/at/en/e/support/additional-services/energy-saving-services-id_21636/, abgerufen am 16.01.2026
[23] Compressed Air System Leak Detection, Kaeser Kompressoren SE, https://us.kaeser.com/services/leak-detection/, abgerufen am 16.01.2026
[24] Optimize your installation with air compressor control systems, Atlas Copco Ltd., https://www.atlascopco.com/en-uk/compressors/service/efficiency/compressor-control-system, abgerufen am 16.01.2026
[25] Fluke ii500 Acoustic Imager, Fluke Deutschland GmbH, https://www.fluke.com/en-us/product/industrial-imaging/fluke-ii500, abgerufen am 16.01.2026
[26] FLIR Si2-LD, Teledyne FLIR LLC, https://www.flir.com/products/si2-ld/, abgerufen am 16.01.2026
[27] White paper, Reduce energy costs in compressed air systems by up to 60%, Festo AG & Co. KG, https://media.festo.com/media/247729_documentation.pdf, abgerufen am 16.01.2026
[28] Flow sensors SFAM, Festo SE & Co. KG, https://media.festo.com/media/153392_documentation.pdf, abgerufen am 16.01.2026
[29] Energy efficiency module MSE6, Festo SE & Co. KG, https://www.festo.com/media/catalog/203348_documentation.pdf, abgerufen am 16.01.2026
[30] Air Management System – AMS20/30/40/60, SMC Corporation, https://www.smc.eu/en-gb/products/air-management-system-ams20-30-40-60~171720~nav?partNumber=AMS, abgerufen am 16.01.2026
[31] Energy analysis and optimization of industrial installations, E2PME, https://e2pme.ch/en/, abgerufen am 16.01.2026
[32] Engineering tools – pneumatic simulation, Festo SE & Co. KG, https://www.festo.com/de/en/s/pneumatic-simulation/, abgerufen am 16.01.2026
[33] M. Doll, R. Neumann, W. Gauchel, Sizing of pneumatic drives under energy efficiency aspects, 14th International Fluid Power Conference (14. IFK), 2024
[34] V. Boyko, F. Nazarov, W. Gauchel, R. Neumann, M. Doll, J. Weber, 'Comprehensive Application-Based Analysis of Energy-Saving Measures in Pneumatics', International Journal of Fluid Power, Vol. 25 1, 27–58, accepted 02 April 2024
[35] App World ECO drive, Festo SE & Co. KG, https://www.festo.com/de/en/app/eco-drive.html, abgerufen am 16.01.2026
[36] Valve terminal VTUX – Platform for the future, Festo SE & Co. KG, https://media.festo.com/media/116224_documentation.pdf, abgerufen am 16.01.2026

[37] Controlled Pneumatics – intelligent air control, Festo SE & Co. KG, https://www.festo.com/de/en/e/solutions/controlled-pneumatics-id_1677430/, abgerufen am 16.01.2026
[38] Compact Ejectors, J. Schmalz GmbH, https://www.schmalz.com/en/products/vacuum-technology-for-automation-301607/vacuum-components-301608/vacuum-generators-307617/compact-ejectors-307841, abgerufen am 16.01.2026
[39] Vacuum generators OVEM, Festo SE & Co. KG, https://www.festo.com/media/catalog/203487_documentation.pdf, abgerufen am 16.01.2026
[40] Product Life Cycle Accounting and Reporting Standard, Greenhouse Gas Protocol, https://ghgprotocol.org/sites/default/files/standards/Product-Life-Cycle-Accounting-Reporting-Standard_041613.pdf, abgerufen am 16.01.2026
[41] Wikipedia, Historical cost, https://en.wikipedia.org/wiki/Historical_cost, abgerufen am 16.01.2026
[42] Wikipedia, Total cost of ownership, https://en.wikipedia.org/wiki/Total_cost_of_ownership, abgerufen am 16.01.2026
[43] Greenhouse gases - Carbon footprint of products - Requirements and guidelines for quantification, EN ISO 14067:2018
[44] VDMA Guideline "Calculation of the Product Carbon Footprint in Mechanical and Plant Engineering", https://www.vdma.eu/documents/34570/83902983/VDMA-Guideline_PCF-Berechnung_%C3%9Cberarbeitung_16.02.23+en-GB.cleaned.pdf, abgerufen am 16.01.2026
[45] What are CO_2 certificates?, Foundation myclimate Switzerland, https://www.myclimate.org/en/information/faq/faq-detail/what-are-co2-certificates/, abgerufen am 16.01.2026
[46] Y. Bessas, ‘SBTi 2025: Die wichtigsten Neuerungen für Unternehmen‘, https://www.globalchanger.com/blog/sbti-aenderungen-2025, abgerufen am 16.01.2026
[47] SBTi CORPORATE NET-ZERO STANDARD, Version 1.3, September 2025, https://files.sciencebasedtargets.org/production/files/Net-Zero-Standard.pdf, abgerufen am 16.01.2026
[48] Carbon Border Adjustment Mechanism, European Commission, https://taxation-customs.ec.europa.eu/carbon-border-adjustment-mechanism_en, abgerufen am 16.01.2026
[49] Wikipedia, Circular economy, https://en.wikipedia.org/wiki/Circular_economy, abgerufen am 16.01.2026
[50] International Organization for Standardization ISO, Circular Economy, https://www.din.de/resource/blob/1100034/6ec27578fca8bae15b2ad068466f2415/iso-circular-economy-ppt-data.pdf, abgerufen am 16.01.2026
[51] The EU circular economy action plan, European Parliament, https://www.europarl.europa.eu/topics/en/article/20210128STO96607/how-the-eu-wants-to-achieve-a-circular-economy-by-2050#the-eu-circular-economy-action-plan-1, abgerufen am 16.01.2026
[52] A new Circular Economy Action Plan For a cleaner and more competitive Europe, European Parliament and Council, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1583933814386&uri=COM:2020:98:FIN, abgerufen am 16.01.2026
[53] Ecodesign for Sustainable Products Regulation, European Commission, https://commission.europa.eu/energy-climate-change-environment/standards-tools-and-labels/products-labelling-rules-and-requirements/ecodesign-sustainable-products-regulation_en, abgerufen am 16.01.2026

[54] Framework for the setting of ecodesign requirements for sustainable products, European Parliament and Council, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1781&qid=1719580391746, abgerufen am 16.01.2026
[55] Circular economy — Vocabulary, principles and guidance for implementation, ISO 59004, First edition 2024-05
[56] Erzeugnisse aus Kautschuk und Gummi – Anforderungen an die Lagerung, Reinigung und Wartung, DIN 7716, Mai 1982
[57] Zentralverband Elektrotechnik und Elektronikindustrie ZVEI, ‘Leitfaden Langzeitlagerfähigkeit von Bauelementen, Baugruppen und Geräten‘, September 2013, https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2013/september/Langzeitlagerfaehigkeit_von_Bauelementen_Baugruppen_und_Geraeten/Lagerfaehigkeit-Bauelemente-Baugruppen-Geraete.pdf, abgerufen am 16.01.2026
[58] Factory-X - The digital ecosystem, https://factory-x.org/, abgerufen am 16.01.2026
[59] Spare Parts Catalog, Festo SE & Co. KG https://www.festo.com/spc/en_de/SparePartsCatalogue/, abgerufen am 16.01.2026
[60] Festo Service, service2see: commissioning and repair, Festo SE & Co. KG, https://www.youtube.com/@FestoService, abgerufen am 16.01.2026
[61] Tools and repair equipment, Festo SE & Co. KG, https://www.festo.com/spc/EN_DE/data/SPC/0/PDF_SAFE/FittingAids.pdf, abgerufen am 16.01.2026
[62] Wikipedia, Reparatur, https://de.wikipedia.org/wiki/Reparatur, abgerufen am 16.01.2026
[63] Compressed air and cooling water supply for robots in the automotive industry, Festo SE & Co. KG, https://www.festo.com/de/en/e/solutions/compressed-air-and-cooling-water-supply-for-robots-id_639444/, abgerufen am 16.01.2026
[64] Modification Aid, Festo SE & Co. KG, https://www.festo.com/spc/en_de/SparePartsCatalogue/?mode=modification-aid, abgerufen am 16.01.2026
[65] https://www.ebay.com/, abgerufen am 16.01.2026
[66] https://www.resale.info/en/, abgerufen am 16.01.2026
[67] Wikipedia, Prognostics, https://en.wikipedia.org/wiki/Prognostics, abgerufen am 16.01.2026
[68] Remanufacturing: a second life for your electric drives?, Bosch Rexroth AG, https://www.boschrexroth.com/en/de/blog/remanufacturing-giving-electric-drives-a-second-life/, abgerufen am 16.01.2026
[69] What does our Repair service offer you?, SEW-EURODRIVE GmbH & Co KG, https://www.sew-eurodrive.de/services/life-cycle-services/operation/services/repair/repair.html, abgerufen am 16.01.2026
[70] Spares & Repairs. Drive circularity by refurbishing and remanufacturing, Siemens Aktiengesellschaft, https://www.siemens.com/global/en/products/services/digital-enterprise-services/topics/circular-spares-repairs.html, abgerufen am 16.01.2026
[71] Rebuilds – economical and sustainable, Uhlmann Pac-Systeme GmbH & Co. KG, https://www.uhlmann.de/en/products-and-solutions/consulting-and-development/rebuilds, abgerufen am 16.01.2026
[72] Pre-owned, TRUMPF SE + Co. KG machines, https://www.trumpf.com/en_INT/products/services/services-machines-systems-and-lasers/pre-owned-machines/, abgerufen am 16.01.2026

[73] Used Machines, DMG MORI EMEA Holding GmbH, https://uk.dmgmori.com/products/stock-and-used/used-machines, abgerufen am 16.01.2026
[74] W. Gauchel, T. Streichert, Y. Wilhelm, 'Predictive maintenance with a minimum of sensors using pneumatic clamps as an example' 12th International Fluid Power Conference (12. IFK). Dresden, October 12 – 14, 2020
[75] Joint research project Fluid 4.0 - Revolutionizing Fluid Engineering, https://fluid40.de/, abgerufen am 16.01.2026
[76] Climate-neutral Circular Economy enabled by Digital Product Passport, CliCE-DiPP, https://clice-dipp.de/, abgerufen am 16.01.2026
[77] C-Teile. Kleine Teile – Große Wirkung!, Würth Industrie Service GmbH & Co. KG, https://www.wuerth-industrie.com/web/de/wuerthindustrie/cteile_management/start_cteilemanagement/was_sind_cteile/cteileeigenschaften.php, abgerufen am 16.01.2026
[78] Round cylinder EG-2,5-5-PK-2, Festo SE & Co. KG , https://www.festo.com/de/en/a/15883/, abgerufen am 16.01.2026
[79] Quarter turn actuator DFPD-HD-32000-, Festo SE & Co. KG, https://www.festo.com/de/en/a/8042203, abgerufen am 16.01.2026
[80] EU Commission, Critical Raw Materials Resilience, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0474, Brussels, 3.9.2020, abgerufen am 16.01.2026
[81] EU Commission, Fifth list 2023 of critical raw materials for the EU, https://single-market-economy.ec.europa.eu/sectors/raw-materials/areas-specific-interest/critical-raw-materials_en#fifth-list-2023-of-critical-raw-materials-for-the-eu, abgerufen am 16.01.2026
[82] SCREEN, Solutions for Critical Raw Materials, ALUMINIUM AND BAUXITE, https://scrreen.eu/wp-content/uploads/2024/01/SCRREEN2_factsheets_ALUMINIUM-update2.pdf, abgerufen am 09.12.2025
[83] Wikipedia, Critical Raw Materials Act, https://en.wikipedia.org/wiki/Critical_Raw_Materials_Act. abgerufen am 16.01.2026
[84] Recycling Rates for Metals, 2012 EUROMETAUX EUROFER, https://www.eurofer.eu/assets/publications/archive/archive-of-older-eurofer-documents/2012-Eurometaux-EUROFERRecyclingRatesforMetals.pdf, abgerufen am 16.01.2026
[85] Recyclingatlas für die Metallerzeugung, Deutsche Rohstoffagentur DERA, https://www.deutsche-rohstoffagentur.de/DERA/SharedDocs/Downloads/Factsheets/factsheets_zusammenfassung.pdf?__blob=publicationFile&v=1 Mai 2025, abgerufen am 16.01.2026
[86] The Circular Economy for Plastics, A European Analysis, Plastics Europe AISBL, https://plasticseurope.org/wp-content/uploads/2024/05/Circular_Economy_report_Digital_light_FINAL.pdf, MARCH 2024, abgerufen am 16.01.2026
[87] A. Werner, H. Best, Rezyklate in Dichtungswerkstoffen – ja, nein – oder irgendwas dazwischen? DICHT! Dichten. Kleben. Polymer. Verstehen. 19. Jahrgang, 4.2025
[88] A World First Precious Metals Recovery Factory, The Royal Mint Limited, https://www.royalmint.com/gold-recovery/, abgerufen am 16.01.2026
[89] MAHLE develops highly efficient magnet-free electric motor, MAHLE GmbH, https://newsroom.mahle.com/press/en/press-releases/mahle-develops-highly-efficient-magnet-free-electric-motor--82368, abgerufen am 16.01.2026
[90] Wikipedia, Rare-earth magnet, https://en.wikipedia.org/wiki/Rare-earth_magnet, abgerufen am 16.01.2026

[91] Wikipedia, ‘Extra-low voltage‘, https://en.wikipedia.org/wiki/Extra-low_voltage, abgerufen am 16.01.2026
[92] Wikipedia, ‘Richtlinie 2001/95/EG über die allgemeine Produktsicherheit‘, https://de.wikipedia.org/wiki/Richtlinie_2001/95/EG_%C3%BCber_die_allgemeine_Produktsicherheit, abgerufen am 16.01.2026
[93] Wikipedia, ‘Low Voltage Directive‘, https://en.wikipedia.org/wiki/Low_Voltage_Directive, abgerufen am 16.01.2026
[94] Deutsche Gesetzliche Unfallversicherung DGUV, DGUV Vorschrift V3, ‘Unfallverhütungsvorschrift Elektrische Anlagen und Betriebsmittel‘, https://publikationen.dguv.de/widgets/pdf/download/article/1052, abgerufen am 16.01.2026
[95] D. Ditterich, Festo AG & Co. KG, White paper: Pneumatics or electrics – important criteria when choosing technology, https://media.festo.com/media/2725_documentation.pdf, abgerufen am 16.01.2026
[96] A. Hildebrandt, S. Hülsmann, D. Ditterich, Pneumatik oder Elektrik – Optimale Antriebsauswahl in der Fabrikautomation, 22. AALE-Konferenz, Rosenheim, 25.-27.02.2026

Biographies





Wolfgang Gauchel received the diploma in mechanical engineering from RWTH Aachen University in 2001, where he also was graduated with a doctorate in 2006. Since then, he is working in the research Department of Festo. He is on the board of the VDMA Fluid Power Research Fund since 2021.

Victoria Pohl received the master's degree in mechanical engineering from the University of Stuttgart in 2021. She then worked as a research associate at the Institute of Industrial Manufacturing and Management (IFF). Since 2024, she has been working in the research department of Festo in the field of circular economy.

D**aniel Ditterich** received his diploma in electrical engineering from the University of Stuttgart and subsequently worked at the Institute of Industrial Automation and Software Engineering (IAS). Since 2010, he has held various positions at Festo, focusing on energy efficiency, digital products, and sustainable automation.

Real-Time Fault Detection in Pneumatic Drive Applications Through Pattern Classification

Vladimir Boyko and Jürgen Weber

Chair of Fluid-Mechatronic Systems (Fluidtronics) – TUD Dresden University of Technology, Helmholtzstraße 7a, 01069 Dresden, Germany, vladimir.boyko@tu-dresden.de

Abstract.

Early fault detection in pneumatic systems is crucial for the prevention of unnoticed energy losses, high energy costs and system failures. According to various estimates, up to 40% of current compressed air consumption can be saved by locating and eliminating leaks. An increasing number of solutions for leakage detection have therefore been proposed in recent decades, which, however, are often cost-intensive, require the use of several additional sensors and cannot be transferred to other machines without extensive customization, leading to their low prevalence in industrial practice of cost-sensitive pneumatics.

This paper proposes a transferable fault detection method for pneumatic linear drives using only one pressure sensor per drive. By integrating neural networks and feature-based classification techniques, the method achieves up to 92.5% classification accuracy on the test dataset. For this purpose, extensive measurement data was generated for various systems and fault scenarios. Data features were extracted in the time and frequency domain using Python's “tsfresh” library and ranked using various feature selection methods such as recursive feature elimination or mutual information. Several classification methods, including Naive Bayes and k-nearest neighbors, were examined and compared with other classification methods such as neural network and cycle time-based fault detection. In this way, a real-time fault detection and localization method can be developed without interrupting machine operation to perform test cycles.

Keywords. Energy efficiency, compressed air systems, pneumatics, pneumatic actuators, fault detection, condition monitoring, energy loss, leakage, friction

1. INTRODUCTION

1.1. Current State

Compressed air is widely used as a practical and easily controllable energy source across many industrial applications. Pneumatic systems are regarded as cost-effective, durable, and environmentally sustainable automatization solutions capable of operating in demanding environments and relatively simple to install and maintain. However, one of the main drawbacks of pneumatic systems is their relatively high energy consumption. This has led

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

to increased research efforts focused on improving their energy efficiency and developing reliable fault detection methods. Studies have shown that up to 40% of the energy used in pneumatics could be saved through more efficient system design and operation [1].

To realize this potential, several strategies are being pursued in science and industry. Primarily, these include demand-based sizing of components such as actuators, pipes, and compressors, as well as energy-saving circuits [2-7]. In parallel, there is growing attention on condition monitoring and fault detection systems. Recent studies [8,9] indicate that research on fault detection in pneumatic systems has grown considerably in recent years.

However, while many of the proposed approaches achieve reliable fault detection, they often require the integration of multiple additional sensors, making them economically impractical for pneumatic systems with numerous actuators [10]. Around 60% of the fault detection methods analyzed in [8,9] depend on the integration of multiple sensors to detect faults, e.g. several pressure and volume flow signals. Such solutions might provide more information and enable fault localization, but they are often not cost-effective for many companies, which limits their use in industry. Moreover, proposed monitoring concepts are often designed for a specific machine, requiring the methods and identified signal patterns to be extensively adjusted before being applied to a different system.

1.2. Commercially Available Fault Detection Solutions

In pneumatic drive systems, leakage and increased friction are the main sources of energy loss [11]. Detecting faults is challenging due to the robustness of pneumatic drives, which often remain functional even with severe issues like major leaks. Increased friction, caused by seal wear or uneven piston loads [12], is also hard to detect as it does not directly increase air consumption. At the same time, addressing leakages can save up to 40% of air consumption [1,13]. For this reason, numerous fault detection systems for pneumatic drives are available on the market; however, they still present several limitations.

Festo's Motion Terminal (VTEM) offers a "Leakage diagnosis" application that performs a diagnostic test cycle to measure leakage volume flow based on cylinder volume, hose length, and pressure drop, halting operation during the test [14]. Similarly, the Festo MSE6-E2M device [15] and other systems such as [16] detect leaks by monitoring pressure drops in a blocked state or by using flow rate limits. However, such solutions lack fault localization without additional data such as valve actuation signals.

Ultrasonic leak detection allows testing without interfering with the system by using portable devices to sense air turbulences, but is less effective in large or hard-to-reach networks and cannot be used in operating systems due to exhaust noise [17-19]. Thermography is also conceivable, but requires expensive cameras and fails in complex machines where the pneumatic actuators are covered by other parts [20].

Friction from piston or guide wear is usually detected by monitoring pressure changes (which in general can also indicate leaks) or using special test rigs, which are not suitable for live systems [21]. High-frequency vibration analysis can identify increased friction during operation, but requires advanced equipment and protection [22,23].

Another monitoring concept that does not incur any additional costs, can be used during operation, and is therefore potentially suitable for industrial use is known as cycle time-based fault detection. The concept, first presented in [24] and expanded in [25,26], involves

using proximity (reed) switches as soft sensors to detect deviations in the extension and retraction times of the pneumatic drive and to draw conclusions about possible fault sources. This allows faults within the pneumatic drive to be precisely localized. However, this approach is partially limited when monitoring highly dynamic drives with a pneumatic frequency ratio of $\Omega < 1.5$ according to [27].

1.3. Pressure-Based Fault Detection and Summary

The outlined state of the art indicates the demand for monitoring methods in pneumatic drive technology that can be applied across different systems and detect and localize faults during operation with minimum number of sensors. Such a low-cost and widely applicable approach based on cycle time was presented in [24,25], but it still has limitations in terms of distinguishability in some use cases. This paper therefore introduces another online method that uses the differential pressure signal (i.e., the difference between current and reference pressure) captured by a single sensor per drive and evaluated with various pattern classification techniques.

Pressure-based fault detection in pneumatic cylinders has been previously studied before (e.g., [23,28]), focusing mainly on analyzing pressure signals and a limited set of features such as min-max values or kurtosis. Moreover, the transferability of classifiers to different cylinder geometries or tasks as well as the influence of the sensor placement have rarely been examined. In the dissertation [11], the effects of three fault types (external/internal leakage and friction) on 32 algebraic features defined by the author, such as pressure difference and mean value, were analyzed using a pneumatic welding gun. This method requires prior training using a validated simulation model of the target system, which limits its direct transferability to other applications.

In his dissertation [10], Dierolf uses pressure, pressure difference, and flow signal and examines 49 features, selecting 4 for online fault detection in a pneumatic machine at 18 Hz sampling rate for pressure sensors. While the method can be applied to other machines without retraining, it has limitations, such as requiring non-overlapping actuator actions. The classification accuracy in the tested example averaged 39%. Another dissertation by Klein [29] demonstrates the potential of using novelty scores and machine learning for fault detection in a single servopneumatic welding gun with high accuracy (up to 100%). Unlike the standard pneumatic drives examined in this paper, such systems typically include multiple sensors (e.g., several pressure sensors and a position sensor), which simplifies fault detection. Nevertheless, significant challenges were observed, including difficulty detecting increased friction and pronounced domain shifts when operating parameters change. These shifts can potentially make trained classification models ineffective, requiring recalibration of the fault detection system.

The aim of this paper is therefore to explore a broader range of time- and frequency-domain features for fault diagnosis of pneumatic cylinders, compare feature-based classification with neural networks and the cycle time-based method, and evaluate their applicability across different pneumatic drives and for different positions of a single pressure sensor per drive.

2. FAULT LOCALIZATION

Table 1 outlines the typical fault locations in pneumatic linear drive systems considered in the paper.

Table 1. Typical faults in pneumatic drive systems.

Fault	Definition and examples
External leakage	Air leakage from the actuator into the environment (locations a)-b) in Figure 1). Possible causes: ▪ Mechanical damage to the hose or hose aging; ▪ Damaged/worn cylinder wiper; ▪ Loosened/damaged push-in fittings.
Internal leakage	Interchamber leakage (location c) in Figure 1). Possible causes: ▪ Damaged/worn cylinder piston sealing. ▪ Piston detachment from the rod
Increased friction	Additional external force (location d) in Figure 1). Possible causes: ▪ Damaged/worn cylinder piston sealing and/or wiper; ▪ Changed friction conditions of mechanical components (e.g., guide wear or contamination).

Figure 1 illustrates potential fault locations in a standard pneumatic drive with meter-out throttling. To detect the faults from Table 1 with minimal amount of sensors, different positions for the single pressure sensor position shown in Figure 1 (placements 1) and 2)) are proposed.

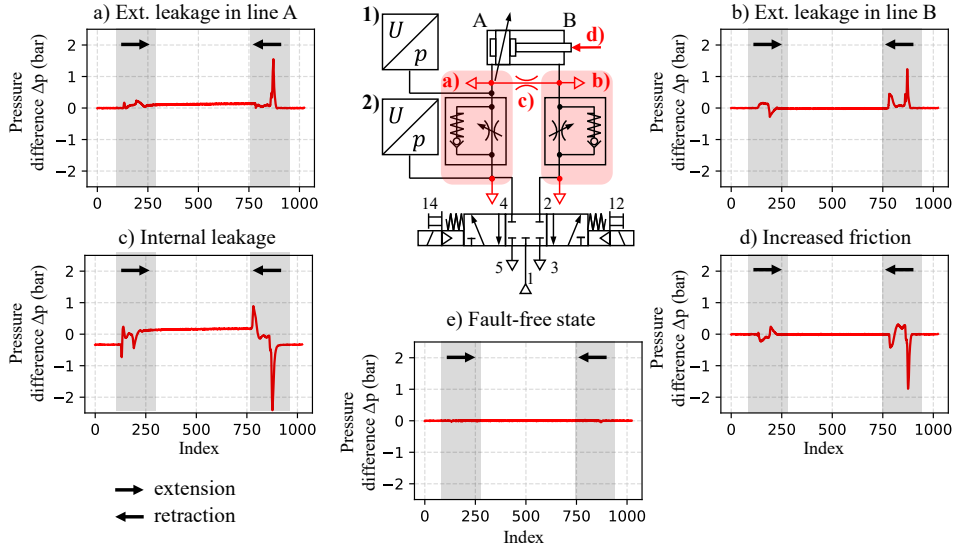


Figure 1. Considered single pressure sensor positions 1) and 2) within a pneumatic drive and potential fault locations a)-d) with corresponding exemplary patterns in pressure difference signals when measured at sensor position 1)

The basic idea behind monitoring the linear drive presented here is to observe the piston-side pressure p_A continuously or periodically. This involves comparing the reference

pressure profile recorded once after commissioning and setup with the repeatedly measured current pressure profile. This results in differential pressure profiles $\Delta p = p_{A,Curr} - p_{A,Ref}$ exemplary shown in Figure 1a-e, the shape of which is to be classified for fault detection and localization.

In the case of leakage in line A, a clear permanent deviation from zero is observed during the standstill in the cylinder's extended state. Corresponding deviations also occur during extension and retraction due to the changed movement time. The deviation is more pronounced during the return stroke, where leakage A accelerates motion due to the quick exhaust effect (see [25]).

Leakage in line B and increased friction also lead to a deviation in extension and retraction times, visible as distinct peaks in the differential pressure signal. Internal leakage alters the pressure profile more severely, leading to significant distortions in the differential pressure signal. In the reference case, by contrast, deviations from zero are absent or minimal, occurring only due to sensor noise or operating fluctuations.

3. USE CASES

For the classification training, measurement data corresponding to fault-free and faulty scenarios was generated on the test bench shown in Figure 2 using the Festo DSBC-32-200-PPV-A (piston diameter 32 mm, stroke 200 mm) and the Hoerbiger R6025/50 (piston diameter 25 mm, stroke 50 mm) double-acting differential cylinders. Leaks were introduced at the locations shown in Figure 1 using additional throttles, as illustrated in Figure 2. Increased friction was simulated with a Festo DSBC-L-50-200-PPV-A pneumatic cylinder (without compressed air supply) mechanically coupled to the guide of the test cylinder.

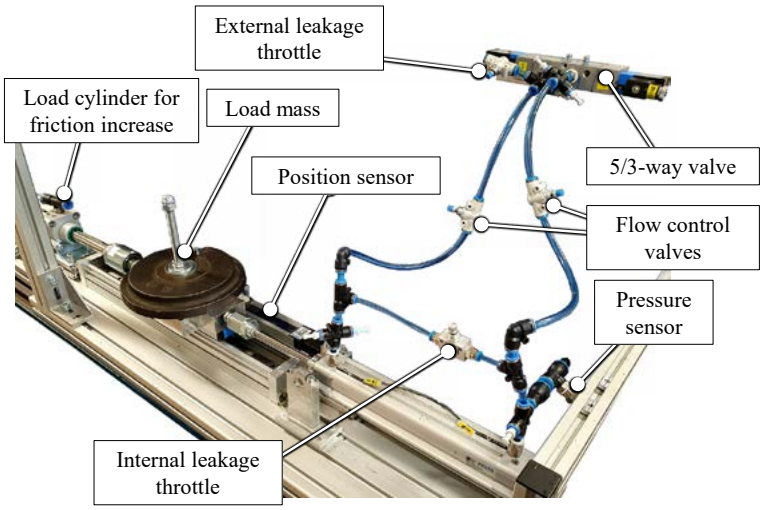


Figure 2. Test bench for pneumatic linear drives

To cover a broad range of operating conditions, the training data parameters were varied according to Table 2. The leakage throttle openings were adjusted to generate two leakage rates (20 l/min and 40 l/min under normal conditions). Throttles connected to the atmosphere at the load cylinder ports were used to generate additional counterforces, representing fault

conditions equivalent to increased friction in the test cylinder. The applied throttle settings resulted in two mean counterforce levels of approximately 20 N and 40 N. These additional force levels were not determined for every operating point of the test cylinder, since the objective was to impose two distinctly different load magnitudes.

The examined drive tasks represent a classic motion task in which a load mass is moved between two positions (load cases 1-1 and 1-2 according to [7]). No pressing or mixed motion-pressing tasks were considered. Approx. 120 pressure profiles per data set were recorded, evenly split between fault-free and faulty cases shown in Figure 1, resulting in about 3000 pressure profiles for training at a 1000 Hz sampling rate. The effect of sampling rate on classification is discussed in Chapter 5.

Table 2. Training data set parameters for pneumatic linear drives

#	Piston diameter (mm)	Stroke (mm)	Hose length (m)	Operating pressure (bar abs.)	Transition time (ms)	Load mass (kg)	Pneum. frequency ratio (-)
1	25	50	5	7	450	11	1.0
2			0.5	7	250	12	1.2
3			0.5	6	500	12	2.3
4			5	6	470	10	1.0
5			0.5	6	500	6	3.2
6			0.5	7	500	12	2.5
7			5	7	800	6	2.4
8			0.5	6	250	12	1.1
9			5	7	250	0.25	3.7
10			5	6	800	6	2.2
11			5	6	250	0.25	3.4
12			0.5	7	500	6	3.5
13	32	200	0.5	6	300	12	1.2
14			5	7	800	6	3.3
15			5	7	600	6	2.5
16			0.5	7	600	12	2.6
17			0.5	7	600	6	3.6
18			5	6	690	8	2.2
19			5	6	800	6	3.0
20			5	6	700	25	1.3
21			0.5	6	600	6	3.4
22			0.5	7	300	12	1.3
23			5	7	700	27	1.4
24			0.5	6	400	6	2.2

The test data in Table 3 was collected to evaluate the classification methods on tasks involving different cylinder geometries and previously unseen operating parameters. Four leakage levels (15, 25, 30, and 40 l/min) and two friction levels (20 and 40 N) were applied. In addition to the pressure signal, cylinder motion times were recorded in the PLC to enable comparison with the cycle time-based method in the final evaluation step.

Table 3. Test data set parameters for pneumatic linear drives

#	Piston diameter (mm)	Stroke (mm)	Hose length (m)	Operating pressure (bar abs.)	Transition time (ms)	Load mass (kg)	Pneum. frequency ratio (-)
1*	50	200	1.5	7	470	31	3.8
2**	40	400	1.5	7	560	0.85	18.2
3	25	50	1.5	7	380	21	1.0

* Festo DSBC-L-50-PPV-A low-friction cylinder with adjustable pneumatic cushioning
** Festo DSBC-40-400-PPSA-N3 cylinder with self-adjusting pneumatic cushioning

4. PATTERN CLASSIFICATION

4.1. Feature-Based Classification Models

The basic procedure for training machine learning models for pattern classification is shown in Figure 3.

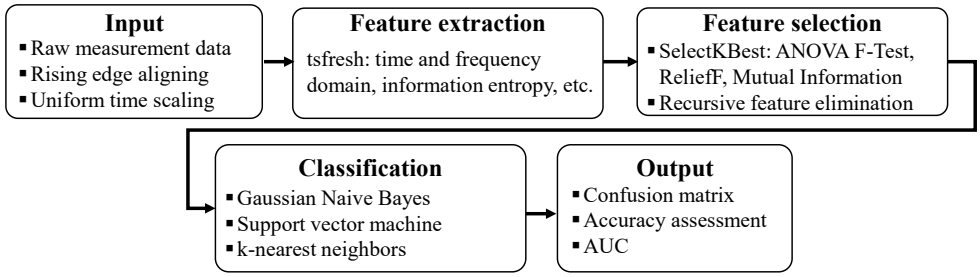


Figure 3. Pipeline for feature-based machine learning in Python

The collected raw measurement data series were singularized so that one file corresponds to a double stroke of the pneumatic cylinder, aligned and scaled to a uniform signal length of 1024 points. Next, the features of the data sets were extracted using the Python library "tsfresh", which extracts 783 features from the time and frequency domains for each profile.

In a further step, four classification models with different combinations of feature selection and classification methods shown in Table 4 were trained and evaluated. The combinations were selected based on established best practices in pattern recognition and machine learning. Filter-based feature selection methods such as ANOVA F-test, mutual information, and ReliefF are widely used with classifiers that do not perform intrinsic feature weighting, including Gaussian Naive Bayes and k-nearest neighbors [30-32]. Recursive feature elimination was combined with a linear support vector machine for classification, following the methodology proposed by Guyon et al. [33].

The training was done using a stratified group k-fold cross-validation with five folds, in which folds were defined by the experimental datasets from Table 2. This ensures that all measurements from the same datasets are assigned to either the training or the validation set, but never split across both (e.g. 1st fold contains datasets 1-5, 2nd fold contains datasets 6-10 etc.) to realistically estimate the performance of the classifier during the training. Within each fold, preprocessing included normalization of all extracted features to the interval [0;1] using min-max scaling and removal of non-informative features via variance thresholding.

Table 4. Combinations of feature selection and classification methods applied in Python for the classification models.

Model #	Feature selection method	Classification method
1	ANOVA F-test from SelectKBest	Gaussian Naive Bayes (GNB)
2	Recursive feature elimination (RFE) with feature ranking via linear support vector machine (SVM)	Linear support vector machine classifier (SVM)
3	ReliefF from SelectKBest	k-nearest neighbors (kNN) with 20 neighbors
4	Mutual information-based feature selection from SelectKBest (MI)	k-nearest neighbors (kNN) with 20 neighbors

The predictive performance of each combination was evaluated on the test data set presented in Table 3, which was not used during training and is therefore unknown to the classifier. Performance was assessed using accuracy and the area under the receiver operating characteristic curve (AUC). The average AUC across all classes indicates how effectively the classifier distinguishes between the different classes.

The described procedure was conducted once for the pressure sensor placement at the cylinder (variant 1) and once at the way valve (variant 2) as shown in Figure 1. In each case, the extension part of the pressure profile and the entire pressure profile (extension and retraction) were examined separately in order to compare performance.

4.2. Neural Network

Another method examined for classifying the pressure profile of the component involves neural networks. Similar to Chapter 4.1, the training data was aligned, scaled to a uniform signal length of 1024 points, filtered, and fed into a neural network for training in the MATLAB environment. The proposed neural network comprises 13 layers in total, including 5 learnable layers: three convolutional layers (1D-CNN) followed by a bidirectional long short-term memory layer (BiLSTM) and a fully connected layer. This allows both local patterns and their global temporal ordering to be captured in the pressure difference signal. The structure of the neural network is presented in Table 5.

The 1D convolutional layers are used as a feature extractor, enabling the detection of characteristic local patterns such as pressure slopes, inflection points, and short-term oscillations that arise from fault events. Their dilation by applying dilation factors of 1, 2, and 4 is used to progressively increase the temporal receptive field of the network while preserving the original time resolution of the signal. This three-stage design represents the minimal depth required to capture local, intermediate, and extended temporal structures without excessive model complexity. Further convolutional layers were avoided, as global temporal dependencies are subsequently modeled by a bidirectional LSTM layer.

The bidirectional LSTM layer models long-term temporal dependencies and explicitly encodes the temporal order of signal components by processing the sequence in both forward and backward directions. This is particularly important for fault patterns that are similar in amplitude but differ in the order of positive and negative pressure lobes. The final fully connected and softmax layers map the learned representation to class probabilities, while

the dropout layer mitigates overfitting. Leaky ReLU activations are employed to preserve information contained in negative signal components and to ensure stable gradient propagation.

The final network depth and parametrization was determined through empirical validation. Investigations on the test dataset confirmed that a three-stage convolutional block provided the optimal balance between sufficient receptive field coverage and generalization capability.

Table 5. Neural network structure

#	Layer	Description and Parametrization
1	Sequence input layer	6 input values
2	1D convolutional layer	Local temporal patterns extraction with unchanged time length; 64 filters, kernel size 7, dilation factor 1
3	Layer normalization	Normalization of activations for training stabilization
4	Leaky ReLU	Activation function with negative slope 0.1
5	1D convolutional layer	64 filters, kernel size 7, dilation factor 2
6	Leaky ReLU	Activation function with negative slope 0.1
7	1D convolutional layer	64 filters, kernel size 7, dilation factor 4
8	Leaky ReLU	Activation function with negative slope 0.1
9	Bidirectional LSTM layer	128 neurons, forward and backward detection of time dependencies, output taken from the last time step
10	Dropout layer	Dropout rate 0.5 for regularization against overfitting
11	Fully connected layer	Maps the 256-dimensional feature vector to 5 output nodes
12	Softmax layer	Output of class probabilities
13	Classification layer	Loss function calculation and final classification

The following six values or features were used as input:

- Normalized pressure difference signal: The original signal is normalized by the global standard deviation σ of all training data
- First derivative of the normalized signal
- Positive portion of the normalized signal.
- Negative portion of the normalized signal
- Order feature: A binary feature indicating whether the first positive peak of the signal occurs before the first negative peak.
- Area ratio: The ratio of the area under the positive signal portions to the area under the negative signal portions

In addition to the normalized pressure difference signal, five auxiliary inputs are provided to embed prior pattern knowledge directly into the learning process of the neural network. The first derivative is intended to improve sensitivity to rising and falling edges. The positive and negative signal portions explicitly encode asymmetric pressure behavior, which is characteristic of different fault locations shown in Figure 1. The order feature and area ratio also directly represent physically meaningful distinctions observed in the patterns, such as a pressure increase followed by a decrease during extension vs. the inverse sequence demonstrated in Figure 1b) and d). Their direct specification through auxiliary inputs simplifies training and classification, as such relationships do not have to be implicitly

recognized or derived by the neural network. Together, this multi-channel representation improves class separability, accelerates convergence, and enhances robustness against variability in real pneumatic systems.

Similar to the classification models in Chapter 4.1, the performance of the neural network was evaluated based on accuracy and AUC for two different pressure sensor positions (at the cylinder vs. at the valve) as well as for two different sections of the pressure profile (extension vs. full stroke).

5. RESULTS

5.1. Performance of Feature-Based Classification Models and the Neural Network

After applying the fault classification methods described in Chapters 4.1 and 4.2 to the test data set from Table 3, which was previously unknown to the classifiers, their performance was compared for the different sensor positions and pressure signal sections. Figure 4 and Figure 5 provide an overview of the accuracy and AUC of the feature-based classification models depending on the number of features in comparison to the neural network.

The neural network and feature-based classifiers showed comparable performance. The highest test accuracy (92.5%) was achieved by model #3 (ReliefF with k-NN) using full-stroke signals with 50 features and pressure measured at the cylinder (Figure 4c). The accuracy drop above 100 features across different classifiers indicates overfitting due to feature redundancy. Model #2 (RFE with SVM) was also highly effective, achieving neural-network-level accuracy and AUC across all settings with only five features. Using RFE, the following “tsfresh” features describing global and local changes in amplitude, periodicity, and frequency content were selected for full-stroke classification with pressure measured at the cylinder:

- `longest_strike_below_mean`: Length of the longest continuous sequence of values below the mean value.
- `sum_of_reoccurring_data_points`: Sum over all repeated observations in the time series (i.e., data points that occur more than once) in the time series.
- `fft_coefficient__attr_"angle"__coeff_77`: Phase angle of the Fourier coefficient at coefficient index 77.
- `fft_aggregated__aggtype_"skew"`: Skewness of the frequency spectrum.
- `agg_linear_trend__attr_"intercept"__chunk_len_10__f_agg_"var"`: Measures how much the baseline level of the signal (intercept of a linear trend) fluctuates between consecutive segments of length 10.

The highest accuracy was obtained when measuring the pressure directly at the cylinder. Measuring pressure at the valve reduced accuracy on average by about 11% for the extension stroke and 8% for the combined extension and retraction. This reduction is likely due to varying hose lengths, which cause more diverse pressure signals for the same fault scenarios, whereas cylinder measurements lead to more uniform patterns.

The inclusion of the return stroke (i.e., evaluation of the full stroke) increased the accuracy of the pattern classification by approx. 4.2% for the pressure measurement at the cylinder and by 6.4% for the pressure measurement at the valve the compared to the variant with extension only.

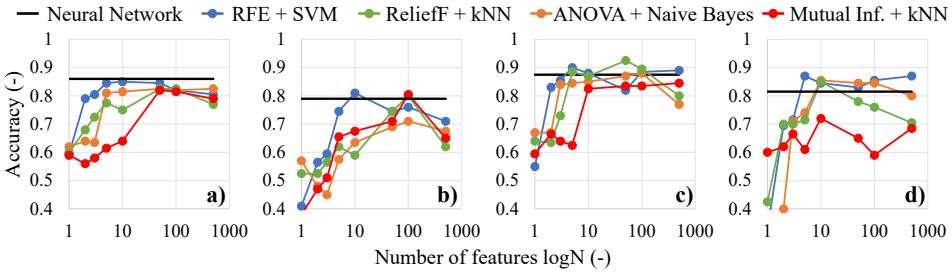


Figure 4. Classification accuracy of the neural network and feature-based models as a function of the number of features, pressure sensor location, and pressure signal segment: (a) cylinder, extension; (b) valve, extension; (c) cylinder, extension + retraction; (d) valve, extension + retraction.

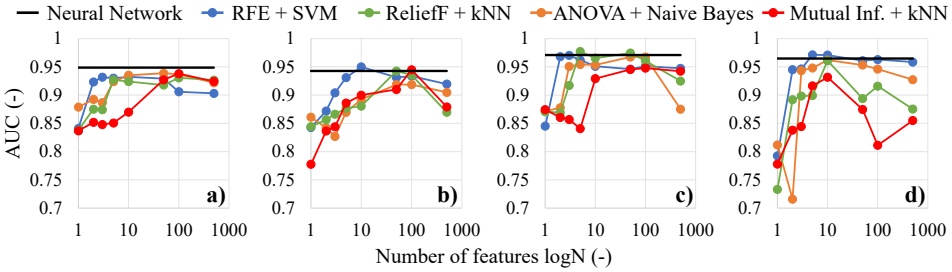


Figure 5. Area under the receiver operating characteristic curve (AUC, average value for all classes) of the neural network and feature-based models as a function of the number of features, pressure sensor location, and pressure signal segment: (a) cylinder, extension; (b) valve, extension; (c) cylinder, extension + retraction; (d) valve, extension + retraction.

Figure 6 shows the influence of the sensor's sampling rate on accuracy and AUC for two selected classification methods (RFE+SVM with 5 features vs. neural network) when measuring extension and retraction pressure profiles at the cylinder. As can be seen, it is possible to reduce the sampling rate from the original 1000 Hz to 100 Hz without significant loss of accuracy. However, further reduction of the sampling rate leads to reduced accuracy.

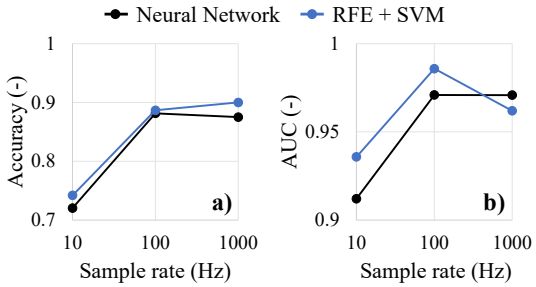


Figure 6. a) Accuracy and b) area under the receiver operating characteristic curve (AUC, average value for all classes) of neural network and classifier RFE+SVM with 5 features depending on the sampling rate of the pressure sensor; pressure measurement at the cylinder, extension and retraction

5.2. Comparison with Cycle Time-Based Method

Finally, the cycle time-based method described in [25] was also applied to the test data set from Table 3. Figure 7 presents the confusion matrices for the selected feature-based classification method and neural network when measuring extension and retraction of the cylinder and compares them with the confusion matrix of the cycle-based method.

The differentiability between the classes is comparably good both for the feature-based method and for neural network methods, with the feature-based classifier in Figure 7a being better at recognizing the error-free state of the cylinder. The cycle time-based method in Figure 7c, on the other hand, encountered difficulties when detecting and localizing minor external leaks of 15 l/min and/or external leaks occurring in the hose between the directional control valve and the throttle valve. Such leaks are particularly difficult to localize, as they often result only in a slight deviation in the cylinder's movement time. Detecting the smaller counterforce increase of 20 N was also problematic. This results in a lower accuracy of the cycle time-based method (74%) compared to the feature-based classifier (90%) and neural network (87.5%).

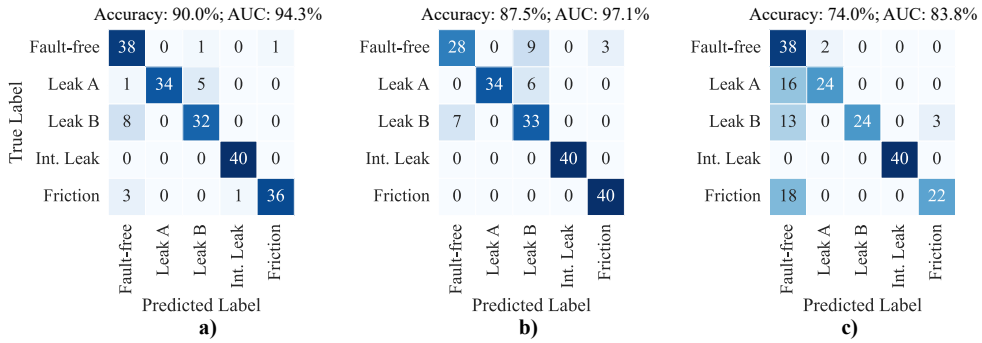


Figure 7. Confusion matrices for test data using different classification methods: a) RFE + SVM (5 features), cylinder pressure, extension + retraction; b) neural network, cylinder pressure, extension + retraction; c) cycle time-based method

6. CONCLUSIONS

This paper presents a fault detection method based on a single pressure sensor per drive combined with various classification techniques. The results show that only a few features are sufficient to detect and localize different fault scenarios with high accuracy (up to 92.5% on the test dataset). Using five features selected by recursive feature elimination and classified with the k-nearest neighbors algorithm, 90% accuracy was achieved. The developed neural network, consisting of three convolutional layers and a bidirectional long short-term memory layer, delivered comparable results. This represents a significant improvement over the cycle time-based soft-sensor method from [25], which achieved 72.4% accuracy on the same test data set. The broad applicability of the method was also confirmed: while training was conducted on cylinders with piston diameters of 25 and 32 mm, the test was performed on other cylinder sizes (40 and 50 mm) and/or under other operating conditions.

However, the method described here still has certain limitations. The current study is restricted to standard horizontal motion tasks. For broader applicability, future work should examine further load cases (e.g., considered in [7]) such as pressing tasks. The training data for various load scenarios and operating conditions can also be obtained through simulation to increase generalizability. Moreover, the method currently requires a pause between extension and retraction cycles and should be extended to continuous-motion applications. High pressure-sensor sampling rates between 100 and 1000 Hz are also necessary to achieve the reported accuracy; reducing the rate to 10 Hz decreased accuracy from 90% to 74.2%. The retrofit with pressure sensors also incurs additional costs that may exceed the cost of the proximity switches. This can be mitigated e.g. by using lower-cost sensors such as Micro-Electro-Mechanical Systems (MEMS) pressure sensors or by their mass production for integration into pneumatic actuators or valves.

Despite these constraints, the proposed method shows strong potential for deployment in typical industrial drive applications. It could be implemented in retrofittable monitoring modules at the cylinder using microcontrollers or mini-PCs, or in valve terminals already equipped with pressure sensors, such as the Festo VTEM, enabling fault detection during operation without dedicated diagnostic cycles.

7. FUNDING

This research was funded by the German Federal Ministry for Economic Affairs and Energy (BMWE), grant number 03EN4050A (OekEffDatA: Data-driven System Monitoring and Design for Ecological Efficiency in Operation).

8. REFERENCES

[1] T. Radermacher, M. Merx, A. Sitte, V. Boyko, ‘Potenzialstudie Energie-/Kosteneinsparung in der Fluidtechnik’, final report No. 37EV181030, TU Dresden, University of Stuttgart, Germany, 2021.

[2] C. R. Jimenez, O. Reinertz, K. Schmitz, ‘Dimensioning Method of Pneumatic Cylinders Based on the Pneumatic End-cushion’, Int. J. Fluid Power, 26(03), pp. 381-410, 2025. DOI:10.13052/ijfp1439-9776.2632

[3] W. Gauchel, ‘Energiesparende Pneumatik. Konstruktive sowie schaltungs- und regelungstechnische Ansätze’, O+P Fluidtechnik, pp. 33-39, 1/2006.

[4] R. Saidur, N. A. Rahim, M. Hasanuzzaman, ‘A review on compressed-air energy use and energy savings’, Renewable and Sustainable Energy Reviews, 14, pp. 1135-1153, 2010. DOI:10.1016/j.rser.2009.11.013

[5] D. D. Šešlija, I. M. Milenković, S. P. Dudić, J. I. Šulc, ‘Improving Energy Efficiency in Compressed Air Systems. Practical Examples’, THERMAL SCIENCE, Vol. 20, Suppl. 2, pp. 335-370, 2016. DOI: 10.2298/TSCI151110022S

[6] D. Gryboś, J. S. Leszczyński., ‘A Review of Energy Overconsumption Reduction Methods in the Utilization Stage in Compressed Air Systems’, Energies, 17, 1495, 2024. DOI:10.3390/en17061495

[7] V. Boyko, F. Nazarov, W. Gauchel, R. Neumann, M. Doll, J. Weber, ‘Comprehensive Application-Based Analysis of Energy-Saving Measures in Pneumatics’, Int. J. Fluid Power, 25(01), pp. 27-58, 2024. DOI:10.13052/ijfp1439-9776.2512

[8] J. Mallia, E. Francalanza, P. Xuereb, P. Refalo, ‘Intelligent Approaches for Anomaly Detection in Compressed Air Systems: A Systematic Review’, Machines, 11, 750, 2023. DOI:10.3390/machines11070750
[9] M. Borg, P. Refalo, E. Francalanza, ‘Failure Detection Techniques on the Demand Side of Smart and Sustainable Compressed Air Systems: A Systematic Review’, Energies, 16, 3188, 2023. DOI:10.3390/en16073188
[10] C. Dierolf, ‘Identifikation und Klassifikation von Druckluft-Leckagen mit Methoden des maschinellen Lernens in Maschinen’, PhD dissertation, Stuttgart, Germany, 2022. DOI:10.18419/opus-12124
[11] S. Fritz, ‘Verfahren zur Erkennung sowie Diagnose von Fehlern in pneumatischen Systemen und Komponenten’, PhD dissertation, Aachen, Germany, 2011.
[12] G. Belforte, L. Mazza, T. Raparelli, ‘Analysis of Typical Failure Situations in Pneumatic Cylinders under Load’, Lubrication Eng., 48(11), pp. 840–845, 1992.
[13] Festo SE & Co. KG, ‘Leckageortung und -beseitigung. Energy Saving Services von Festo’, Esslingen, Germany, 2018.
[14] Festo SE & Co. KG, ‘Weltneuheit Festo Motion Terminal VTEM’, Esslingen, Germany, 2017.
[15] Festo SE & Co. KG, ‘Energieeffizienzmodul MSE6’, Esslingen, Germany, 2018.
[16] AVENTICS, ‘Smart Pneumatics Opportunities for Industry Through Intelligent Pneumatics’, Esslingen, Germany, 2024.
[17] H. Mahmoud, F. Vlašic, P. Mazal, M. Jana, ‘Leakage analysis of pneumatic cylinders using acoustic emission’, Insight-Non-Destr. Test. Cond. Monit, 59, pp. 500–505, 2017. DOI:10.1784/insi.2017.59.9.500
[18] H. Mahmoud, ‘Diagnosis of Pneumatic Cylinders Using Acoustic Emission Methods’, PhD thesis, Brno, Czech Republic, 2017.
[19] D. Czopek, D. Gryboś, J. Leszczyński, J. Wiciak, ‘Identification of energy wastes through sound analysis in compressed air systems’, Energy, 239, 122122, 2021. DOI:10.1016/j.energy.2021.122122
[20] Y. Shi, J. Chang, L. Li, Y. Wang, S. Xu, Y. Niu, ‘Improved object detection network for pipeline leakage localization in the pneumatic system based on thermal images’, Measurement, 242, 116225, 2025. DOI:10.1016/j.measurement.2024.116225
[21] A. Muth, ‘Reibkraftermittlung an pneumatischen Ventilen und Zylindern’, PhD thesis, Dresden, Germany, 1997.
[22] A. Vygantas, M. Saunoris, ‘Investigation of high frequency vibrations of pneumatic cylinders’, Ultragarsas, 51, pp. 21–26, 2004.
[23] M. Tiboni, C. Remino, ‘Condition Monitoring of Pneumatic Drive Systems Based on the AI Method Feed-Forward Backpropagation Neural Network’, Sensors, 24(6), 1783, 2024. DOI:10.3390/s24061783
[24] W. Gauchel, T. Streichert, Y. Wilhelm, ‘Predictive maintenance with a minimum of sensors using pneumatic clamps as an example’. Proc. of the 12th International Fluid Power Conference, Dresden, Germany, 9–11 March 2020; pp. 175–183, 2020. DOI:10.25368/2020.81
[25] V. Boyko, J. Weber, ‘Cycle Time-Based Fault Detection and Localization in Pneumatic Drive Systems’, Actuators, 13, 447, 2024. DOI:10.3390/act13110447
[26] J. Zubieta, U. Izagirre, L. Eciolaza, A. S. De Buruaga, L. Galdos, ‘Step-time measurement: A scalable sub-cycle time defining methodology for anomaly detection and predictive maintenance in sequential production lines’, J. of Manufacturing Systems, 83, pp. 1-11, 2025. DOI:10.1016/j.jmsy.2025.08.011

[27] M. Doll, R. Neumann, O. Sawodny, ‘Dimensioning of pneumatic cylinders for motion tasks’, Int. J. of Fluid Power, 16:1, pp. 11-24, 2015. DOI: 10.1080/14399776.2015.1012437

[28] Ž. Nakutis, P. Kaškonas, ‘An approach to pneumatic cylinder on-line conditions monitoring’, Mechanika, 4(72), pp. 41-27, 2008.

[29] S. Klein, ‘Kombination von überwachten und teilüberwachten Methoden des maschinellen Lernens für eine adaptive und robuste Zustandsüberwachung industrieller Applikationen’, PhD dissertation, Saarbrücken, Germany, 2025. DOI:10.22028/D291-44645

[30] I. Guyon, A. Elisseeff, ‘An Introduction to Variable and Feature Selection’, J. of Machine Learning Research, 3, pp. 1157-1182, 2003. DOI: 10.1162/153244303322753616

[31] P. Schumann, M. Scholz, K. Trentzsch, T. Jochim, G. Śliwiński, H. Malberg, T. Ziemssen, ‘Detection of Fall Risk in Multiple Sclerosis by Gait Analysis—An Innovative Approach Using Feature Selection Ensemble and Machine Learning Algorithms’, Brain Sci, 12, 1477, 2022. DOI: 10.3390/brainsci12111477

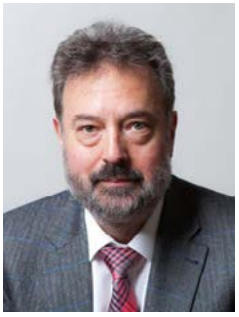
[32] I. Jebadurai, G. Paulraj, J. Jebadurai, S. Silas, ‘Experimental analysis of filtering-based feature selection techniques for fetal health classification’, Serbian J. of Electrical Engineering, 19, 2, pp. 207-224, 2022. DOI: 10.2298/SJEE2202207J

[33] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, ‘Gene Selection for Cancer Classification using Support Vector Machines’, Machine Learning, 46, pp. 389-422, 2002. DOI:10.1023/A:1012487302797

Biographies



Vladimir Boyko holds a M.Eng. degree in Mechanical Engineering from the University of Applied Sciences Jena, Germany. Since 2018 he has been working as a research assistant at the Chair of Fluid-Mechatronic Systems (Fluidtronics), Institute of Mechatronic Engineering, TUD Dresden University of Technology, Germany. His current research fields include pneumatics, fault detection, and energy efficiency.



Jürgen Weber has been appointed in 2010 as a University Professor and the Chair of Fluid-Mechatronic Systems as well as the Director of the Institute of Fluid Power at the TUD Dresden University of Technology, Germany, and took on the leadership of Institute of Mechatronic Engineering in 2018. He finished his doctorate in 1991 and was an active Senior Engineer at the former Chair of Hydraulics and Pneumatics until 1997. This was followed by a 13-year industrial phase. Besides his occupation as the Head of the Department Hydraulics and Design Manager for Mobile and Tracked Excavators, starting in 2002, he took on responsibility for the hydraulics in construction machinery at CNH worldwide. From 2006 onwards, he was the Global Head of Architecture for hydraulic drive and control systems, system integration and advance development CNH construction machinery.

Quasi-Isobaric Compressed Air Reservoirs

Prof. Dr.-Ing. Olivier Reinertz

RWTH Aachen University, Institute for Fluid Power Drives and Systems (ifas), olivier.reinertz@ifas.rwth-aachen.de

Abstract

State of the art compressed air reservoirs are designed in form of isochoric cavities. This implies that all air mass changes, either due to mass demand from the consumers or delivered air from compressors, and temperature changes lead to inherent pressure changes. Therefore, huge reservoir volumes are installed, which necessitate regular and expensive inspection works, and/or significant pressure fluctuations are tolerated, which require inefficient compressor operation at higher supply pressures.

This paper focuses on alternative compressed air storage systems enabling significantly higher capacity to volume ratios compared to conventional reservoirs. The analytically investigated concepts rely on air storage in different sorts of customized volume variable chambers acting against application-optimized arrangements of (hydro-)mechanical loading systems. The resulting benefits and weaknesses of the different variants are discussed with respect to their operational behavior, building space, as well as the requirement for recurring inspections. The paper concludes with an assessment of the required steps for realization of the best suited concept and the opportunities which arise from industrializing the investigated storage systems.

Keywords. Compressed Air Reservoir, Isobaric Storage.

1. INTRODUCTION

1.1. Usage and dimensioning of state-of-the-art reservoirs

Pneumatic reservoirs are widely used in industrial and mobile applications. In literature, two different usage scenarios are commonly distinguished: Firstly, application downstream of air compressors to stabilize network pressure for quasi continuous air demand, especially in industrial applications, and secondly, local application in pneumatic machines and networks to prevent pressure drops during short-term increased air demands.

In the first case, main task of the reservoir is to allow the pressure of the air distributing network to stay in a predefined margin despite fluctuating air supply and demand. For simple compressor controls, which only allow for switching selected compressors on and off, a

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

minimally required cycle time for switching of the compressors usually exists to prevent the machines from inefficient operation, overheating, and excessive wear.

Varying sizing rules for the required reservoir volume can be found in different handbooks on the design of compressed air systems, e.g. in [1, 2]. These are based on the ideal gas law and a mass balance of inflowing and outflowing air of the reservoir. While the equation in [1] allows sizing for a defined compressor and specific air demand, the worst case of a 50% duty cycle is considered in most cases, as usually, the air demand of a production plants is time variant, and the system has to operate correctly under all circumstances. The equation in [2] includes this assumption of a 50% duty cycle and additionally allows to take varying inlet pressures and temperatures of the compressor and reservoir temperatures into account. By setting these values to standard conditions, both equations are identical and equal equation 1.1. The sizing of the reservoir volume V_{res} is given as a function of the allowable pressure fluctuation Δp and cycle time t_{cycle}, with the pressure at standard conditions p_0, and the volumetric flow of the biggest installed compressor at standard conditions Q_{Comp}.

$$V_{res} = \frac{p_0}{4\,\Delta p} \cdot Q_{comp} \cdot t_{cycle} \tag{1.1}$$

An alternative sizing approach is used, when pneumatic reservoirs are implemented to allow for peak flow Q_{peak} delivery over a defined (short) time Δt [1].

$$V_{res} = \frac{p_0}{\Delta p} \cdot Q_{peak} \cdot \Delta t \tag{1.2}$$

In both cases, pressure fluctuations occur, and the system pressure needs to be selected to allow for correct function of all systems downstream the reservoirs under all operational conditions. In general, this means that the maximum outlet pressure of the compressor must be higher than the required pressure in the application by the sum of all permissible pressure fluctuations of the reservoirs connected in series and the line losses. The impact of this additional pressure is discussed in the next subchapter.

1.2. Influence of pressure margins on system efficiency

It can be assumed as a rule of thumb that well-dimensioned and well-maintained compressed air stations require approximately 145 % of the energy that would be necessary for isentropic compression to generate compressed air [3]. Hence, an estimate for the mass specific energy demand w of a compressed air station, which delivers the pressure p_{out} from ambient air with pressure p_{in} and temperature T_{in} is given by equation 1.3, wherein c_p denotes the isobaric heat capacity and κ the isentropic number [4].

$$w(p_{out}) = 1.45 \cdot c_p \cdot T_{in} \cdot \left[\left(\frac{p_{out}}{p_{in}}\right)^{\frac{\kappa-1}{\kappa}} - 1\right] \tag{1.3}$$

For the simplified case of a compressor with on/off control, which delivers, when switched on, a pressure independent mass-flow into a downstream reservoir with time invariant outflow, the cyclic pressure increases and decreases in the reservoir are linear over time. The average energy demand of the compressor $\overline{w}$ at a pressure between $p_{out} = p_{req}$ and $p_{out} = p_{req} + \Delta p$ can be calculated by equation 1.4.

$$\overline{w} = \frac{\int_{p_{req}}^{p_{req}+\Delta p} w(p)\ dp}{\Delta p} \tag{1.4}$$

The following figure shows the impact of the pressure control band on specific energy demand for compressed air delivery. It becomes evident that several percent in efficiency can be gained simply by reducing the pressure control band, which is aspired by actual more complex compressor control schemes.

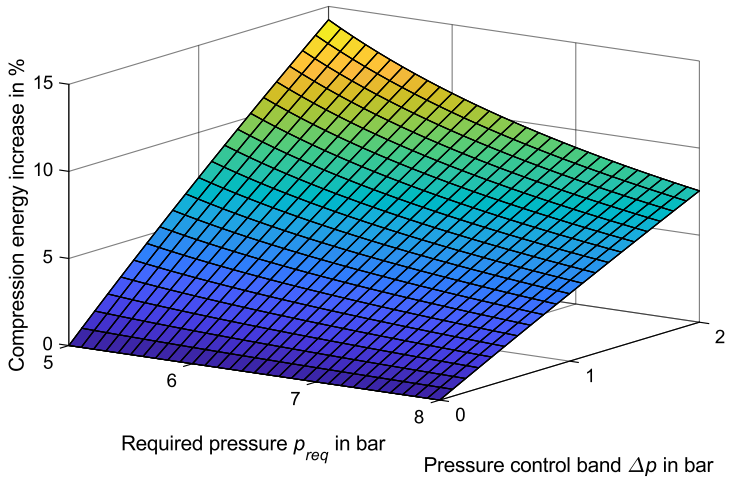


Figure 1.1. Influence of pressure control margin on energy demand for compressed air generation

Despite the derivation of the preceding equation under the assumption of a constant reservoir outflow, the above stated results are generally applicable to all kinds of systems with time-variant compressed air demand as long as no timewise coupling of reservoir outflow and compressor cycles exists, which is rarely the case in industrial systems.

1.3. Normative demands on certification and inspection of conventional reservoirs

The stored energy and associated risks of compressed air reservoirs grow with pressure and volume. Therefore, reservoir classification by means of the product of rated pressure and volume is used to specify design, manufacturing, documentation, commissioning, and recurring inspection procedures and requirements. Figure 1.2 depicts the categorization of reservoirs according to Directive 2014/68/EU by reservoir volume V and rated pressure PS. The categories are numbered in order of increasing demand in all aforementioned areas. [5]

Additionally, directive 2014/29/EU handles the making available on the market of simple pressure vessels, which cover welded pneumatic reservoirs manufactured in series for stationary use up to a PS·V product of the vessel of 10,000 bar·l and the resulting obligations of manufacturers, importers, and distributors. [6]

In the scope of this paper, the impact of quasi-isobaric reservoirs on system integrator and machine user duties are focused. In Germany, the required inspection tasks and intervals at commissioning and during the use phase are regulated in Annex 2, Table 4 of the *Betriebssicherheitsverordnung* (Operational Safety Regulation) [7]. Therein, it is distinguished between internal inspections, which require the reservoir to be depressurized and accessible, and strength tests, which require all safety valves to be dismounted and the

reservoir to be entirely filled with water for pressure and leak tightness testing. Obviously, such tests are time consuming and thus expensive.

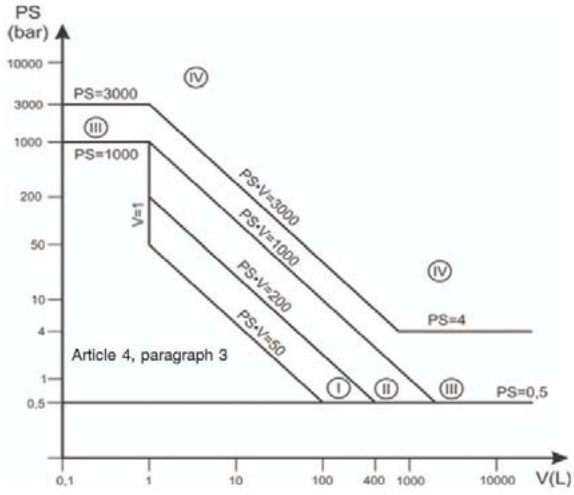


Figure 1.2. Classification of compressed air reservoirs according to [5]

The required inspection tasks for a specific reservoir depend on the categories of Directive 2014/68/EU, with the sole exception that reservoirs, which are designed to store internal relative pressures of less than 1 bar, are exempt from inspections regardless of their size. Table 1 provides an overview of mandatory inspection tasks during commissioning and in operation. Herein, it is distinguished between the inspection by notified bodies (NB) and user inspectorates (UI).

Category	**Commissioning**	**Periodic internal inspection**	**Periodic strength test**
Article 4, paragr. 3	-		
I	UI	UI (all 10 years)	UI (all 10…15 years)
II	NB		
III & IV		NB (all 5 years)	NB (all 10 years)

Table 1: Inspection duties of compressed air reservoirs

It becomes obvious that sizes, which are covered by article 4, paragraph 3 are very favorable, as no inspection duties result. As the product of pressure and volume for this category is limited to 50 bar·l, only small reservoirs fulfill this requirement, e.g. for a rated pressure of 16 bar, reservoir size is limited to 3.125 l. In case of higher required volumes, categories I and II at least do not require an external supplier for inspection. Nevertheless, training the user inspectorate and shutdown of the reservoir are required for inspection, still resulting in significant costs. For reference, pneumatic reservoirs with up to 12.5 l and 16 bar pressure ratings are covered by category I, while category II allows for up to 62.5 l.

1.4. Dimensioning of quasi-isobaric reservoirs

Quasi-isobaric storage of compressed air demands for a volume variable reservoir. Therewith, pressure fluctuations can be minimized, leading to higher efficiency of the compressed air supply or better system performance, and the installed reservoir volume can be drastically reduced. The volume change ΔV_{QIB} of a quasi-isobaric reservoir with comparable operational behavior to a conventional reservoir with Volume V_{res} is given by equation 1.5.

$$\Delta V_{QIB} = V_{res} \cdot \frac{\Delta p}{p_{req}} \tag{1.5}$$

Especially for small allowable pressure changes, huge differences in volume occur. If a quasi-isobaric reservoir is used for optimization of compressor cycling, equation 1.5 can be inserted into equation 1.1 and the variable reservoir volume in equation 1.6 results.

$$\Delta V_{QIB} = \frac{1}{4} \cdot \frac{p_0}{p_{req}} \cdot Q_{comp} \cdot t_{cycle} \tag{1.6}$$

For peak flow demands, equation 1.2 and 1.5 are combined and lead to equation 1.7.

$$\Delta V_{QIB} = \frac{p_0}{p_{req}} \cdot Q_{peak} \cdot \Delta t \tag{1.7}$$

1.5. Hysteresis losses on volume variable cavities

Several designs of volume variable cavities are used in actual pneumatic actuators. While cylinders with sliding seals are the common solution, also diaphragm arrangements exist. Seal friction or plasticity of sealing diaphragms leads to hysteresis of such actuators, which ultimately results in pressure deviations in the investigated reservoirs between filling and discharge. Moreover, the breakaway force of friction loaded sealing systems is especially relevant for the application, as it leads to dynamic excitation.

Multiple investigations demonstrate the strong impact of pressure and standstill time on the breakaway force of pneumatic cylinders [8, 9]. With the assumption of rare and short standstill periods of the accumulator piston in operation, a standstill time of less than 10 s can be assumed. In this case, *Meuser's* measurements on a cylinder with 25 mm diameter show a breakaway force per seal length of up to F_{FrL} = 0.28 N/mm at a differential pressure of 6 bar [8]. Slightly lower values have been measured by *Wakasawa et al.* for a cylinder with 32 mm in diameter [9]. Equation 1.8 estimates the friction force induced pressure hysteresis Δp_{Fr} of a cylinder based quasi-isobaric reservoir with piston diameter D.

$$\Delta p_{Fr} = 2 \frac{D \cdot \pi \cdot F_{FrL}}{D^2 \cdot \frac{\pi}{4}} = 8 \frac{F_{FrL}}{D} \tag{1.8}$$

The reducing effect of seal friction with increasing diameter is based on the quadratic increase in pressure loaded surface counteracting the only linearly increasing sealing length. Based on *Meuser's* data, for a piston diameter of 200 mm, a hysteresis in pressure of Δp_{Fr} = 112 mbar between filling and discharge occurs.

Even more relevant for the application is the reduction in friction force over speed at small velocities and the change in friction force between acceleration and deceleration of the piston in the mixed lubrication regime, which lead to stick-slip behavior and corresponding velocity and pressure oscillations. According to measurements from *Fujita et al.*, the friction

force at deceleration to standstill can be as low as half of the breakaway force [10]. Therewith, pressure fluctuates during charging and discharging a quasi-isobaric piston reservoir by Δp_{Fr} / 4, which equals ca. 28 mbar for the before stated example of the 200 mm piston, as shown in figure 1.3.

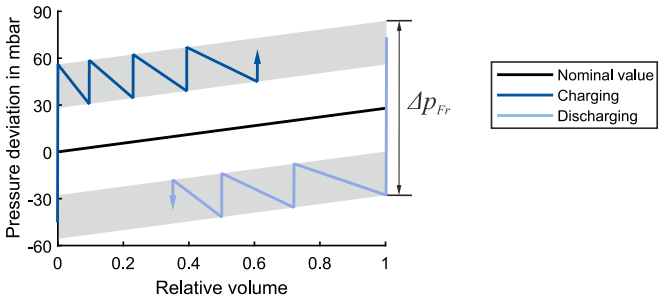


Figure 1.3. Schematic pressure cycle of a seal friction subjected quasi-isobaric reservoir

By using diaphragms for sealing of the quasi-isobaric reservoir, stick-slip effects and pressure deviations can be significantly reduced. Nevertheless, deformation of the elastomeric material with its visco-elastic material properties and possible friction between the deforming diaphragm and housing still contribute to some deviations in pressure.

A quantification of these effects in terms of pressure deviations in the aspired application is hindered by the strong impact of material selection and geometric design of the diaphragm. Therefore, only a rough estimate from comparable devices in literature can be achieved. *Bedarff* models in his PhD-thesis an active air spring for passenger cars. For model validation, a cyclic motion at constant pressure is experimentally investigated, which corresponds exactly to the application in a quasi-isobaric reservoir. The hysteresis amplitude in force amounts to ca. 0.6 % of the nominal value at an outer diameter of ca. 150 mm [11]. Hence, by using a rolling diaphragm, pressure hysteresis of the quasi-isobaric reservoir can be reduced by half in comparison to a conventional cylinder if operated at 6 bar_{rel}. Also, in the case of diaphragms, the use of higher diameters leads to reduced hysteresis.

1.6. Intermediate conclusion – Potential of quasi-isobaric reservoirs

Increasing efficiency of industrial compressed air systems requires lower pressure control bands to be realized. Thereof, the problem arises, that conventional reservoirs in compressor stations have reduced effectiveness in storing and providing compressed air to the systems, which can only be counteracted by raising reservoir size. Usage of quasi-isobaric reservoirs decouples the pressure control band from reservoir size and allows significant reduction in reservoir volume in comparison to actual design rules. Additionally, using quasi-isobaric reservoirs enables downsizing pressure loaded volumes, which can have significant effect on commissioning and inspection by the classification in lower categories according to Directive 2014/68/EU. Despite seal friction and material losses of diaphragms hinder true isobaric behavior, increasing the piston or diaphragm diameter reduces these effects to a reasonable value, thus allowing for quasi-isobaric systems.

2. CONCEPT DEVELOPMENT AND ANALYSIS

Concepts, which allow quasi-isobaric conditions in pneumatic reservoirs, require the reservoir volume to increase drastically with raising pressure in the aspired operating point. Meanwhile, the mechanical work performed by the moving walls of the reservoir under pressure needs to be stored, to allow for later deflation of the reservoir against system pressure. Out of this simple system description, the system structure for all kinds of isobaric reservoir concepts gets obvious:

Grid ↔ Variable volume with force takeoff ↔ Energy transmission ↔ Energy storage

Thereof, different concepts can be deduced and evaluated. The following table shows an overview of all possible solutions investigated in this paper.

Concept	Reservoir volume	Transmission	Energy storage
1	Pneumatic cylinder	---	Lifting of a weight
2	Pneumatic cylinder	---	Mechanical spring
3	Pneumatic cylinder	Hydraulic cylinder	Hydraulic accumulator
4	Nonlinear actuator	---	Mechanical spring

Table 2: Investigated concepts

In the following subchapters, the different concepts are presented and discussed in terms of filling state induced pressure deviations, adjustability in pressure, and building space. These findings are the basis for concept comparison and decision in the subsequent chapter 3.

2.1. Concept 1: Weight loaded pneumatic cylinder

The simplest concept of a constant pressure reservoir for pneumatic systems consists of a volume variable pneumatic chamber, e.g. a cylinder or bellow, which is loaded by a significant weight. Such arrangements can be found e.g. in the air supply of pipe organs to allow for constant pressure in the pneumatic system [12]. Also in process technology, e.g. for storage of biogas for power generation, comparable arrangements are used with huge storage volumes [13]. Both systems have in common, that the load is integrated into the reservoir piston. This approach is not feasible for typical pressures of pneumatic systems, as even for a steel weight with a density of approx. 7,800 kg/m³, a required pressure of 6 bar_{rel} would result in a height of the weight of approx. 7.8 m. Therefore, an alternative configuration as shown in figure 2.1 is further investigated. It is characterized by a weight, which surrounds a pneumatic cylinder with stroke dependent volume $V_{QIB}(s)$, thus enabling long strokes and compact design. The maximum displacement volume of the reservoir ΔV_{QIB} hence equals the difference between $V_{QIB}(s_{max})$ and $V_{QIB}(s = 0)$ with s_{max} beeing the stroke length. For this case, the width a of the weight with quadratic cross-section and length l can be estimated by equation 2.1 under the simplifying assumption, that the length of the weight l and the cylinder length s_{max} are equal. For given pneumatic parameters, weight density ρ, and gravitational constant g, the length l is the only remaining optimization parameter. For a constant displacement, the assembly gets significantly thinner and lighter with increasing stroke length while material demand and building space are reduced. On the other hand, higher strokes and smaller piston diameters lead to increased velocities, friction, and wear on the piston sealings.

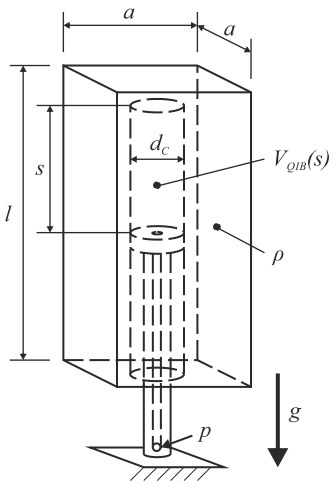


Figure 2.1. Investigated weight loaded cylinder

Equation 2.2 gives a rough approximation of the building space $V_{BS,1}$ required.

$$a = \sqrt{\Delta V_{QIB} \cdot \frac{p - p_0 + \rho g l}{\rho g l^2}} \tag{2.1}$$

$$V_{BS,1} = 2a^2 \cdot l \approx 2 \cdot \Delta V_{QIB} \cdot \frac{p - p_0}{\rho g l} \tag{2.2}$$

Regarding system dynamics, the proposed system allows to deliver constant pressure over the stroke in steady state but represents a dynamic first-order system with very low damping and small eigenfrequency. The undamped eigenfrequency in the highest position, which is the lowest eigenfrequency in operation, can be calculated by equation 2.3 and results directly from the stroke length of the system for given operational parameters [4]. For strokes of 1 to 10 m and usual pneumatic parameters, natural frequencies of the system are in the range of 0.15 to 0.7 Hz. These values are further reduced by the capacity of the connected pipework in an application.

$$f_0 = \frac{1}{2\pi} \cdot \sqrt{\frac{g}{l}} \cdot \sqrt{\frac{n \cdot p}{p - p_0}} \tag{2.3}$$

The usability of passive damping measures to optimize oscillation stability is limited, as these cause speed and thus flow dependent deviations from isobaric conditions. Therefore, weight loaded isobaric reservoirs are only applicable to systems with significantly lower dynamics than their own, which can be the case for central reservoirs in compressed air supply. An application for smoothing dynamic load peaks is not feasible. Furthermore, the huge weight of the system prevents usage in mobile compressors.

2.2. Concept 2: Pneumatic cylinder with mechanical springs

An alternative to the before-mentioned weight loading of the piston is the use of a pretensioned mechanical spring. This enables significantly higher eigenfrequencies due to the strong reduction in moving masses. The downside of this approach is that it leads to increasing pressure with reservoir filling due to the spring stiffness. In analogy to

conventional reservoirs, this pressure fluctuation needs to stay in the allowable pressure margin of the application, which limits the usable spring stiffness. As the pressure margin is usually small in comparison to the absolute pressure, for conventional helical springs, enormous tensioning strokes result. These make the spring assembly with its tensioning mechanism very complex and bulky and tend to fatigue issues on the springs.

Disc springs with their varying stiffness over the stroke can be used for size and complexity reduction. The dimensions of disk springs can be tuned to allow for high stiffness at stroke begin, which is used to generate sufficient pretension force, while the force gets mostly stroke independent over a second stroke regime, which is the working stroke of the quasi-isobaric reservoir. Figure 2.2 shows the force of such an optimized disc spring over its stroke on the right-hand side, based on the force equations in [14]. This optimization is achieved by means of sheet thickness *t* being 0.707 times the height of the truncated cone *h*, which is formed by the disc in unloaded state. In the diagram, the stroke is scaled to the height *h*, while the force on its value at constant stiffness.

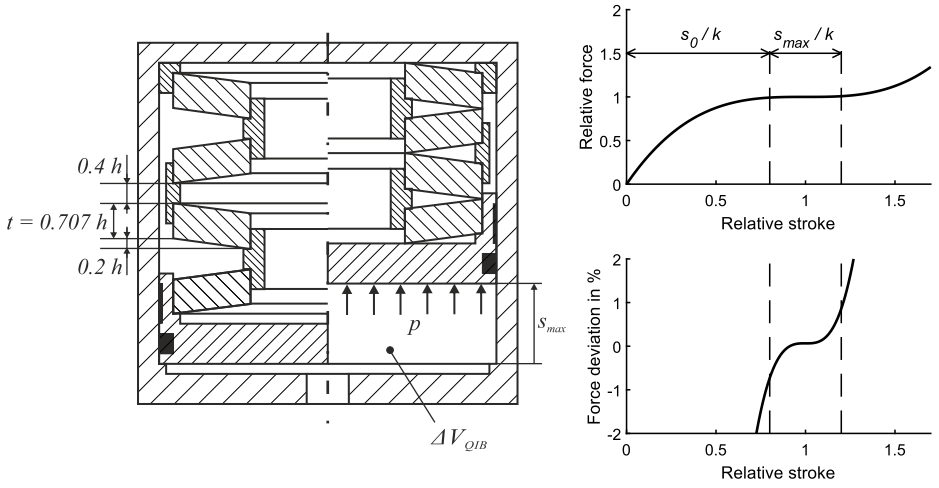


Figure 2.2. Disk spring reservoir concept and force characteristics

For disc spring compressions between 80 % and 120 % of *h*, the force stays constant with an accuracy of better than ±1 %. Thereof, it can be concluded that if using *k* disc springs, the pretension stroke s_0 of the spring assembly has twice the size of the piston stroke s_{max}. Meanwhile, very small pressure variations are ensured, which is a huge optimization in comparison to piston loading by constant stiffness helical springs.

The building space of the disk spring loaded reservoir $V_{BS,2}$ can be approximated from the volume change in operation ΔV_{QIB} by scaling it with the pretensioned length of the spring assembly divided by the stroke length s_{max}, as shown in equation 2.4. The numbers result from the schematics in figure 2.2, left and the disk spring design for constant force behavior.

$$V_{BS,2} \approx \frac{(0.4 + 0.2 + 0.707) \cdot h \cdot k}{0.4 \cdot h \cdot k} \cdot \Delta V_{QIB} \approx 3.3 \cdot \Delta V_{QIB} \qquad (2.4)$$

However, the design and fabrication of the disc springs need to be very exact, as small tolerances in the design parameters, e.g. in sheet thickness, can lead to decreasing force over the stroke of the disc spring in the operational regime. This will ultimately result in unstable behavior of the arrangement, as pressure drops with increasing filling state. Furthermore, such tolerances can lead to a deviation in operating pressure of the reservoir, as the generated force in the constant force regime of the disc springs deviates from its design value.

A further drawback of the design consists in the inability to set pressure adjustments. By changing pretension, stroke regimes with significant pressure increase over the stroke are part of the operational range and the quasi-isobaric properties are lost.

2.3. Concept 3: Hydro-pneumatic reservoir

The third concept is based on a hydraulically loaded pneumatic piston. A significant reduction in building space is aspired by increasing the internal pressure of the hydraulic system in comparison to the pneumatic side and therewith its energy density. In contrast to the hydraulic bootstrap accumulator, discussed e.g. by *Ketelsen at al.* [15], the hydraulic loading is deemed to be passive. Hence, a hydraulic accumulator, which loads a hydraulic piston to act against the pneumatic cavity is investigated. The precharge level of the accumulator results in sufficient air pressure, while its size and therewith its capacity allows control of the pressure increase over the stroke. Figure 2.3 shows a schematic representation of the investigated concept. Compressed air is marked in blue, hydraulic fluid in orange.

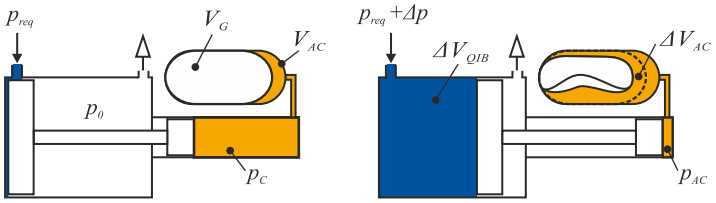


Figure 2.3. Concept of the hydro-pneumatic reservoir; left: discharged, right: charged

As a strongly simplified thermodynamic model of the hydraulic accumulator, a polytropic change of state with polytropic number n is assumed. When the pneumatic side is loaded, the liquid volume ΔV_{AC}, which is pressed by the amplifier piston into the hydraulic accumulator leads to a decrease in its gas volume and thus to a pressure increase. The accumulator pressure p_{AC} is described by equation 2.7 with the precharge pressure of the gas side p_C, and the gas volume of the accumulator V_G, which is present if the pistons stand at their left end stop. This volume is used as a reference for the precharge pressure adjustment.

$$p_{AC} = p_C \left(\frac{V_G}{V_G - \Delta V_{AC}}\right)^n \tag{2.7}$$

The piston amplifier acts as a transformer with area ratio α between the hydraulic and pneumatic system. Herein area ratio $\alpha << 1$ describes the quotient of the hydraulic to the pneumatic piston surface. The force equilibrium and energy balance lead to equations 2.8 and 2.9, which both describe the energy transmission through the amplifier piston:

$$p_C = \frac{p_{req} - p_0\,(1-\alpha)}{\alpha} \approx \frac{p_{req} - p_0}{\alpha} \tag{2.8}$$

$$\Delta V_{AC} = \Delta V_{QIB} \cdot \alpha \tag{2.9}$$

Thereof, for a given required system pressure, pressure margin, and required air volume of the quasi-isobaric accumulator, the following equation 2.10 defines the accumulator dimension, as the nominal accumulator volume V_{AC} should be slightly bigger than the calculated gas volume in reference position.

$$V_{AC} > V_G \approx \frac{\Delta V_{QIB} \cdot \alpha}{1-\sqrt[n]{\frac{p_{req}-p_0}{p_{req}+\Delta p-p_0}}} \tag{2.10}$$

The resulting building space $V_{BS,3}$ can be roughly approximated by the sum of the pneumatic and hydraulic displacement of the piston amplifier and the volume of the hydraulic accumulator, as shown in equation 2.11.

$$V_{BS,3} = \Delta V_{QIB}\left(1+\alpha+\frac{\alpha}{1-\sqrt[n]{\frac{p_{req}-p_0}{p_{req}+\Delta p-p_0}}}\right) \tag{2.11}$$

For an isothermal accumulator, this simplifies to equation 2.12. It is obvious that for small area ratios of the amplifier, relevant reductions in space occur.

$$V_{BS,3} = \Delta V_{QIB}\left(1+2\alpha+\alpha\frac{p_{req}-p_0}{\Delta p}\right) \tag{2.12}$$

A further benefit of hydro-pneumatic systems is the ability to simply adjust the pretension and thus working pressure of the system by the precharge pressure of the accumulator. This allows for pressure adjustments on a running system, e.g. to incrementally reduce operation pressure of a compressed air network for energetic optimization.

2.4. Concept 4: Nonlinear pneumatic actuator with mechanical spring

A further solution for quasi-isobaric behavior can be achieved by using unconventional pneumatic actuators for the realization of the reservoir volume with increasing active area over the stroke. As has been shown in [16] actual designs for pneumatic actuators realize either reducing forces or constant forces over the stroke at constant pressure. But, by implementing novel actuator design schemes, as shown e.g. in [17] for machine suspension, rising force over the stroke arrangements can be achieved. In the case of quasi-isobaric reservoirs, the use of such actuators as reservoir volume has the big advantage that the force characteristics of the reservoir volume can be tailored to the force stroke behavior of the loading spring. Therefore, in the following, the direct coupling of an unconventional actuator with mechanical springs, as shown in figure 2.4 is investigated.

For the quadratic correlation between volume *V(s)* and stroke *s* of the reservoir in equation 2.13, the force *F* on the loading device is given in equation 2.14. The force balance with a loading spring with stiffness *C* and pretension stroke s_0 results in equation 2.15, which describes the reservoir pressure as a function of the stroke.

$$V(s) = as^2 + bs + V_0 \tag{2.13}$$

$$F(s,p) = \frac{\partial V}{\partial s}\cdot(p(s) - p_0) = (2as+b)\cdot(p(s) - p_0) \tag{2.14}$$

$$p(s) = p_0 + \frac{C}{2a}\cdot\frac{s+s_0}{s+\frac{b}{2a}} \tag{2.15}$$

For a given volume stroke characteristics with parameters a, b, and V_0 and spring stiffness C, a distinct pretension stroke exists, which allows the system pressure to stay constant over the entire volume change of the reservoir. This pressure is achieved with a pretension stroke of $s_0 = b/(2a)$. Moreover, especially for small values of b, and hence small forces at stroke begin, pretension of the springs can be very low and almost the entire working stroke of the springs can be used for energy storage, which makes this arrangement particularly material efficient regarding the mechanical spring.

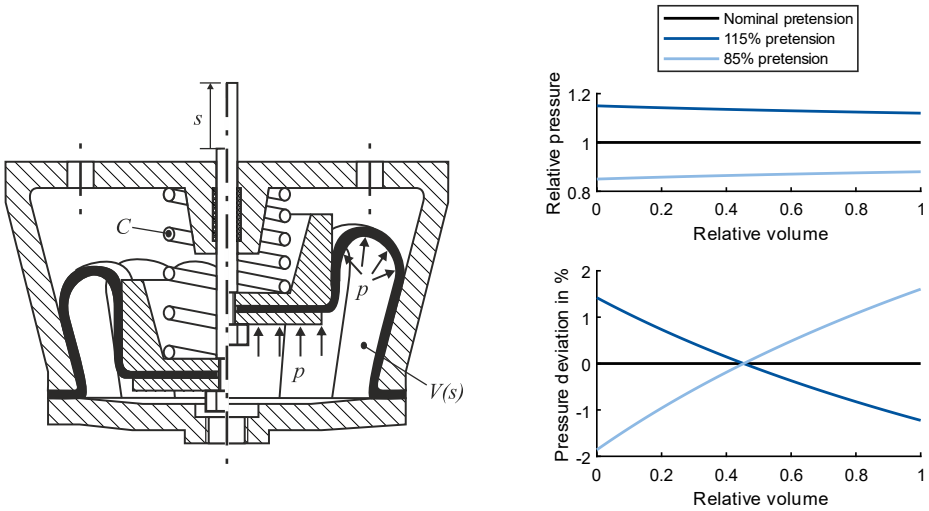


Figure 2.4. Investigated nonlinear pneumatic actuator with mechanical spring and its force deviation over the stroke for $b = 8\ a\ s_{max}$

On the other hand, if pressure adjustments by changing spring pretension are considered, the pretension stroke should be maximized to allow for small deviations in pressure over the stroke and thus stored volume. To achieve a pressure fluctuation over the stroke of less than ± 2 % at a pressure adjustment range of ± 15 %, the pretension stroke of the spring must be equal or higher than four times the piston stroke s_{max}. Hence, significant spring length and pretension stroke are required to allow for adjustability.

The change in pressure over the filling state is shown in figure 2.4 for the nominal pressure and deviating pressure settings by ± 15 %. The loss of the constant pressure over the stroke operation is clearly visible at varying pretension. Moreover, lower pressure settings can cause the system to have negative fluidic impedance, and thus to be unstable if connected to sufficient grid capacity. The impact of the slight negative impedance in the depicted case on system stability is to be further analyzed. Due to its small amount, it can probably be stabilized by sufficient viscose friction of the arrangement.

Finally, also this design has the major advantage in that it is very compact. Its required building space is estimated from figure 2.4 to be approximately four times its displacement volume, as shown in equation 2.16.

$$V_{BS,4} \approx 4 \cdot \Delta V_{QIB} = 4 \cdot (V(s_{max}) - V_0) \tag{2.16}$$

3. CONCEPT EVALUATION

First, the different concepts are compared by means of their general characteristics, as shown in table 3.

The weight loaded concept 1 has its only major benefit in being simple in realization. The required small piston diameter and long stroke and the additional guiding elements to prevent the system from tilting lead to high hysteresis and seal wear and poor dynamics.

The disc spring loaded concept 2 has advantageous dynamics, higher piston diameter, and small pressure increase over the stroke but lacks adjustability while being difficult in maintenance due to the long pretension stroke and high stored energy of the springs. Also, realization has high demands on accuracy of the disc springs to allow all sequential springs of the system to work exactly in the required working regime.

The hydro-pneumatic concept 3 is very advantageous in adjustability and dynamics, while complicating maintenance by requiring additional seals, a pressure medium, a system for flushing the fluid and for filling the gas into the accumulator, and multiple hydraulic components.

Lastly, the spring-loaded nonlinear actuator concept 4 is beneficial in all categories despite its simplicity in implementation, as this new kind of actuator first needs to be developed. Maintenance therefore is difficult to evaluate, as it is strongly design dependent.

	Concept	1	2	3	4
Criterion	Adjustability	+	- -	+ +	+
	Small hysteresis	- -	+	-	+ +
	Dynamics	- -	+ +	+ +	+ +
	Simplicity	+ +	0	+	- -
	Maintenance	0	0	-	0

Table 3: General concept comparison (Rating from poor to good: -- / - / 0 / + / ++)

In the following, the concepts are further investigated by sizing and comparing the capacity of exemplary solutions for two different applications. First, the concepts are evaluated for use in compressed air supply, with huge volume changes at lower speed. Subsequently, decentralized implementation to smoothen peak flow demands and associated pressure drops is evaluated.

3.1. Usage in compressed air supply

Reservoirs in compressed air supply are used to smoothen the grid pressure and therewith to allow the compressor control to react on differences between compressed air generation and consumption. Meanwhile, it is aspired to keep the reservoir pressure as constant as possible slightly above the required system pressure, which should be adjustable in a defined margin. By means of the quasi-isobaric reservoir, inspection duties can be minimized, which means, that in best case, a category II classification ($PS \cdot V < 1000$ bar·l) is targeted.

To compare the different concepts, their realizable displacement volumes as a function of the allowed pressure margin are calculated and depicted in figure 3.1, left. A required pressure of 7 bar_{abs} and a building space with a ground surface of 1 m² and 2 m in height are

assumed. The main components and their arrangement were sized for all concepts to fit in the building space and to the allowed pressure margin. From this figure, several conclusions can be drawn:

The weight loaded system consists of one or several parallel arranged cylinders with a stroke of 1 m to fit in the height limitation and a lumped piston surface of ca. 0.1275 m². The weight is assumed to be fabricated from steel and has a mass of ca. 7.8 tons. This makes the system impractical while its displacement is rather limited compared to the other systems. The disc spring loaded system has the highest capacity at small pressure margins but does not allow any correction in system pressure. Therefore, it is also not well suited for the application.

The hydro-pneumatic reservoir achieves the highest capacity at bigger pressure margins. For smaller values, its displacement capability increases with pressure margin. In this parameter region, the restrictive parameters are the limited hydraulic accumulator size, which was set to 300 l, and its assumed maximum pressure of 300 bar. For higher pressure margins, the displacement volume of the compressed air piston limits the achievable effect, as a displacement of 1 m³ cannot be overcome in the building space under the assumption of an air piston with coaxial hydraulic loading piston and circumferential hydraulic accumulators.

By using a spring-loaded nonlinear actuator, the second-best usage of the building space at small pressure margins results. Very likely this concept has the best weight to displacement ratio, as it does not rely on heavy disk springs, a weight, or heavy hydraulic components.

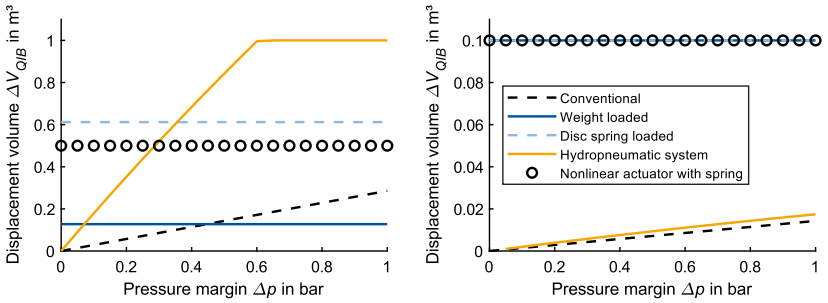


Figure 3.1. Comparison of quasi-isobaric concepts with a conventional reservoir, left: for 2 m³ build space, right: for up to 2 m³ build space and category II classification

If it is required to fulfill category II to allow for longer inspection intervals and to circumvent regular inspections by notified bodies, the rating of the different concepts changes drastically. To calculate the allowable dimensions for category II, the rated pressure of the pneumatic cavity and in case of the hydro-pneumatic system also of the hydraulic accumulator need to be defined. These were set to 10 bar_{rel} for the pneumatic and 300 bar for the hydraulic parts. The results are shown in figure 3.1, right.

To allow for category II classification, the conventional reservoir needs to stay in the PS·V = 1000 bar·l margin and diminishes in size to 100 l. Thus, it can only use a small part of the available building space, and its effectiveness drastically reduces. The cylinder chambers of the weight loaded piston, the disc spring loaded piston, and the nonlinear actuator need also to stay within this margin, which reduces their displacement capacity ΔV_{QIB} to 100 l.

The hydro-pneumatic system in contrast strongly loses in performance, when built to fit into category II. This is because, in all investigated configurations, the hydraulic accumulator needs to be strongly downsized to a nominal volume of only 3.33 l to reach category II. This significantly diminishes its feasible volumetric change of air capacity.

Finally, the missing adjustability in pressure of the disc spring device and the huge mass of the weight loaded system with its associated issues in implementation and dynamics hinder their industrial usage. Therefore, it can be concluded that for practical usage, the nonlinear actuator with spring is the most promising solution despite its required development effort.

3.2. Usage for peak demand

Usually, reservoirs for peak demand supply are smaller sized. Therefore, it is aspired to reduce the pressurized volumes to the point that an article 4, paragraph 3 classification is achieved in all cases. Meanwhile, a maximum ground surface of 0.01 m² and height of 0.5 m are assumed, which enable the conventional reservoir to use the entire building space and categorization limit in parallel. The resulting displacement volumes are shown in figure 3.2.

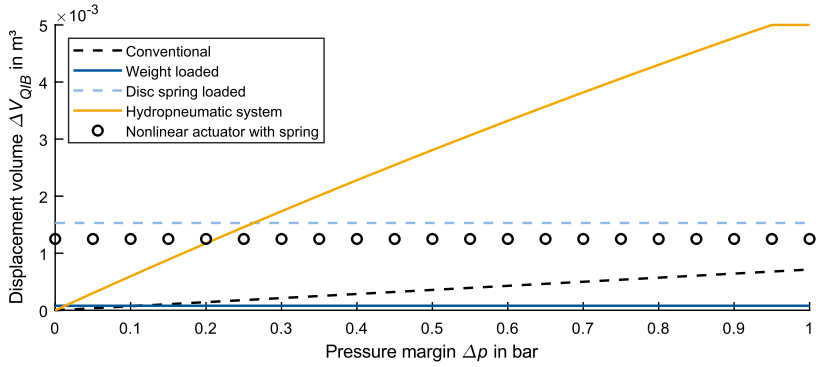


Figure 3.2. Comparison of quasi-isobaric concepts with a conventional reservoir in accordance with article 4, paragraph 3

While all kinds of quasi-isobaric reservoirs can reach a displacement volume of 5 l while falling under article 4, paragraph 3 with sufficient building space, the additional restriction in available space leads to a clear differentiation between the different designs. The hydro-pneumatic system is best suited for higher pressure margins, while the disc spring loaded cylinder and the spring-loaded nonlinear actuator are best suited for small pressure changes. In general, these three kinds of quasi-isobaric reservoirs allow for significant optimization in comparison to regular designs, while the disc spring loaded device lacks adjustability.

4. Conclusion and Outlook

Quasi-isobaric reservoirs have the potential to decrease the building space requirements in comparison to state-of-the-art designs in combination with their demands for regular inspections. Furthermore, the increased capacity of quasi-isobaric reservoirs at low pressure margins can be used to reduce pressure control bands of actual compressor stations to gain several percent in efficiency with corresponding cost savings in comparison to actual solutions. Different concepts for such reservoirs have been developed and compared by means of two different applications. Thereof, it can be concluded that the spring-loaded

nonlinear actuator has the biggest potential of all designs investigated, especially for small pressure margins. In contrast, hydro-pneumatic systems have the biggest potential for higher pressure amplitudes at small scales. In future research, these findings should be experimentally validated. While the hydro-pneumatic system can be built from off-the-shelf components, the nonlinear actuator itself is part of actual research. Furthermore, system and compressor control integration of the novel reservoirs need to be investigated.

5. REFERENCES

[1] Boge, 'Compressed Air Compendium', 6th edition, Hoppenstedt, Darmstadt, 2004, ISBN: 3-935772-12-2.

[2] Atlas Copco, 'Compressed Air Manual', 9th edition, 2019.

[3] N.N., 'Druckluft Effizient - Thermodynamik', www.druckluft-effizient.de, 2003.

[4] H. Murrenhoff, O. Reinertz, 'Fundamentals of Fluid Power – Part 2: Pneumatics', 1st edition 2014, Shaker, Aachen, ISBN: 978-3-8440-3213-0.

[5] N.N., 'Directive 2014/68/EU of the European Parliament and of the Council of 15 May 2014 on the harmonization of the laws of the Member States relating to the making available on the market of pressure equipment', 2014.

[6] N.N., 'Directive 2014/29/EU of the European Parliament and of the Council of 26 February 2014 on the harmonization of the laws of the Member States relating to the making available on the market of simple pressure vessels', 2014.

[7] Berufsgenossenschaft Rohstoffe und chemische Industrie (BG RCI), 'Prüfungen Druckbehälter gefertigt nach RL 2014/68/EU (DGRL)', presentation, 12/2020.

[8] M. Meuser, 'Nichtlineare Regelung pneumatischer Antriebe', Dissertation, Shaker, Aachen, 2010, ISBN: 978-3-8440-0416-8.

[9] Y. Wakasama, Y. Ito, H. Yanada, 'Dynamic Behaviors of Pneumatic Cylinder', JFPS International Journal of Fluid Power System 8-1, 60/65, 2015.

[10] T. Fujita, L- R- Tokashiki, T. Kagawa, 'Stick-Slip Motion in Pneumatic Cylinders Driven by Meter-Out Circuit', Fluíd Power - 4th JFPS International Symposium, 1999, ISBN: 4-931070-04-3.

[11] T. Bedarff, 'Grundlagen der Entwicklung und Untersuchung einer aktiven Luftfeder für Personenkraftwagen', Dissertation, Darmstadt, 2016.

[12] J. Angster, S. Pitsch, A. Miklós, 'Design of new Wind Systems for Pipe Organs', Proceedings of the International Symposium on Music Acoustics (ISMA), Barcelona, Spain, 2007.

[13] Eisenbau Heilbronn, 'Produktpräsentation Gasspeicher', Eisenbau Heilbronn, Heilbronn, 2023.

[14] J. O. Almen, A. Laszlo, 'The Uniform-Section Disk Spring', Transactions of the American society of mechanical engineers, 1936, doi: 10.1115/1.4020233.

[15] S. Ketelsen, D. Padovani, M. K. Ebbesen, T. O. Andersen, L. Schmidt, 'A Gasless Reservoir Solution for Electro-Hydraulic Compact Drives with Two Prime Movers', Proceedings of the BATH/ASME 2020 Symposium on Fluid Power and Motion Control (FPMC), Sept. 2020, doi: 10.1115/FPMC2020-2773.

[16] O. Reinertz, 'Static Properties Based Classification of Pneumatic Actuators - Potential of Progressive Effective Area Over the Stroke', Proceedings of Fluid Power 2025 (FT2025), Slovenia, Sept., 2025, doi: 10.18690/um.fs.7.2025.9.

[17] M. Sprengholz, C. Hühne, 'Zero-stiffness in Rolling-Lobe Air Springs for Passive, Load Adaptable and Low-Frequency Vibration Isolation', Journal of Sound and Vibration 608 (2025) 119061, doi: 10.1016/j.jsv.2025.119061.

Biographies



Prof. Olivier Reinertz received his diploma and his doctoral degree in mechanical engineering from RWTH Aachen University, Germany. He is currently Professor and Scientific Director at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. His research focuses on the model-based analysis of fluid power components and systems and the derivation and validation of innovative strategies for efficiency and performance optimization with an emphasis on compressed air and gas-driven systems.

Integration of Pneumatic Systems in Training Devices for Combat Sports

Željko Šitum, Tin Stanišak and Dominik Šarić

University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture, I. Lučića 5, HR-10000 Zagreb, Croatia, zeljko.situm@fsb.unizg.hr, tin.stanisak@hotmail.com, dominik.saric17@gmail.com

Abstract.

The inherent compressibility of air, characteristic of all pneumatic systems, can serve as an advantageous feature in the development of conceptual solutions for training devices aimed at enhancing motor reactions required in combat sports. These devices typically consist of a supporting structure and movable components equipped with striking surfaces. Combat sports such as boxing and taekwondo have historical roots extending far into the past. Although their roles and cultural significance have evolved over time, these disciplines have gained substantial popularity in recent decades, largely due to their commercial appeal and adaptation into various forms of competitive entertainment for broad audiences. Enhancing specific muscle movements, manifested through attributes such as speed, agility, strength, precision, and endurance, is essential in combat sports and can be facilitated through specialized training devices. Pneumatic systems offer a highly efficient and adaptable platform for developing such devices, owing to their operational speed, reliability, and programmability in motion control. Their simple mechanical design and rapid actuation capabilities make them particularly suitable for sports training scenarios, where high dynamic performance and user safety are critical. Accordingly, the development of a system designed to reduce athletes' reaction times to visual stimuli would be of significant benefit. To achieve this, the pneumatic system must be equipped with appropriately selected actuators, a stable load-bearing structure, and a component arrangement tailored to match the user's physical stature. The training apparatus should also feature adjustable components, enabling vertical repositioning to align with the athlete's height, and allow programming adjustments to modulate the speed and pattern of movements, thereby providing a high degree of application flexibility. This paper presents the conceptual design and practical implementation of two devices developed for training defensive techniques and improving reaction times in martial arts disciplines such as boxing and taekwondo. The movement of the pneumatic actuators is randomized, requiring the practitioner to react quickly and appropriately, either by striking the target surface, as in taekwondo, or by dodging or blocking simulated punches, as in boxing. The system includes pneumatic valves and a control unit capable of generating random activation sequences for the pneumatic cylinders, thereby simulating unpredictable movements typical of combat scenarios. A programmable control algorithm was implemented using a Controllino device, and the

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

developed control software was tested on the fabricated training equipment within sports hall environments.

Keywords. Pneumatic systems, combat sports, reaction time training, martial arts equipment, programmable control.

1. INTRODUCTION

Combat sports, including boxing and taekwondo, require exceptional neuromuscular coordination, speed, precision, and agility. In recent decades, the rising popularity of these disciplines, both as professional sports and recreational activities, has driven the development of advanced training methodologies [1], [2]. A key factor in improving athletic performance in these sports is the enhancement of reaction time, particularly in response to unpredictable stimuli that closely mimic real combat scenarios [3]. Modern training equipment increasingly incorporates technologies such as electromechanical actuators, robotic systems, and sensor-based feedback mechanisms [4]. However, despite their proven effectiveness in industrial automation for tasks demanding speed, precision, and repeatability, pneumatic systems remain underutilized in sports training applications [5]. Pneumatic actuators offer several distinct advantages for athletic training devices: rapid response, smooth motion control, natural impact absorption due to air compressibility, and seamless integration with programmable control systems [6]. This study introduces a novel approach by implementing pneumatic actuation to simulate dynamic and unpredictable combat situations in a safe and controlled environment. Unlike conventional training tools, the proposed devices feature randomized movement patterns of striking or blocking elements, effectively replicating the variability of real-life sparring. Through adjustable air pressure and actuation speed, the system can be customized to suit different user profiles, enabling both novice and elite athletes to train at appropriate intensity levels [7]. Furthermore, the integration of sensors and programmable controllers (e.g., Controllino PLC) allows for real-time monitoring of reaction time, movement precision, and timing accuracy. This feedback not only enhances training effectiveness but also provides objective metrics for tracking athletic progress [8].

This paper presents the conceptual design, development, and practical implementation of two pneumatic-based training devices tailored for martial arts disciplines. The systems were tested in sports hall environments to assess usability and performance. The overarching goal is to contribute to the advancement of intelligent and adaptable sports training equipment that improves reflexive responses and overall athletic readiness in combat sports.

2. PNEUMATIC SYSTEM FOR TRAINING REACTION SPEED IN TAEKWONDO

Martial arts originated as systems of unarmed self-defence and have evolved over centuries into complex disciplines encompassing both physical and philosophical dimensions. Among these, taekwondo stands out as one of the most dynamic and globally popular sports, particularly following its inclusion in the Olympic Games in 2000. The sport emphasizes speed, agility, and precision (especially in foot techniques) and demands rapid responses to visual stimuli during combat. Traditional methods for improving reaction speed often rely

on partner-based drills, where one participant initiates a movement and the other responds. While effective, these exercises require constant human interaction, limiting opportunities for individual training. To address this challenge, a pneumatic training system was developed to provide visual signals that trigger athlete reactions. This chapter outlines the design, implementation, and testing of an experimental system aimed at enhancing reaction speed in taekwondo practitioners.

2.1. System Architecture and Mechanical Design

The proposed system consists of two independently controlled double-acting pneumatic cylinders, each equipped with a circular impact target suitable for kicks or punches. Cylinder movement is regulated by 5/3 solenoid valves, controlled via a Geekcreit UNO R3 microcontroller. All components are mounted on a steel frame that ensures mechanical stability and allows for height adjustments.

The frame is constructed from 25 × 25 mm steel square tubing with a 2 mm wall thickness. Its H-shaped base, formed from four 900 mm tubes, provides a stable foundation. Vertical supports (1900 mm in height) feature machined grooves for adjusting the cylinder mounting height to accommodate different training levels.

Both pneumatic actuators are SMC CP96 cylinders with a 100 mm stroke and a 16 mm piston rod diameter. These cylinders are selected for their high actuation speed (less than 0.5 seconds) and mechanical robustness, critical for withstanding repeated impacts without deformation. Each piston rod is fitted with a 200 mm diameter, 15 mm thick circular board made of laminated wood, topped with a soft impact layer sourced from modified taekwondo armor and focus pads (Figure 2.1). To enhance safety and simulate realistic contact, the piston is depressurized after extension, reducing resistance upon impact.

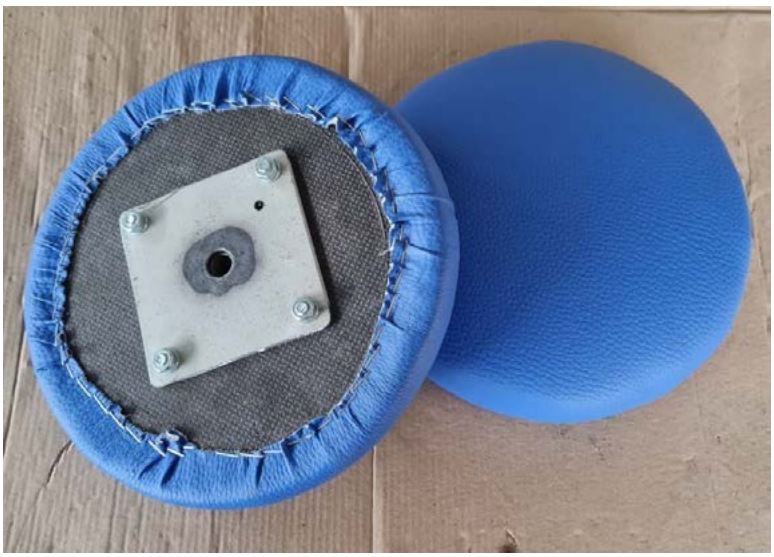


Figure 2.1. Striking Surfaces Mounted on Piston Rods for Taekwondo Training

2.2. Electronic Control System and Software

The control system is based on a Geekcreit UNO R3 microcontroller (ATmega328P), programmed via the Arduino IDE. It manages two SMC SY5000 5/3 solenoid valves that control airflow to the pneumatic cylinders.

Power is supplied by a Schrack LP 746201 AC/DC converter, delivering 24 V DC at 1.5 A, which is sufficient for valve operation and LED indicators. The microcontroller outputs are interfaced through transistors on a custom development board, ensuring safe and efficient switching of the solenoids (Figure 2.2).

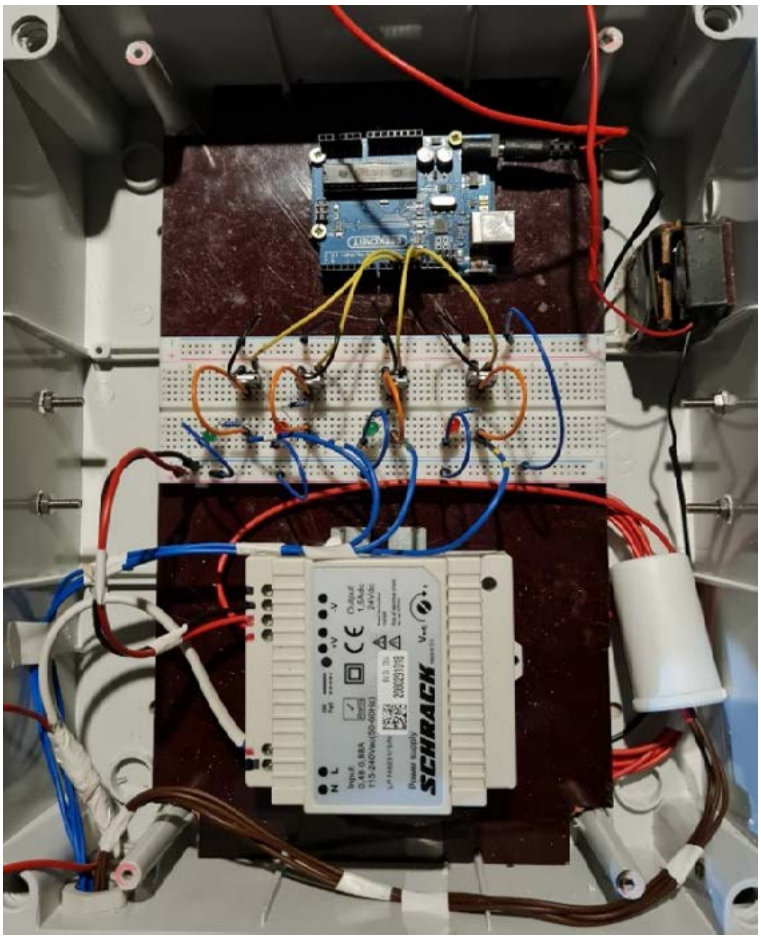


Figure 2.2. Control Box: Microcontroller, Power Supply Unit, and Interface Board

The software initializes all input/output pins and runs a continuous loop that triggers randomized delays and piston activations. This unpredictability simulates real combat scenarios and trains athletes to react quickly and make decisions under pressure. The system supports multiple training modes, including sequential activation, dual-piston challenges, and time-based drills.

2.3. System Assembly and Field Testing

Following component-level validation, the complete system was assembled and tested in its final configuration (Figure 2.3). Initial tests were conducted without user interaction to verify sequence logic and ensure reliable autonomous operation over ten repetitions. Subsequent tests involved athletes responding to piston activations.



Figure 2.3. Final Design of the Taekwondo Reaction Speed Training Device in Operational Setting [9]

After minor adjustments (primarily to the height of the impact targets) the system was deployed in a school gymnasium with primary school students. The device generated strong interest and engagement, indicating high potential for youth training. Further evaluation was conducted at Taekwondo Club Zaprešić, where junior and senior athletes, including European Championship participants, used the system during regular training sessions. Various techniques were practiced, including roundhouse kicks, sidekicks, and punches at different target heights (body and head level). Reaction time was manually recorded using a stopwatch, providing a qualitative baseline for future improvements and confirming the system's effectiveness in enhancing performance. The system achieved a level of stability and safety suitable for regular integration into taekwondo training. Users expressed satisfaction with the innovative experience, especially the chance to train with unpredictable visual cues. The programmable architecture also enables future upgrades, such as:

- Integration of impact-sensitive plates for reaction time measurement;
- Monitor display for real-time feedback and performance tracking;
- Accuracy sensors for hit precision analysis.

These enhancements would transform the device from a basic reaction trainer into a comprehensive smart training platform, offering data-driven insights for athletes and coaches.

3. PNEUMATIC DEVICE FOR TRAINING DEFENSIVE TECHNIQUES AND REACTION SPEED IN BOXING

This chapter presents the development and implementation of a training device designed to improve defensive techniques and reaction speed in boxing. The system uses pneumatic cylinders to replicate incoming punches, prompting athletes to respond quickly and accurately. To ensure reliable operation, the device integrates sensors and programmed control logic that monitor each cylinder's movement, ensuring it returns to its initial position before triggering the next simulated strike. The following sections describe the mechanical design, pneumatic integration, control system, and operational behaviour of the device.

3.1. Mechanical Design and Pneumatic System Integration

The foundation of the training device is a 3D model that incorporates all mechanical and pneumatic components (Figure 3.1). The frame is built using aluminium profiles connected with angular brackets, providing a lightweight yet stable structure. Pneumatic cylinders are mounted on custom-designed plates, which are securely attached to the frame.



Figure 3.1. Designed 3D Model of the Training Device

The primary function of the pneumatic system is to move a focus pad from its resting position to a strike position, simulating a punch. This is achieved using double-acting pneumatic cylinders. Through CAD simulations, optimal stroke lengths were determined. The selected actuators are SMC model C(D)85N16-140C-B. System feasibility was evaluated using SMC's Model Selection Software, which simulated actuator behaviour and confirmed that the chosen cylinders meet the required performance criteria (Figure 3.2). Actuation speed is adjustable via throttle-check valves.

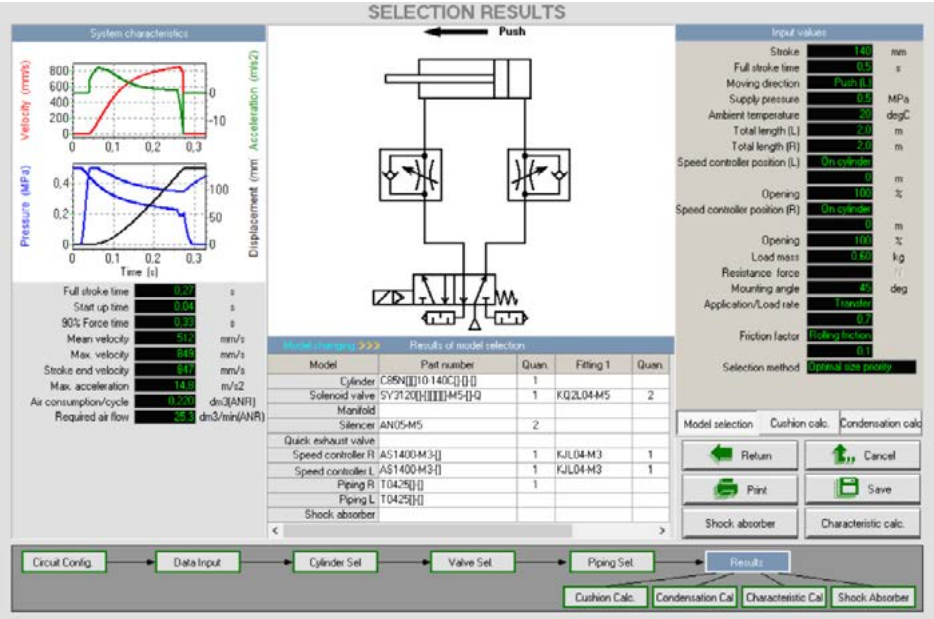


Figure 3.2. Simulation Outputs of Pneumatic Cylinder



(a) (b)

Figure 3.3. Mounting Components onto the Device Frame, a) Joint Connection to the Frame, b) Connection Between Pneumatic Cylinder and Focus Pad

The device uses four pneumatic cylinders, each paired with an angular connector to move a focus pad. These connectors are mounted on axles and rotate using sliding bearings (Figure 3.3a). The core of each pad is made from PPR pipes, chosen for their strength, affordability, and ease of customization. For safety and impact absorption, thermal insulation foam is used as the outer layer, tightly fitted without cutting. Each pad is attached via a 20 mm diameter joint connection, secured with screws (Figure 3.3b).

3.2. Control System and Electronics Integration

The pneumatic actuators are controlled by a Controllino MINI microcontroller, based on the ATMega328 platform and compatible with Arduino. Power is supplied by a 24V Mean Well MDR-20-24 unit, which drives the solenoid valves responsible for airflow control. Controllino MINI simplifies the hardware setup by integrating relay outputs used to control four SMC SY3120-5LOU-C6-Q solenoid 5/2 valves. These valves are compactly arranged on an SS5Y3-20-04-004-Q manifold. Air pressure is provided by a compressor capable of delivering up to 8 bar, connected to the pneumatic circuit via hoses and fittings. Each cylinder's speed and return flow are fine-tuned using AS1201F-M5-06A throttle-check valves with push-lock adjustment. Position sensors (D-M9PL) are powered by Controllino's 5V supply and provide real-time feedback on cylinder status, ensuring safe and accurate operation. All electronic components are housed within a protective control box, as depicted in Figure 3.4.

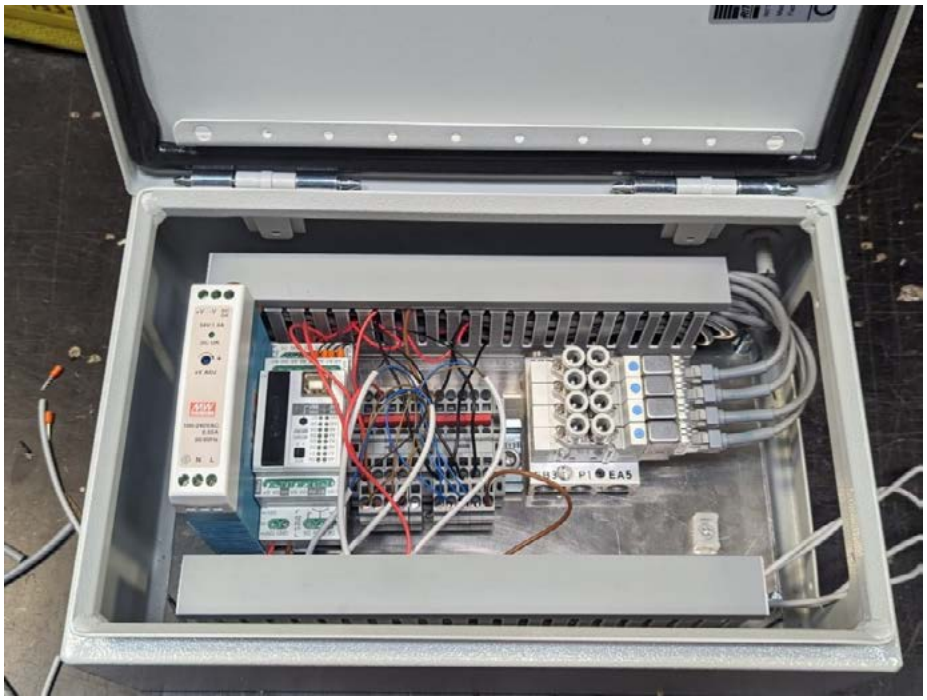


Figure 3.4. Control Unit with Integrated Components

3.3. Operating Principle and System Behavior

The device operates by randomly activating pneumatic cylinders to simulate unpredictable strikes. The control logic ensures that only one cylinder is active at a time and that each activation occurs only after the previous cylinder has returned to its resting position. System initialization involves setting air pressure and adjusting throttle valves to control actuation speed. Training parameters such as strike frequency and response intervals are configured via the Arduino interface.

During operation, the microcontroller continuously monitors sensor inputs. When a randomly selected cylinder is confirmed to be idle, a control signal activates the corresponding valve, initiating the strike. After a brief dwell time to simulate impact, the valve is deactivated, allowing the cylinder to return via spring force or reverse pressure. This cycle repeats, creating a dynamic and reactive training environment.
To ensure safety, the system halts if a sensor fails to confirm a cylinder's position, preventing unintended activation. The fully assembled device is shown in Figure 3.5, illustrating both idle and active states.



(a) (b)

Figure 3.5. Completed Training Device, a) Idle State, b) During Simulated Strike

This device successfully integrates mechatronic and pneumatic technologies to deliver a responsive, safe, and realistic training platform for boxing. Its modular design and programmable features provide a solid foundation for future enhancements, including adaptive difficulty levels, advanced strike pattern algorithms, and real-time performance analytics.

4. CONCLUSION

This study has demonstrated the practical feasibility and benefits of integrating pneumatic systems into training devices for combat sports, specifically taekwondo and boxing. By harnessing the core advantages of pneumatic actuation, such as rapid response, impact absorption through air compressibility, and seamless integration with programmable control systems, two distinct training prototypes were successfully designed, built, and tested in real-world sports environments. The taekwondo-focused device aimed to improve reaction time by presenting randomized visual targets through the extension of striking surfaces.

Athletes were challenged to respond with speed and precision to unpredictable stimuli. Field testing with both youth and professional practitioners revealed strong engagement and confirmed the system's effectiveness in enhancing reflexes and decision-making under pressure. The boxing training device simulated incoming punches using four independently controlled pneumatic cylinders. Athletes practiced defensive manoeuvres such as dodging and blocking, guided by randomized actuation sequences managed by a Controllino microcontroller. Sensor feedback and mechanical safeguards ensured safe operation, while the modular design allowed for personalized adjustments based on user height and skill level. Both devices proved to be stable, adaptable, and safe for use in various training scenarios. They offer a new paradigm in combat sports training, where mechanical systems are used not only to mimic combat situations but also to enhance athletic performance through precise control, repeatability, and data-driven feedback mechanisms.

Future improvements may include the integration of sensor-based performance tracking (e.g., impact sensors, motion detection), real-time feedback through visual or auditory interfaces, and machine learning algorithms for adaptive training programs. Such upgrades would elevate these devices from experimental prototypes to intelligent training platforms, aligned with the broader digital transformation in sports science and athlete development. In summary, the integration of pneumatic systems into combat sports equipment presents a promising and underexplored avenue for improving reaction speed, technical proficiency, and training efficiency. This work contributes to the growing intersection of mechatronics and sport, opening new possibilities for innovation in athletic training technologies.

5. REFERENCES

[1] A. Slimani, M. Chaabène, T. Miarka, H. Del Vecchio, and K. Chamari, 'Kickboxing review: anthropometric, psychophysiological and activity profiles and injury epidemiology', Biology of Sport, vol. 34, no. 2, pp. 185–196, 2017.

[2] P. Chaabene et al., 'Physical and physiological attributes of taekwondo athletes: A brief review', Journal of Strength and Conditioning Research, vol. 28, no. 6, pp. 1714–1721, 2014.

[3] M. Dabnichki and A. J. Crocombe, 'Computer simulations of martial arts techniques', Sports Engineering, vol. 3, no. 4, pp. 189–194, 2000.

[4] H. Shiratori, M. Mizuguchi, and T. Ohtsuki, 'Development of a measurement and feedback system for sport training', in Proc. IEEE International Conference on Systems, Man, and Cybernetics, pp. 2574–2578, 2007.

[5] R. Wood and G. Bush, 'Design of high-speed pneumatic actuators for automation applications', Journal of Manufacturing Processes, vol. 12, no. 2, pp. 130–138, 2010.

[6] F. Bonivento, S. Gentili, L. Righetti, and L. Zaccarian, 'Modeling and control of a pneumatic actuator', IEEE Transactions on Control Systems Technology, vol. 12, no. 3, pp. 469–474, 2004.

[7] M. Mori et al., 'Development of a robot-based system for reaction time training in sports', in Proc. IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 912–917, 2019.

[8] K. Isshiki, K. Kusaka, and H. Tokuda, "Development of a reflex training device using pneumatic drive system," in Proc. SICE Annual Conference, pp. 224–229, 2018.

[9] T. Stanišak, ‘Pneumatic system for training reaction speed in combat sports’ Master's thesis, (in Croatian), University of Zagreb, Fac. of Mech. Eng. and Naval Arch., https://repozitorij.fsb.unizg.hr/islandora/ object/fsb:7417, 2021.
[10] D. Šarić, ‘Device for training defence techniques and reaction speed in boxing’ Undergraduate thesis, (in Croatian), University of Zagreb, Fac. of Mech. Eng. and Naval Arch., https://repozitorij.fsb.unizg.hr/islandora/ object/fsb:11073, 2024.

Biographies



Željko Šitum received his M.Sc. degree in Mechanical Engineering from the University of Zagreb, Croatia, in 1997, and his Ph.D. degree in the same field in 2001. He is currently a Full Professor at the Department of Control Engineering, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb. His research interests include digital control of dynamic systems, fluid power systems control, mechatronics, and computer simulations. His current focus is primarily on modelling and control of hydraulically and pneumatically actuated robotic systems.



Tin Stanišak graduated in 2021 from the Automation in Production study program at the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatia. He currently works as a Metrology Coordinator at the pharmaceutical company Pfizer Croatia. He is a professional leader with extensive experience and serves as the head coach of the Plamen Pušća Taekwondo Club.



Dominik Šarić is a fifth-year student in the Mechatronics and Robotics program at the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatia. During his high school years, he was an active boxing competitor, achieving notable success, including the title of vice-champion of Croatia. In addition to his academic commitments, he currently works as a boxing coach and has trained numerous national champions across various boxing categories.

Estimating the Global Emissions by Fluid Power Machinery

[1]Lasse Schmidt, [2]Katharina Schmitz, [3]Tatiana Minav & [4]Adolfo Senatore

[1]AAU Energy, Aalborg University, Denmark, lsc@energy.aau.dk

[2]IFAS, RWTH Aachen University, Germany, katharina.schmitz@ifas.rwth-aachen.de

[3]IHA, Tampere University, Finland, tatiana.minav@tuni.fi

[4]University of Naples Federico II, Italy, adolfo.senatore@unina.it

Abstract.

Faced with the ongoing climate crisis and increased regulatory focus on emission reductions, it is especially relevant to identify key loss mechanisms across industries, responsible for energy consumption and associated emissions not producing any work. Minimizing losses may directly and significantly reduce emissions, fossil fuel dependence, the need for new renewable energy sources , and potential changes to the electric grid. Given the principal losses associated with many fluid power systems, fluid power machinery sectors are likely a source of high emissions when compared to emissions on a global scale. Therefore, this study aims to estimate the global emissions by fluid power machinery. Literature on this topic is generally fragmented, and the most recent and most exhaustive study on this subject was conducted for the United States in 2012, which encompassed both mobile hydraulic, industrial hydraulic, pneumatic, and aviation industries. This 2012 survey was conducted by the Oak Ridge National Laboratory in collaboration with the National Fluid Power Association and major U.S. and European-based industry players. Combining these 2012 findings with global emissions by relevant industries, such as construction machinery and energy production, this paper estimates the global emissions by fluid power machinery, sans aviation hydraulics; these are then grouped according to hydraulic mobile machinery, hydraulic industrial machinery, and pneumatic machinery. The estimation results suggest emissions by fluid power machinery exceed twice that of global aviation, amounting to roughly 5% of the global energy-related emissions. The findings of this study present a novel state-of-the-art assessment of the global CO_2 emissions generated by this sector. Finally, it suggests remarkable potential among multiple decarbonization pathways.

Keywords. Fluid Power Machinery, Energy Consumption, Emissions, Sustainability.

1. INTRODUCTION

Fluid power is a means of power transmission through pressurized working fluids. It includes hydraulics, using liquid as the working fluid, and pneumatics using air/gas as the working fluid, and plays a vital role in modern machinery, automation and transportation. Fluid power technologies are indispensable in applications demanding high power and/or

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

force density, precision and reliability. Fields of application are virtually countless and Fig. 1.1. illustrates a few examples of the wide application range where fluid power technologies are key enablers, emphasizing its broad versatility and vital role in modern technology.



Figure 1.1. Examples of machinery where fluid power is indispensable.

Fluid power technologies improve resilience, functionality and productivity across industries. An example known to many people is an excavator as shown in Fig. 1.2., where fluid power is an indispensable enabler for its functionality.

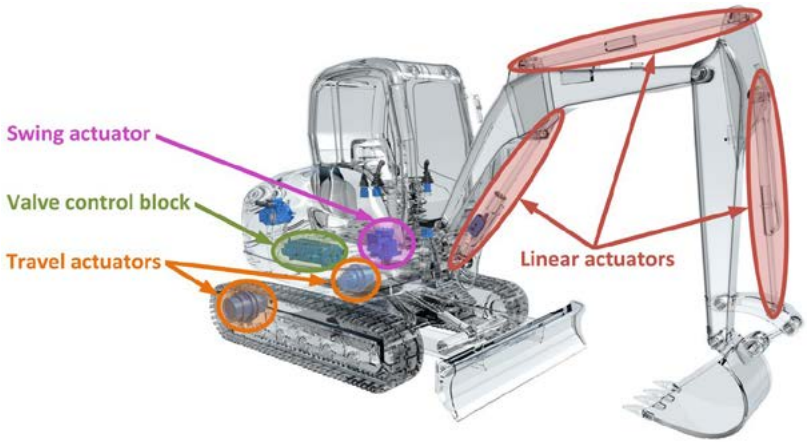


Figure 1.2. Examples of indispensable functions in an excavator enabled by fluid power.

Facing climate change, key challenges for fluid power technology include low system level energy efficiencies, resulting in excessive energy use and associated emissions. Whereas todays fluid power components are rather efficient, system level working principles have largely remained unchanged for decades. To unlock the decarbonization potential of fluid power technologies, energy efficient fluid power system architectures must be at the heart of research efforts, enabling competitive and sustainable technologies.

1.1. Starting Point for Emission Estimates

Studies related to emissions by fluid power machinery in literature is generally fragmented and often related to single machines, single machine categories and so forth, see e.g. [15-18, 26]. The most recent and most exhaustive study on this subject was conducted in 2012 for the U.S. encompassing both mobile hydraulic, industrial hydraulic, pneumatic and aviation industries [1]. The study was conducted by the Oak Ridge National Laboratory in collaboration with the U.S. National Fluid Power Association (NFPA) and 31 industrial partners, including major technology players in both manufacturing as well as OEMs such as Bosch Rexroth, Danfoss, Festo, Poclain Hydraulics, ABB, Siemens, Parker Hannifin, Caterpillar and Sun Hydraulics. The study was targeting the collection and consolidation of specific measurements related to energy consumption, associated emissions and energy efficiencies of deployed fluid power systems [1]. In this context the results may be considered the most broad and trustworthy information on the subject available and is hence used as a reference case for estimating the global emissions by fluid power machinery. As the primary aim here is to estimate the emissions related to mobile and industrial hydraulic and pneumatic machinery, the findings related to aviation in [1] are left out in the following.

1.2. Considerations & Assumptions Applied

The study presented in this paper aims to estimate the global emissions by fluid power machinery in a 2022 scenario, with the reason for this specific year being reference availability. The emission estimates presented uses the 2012 U.S. energy consumption and emission estimates by fluid power machinery [1] as a reference, taking offset in the energy consumption reported here. A main difference from [1] is the emissions by electric energy generation, which generally was lower in 2022 compared to 2012, partly due to a different mix of energy sources, both in the U.S., and globally. Hence, the choices of the emissions per energy unit used in both the 2012 U.S. reference case and the 2022 scenario are considered below for clarity. Note that "CO_2" in the following refers to CO_2 equivalents.

1.2.1. Emissions Related to Diesel Fuel Consumption

Mobile hydraulic machinery remains predominantly powered by internal combustion engines (ICEs) using diesel fuel. According to [1], the emissions per energy unit is equal to 0.249 MMT CO_2/TWh (73 MMT CO_2/Quad), which is in close correspondence with estimates obtainable via other sources, e.g. [5-8]. For this reason, both the 2012 U.S. reference case and the 2022 global emissions scenario uses 0.249 MMT CO_2/TWh for ICE powered fluid power machinery, not accounting for the emissions related to production of the fuel itself.

1.2.2. Emissions Related to Electric Energy Generation

Industrial hydraulic and pneumatic machines are generally electrically powered, and compared to the 2012 study [1], the 2022 emissions/electric energy unit are much different both in the U.S. and globally. The 2012 estimate for the U.S. in [1] suggests emissions related to electric energy generation to be 0.607 MMT CO_2/TWh (178 MMT CO_2/Quad). According to [3], the average 2022 emissions/electric energy unit were for example:

- 0.586 kg CO_2/kWh in China [3]
- 0.292 kg CO_2/kWh in the EU [3]
- 0.386 kg CO_2/kWh in the U.S. [3]
- 0.486 kg CO_2/kWh globally [3]

In addition, [2] suggests that emissions from electric energy generation in the EU in 2022 were even lower than the estimate of [3], namely with emissions of 0.258 kg CO_2/kWh. To be conservative in the emission estimates the 2022 EU emissions per electric energy unit generated as presented in [3] are assumed globally valid.

In addition, it is notable that the emissions per electric energy unit = 0.258 kg CO_2/kWh = 0.258 MMT CO_2/TWh [3] is only 3.6 % higher than the emissions per diesel energy unit = 0.249 MMT CO_2/TWh [1]. Hence, for conservativeness and transparency, **the 2022 global emission estimates assumes that emissions per electric energy unit are identical to those of diesel fuel**, i.e. = 0.249 MMT CO_2/TWh.

1.2.3. Relative Impact of Emissions in the Context of Efficiency

Considering especially the emissions per electric energy unit presented in [1], with 0.607 MMT CO_2/TWh and 0.249 MMT CO_2/TWh for diesel energy, it may seem like diesel fuel being the preferred choice if emissions should be as low as possible. However, here the energy efficiency plays a significant role, with this being highly dependent on the load conditions. As discussed in section 3.2. in the following, the average efficiency of ICEs and electric systems (prime movers) can be assumed to be 45% and 90%, respectively. In this case an ICE producing 1.0 unit of mechanical energy at the output shaft would require 2.2 units of fuel energy, whereas an electric motor/drive unit would require 1.1 units of electric energy to produce 1.0 unit of mechanical energy at the shaft.

Using the emissions per energy unit presented in [1], the corresponding emissions related to producing 1.0 TWh on the shaft would be 0.55 MMT CO_2 for an ICE, and 0.67 MMT CO_2 for an electric motor/drive unit, suggesting that ICE powered machinery emits less CO_2 than electric powered machinery in the 2012 U.S. reference case.

However, whereas emissions by diesel fuel remains the same, the 2022 emissions related to electric energy in the EU was 0.258 MMT CO_2/TWh as discussed in section 1.2.2. Using this measure, and an electric motor/drive unit producing 1.0 TWh on the shaft would emit 0.29 MMT CO_2, corresponding to 53% of the emissions by ICEs with the same output.

It is notable that with an increasing implementation of renewable energy sources, the emissions associated with electric powered machinery will decrease, whereas it for ICE powered machinery will stay the same, assuming fluid power system efficiencies unaltered.

1.2.4. Summary of Emissions per Energy Unit used in the Study

From the considerations above the emissions per energy unit used in the 2012 reference case and the assumed emissions per energy unit for the 2022 global emission scenario, are summarized below for clarity.

2012 reference case for the U.S. emissions by fluid power machinery

- 0.249 MMT CO_2/TWh for diesel energy
- 0.607 MMT CO_2/TWh for electric energy

2022 global emission scenario for fluid power machinery

- 0.249 MMT CO_2/TWh for diesel energy
- 0.249 MMT CO_2/TWh for electric energy

The emissions per energy unit used in the 2022 global emission estimates are indeed conservative from the considerations above, and likely somewhat higher.

Finally, from the above considerations and assumptions as well as the limited sources available, the estimated emissions by fluid power machinery cannot be considered subject to scientific rigour, but rather *best estimates*.

2. ENERGY USE & EMISSIONS BY FLUID POWER MACHINES – THE 2012 U.S. REFERENCE CASE

Following [1], the energy consumption by fluid power machinery in the U.S. in 2012 are summarized in Table 1. For mobile fluid power machinery, the maximum TWh/year measures for *Construction machinery*, *Agriculture machinery* and *Other mobile machinery* are extrapolated from their minimum values using the ratio of maximum and minimum TWh/year for *Total mobile hydraulic machinery* in Table 1.

As the maximum and minimum energy consumption covers estimate ranges, it is assumed that the **average of the minimum and maximum values is representative for the *actual* energy consumption** as specified in Table 1. From this assumption, the % distribution of energy consumption of fluid power machinery is obtained as also specified in Table 1.

Table 1: Energy consumption estimates by fluid power machinery in the US in 2012. Numbers in **bold** font are obtained or directly deduced from [1] and converted from Quads to TWh. Remaining numbers are own calculations.

2012 US Energy Consumption by Fluid Power Machinery	Minimum TWh/yr	Maximum TWh/yr	Avg. TWh/yr	% Avg. TWh/yr
Mobile hydraulic machinery				
- Construction machinery	**68.3**	237.7	153.0	21.7
- Agriculture machinery	**4.7**	16.3	10.5	1.5
- Other mobile machinery	**33.1**	115.3	74.2	10.5
Total mobile hydraulic machinery	**106.1**	**369.3**		
Total industrial hydraulic machinery	**321.2**	**321.2**	321.2	45.5
Total pneumatic machinery	**146.5**	**146.5**	146.5	20.8
Total	**573.8**	**837.0**	705.4	100.0

The estimated CO_2 emissions associated with the energy consumption measures in Table 1 are outlined in Table 2. Here the estimated CO_2 emissions for mobile hydraulic machines are estimated from using 0.249 kg CO_2/kWh = 0.249 MMT CO_2/TWh. As discussed, industrial hydraulic and pneumatic machines are primarily electrically powered, and [1] suggests related emissions of approximately 0.607 MMT CO_2/TWh in 2012 as discussed in section 1.2. Applying these emissions/energy unit measures, the corresponding emissions in Table 2 are obtained, which are nearly identical to the estimates presented in [1].

Table 2: 2012 emission estimates by fluid power machines in the US deduced from [1].

2012 US Emissions by Fluid Power Machinery	Minimum MMT CO_2/yr	Maximum MMT CO_2/yr	Avg. MMT CO_2/yr	% Avg. CO_2/yr
Mobile hydraulic machinery (diesel)				
- Construction machinery	17.0	59.1	38.0	11.1
- Agriculture machinery	1.2	4.1	2.6	0.8
- Other mobile machinery	8.2	28.7	18.4	5.4
Total mobile machinery	26.4	91.8		
Total industrial hydraulic machinery	195.1	195,1	195.1	56.8
Total pneumatic machinery	89.0	89.0	89.0	25.9
Total	310.5	375.9	343.2	100.0

3. GLOBAL EMISSION ESTIMATES BY FLUID POWER MACHINERY – A 2022 SCENARIO

Fluid power system technologies remain virtually unaltered since 2012, i.e. primarily related to throttle control. Similarly mobile hydraulic machines remain primarily powered by diesel fuelled ICEs, and industrial hydraulic and pneumatic systems remain predominantly electrically powered. It has not been possible to account for the % distribution of deployed fluid power machines in different geographic regions. Hence, for conservativeness, the following assumes that the percentage distribution of mobile hydraulic, industrial hydraulic and pneumatic machinery outlined in Table 1 is valid globally.

In 2022, [9] reported that global emissions by construction machines amounted to 400 MMT CO_2/year, a number that has been referenced in several publications, e.g. [10-13]. Applying this measure, the assumptions from section 1.2.4, and the % energy consumption by construction machines in Table 1, emissions are estimated to 18.44 MMT CO_2/ % energy consumption. From this, the global emissions are estimated as in Table 3.

Table 3: Global CO_2 emission estimates by fluid power machines.

Global CO_2 Emission Estimates based on average US Energy % Distribution	% Average TWh/yr	Average MMT CO_2/yr
Mobile hydraulic machinery (diesel)		
- Construction machinery	21.7	400.0
- Agriculture machinery	1.5	27.5
- Other mobile machinery	10.5	194.0
Total mobile machinery		621.5
Total industrial hydraulic machinery	45.5	839.9
Total pneumatic machinery	20.8	383.1
Total	100.0	1,844.5

3.1. Emission Impact of Fluid Power Machinery

To put the estimated emissions by fluid power machinery in perspective, these are compared to the much-scolded aviation sector and the global energy related emissions.

In 2022 international aviation was estimated to be responsible for emitting approximately 800 MMT CO_2/year [14], distributed by 331 MMT CO_2/year for domestic aviation and 470 MMT CO_2/year for international aviation. In comparison fluid power machinery is the same year estimated to emit 2.3 times more than global aviation.

Similarly, the 2022 global energy related emissions are estimated to 36,800 MMT CO_2/year [2], hence the total emission estimate above suggests that fluid power machinery was responsible for 5% of the global energy related emissions.

3.2. Emissions of Fluid Power Machinery related to Losses & Useful Work

Having established the significant emission impact of fluid power machinery, the following aims to quantify which part of emissions that are related to prime mover losses, fluid power system losses and useful work. To investigate this, the average efficiencies related to fluid power machinery segments proposed in [1] are used. Considering the portion of energy use related to mobile hydraulic, industrial hydraulic and pneumatic machinery when combined with the respective average efficiencies, the weighted efficiencies in Table 4 are obtained.

Table 4: Weighted energy efficiencies based on percentage energy use.

Fluid Power Machinery	% Average TWh/yr	Average Energy Efficiency in % [1]	Weighted Efficiency
Mobile hydraulic machinery (diesel)	33.7	**21**	7.1
Total industrial hydraulic machinery	45.5	**50**	22.8
Total pneumatic machinery	20.8	**15**	3.1
Total	100.0		33.0

From Table 4 the overall weighted efficiency of fluid power machinery is found to 33%, meaning that 67% is lost in power transmission processes. Considering now the underlying efficiencies, several reports suggest ICE brake thermal efficiencies in ranges 42-43% [22], 43-44% [23] and 45–46% [24]. Similarly, total efficiencies electric motor plus electric drives and wiring have been estimated to 88.0% [19], 89.8% [20] and [21] suggest efficiencies in the range 80-90%, which is supported by the findings in e.g. [25]. Based on this, the following assumes an average fuel-to-shaft efficiency for ICEs of 45% and an average grid-to-shaft efficiency for electric prime movers of 90%.

Table 5: Energy efficiency estimates on machine subsystem level.

Fluid Power Machinery	ICE/Electric Subsystem Efficiency in %	Hydraulic/Pneumatic Subsystem Efficiency in %	Total Efficiency
Mobile hydraulic machinery	45	47	21
Industrial hydraulic machinery	90	56	50
Pneumatic machinery	90	17	15

Based on these *prime mover* efficiencies the hydraulic and pneumatic (sub)system efficiencies outlined in Table 5 are obtained, such that the total machine efficiencies are in correspondence with those specified in [1]. Hence, these efficiency estimates are used to estimate the associated emissions. Defining energy consumption as the input energy E_{in}, the output energy of a prime mover (ICE or electric) as E_{pm} and the output energy of a fluid power systems E_{fp}, the energy flow in fluid power machinery may be depicted as in Fig. 3.1.

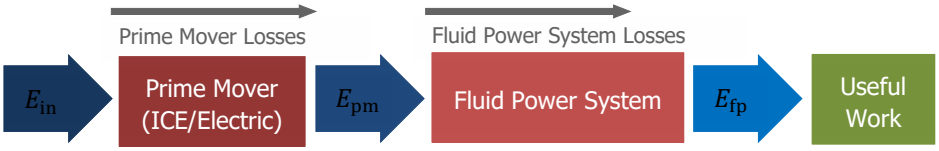


Figure 3.1. Energy flow in fluid power machinery.

Furthermore, defining the prime mover and fluid power system efficiencies as η_{pm} and η_{fp}, respectively, the prime mover losses $E_{\text{pm,loss}}$, fluid power system losses $E_{\text{fp,loss}}$ and useful work E_{uw} may be obtained through Eqs. (3.1)-(3.4).

$$E_{\text{pm}} = \eta_{\text{pm}} E_{\text{in}} \quad , \quad E_{\text{fp}} = \eta_{\text{fp}} E_{\text{pm}} = \eta_{\text{fp}} \eta_{\text{pm}} E_{\text{in}} \tag{3.1}$$

$$E_{\text{pm,loss}} = E_{\text{in}} - E_{\text{pm}} = (1 - \eta_{\text{pm}}) E_{\text{in}} \tag{3.2}$$

$$E_{\text{fp,loss}} = E_{\text{pm}} - E_{\text{fp}} = \eta_{\text{pm}} E_{\text{in}} - \eta_{\text{fp}} E_{\text{pm}} = (1 - \eta_{\text{fp}}) \eta_{\text{pm}} E_{\text{in}} \tag{3.3}$$

$$E_{\text{uw}} = E_{\text{in}} - E_{\text{pm,loss}} - E_{\text{fp,loss}} = \eta_{\text{fp}} \eta_{\text{pm}} E_{\text{in}} \tag{3.4}$$

Noting that emissions/energy unit are assumed identical for diesel and electric powered machinery as discussed in section 1.2., the emissions may be considered proportional to the associated energy consumption. Hence, prime mover loss related emissions $\varepsilon_{\text{pm,loss}}$, fluid power system loss related emissions $\varepsilon_{\text{fp,loss}}$ and the emissions related to useful work ε_{uw}, may be estimated like the associated energy losses Eqs. (3.4), as Eqs. (3.5).

$$\varepsilon_{\text{pm,loss}} = (1 - \eta_{\text{pm}}) \varepsilon_{\text{in}} \ , \ \varepsilon_{\text{fp,loss}} = (1 - \eta_{\text{fp}}) \eta_{\text{pm}} \varepsilon_{\text{in}} \ , \ \varepsilon_{\text{uw}} = \eta_{\text{fp}} \eta_{\text{pm}} \varepsilon_{\text{in}} \tag{3.5}$$

Using the global emission estimates in Table 3, the efficiency estimates in Table 5 and Eqs. (3.5), emissions related to prime mover losses, fluid power system losses and useful work appear as depicted in Table 6.

Table 6: Emission estimates on machine subsystem level.

Emissions related to losses and useful work	Mobile hydraulic machinery in MMT CO_2/yr	Ind. hydraulic machinery in MMT CO_2/yr	Pneumatic machinery in MMT CO_2/yr	Total
Emissions related to ICE/electric subsystem losses	341.8	84.0	38.3	464.1
Emissions related to fluid power subsystem losses	148.2	332.6	286.2	767.0
Emissions related useful work	131.4	423.3	58.6	613.3
Total	621.5	839.9	383.1	**1,844.5**

Due to the proportionality assumption of energy consumption and emissions across machine types, 33% of the emissions are related to useful work, corresponding to the weighted energy efficiency as expected. Furthermore, 25% is associated with ICE/electric system losses and 42% is associated with losses in fluid power systems.

4. DECARBONIZATION PATHWAYS & PERSPECTIVES

From the distribution of emissions related to losses and useful work estimated above, it is evident that there is a significant emission reduction potential for fluid power machinery. Indeed, if all electric energy was generated by renewable energy sources like wind turbines and solar panels, the emission reduction potential would be related to electrification of mobile fluid power machines alone. However, a broadly electrified world will likely not be a reality for decades. In any case, the use of electric energy should be reduced as much as possible to limit the requirement for renewable energy sources, rare earth materials, allow for realistic and commercially feasible battery sizes in mobile machinery etc. Hence, the increase of energy efficiencies to reduce energy consumption should be at the core of future research, development and innovation efforts.

To evaluate the emission reduction potentials by increased energy efficiencies and electrification viewed in a 2022 context as the for the above estimates, Eqs. (4.1)-(4.4) are applied taking offset in useful work, E_{uw}, and hence the reverse calculation process compared to Eqs. (3.1)-(3.4).

$$E_{\mathrm{fp}} = \frac{E_{\mathrm{uw}}}{\eta_{\mathrm{fp}}} \quad , \quad E_{\mathrm{pm}} = \frac{E_{\mathrm{fp}}}{\eta_{\mathrm{pm}}} = \frac{E_{\mathrm{uw}}}{\eta_{\mathrm{pm}}\eta_{\mathrm{fp}}} \tag{4.1}$$

$$E_{\mathrm{fp,loss}} = \frac{E_{\mathrm{uw}}}{\eta_{\mathrm{fp}}} - E_{\mathrm{uw}} = \frac{1-\eta_{\mathrm{fp}}}{\eta_{\mathrm{fp}}} E_{\mathrm{uw}} \tag{4.2}$$

$$E_{\mathrm{pm,loss}} = \frac{E_{\mathrm{fp}}}{\eta_{\mathrm{pm}}} - E_{\mathrm{fp}} = \frac{E_{\mathrm{uw}}}{\eta_{\mathrm{pm}}\eta_{\mathrm{fp}}} - \frac{E_{\mathrm{uw}}}{\eta_{\mathrm{fp}}} = \frac{1-\eta_{\mathrm{pm}}}{\eta_{\mathrm{pm}}\eta_{\mathrm{fp}}} E_{\mathrm{uw}} \tag{4.3}$$

$$E_{\mathrm{in}} = E_{\mathrm{uw}} + E_{\mathrm{pm,loss}} + E_{\mathrm{fp,loss}} = \left(1 + \frac{1-\eta_{\mathrm{pm}}}{\eta_{\mathrm{pm}}\eta_{\mathrm{fp}}} + \frac{1-\eta_{\mathrm{fp}}}{\eta_{\mathrm{fp}}}\right) E_{\mathrm{uw}} = \frac{1}{\eta_{\mathrm{pm}}\eta_{\mathrm{fp}}} E_{\mathrm{uw}} \tag{4.4}$$

Again, noting the assumption that emissions/energy unit are identical for diesel and electric powered machinery (section 1.2), the prime mover loss related emissions $\varepsilon_{\mathrm{pm,loss}}$, the fluid power system loss related emissions $\varepsilon_{\mathrm{fp,loss}}$ and the emissions related to the input energy $\varepsilon_{\mathrm{in}}$ may be estimated as Eqs. (4.5).

$$\varepsilon_{\mathrm{fp,loss}} = \frac{1-\eta_{\mathrm{fp}}}{\eta_{\mathrm{fp}}} \varepsilon_{\mathrm{uw}} \ , \varepsilon_{\mathrm{pm,loss}} = \frac{1-\eta_{\mathrm{pm}}}{\eta_{\mathrm{pm}}\eta_{\mathrm{fp}}} \varepsilon_{\mathrm{uw}} \ , \varepsilon_{\mathrm{in}} = \frac{1}{\eta_{\mathrm{pm}}\eta_{\mathrm{fp}}} \varepsilon_{\mathrm{uw}} \tag{4.5}$$

The emission estimates presented in the following are hence based on Eqs. (4.5).

4.1. Emission Reduction Potential of Fully Electrified Fluid Power Machinery

As previously discussed, industrial hydraulic and pneumatic machines are already largely electrified [1], hence the main emission reduction potentials of electrification are related to mobile hydraulic machinery. These machines remain predominantly powered by ICEs, however the process of electrifying these is ongoing with a the main focus being the replacement of ICEs by battery powered electric motors and drives, while hydraulic systems

remain based of traditional design principles[1]. Hence, following the current trends described, assuming that electrification of such machinery increases the prime mover efficiencies from an average of 45% (Table 5) to an average of 90%, and using Eqs. (4.5.), the emissions appear as in Table 7 for a full electrification of mobile hydraulic machinery.

Table 7: Emission estimates on machine subsystem levels assuming full electrification of fluid power machinery.

Emissions related to losses and useful work	Mobile hydraulic machinery in MMT CO_2/yr	Ind. hydraulic machinery in MMT CO_2/yr	Pneumatic machinery in MMT CO_2/yr	Total
Emissions related to ICE/electric subsystem losses	31.1	84.0	38.3	153.4
Emissions related to fluid power subsystem losses	148.2	332.6	286.2	767.0
Emissions related useful work	131.4	423.3	58.6	613.3
Total	310.7	839.9	383.1	**1,533.7**

Table 7 suggests that a full electrification of fluid power machinery, in a 2022 scenario, would reduce emissions from 1,844.5 MMT CO_2/year to 1,533.7 MMT CO_2/year, amounting to approximately a 17% emission reduction for fluid power machinery.

4.2. Emission Reduction Potential by 50% Fluid Power System Efficiency Increase

Consider now the case of broadly increasing the energy efficiency of fluid power systems with prime mover efficiencies remaining the same as in Table 5. Various works have been reported in literature in the past decades comprising more efficient component designs and especially novel system architectures that reduces conceptual losses, increases energy recuperation capabilities etc. Such developments suggest that energy efficiencies may be significantly increased. It is difficult to establish the exact energy efficiency potential, however, in view of the mentioned development the authors believe that a 50% increase in fluid power system efficiencies is achievable on the average. Using this measure and Eqs. (4.5), the emission estimates for fluid power machinery appears as outlined in Table 8.

While the assumed efficiency increase reduces the energy consumption and the associated emissions related to fluid power systems, the reduced fluid power system losses also reduce the energy throughput of the prime movers, hence the prime mover energy losses and emissions. The results in Table 8 suggest that a 50% increase in fluid power system efficiencies would result emission reductions from 1,844.5 MMT CO_2/year to 1,229.6 MMT CO_2/year, i.e. a reduction of 614.9 MMT CO_2/year corresponding to 33.3%.

It is notable that the emission reduction potential of broadly increasing energy efficiencies in fluid power systems by 50%, is roughly twice that of fully electrifying fluid power machinery while maintaining traditional hydraulic system architectures. This emphasizes the necessity of the fluid power research community and industry to focus on research and development of more efficient components and especially system architectures.

[1] Experiences by the authors at BAUMA fairs in 2022 and 2025 in Munich, Germany.

Table 8: Emission estimates on machine subsystem levels assuming a 50% efficiency increase in all three fluid power machinery groups.

Emissions related to losses and useful work	Mobile hydraulic machinery in MMT CO_2/yr	Ind. hydraulic machinery in MMT CO_2/yr	Pneumatic machinery in MMT CO_2/yr	Total
Emissions related to ICE/electric subsystem losses	227.9	56.0	25.5	309.4
Emissions related to fluid power subsystem losses	55.0	80.6	171.3	306.9
Emissions related useful work	131.4	423.3	58.6	613.3
Total	414.3	559.9	225.4	**1,229.6**

4.3. Emission Reduction Potential of a 50% Efficiency Increase & Electrification

Having considered the emission reduction potentials separately related to a full electrification of fluid power machinery and to a 50% increase in fluid power system efficiencies, the impact of doing both is considered.

Applying efficiencies of 90% for both prime mover types, corresponding to fully electrified machinery, and a 50% increase in fluid power system efficiencies to Eqs. (4.5), the emissions presented in Table 9 are obtained.

Table 9: Emission estimates assuming a 50% efficiency increase broadly, and a full electrification of fluid power machinery.

Emissions related to losses and useful work	Mobile hydraulic machinery in MMT CO_2/yr	Ind. hydraulic machinery in MMT CO_2/yr	Pneumatic machinery in MMT CO_2/yr	Total
Emissions related to ICE/electric subsystem losses	20.7	56.0	25.5	102.2
Emissions related to fluid power subsystem losses	55.0	80.6	171.3	306.9
Emissions related useful work	131.4	423.3	58.6	613.3
Total	207.2	559.9	225.4	**1,022.5**

From Table 9 it is found that the emissions may be reduced from 621.5 MMT CO_2/year to 207.2 MMT CO_2/year for mobile hydraulic machinery, from 839.9 MMT CO_2/year to 559.9 MMT CO_2/year for industrial hydraulic machinery and from 383.1 MMT CO_2/year to 225.4 MMT CO_2/year for pneumatic machinery. These reduction measures correspond to 66.7%, 33.3% and 41.2% for mobile hydraulic, industrial hydraulic and pneumatic machinery, respectively.

In total, emissions by fluid power machinery may be reduced from 1,844.5 MMT CO_2/year 1,022.5 MMT CO_2/year corresponding to 44.6%.

5. SUMMARY

A study on the emission reduction potential by fluid power machinery is presented. With offset in previous studies on energy consumption and energy efficiencies of fluid power machinery conducted for the U.S. combined with conservative estimates for emissions by especially electric energy generation and recent emission estimates for construction machinery, the emissions by fluid power machinery are conservatively estimated to approximately 1,844 MMT CO_2/year in a 2022 scenario. It is found that this emission level corresponds to 2.3 times that of global aviation and approximately 5% of the global energy related emissions.

In addition, the emission reduction potentials offered by a full electrification of fluid power machinery, respectively a 50% increase in fluid power system efficiencies, is investigated, as well the combination of a full electrification and a 50% efficiency increase. The results are shown for comparison in Fig. 5.1., emphasizing the significant emission reduction potentials present under these conditions, separately, and combined.

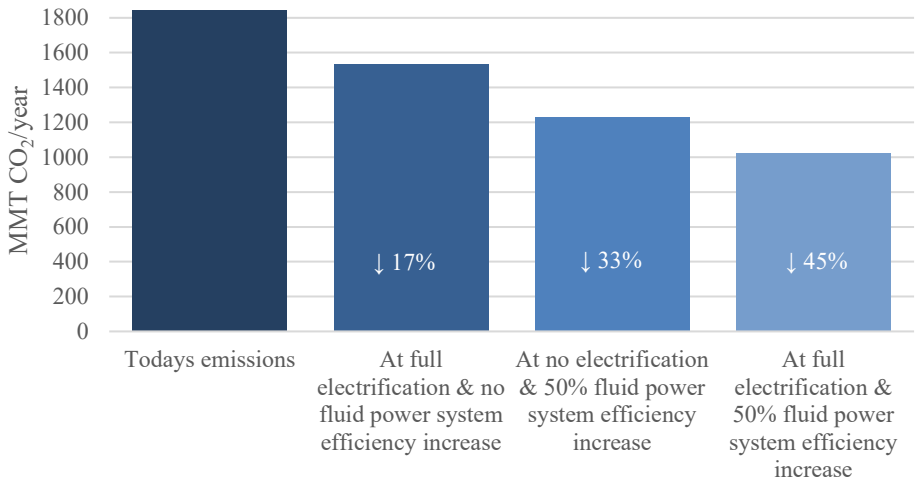


Figure 5.1. Comparison of emission reduction potentials.

In context of the global energy related emissions, the emission reduction potential for a full electrification of fluid power machinery is 0.8%, whereas the emission reduction potential of a 50% fluid power system efficiency increase is a 1.7% reduction in the global emissions. These results emphasize that the emission reduction potentials related to increasing fluid power system energy efficiencies by far exceeds that of electrifying fluid power machinery. However, the potential of a 50% fluid power system efficiency increase and a full electrification has the potential to reduce global emissions by 2.2%. This corresponds to an emission reduction of 822 MMT CO_2/year, exceeding the total emissions by global aviation.

From these estimates of emission reduction potentials, it is found that it is not the electrification of fluid power machinery, which is currently ongoing, that individually possess the largest emission reduction potential, but rather the increase of fluid power system efficiencies. Naturally, however, both increasing system level fluid power efficiencies while electrifying fluid power machinery should be key focus areas for research, development and innovation in the years to come.

6. REFERENCES

[1] Oak Ridge National Laboratory, Estimating the Impact (Energy, Emissions and Economics) of the U.S. Fluid Power Industry, 2012. https://info.ornl.gov/sites/publications/files/Pub28014.pdf

[2] International Energy Agency, CO_2 Emissions in 2022, 2022. https://www.iea.org/reports/co2-emissions-in-2022 (assessed on 18. October 2025)

[3] Our World in Data, Carbon Intensity of Electricity Generation, 2025. https://ourworldindata.org/grapher/carbon-intensity-electricity?tab=chart&time=earliest..2022&country=EU-27%7EOWID_EU27%7ECHN%7EUSA%7EOWID_WRL (assessed on 18. October 2025)

[4] European Environment Agency, Greenhouse gas emission intensity of electricity generation, EU level, 2024. https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emission-intensity-of-1-1751032678/greenhouse-gas-emission-intensity-of-electricity-generation-eu-level (assessed on 18. October 2025)

[5] United States Department of Transportation, Bureau of Transportation Statistics, Energy Consumption by Mode of Transportation, 2025. https://www.bts.gov/content/energy-consumption-mode-transportation-0 (assessed on 18. October 2025)

[6] The Engineering Toolbox, Energy Content in common Energy Sources, 2025. https://www.engineeringtoolbox.com/energy-content-d_868.html (assessed on 18. October 2025)

[7] United States Environmental Protection Agency, Greenhouse Gas Equivalencies Calculator - Calculations and References, 2025. https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator-calculations-and-references (assessed on 18. October 2025)

[8] Forest Research, Carbon Emissions of Different Fuels, 2025. https://www.forestresearch.gov.uk/tools-and-resources/fthr/biomass-energy-resources/reference-biomass/facts-figures/carbon-emissions-of-different-fuels/ (assessed on 18. October 2025)

[9] IDTechEx, Electric Construction Machines Vital for Greener Construction, 2022. https://www.idtechex.com/en/research-article/electric-construction-machines-vital-for-greener-construction/26187 (assessed on 18. October 2025)

[10] Construction Executive CE, Combating Construction Carbon: How Heavy Machinery Can Further Sustainability, 2025. https://constructionexec.com/article/combating-construction-carbon-how-heavy-machinery-can-further-sustainability/?utm_source=chatgpt.com (assessed on 4. January 2026)

[11] Danfoss, Roadmap for decarbonizing cities, Danfoss Impact, Issue no. 3, 2022. https://assets.danfoss.com/documents/latest/461190/BE512836097054en-000101.pdf (assessed on 4. January 2026)

[12] Danfoss, Decarbonizing construction machines, A Case Study from Danfoss, 2023. https://cdn.sanity.io/files/5zabm86v/production/9c5a7b30987d908f56d2c909b03cae1ffffd92cb.pdf (assessed on 4. January 2026)

[13] Nguyen TH, Do TC, Nguyen MH, Ahn JH, Ahn KK, 'Energy consumption reduction technologies in hybrid hydraulic construction machine: A review', Renewable and Sustainable Energy Reviews (225) 2026. https://doi.org/10.1016/j.rser.2025.116150

[14] International Energy Agency, CO_2 Emissions in Aviation in the Net Zero Scenario, 2000-2030, 2025 https://www.iea.org/energy-system/transport/aviation (assessed on 18. October 2025)

[15] Sprink J, Schmitz K, 'Product Carbon Footprint of Hydraulic and Pneumatic Components – Challenges in Accounting and Comparability', Proceedings of the 14th International Fluid Power Conference, 2024. https://doi.org/10.13052/rp-9788770042222

[16] Waerder M, Merkelbach S, Schmitz K, 'Carbon Footprint Determination of an Electro-hydraulic Compact Axle (EHA) based on DIN EN ISO 14067'. Proceedings of the 16th Scandinavian International Conference on Fluid Power, 2019. https://publications.rwth-aachen.de/record/762608 (Assessed on 22. October 2025)

[17] Rückert M, Merkelbach S, Alt R, Schmitz K, 'Online Life Cycle Assessment for Fluid Power Manufacturing Systems – Challenges and Opportunities', Advances in Production Management Systems. Smart Manufacturing for Industry 4.0, Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-99707-0_17

[18] Oak Ridge National Laboratory, Estimating Energy Consumption of Mobile Fluid Power in the United States, 2017. https://docs.nrel.gov/docs/fy18osti/70240.pdf

[19] European Commission, Electric Motors and variable speed drives - Ecodesign requirements apply to these products, 2024 https://energy-efficient-products.ec.europa.eu/product-list/electric-motors_en (assessed on 5. January 2026)

[20] TopMotors, Fact Sheet No. 32 – Optimizing electric motor-driven systems, 2021. https://www.topmotors.ch/sites/default/files/2021-12/E_MB_32_Optimizing_electric_motordriven.pdf (assessed on 5. January 2026)

[21] Control Engineering, Go Beyond Motors to Reduce Energy Use, 2009. https://www.controleng.com/go-beyond-motors-to-reduce-energy-use/ (assessed on 22. October 2025)

[22] U.S. Department of Energy, FY 2008 Progress report for Advanced Combustion Engine Technologies, 2008. https://www.energy.gov/sites/prod/files/2014/03/f8/2008_adv_combustion_engine.pdf (assessed on 22. October 2025)

[23] ICCT20, The Ever-improving Efficiency of the Diesel Engine, 2015. https://theicct.org/the-ever-improving-efficiency-of-the-diesel-engine/ (assessed on 22. October 2025)

[24] Intergovernmental Panel on Climate Change, IPCC Fourth Assessment Report: Climate Change 2007, 2007. https://archive.ipcc.ch/publications_and_data/ar4/wg3/en/ch5s5-3-1-2.html (assessed on 22. October 2025)

[25] Minav, T, Electric-Drive-Based Control and Electric Energy Regeneration in a Hydraulic System. Lappeenranta University of Technology: PhD Thesis, 2011. https://lutpub.lut.fi/bitstream/handle/10024/69968/isbn%209789522651044.pdf?sequence=2&isAllowed=y

[26] Koitto T, Kauranne H, Calonius O, Minav T, Pietola M, 'Experimental Investigation of a Directly Driven Hydraulic Unit in an Industrial Application', Proceedings of the 11th International Fluid Power Conference, 2024. https://publications.rwth-aachen.de/record/726059/files/726059.pdf

Biographies



Lasse Schmidt received the M.Sc. degree in engineering (mechatronics) from Aalborg University, Denmark, in 2008. From 2008 he was with the application engineering group of Bosch Rexroth A/S, Denmark, and from 2010 an industrial Ph.D. fellow also associated with Aalborg University. He received the Ph.D. degree in robust control of hydraulic cylinder drives in 2014. Subsequently, he has been a postdoctoral researcher at AAU Energy while concurrently being with Bosch Rexroth AG. Hereafter he became an Assistant Professor and later an Associate Professor with AAU Energy. He is heading research activities related to electro-hydraulic drive network technology, a field in which he is the founder of the fundamental design and control principles. He has authored more than 70 scientific peer-reviewed publications, and his current research interests are in design and control of electro-hydraulic drives and their integration into both mobile industrial machines.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated in 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is a full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she has held since 2020. Prof. Schmitz's awards and honors include several best paper awards and the 2023 IMechE Joseph Bramah Medal award.



Tatiana Minav (ASME and IEEE member) is Associate professor in IHA –Innovative Hydraulics and Automation lab, Tampere University, Finland. She has over 18 years' experience working on improving the efficiency of non-road mobile working machines. She has authored 140 scientific publications in total. Her current interests include zonal hydraulics, electro-hydrostatic and electro-mechanical actuators, sensorless position control with direct electric drive in hydraulic systems, AI-based condition monitoring and predictive maintenance, thermal management, energy balance and energy recovery systems in non-road mobile working machines.



Adolfo Senatore started his academic career in 1984, and since 2001 he is a Full Professor and director of Fluid Power Laboratory of University of Naple Federico II, Italy. From 2001 to 2016, he served as President of the Degree Course Council in Mechanical Engineering and from 2014 to 2016 as Teaching Coordinator of the Faculty of Engineering at the University of Naples. From 2017 to 2019 and was President of the Campania TEST Interuniversity Consortium. Professor Senatore was and is responsible for various projects financed with Italian ministerial funds and European funds and industrial funds such as Ferrari, Porsche, Duplomatic, Fluid-O-Tech. His research group consists of 3 permanent persons and about 6 PhD students and Post Doc. His research activities, since the beginning of his career, are documented by around 250 scientific publications and monographic works.

Decarbonization in mobile machinery – from reducing the carbon emissions to emission-free drivetrains

Dr. Andreas Schumacher

Dr. Andreas Schumacher, Danfoss Power Solutions GmbH & CO OHG, aschumacher@danfoss.com

Abstract.

Here in Europe, the carbon emissions of components and machinery are getting more in focus. A lot of regulations are driving this, but also the market demand increases for reduced carbon footprint to also emission-free vehicles. As the use-phase is typically in mobile machinery the by far and outmost the biggest contributor to the CO2-footprint of a machine, reducing energy consumption is key. Significant contributions are coming surely from the typical prime energy source, the engine, but also from the drivetrain as well as the work hydraulics. To reduce the energy consumption of these functions, solutions can be applied with a three-level approach:

- The first level is based on the components.
- The next level is on (sub-)systems and software.
- The last level is on the machine and the prime energy source, where e.g. switching from diesel to battery-electric also leads to significant decarbonization.

The paper will show the impact on all three levels which can be done with Danfoss components in more detail. The potential shown will be backed up with measurement data. So it will become clear, that the journey towards zero emission vehicles is possible, it can start today on these different levels on the machines, also with very high impact on reducing the TCO of a machine as highly efficient components and systems not only reduce the CO2e emissions but also reduce operation costs.

Keywords. Datacenter design, energy efficiency of datacenter, energy efficient metrics, datacenter carbon footprint computation.

1. INTRODUCTION

In the current situation, especially in Europe, the carbon footprint of products and their usage is becoming increasingly important. Several drivers, from the regulatory site, reporting site and application requirements demand the labelling, publishing and reducing of the carbon footprint of products. The publicly funded research project Fluid 4.0 [1] develops methodologies to share among further product information also CO_2-footprint data of products [2].

Reducing carbon footprints is in mobile machinery also important. There are different regulatory drivers, requiring the report of CO_2 emissions during production and usage of components and complete machines, for example as part of the reporting obligations CSR-D/ ESRS, there as part of product or company carbon footprint reporting (PCF and CCF). Another level of reporting the product or company carbon footprint takes place, when a public sustainability commitment has been engaged. Good examples are here commitments with the Science-based Targets Initiative SBTi [3] or the Carbon Disclosure Project CDP [4]. An obvious consequence is when reporting the PCF of a product, that there is an intrinsic motivation of optimizing the footprint, meaning reducing the product emissions.

Another aspect leading to the need to reduce the carbon footprint of mobile machinery becomes imminent if external regulations require reduced emissions in specific application cases. Here, good examples are indoor operations like with areal lifts in buildings or farm machinery in stables. Or if emission reductions are prerequisites for working in specific areas, like construction sites in dense cities, requiring emission-reduced or even emission free machinery and equipment.

And last but not the least, the bigger-picture-based motivation for emission reductions of machinery is based on the aspect of doing the right thing for our climate. Reducing the emission of CO_2 is scientifically proven the right thing to do to limit the progress of climate change. As this applies for all emitting machines – and mobile machines often can reach big sizes with big power needs – improving the emissions of mobile machinery is a smart move for climate change prevention.

All these motivations to reduce the carbon footprint of a mobile machinery have the same baselining challenge: Where to start and what is an effective way to reduce emissions, especially in the use phase with all the challenges to determine use-phase emissions [5] ?

2. POWER AND LOSSES IN A TYPICAL HYDROSTATIC DRIVETRAIN

For gaining an inside of what impact the different components in typical mobile machinery drives have, the example of a standard hydrostatic drivetrain is being analyzed. **Figure 1** depicts such a drivetrain in a Sankey diagram, which is here, to represent the most common standard architecture, energized by a combustion engine, typically a diesel engine. The power of the engine as prime mover is distributed to the work functions (not shown in that figure) and to the drivetrain, represented with the hydraulic pump and motor to form the hydrostatic transmission. The transmission supplies the power via a gearbox to the wheels. For the sake of simplification, the wheels are not represented in Figure 1, so the usable energy to propel the machine is emitted at the gearbox, here.

Figure 1 also includes typical maximum efficiency values for the components in the drivetrain, starting from the engine with around 40% total efficiency in best points. With these given values, assuming all components are operated in their best points, from 100 *kW* input power only nearly 35 *kW* are usable at the gearbox output.

It also needs to be considered, that the maximum efficiencies are listed in Figure 1. That means, in other words, in normal operations, when engine and hydrostatic transmission

dynamically sweep through their operation and efficiency maps, lower component and system efficiencies are very likely.

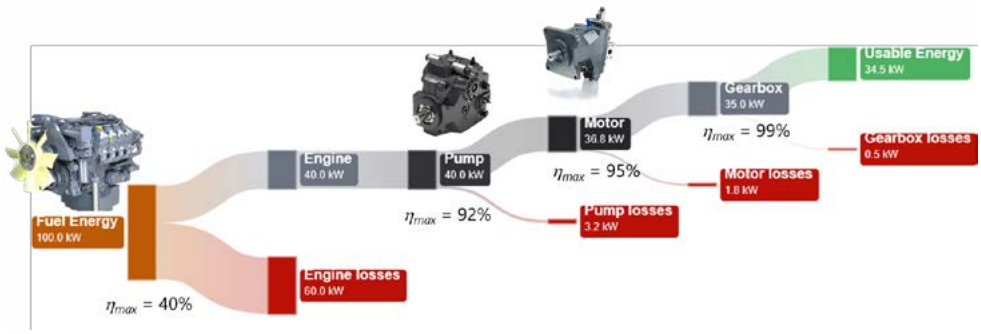


Figure 1: Power- and loss-flow in a standard hydrostatic drivetrain

Based on the illustrative losses of the individual components, Figure 1 makes it clear that going from engine to the wheels has the biggest impact, gaining reductions, but that is also the most complex path from the change-perspective. Hence, depending on the target one has in decarbonization, swapping the prime energy supply might lead to too complex solutions to begin with, especially taking into account that for some bigger mobile machinery the energy demand for a regular working day exceeds what alternative decarbonized energy sources can deliver, like batteries or hydrogen H_2. Their storage would become way too big to support the equivalent energy supply to diesel fuel.

Thus, it is very reasonable to analyze from wheels to the engine, excluding the gearbox and start with the hydrostatic motor. This also gives an estimation of what level of change is sufficient to realize the sought decarbonization level in the drivetrain.

3. DECARBONIZATION ON COMPONENT LEVEL

Starting the impact analysis as mentioned with a hydraulic motor, to derive a better understanding of what is the impact in efficiency and CO_2-emissions, two different types of hydraulic motors used in transmissions will be discussed in this chapter. A typical configuration of a hydrostatic transmission consists of a variable displacement pump in axial piston design paired up with a fixed motor.

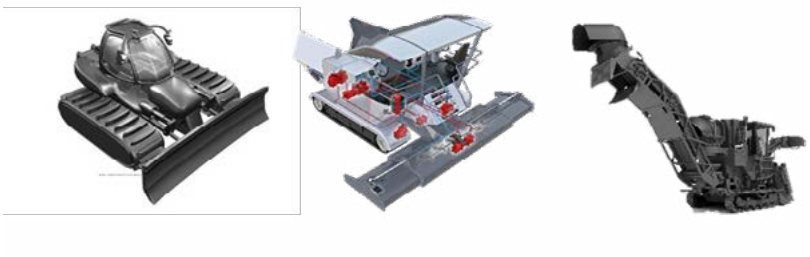


Figure 2: Power and loss-flow in a hydrostatic drivetrain [Source right image: Internet]

This configuration is often used in smaller machines where the cost impact has a high priority or also in more complex machines with a so-called dual path configuration – driving two tracks or wheels in a skid-steering architecture, demonstrated in **Figure 2**. These vehicles are not intended for high speeds and thus operate well with a fixed motor - compared to a variable motor.

To get a fair comparison, the vehicle example used here is a smaller crawler as shown in Figure 2 on the left side. In this setup, the comparison is made between version 1, consisting of a Danfoss S90MF075 swashplate-technology motor with 75 cm^3/rev as the predecessor and a Danfoss H1F080 bent-axis motor with 80 cm^3/rev as the new product with higher efficiency. But just the comparison of the peak efficiency is not giving satisfying information what might be achieved in normal machine operation. To overcome this, two typical operation conditions of the crawler are compared, which are:

- **Working:** Lower speed and high tractive force; approx. 5 *km/h* and ~75 *kN* tractive force. This represents the condition when the crawler moves material by pushing it forward.
- **Transport:** Highest speed and medium tractive force; approx. 13 *km/h* and ~40 *kN* tractive force, representing the machine returning to the starting point after pushing material to continue moving material or relocating to a different location for work.

As this comparison focuses on the impact of motor technology, the pump as well as the further drivetrain components are not shown in this comparison, it will be assumed that the difference in motor displacement is compensated by the gearbox driving the sprocket so that the same speeds and tractive forces are achieved on machine level.

The results of this comparison are shown in the next figure, **Figure 3**, where the resulting advantage of the new H1F is demonstrated in the bar chart of the predecessor-product S90MF. The gray bars are for the S90 motor, the red for the new H1F.

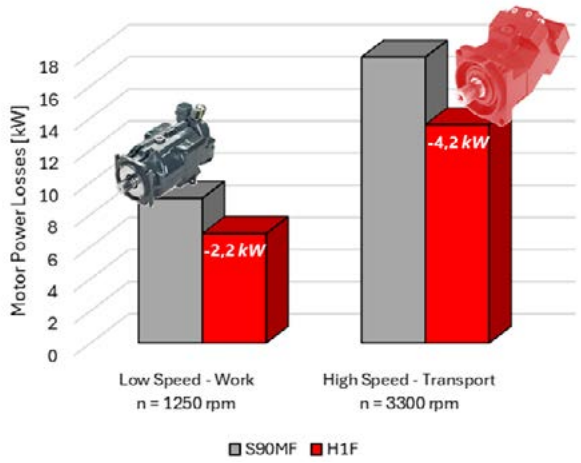


Figure 3: Results of the losses comparison S90 vs H1F-motor in two operation conditions

The diagram shows clearly that the new technology has an impact when analyzed in real operation conditions. The losses for the low-speed working conditions and in the higher-speed conditions are on a similar reduction level close to 25% lesser that using the older product with the swashplate technology – or in other words, the H1F saved 2,2 *kW* in the low-speed and 4,2 *kW* in the high-speed condition. To calculate the impact on the machine's CO_2-footprint, an approximation of the usage of the machine needs to be determined:

- Small Crawler with typical yearly 1000 *hrs* work, 300 *hrs* transport, and 8 *yrs* machine lifetime. That results in typical 10.400 *hrs* total machine lifetime
- In that load cycle, 17,6 *MWh* losses are avoided during Work and 33,6 *MWh* losses avoided in Transport conditions, so 51,2 *MWh* in total.
- Typical diesel engine with 215 *g/kWh* fuel consumption – this value is lower than the best point and thus represents variable load conditions with around 35% engine efficiency.

With these approximations and an emission factor of 2,9 *kg* CO_2e per liter diesel fuel, it can be calculated that the application of an H1F **avoids 38,3 *to* CO_2e** per machine life through a 2% weighted efficiency increase only compared to the S90MF – by just exchanging this single component in a drivetrain.

4. DECARBONIZATION ON SYSTEM LEVEL

If a higher decarbonization level is desired, the system level offers higher reduction potential. System level, that might either comprise of multiple components, making a hydraulic system or it can also take the control software into account. In this chapter, the key element leading to CO_2-reductions is the Danfoss Digital Displacement® Pump DDP. That pump is combining multiple smaller pump units in one housing. The pump units are based on a variable radial piston pump design whereas the variability is achieved by connecting and disconnecting the piston during its stroke to/from the outlet. The pistons of the pump can be actuated all individually, run on different pressure levels and they can also be completely deactivated (not connected), so that Power on Demand is possible without efficiency compromises. **Figure 4** depicts two of these pumps, whereas the left is the single DDP096 and the right is a double pump DDP1X0D. The principle of switchable radial piston design delivers very high efficiency also at part stroke and additionally features very high reaction times in changing the displacement. This allows loading the engine at high levels as stalling can be avoided through quick unloading through the pump.



Figure 4: Digital Displacement pump in single (DDP096) and double version (DDP1X0D)

This quick reaction of the pump being applied to an excavator can – with an adapted control software – deliver higher efficiency, as **Figure 5** is explaining. Normally, to prevent the engine from getting overloaded at sudden load-increase situations, the engine is not loaded to its maximum capability, some safety margin is left. That can be avoided with a DDP, the engine can be loaded higher, close to its maximum and thus also operate in areas with higher efficiency.

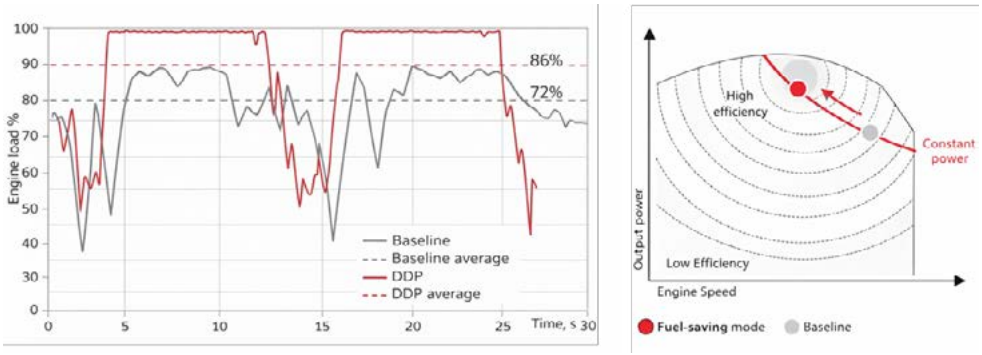


Figure 5: Adapting the control software to the quick reaction characteristics of the DDP

In the left part of Figure 5 the loading of the engine is depicted, in grey the unchanged baseline system with an average loading of 72% of the engine and in red the DDP pump system with adapted control software, where the average load can be increased up to 86%, due to quicker overload prevention. To demonstrate the effect onto the engine, the right diagram is an illustrative example of a fuel consumption shell diagram. The baseline, to reserve some reaction margin through lower loading, operates the engine at lower efficiency conditions (grey point), whereas the optimized system with the DDP moves the operation point of the engine through higher loadings to areas with higher efficiencies. In measurements, published in [6], 21% fuel saving was obtained, although in normal operation conditions, the practical fuel consumption reductions achieve between 10-15%, also dependent on the baseline system design.

The decarbonization potential is further increased when the hydraulic characteristics of the DDP are fully used – the independent use of the different pumping unit outlets. This will now be discussed by taking a classical open-circuit work-function system with a pump,

valves and cylinders as an example, like in an excavator boom. In typical hydraulic circuit designs in mobile machinery, an axial piston pump is combined with a proportional valve to actuate the consumer(s). Those systems are typically very prone to throttling losses, especially if multiple consumers with a different flow and pressure demand are supplied simultaneously.

To overcome these conditions, multiple solutions are possible. One is the direct drive of an actuator with a pump, including changing into closed circuit setup, if the drive needs to be reversable. That means, the actuator – a cylinder or motor – is directly coupled to a pump which controls the speed through changing its displacement – aka displacement controlled [7], [8]. Thus, that avoids the valve losses. The downside of this solution is higher installation costs as per consumer one pump is required.

Another solution, which will be discussed here further, is using the characteristics of the DDP. That pump can also be seen as an enhancement of the just described displacement-controlled principle through combining multiple smaller pump units in one housing. The system impact of an installation with these pumps is explained in **Figure 6**, where a classical architecture ("Baseline") on the left is compared with a DDP-architecture on the right ("System Architecture 2").

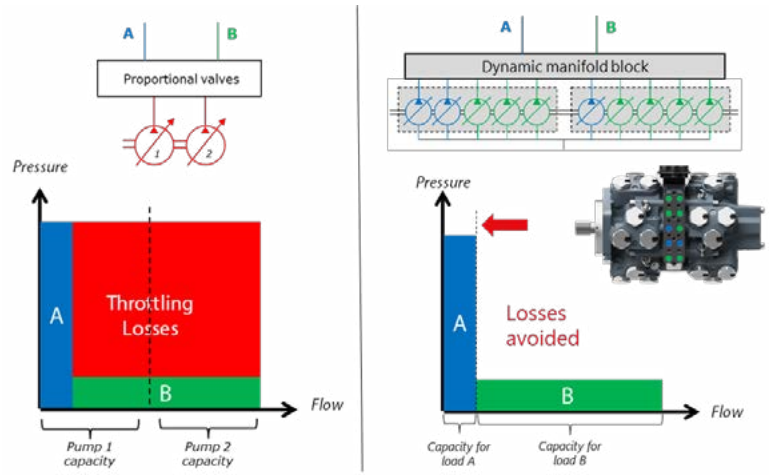


Figure 6: Results of the losses comparison S90 vs H1F-motor in two operation conditions

The baseline system consists of two pumps, delivering flow into a valve block out of which two consumers A and B are supported. The System Architecture 2 uses the double DDP pump with 10 pump units and a dynamic manifold block, out of which the same two consumers A and B are supported. For both systems, the flow and pressure demand is the same, consumer A demands a small flow at a high pressure and B demands a high flow at low pressure. In the classical system, as the flow demand of B is bigger than what only one pump can deliver, both pumps need to be combined and thus run also on the same pressure level – defined by the higher pressure demand from A. For B, the higher pressure needs to be throttled away by the valve block, resulting in the losses indicated by the red area in Figure 6. In the DDP system, with the dynamic manifold, two pistons (blue) will be combined to support the demands of A, the remaining 8 piston suffice the flow demand of B. As the pistons run pressure independently, the throttling losses of the classic system design will be avoided. To

estimate and measure the impact of System Architecture 2, [9] has made simulations and tests with an excavator. The outcome is – in short:

- The simulation of a digging cycle delivered that for a classical unchanged **baseline version** of an excavator a typical average **57*kW*** is transmitted from the engine into the hydraulic system (at engine/pump interface).
- The excavator with the **DDP** architecture consumes in same digging cycle **42,3 *kW*** at that pump/engine interface. That is 26% less than the baseline and is composed of a pump loss reduction of 4,4 *kW* (-60%) and 10,5 *kW* in the main control valve (-45%).
- **Measurements** show a realistic fuel consumption reduction of around **25 – 30%** in typical digging cycles. That validates the simulation results well and took also some further reduction potential into account that, due to lower loads, the engine runs in different, more efficient operation conditions.

The fuel saving potential equates roughly with the CO_2-emission reductions, but the detailed impact will be discussed in the next chapter, where also the decarbonization potential of an electrified system will be discussed.

5. DECARBONIZATION POTENTIAL OF E-POWERED MACHINES

Following the system comparison, where "Baseline" is the unmodified standard system of an excavator und "System Architecture 2" in the modified system with a DDP, the next step in analyzing the decarbonization potential is to compare different prime energy sources and their carbon emissions. The natural comparison is a diesel engine to electrification.

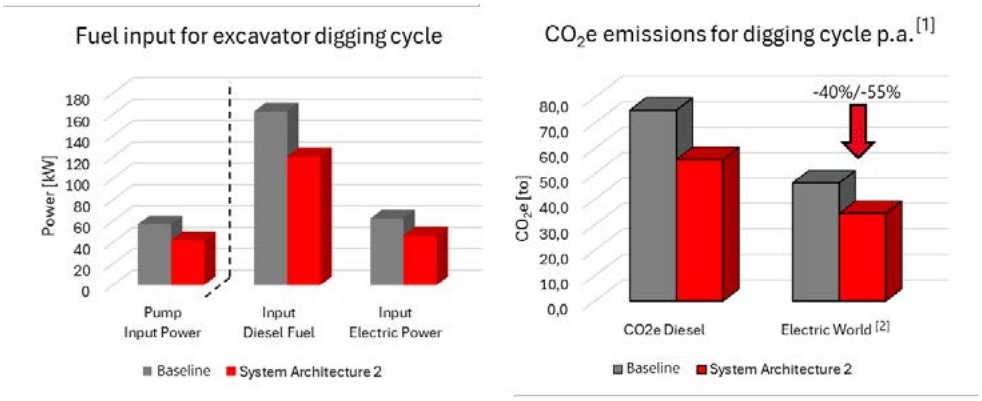


Figure 7: Comparing CO_2 emissions of diesel and electric systems in the excavator [1],[2]

Although there are other fuels available for reducing the CO_2-footprint like HVO, LPG or H_2, the most common alternatives are still diesel and electrification with batteries. Thus, for the above-mentioned excavator, the difference between these both will be further analyzed.

[1] Digging cycle per year based on an 8hrs shift and 220 workdays digging
[2] Source of grid emission factors (world/regions): 2024 IEA data

Figure 7 introduces the results of the comparison: Starting on the far-left side with the power at the engine/pump interface, 57 *kW* for the baseline and 42,3 *kW* for the modified system. To be able to realize a CO_2-comparison, the total fuel/electricity consumption needs to be estimated, which is also shown in the left diagram from Figure 7:

- **Diesel:** 163 *kW*/121 *kW* fuel power for Baseline/System Architecture 2 with 215 $^{g}/_{kWh}$ specific fuel consumption or 35% engine efficiency in the digging cycle. Based on this power consumption, System Architecture 2 could save up to 110 *to* CO_2e on a 10.000 *h* machine life.
- **Electric:** 63 *kW*/46 *kW* with using a Danfoss Editron PMI motor with an average 93% and Editron Inverter with an average 98% efficiency during the same cycle (no battery loading/unloading considered here to incorporate cable-powered machines, too)

The highest prime energy and emissions saving can be obtained if the DDP-system is combined with an electric system. The decarbonization potential of electrified systems significantly depends on the electric power generation. The right diagram of Figure 7 depicts that for a diesel engine 75 *to*/56 *to* CO_2e are emitted in a year's time, based on this digging cycle, versus 46 *to*/34 *to* CO_2e for the Baseline and DDP-system with electricity as prime energy – and the CO_2e calculation based on the IEA world grid factors [10]. While the power saving in an electrified System Architecture 2 is approx. $^{2}/_{3}$ over the diesel-driven baseline, the CO_2e emissions are only 55% lower – explained through the power mix in the world grid factors, as that also factors power generation with coal and gas into account. Still, this equates to a machine lifetime saving of around 230 *to* CO_2e in this use case.

The impact of the regional power generation and their CO_2e-factors is shown in **Figure 8**, only for System Architecture 2 (DDP-system).

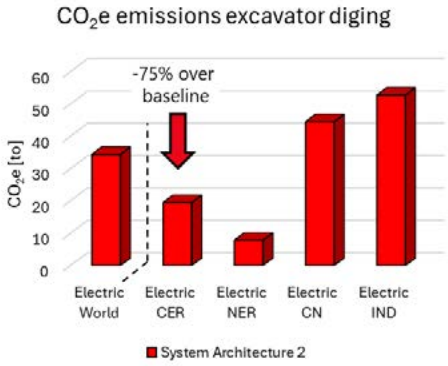


Figure 8: Regional power generation impact of CO_2 emissions

With the world grid emissions equivalent on the far-left side, the other bars in Figure 8 show big the power generation impact is. While in the Northern-European regions NER the renewables in the power generation dominate and thus the CO_2-emissions for the excavator with System Architecture 2 are very small, lower than 10 *to/year*, the same excavator with the same operation cycle would create only 20% less emissions than with a diesel engine in China and nearly the same emissions as a diesel excavator if used in India, as shown in

Figure 7 right part (all with System Architecture 2). This comparison shows that especially in electrification, which is often considered as zero emissions, the emissions during production do have a significant impact on the total emissions of a mobile machine in the use phase. A possibility to capture the information where the machines are used and store for further analyses might be using the emission and energy-reporting functionalities of an Asset Administration Shell, like developed in the Fluid 4.0 project [1].

6. SUMMARY AND CONCLUSIONS

In this paper, several steps towards decarbonization were discussed and presented, in ascending order of the impact. The smallest impact is achieved if only components of a system will be exchanged to higher efficient components. The example of swapping the S90 fixed swashplate motor to the new H1F bent axis fixed motor delivered in the given application a considerable positive potential of saving up to 10% of CO_2e over the machine lifetime. The next solution discussed is optimizing the system. In the given example, the DDP was the key element. It can contribute with its differentiating characteristics on the one side to a different engine control algorithm, allowing to optimize the engine operation points – accounting up to 25% CO_2e saving potential – and on top of that, on the other side, it allows a redesign in the hydraulic system which especially pays off in valve-based work function systems by reducing throttling losses. This would account for an additional efficiency increase in best case up to 30% in a typical excavator system.

If, in a possible last step, the diesel engine will be replaced by a battery-electric system, the impact is potentially even higher as electric systems have in a broad cycle average 90% efficiency to convert the electric energy into mechanical energy, where the engine efficiency is around 40% – only in its best point and in part-load conditions significantly worse. Nonetheless, an additional factor to determine the CO_2-footprint impact is how the electricity used is generated. In a worst-case scenario, like in India (according to the IEA), the energy mix eliminates the efficiency advantages of the electrification and creates a wash on CO_2-emission level, in a best-case scenario when the machine is used in the Northern European Region NER, the decarbonization potential is up to 90% compared to a diesel engine.

A step in between, not requiring changing the machine architecture as drastic as going to a battery-electric machine, would be the use of alternative fuels. A fuel which recently gains attention is HVO, Hydrotreated Vegetable Oil. It has the potential to reduce CO_2e emissions by up to 90% and has a high engine compatibility, but the availability in bigger masses is currently still a challenge, in addition to a higher price challenging TCO. Furthermore, LPG and CNG are also alternatives with lower CO_2e-emissions, but their advantage is compensated by the lesser efficiency of the gas engines and thus equates in a wash on emissions-level. Hydrogen/H_2 is also an emerging alternative, whereof the machine design for storage and depending on the technology used (combustion engine or fuel cell) is also subjected to a radical redesign in its architecture.

Closing the loop to the motivation of machine decarbonization, depending on the goal and the requirements a machine manufacturer wants to fulfill, several steps and their impact have been highlighted in this paper. It can be used as a guide determining what to do to achieve a certain level of decarbonization. Component swap could be sufficient for lower

decarbonization goals or gearing up to electrification to make the vehicle locally emission free, where required. The in Fluid 4.0 developed toolchain [2] for the Asset Administration Shells could be a tool to track the machine's emissions during their use-phase to gain further information about real-time and real-life decarbonization effects.

Danfoss Power Solutions offers products and solutions to accompany mobile machine manufacturers on their decarbonization journey in all steps towards reaching their goals, from application knowledge, components and systems as well as software solutions for classical hydraulic as well as electric machines. Simulative and test-based decarbonization estimation is part of the support, to make future machines ready for a sustainable future and a world worth living in it.

7. REFERENCES

[1] "Fluid 4.0 Research Project." Dec. 01, 2025. [Online]. Available: https://fluid40.de

[2] IDTA, *Submodel Carbon Footprint*, 02023, Mar. 2025. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2025/03/IDTA-02023_Submodel_CarbonFootprint.pdf

[3] Science-based Targets Initiative, Jan. 22, 2026. [Online]. Available https://science-basedtargets.org/

[4] Carbon Disclosure Project, Jan. 22, 2026. [Online]. Available www.cdp.net

[5] Schumacher, A. Hanpantsurova, O., CO_2 use phase calculation – a status update from the (Fluid 4.0) project work, Hannover Industry fair presentation, 2025

[6] Green et al., DEXTER – The Application of a Digital Displacement® Pump to a 16 tonne Excavator, FPMC/Bath, 2018

[7] Rahmfeld, R.; Ivantysynova, M.; Weber, J.: Displacement controlled wheel loader — a simple and clever solution. Dresden, 4th International Fluid-power colloquium, 2004

[8] Schneider, M., Koch, O., Weber, J., Green Wheel Loader – improving fuel economy through energy efficient drive and control concepts, Dresden, 10th International Fluid-power colloquium, 2016

[9] Lagarde et al., Danfoss Digital Displacement & Editron: An efficient electro-hydraulic system for mobile applications, 13th IFK/Aachen, 2022

[10] IEA International Energy Agency, Data and statistics, 2024 dataset, available: https://www.iea.org/data-and-statistics/data-sets

Biography



Senior Director **Dr. Andreas Schumacher** is responsible for Sustainability and Product Compliance for the complete Danfoss Power Solutions segment. He joined Danfoss in 2007 as a simulation engineer after his PhD in mobile machinery tire and driving dynamics simulations at the Technical University of Braunschweig under the supervision of Professor H.-H. Harms. The career development over the nearly 20 years in Danfoss increased his responsibilities and changed from Engineering to the current role. Dr. Schumacher is also author and co-author of over 30 journal and conference papers as well as presentations. As part of his current role, he is representing Danfoss Power Solutions externally in the field of mainly Sustainability in regulatory, scientific/research and business contexts and further publications like publicly available podcasts (e.g. on Spotify).

Calculating use-phase greenhouse gas emissions of hydraulic components

Felix Ruthmann[1,2,*], Prof. Dr.-Ing. Marcus Geimer[1], Prof. Dr.-Ing. Gerhard Geerling[2], Ulrich Lenzgeiger[2]

[1]Instute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), Rintheimer Queralle 2, 76131 Karlsruhe, Germany
[2]Bosch Rexroth AG, Glockeraustrasse 2, 89275 Elchingen, Germany
**Corresponding author: felix.ruthmann@partner.kit.edu*

Abstract.

Climate change, increasing regulation, and heightened customer demand require transparent emissions calculations and sustainable product development. Currently, no standardized procedure to quantify use-phase greenhouse gas emissions of hydraulic components exists. On machine level emissions can be measured at the tailpipe or calculated, but the link to components is missing. This paper addresses this gap by developing and validating a methodology as baseline for further standardization. Using a wheel loader as a reference, two allocation approaches are derived for a hydraulic pump in a travel drive: First an approach based on the component's energy loss and second an extended approach that additionally accounts for primary energy supply efficiency. Meaningful comparisons across energy carriers such as diesel and electricity require qualitative emission results from machine level being reflected on component level. The proposed methodology provides, to the authors knowledge for the first time, such a framework for calculating use-phase greenhouse gas (GHG) emissions of hydraulic components. Further, it establishes the methodological foundation for assessing the impact energy carriers on component emissions and supports sustainability-oriented design decisions in fluid power systems. This work thus contributes to the creation of a standardized emissions assessment method that allows companies to promote sustainable products and differentiate themselves in the market through reduced energy losses and thus lower emissions.

Keywords. greenhouse gas emissions, use phase, hydraulic components, product carbon footprint, non-road mobile machinery, mobile working machines.

1. INTRODUCTION

Hydraulic machinery, including non-road mobile machinery (NRMM), is estimated to contribute nearly four percent to global greenhouse gas (GHG) emissions [1], thereby significantly influencing climate change. In response to increasingly stricter regulatory frameworks – such as the EU Green Deal [2] – industrial stakeholders are under growing pressure to implement credible emission reduction strategies. Compliance with these

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

requirements is not only a legal obligation but also gaining economic importance, as companies seek to avoid penalties and strengthen their market competitiveness. Consequently, many companies are establishing strategic sustainability targets that rely on robust and comparable metrics, particularly for GHG emissions. Customers and investors are also increasingly demanding such transparency, making these metrics a key differentiator.

The concept of a "footprint" as a quantified measure of the ecological impact of human activities has gained significant traction across society. Given the urgent need for climate action, the quantification of GHG emissions has become highly relevant. The Product Carbon Footprint (PCF) refers to the total GHG emissions associated with a product across its entire life cycle, expressed in carbon dioxide equivalents (CO_2e). This measure encompasses all greenhouse gases, not only carbon dioxide. [3]

Several frameworks exist for calculating GHG emissions, including the Greenhouse Gas Protocol (GHG Protocol), DIN EN ISO 14067, and the Product Environmental Footprint (PEF). These frameworks offer standardized guidance for companies to systematically assess, quantify, and reduce GHG emissions [4], [5], [6]. However, while they provide general methodological scaffolding, they leave room for interpretation. To increase comparability and specificity, Product Category Rules (PCRs) have emerged [7]. These are often initiated by industry consortia to define more granular rules for specific product groups. Furthermore, more specific standards like DIN EN ISO 14083 (ISO 14083) can arise. This specific one focusses on emission accounting for supply chains [8]. Currently, no PCRs or specific standards exist for either NRMM or hydraulic components to the authors knowledge. Some companies have disclosed internal procedures for PCF calculation, yet these remain proprietary and are not standardized. The existing more specific standards are mainly used by companies for calculating cradle-to-gate GHG emissions and do not offer companies enough guidance to calculate use-phase emissions.

The present work addresses this gap by developing a methodology as baseline for further standardization in calculating use-phase GHG emissions of hydraulic components in NRMM, while accordance with existing standards is not the focus. The goal is to provide a methodology that enables companies to transparently assess and reduce these emissions of individual hydraulic components including comparison of application in NRMM using different energy carriers. Firstly, section 2 introduces a representative reference machine. Based on this, a method for calculating use-phase GHG emissions at machine level is illustrated. In the end of section 2 the research question is derived describing the challenge of emissions attribution to component level, especially with different energy carriers. In section 3 a methodological approach based on component's energy losses for assigning use-phase GHG emissions is developed, validated and discussed for an exemplary hydraulic component. In section 4 this approach is extended to include the primary energy supply to achieve consistent results between component and machine level emission calculations with different energy carriers.

2. GHG EMISSIONS OF REFERENCE WHEEL LOADER

2.1. Definition of the Reference Wheel Loader

To understand the relevant interdependencies and to ensure comparable emissions results, it is essential to first define system boundaries. A wheel loader serves as the reference machine in this study. Using a system perspective similar to [9], the wheel loader as a machine can

be structured into systems such as the primary energy supply, travel and implement drives, and auxiliary systems, shown in Figure 1.

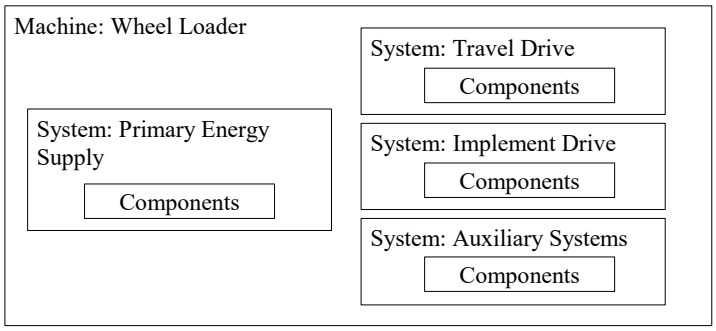


Figure 1: Structuring of wheel loader (based on [9])

The primary energy supply can be realized with various energy carriers, including diesel fuel, electrical energy, or hydrogen [9]. In this study, two configurations are examined:

- A diesel-powered wheel loader.
- A battery-powered wheel loader.

For the diesel variant, the primary energy supply consists of a fuel tank and an internal combustion engine on component level. In contrast, the electric version replaces these components with a battery, an inverter, and an electric motor. Other topologies for electric powertrains are not considered here to maintain conceptual clarity.

This study focuses on component level analysis of an axial piston unit (swashplate type) as pump within the hydrostatic travel drive system (closed circuit). The internal charge pump is excluded from the analysis for simplification. It is further assumed that the operating points of the hydraulic pump remain identical across both energy configurations.

2.2. GHG Emissions Calculation of Wheel Loader

Initially, the reference machine is analysed in both energy configurations. The methodology for calculating use-phase GHG emissions of the reference machine is based on established frameworks and industry practices in automotive and construction equipment sectors like [10].

The cited share of around 4 % of global GHG emissions coming from hydraulic machinery refers to emissions from diesel combustion [1]. These emissions can be measured at the tailpipe, but with the large variety of possible configurations of NRMM, this would lead to high measuring efforts being time- and cost-intensive. As an alternative, the GHG Protocol recommends calculating emissions via Activity Data and Emission Factors, as illustrated in Equation 2.1.

$$\text{GHG Emissions} = \text{Activity Data} \times \text{Emission Factor} \tag{2.1}$$

Activity Data quantify the extent of a process that emits GHG emissions (e.g. energy consumed). An Emission Factor denotes the GHG emissions per unit of activity, expressed in CO_2e. The chemical energy of combusted diesel fuel over an assumed reference lifetime represents the Activity Data of the diesel machine. Since measurement data of the amount

of fuel combusted are available from the diesel-powered wheel loader, it serves as the base case. From this, Activity Data for the electric variant are derived.

A usage profile was needed to obtain this measurement data. Unlike passenger cars, where standardized drive cycles like Worldwide Harmonized Light Vehicle Test Procedure (WLTP) exist, no standardized usage profiles are available for NRMM like wheel loaders [11]. Hence, a representative usage profile was used consisting of working and driving tasks as well as idling. Such a usage profile consists of typical cycles and their time shares within the overall operating period. This work provides no further details concerning the used cycles since they are subject to many parameters and therefore very individual.

During test drives with the wheel loader, the fuel consumption for each cycle was measured. Average values were calculated per cycle and weighted according to their time shares. Concluding, the average fuel consumption for the assumed usage profile of the wheel loader is 7.7 $\mathrm{l/h}$. With a density of 0.743 $\mathrm{kg/l}$ and a lower heating value of 42.5 $\mathrm{MJ/kg}$, according to [16], and an assumed reference lifetime of 10.000 h equates to 675.42 MWh, equaling the Activity Data of the diesel configuration ($\mathrm{E_{In,Diesel}}$).

A diesel engine efficiency of 40 % is assumed similar to [12] and [13]. In the electric configuration, the efficiency of battery, inverter, and electric motor is assumed to be 81 %. To provide the same useful energy, only around 49 % of the diesel input energy is required electrically equating to 330.96 MWh as Activity Data of the electric configuration ($\mathrm{E_{In,Electric}}$).

An **Emission Factor** is now needed to translate Activity Data into GHG emissions. If only tailpipe emissions were considered, the electric loader would appear emission-free. However, GHG emissions are also generated during electricity generation and supply. Hence, system boundaries must be distinguished [14]. For instance, [14] uses well-to-tank and tank-to-wheel which are also used by companies like Volvo Construction Equipment in their emission reporting [10]. ISO 14083 introduces the terms energy provision and operation for these boundaries [8]. ISO 14083 aims to define guidelines for quantifying GHG emissions from passenger and freight transport chains. The term GHG emissions from energy provision refers to the release of greenhouse gases during the production, storage, processing, and distribution of energy carriers for use in transport chains. GHG emissions from operation, on the other hand, refer to the release of greenhouse gases due to the operation of vehicles or locations. This enables ISO 14083 to also make statements about locations in logistics where the term tank-to-wheel would be inappropriate. At the same time, energy supply can serve as a direct substitute for well-to-tank and operation for tank-to-wheel in vehicles, which is why the terms used in ISO 14083 are used for this work. [15]

- Energy provision (well-to-tank) and
- Vehicle operation (tank-to-wheel).

The total use-phase GHG emissions for the diesel configuration ($\mathrm{GHG_{Machine,Diesel}}$) are calculated as:

$$\mathrm{GHG_{Machine,Diesel}} = \mathrm{E_{In,Diesel}} \times \left(\mathrm{EF_{Operation,Diesel}} + \mathrm{EF_{Provision,Diesel}}\right) \quad (2.2)$$

For diesel, Emission Factors from [16] are used within the framework of ISO 14083, which are also used here. According to [16], the Emission Factor for diesel combustion ($\mathrm{EF_{Operation,Diesel}}$) is 0.271 kg $\mathrm{CO_2e/kWh}$. An additional 0.081 kg $\mathrm{CO_2e/kWh}$ is released

during diesel supply chain processes ($EF_{Provision,Diesel}$). According to these Emission Factors, total use-phase GHG emissions from the diesel-powered wheel loader are

- 77 % from operation, and
- 23 % from energy provision.

With Equation 2.2 $GHG_{Machine,Diesel}$ calculates to 237.75 t CO_2e with the earlier calculated Activity Data and introduced Emission Factors, being the total of the left bar in Figure 2.

For the electric variant GHG emissions ($GHG_{Machine,Electric}$) are calculated as:

$$GHG_{Machine,Electric} = E_{In,Electric} \times EF_{Provision,Electric} \quad (2.3)$$

Since no emissions occur during operation, only the energy provision is relevant. The Emission Factor for energy provision ($EF_{Provision,Electric}$) in the German electricity mix is both region- and time-dependent. [16] provides a value for energy supply in the German electricity mix for 2022. The values change every year by a significant amount [17]. In this work, the 2024 factor published by the German Environment Agency (UBA) of 0.427 kg CO_2e/kWh is used [17]. This factor includes all processes mentioned earlier for energy provision. Not taking all the provision processes into account is one reason why varying factors may arise. Since the efficiency of the electric drivetrain is assumed with 81 % instead of the 40 % of the diesel drivetrain and with the different Emission Factors, $GHG_{Machine,Electric}$ equates to 141.32 t CO_2e. This results in total use-phase GHG emissions of the electric wheel loader being approximately 59 % of the diesel variant, a 41 % reduction in use-phase emissions, as shown in Figure 2.

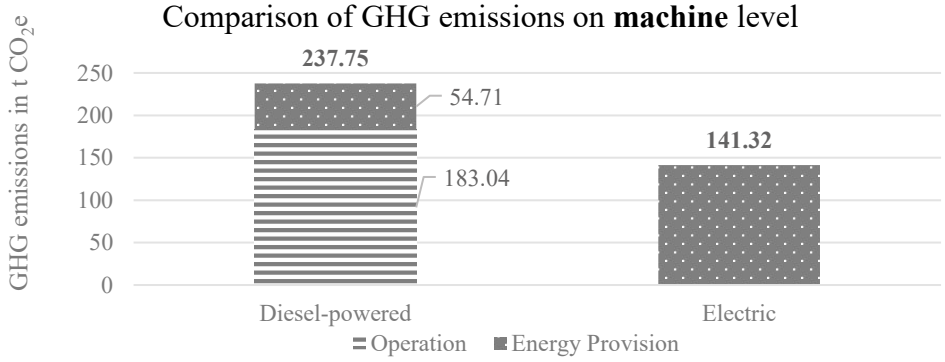


Figure 2: Comparison of wheel loader's use-phase GHG emissions for operation and energy provision with different energy carriers

Emissions from operation are depicted striped in white and grey and energy supply emissions are white dotted on grey background. While the electric loader produces no operational GHG emissions, the energy provision emissions are higher than those of diesel. Nonetheless, the overall emissions are significantly lower in the electric configuration under the given assumptions.

2.3. Research Questions

However, the presented approach offers no resolution at system or component level. While methods for assessing emissions from hydraulic components during production stages exist,

a methodology for evaluating their use-phase emissions remains absent to the authors knowledge. Even though the emissions of the use-phase on machine level are dominant, especially for diesel-powered machines [18]. Given that GHG emissions of hydraulic components cannot be measured, it remains uncertain how their use phase emissions can be quantified. However, for regulatory and customer requirements, companies must be able to disclose and ultimately reduce these emissions transparently.

As used for the machine level approach shown previously, this paper investigates how a component level approach can also rely on Activity Data and Emission Factors. As alternative energy carriers like electricity rise in importance, e. g. [12] [19], a key requirement is that qualitative emission trends at the component level must be consistent with those at machine level, when comparing different energy carriers. In order to attribute use-phase GHG emissions to hydraulic components the following research question must be answered:

How can use-phase GHG emissions of hydraulic components be determined, while allowing consistent comparisons across different energy carriers?

In the following, approaches to answer this question are developed, validated and discussed by focusing on the hydraulic pump of the travel drive of the wheel loader in two different energy carrier configurations.

3. FIRST APPROACH

3.1. Methodology

In the first step, the diesel-powered wheel loader is taken as a reference. Firstly, only GHG emitted during machine operation are considered. The activity that causes these emissions is characterized by energy flow, as illustrated in Figure 3.

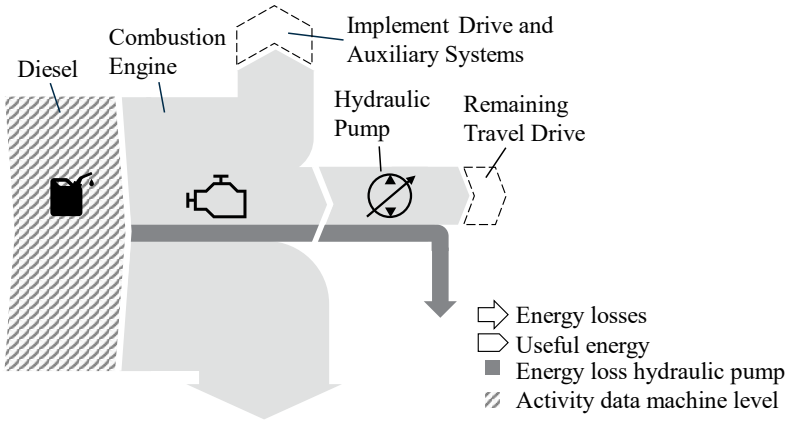


Figure 3: Sankey diagram of diesel-powered wheel loader

In this configuration, diesel serves as input energy. Through combustion in the internal combustion engine, chemical energy is partially converted into mechanical energy, while the remaining portion dissipates, indicated by the light grey arrow pointing downwards. The mechanical energy is distributed among the implement drive and auxiliary systems (indicated by the upward pointing arrow) as well as the travel drive (indicated by the arrow

pointing right), which includes the here focussed hydraulic pump. The pump transforms mechanical energy into hydraulic energy, whereby hydromechanical and volumetric losses occur [20], which are summarized as dark grey arrow.

The combustion results in GHG emissions, which are measurable at the tailpipe. As diesel is both an energy carrier and a source of emissions, the chemical energy supplied to the machine is directly proportional to the resulting GHG emissions. Accordingly, the supplied chemical energy (E_{In}) represents the Activity Data on machine level, depicted with a grey hatch, for use-phase emissions, as used in Equation 2.2.

The engine supplies this energy to downstream systems and components, which in turn dissipate part of it and pass the remainder along. The sum of energy loss from each component ($E_{Loss,Component}$) and the useful energy output, for working or driving tasks, (E_{Task}) equals the input energy:

$$E_{In} = \sum E_{Loss,Component} + E_{Task} \tag{3.1}$$

With Equation 2.2 this leads to Equation 3.2 with EF representing Emission Factors for energy provision and operation.

$$GHG_{Machine} = \left(\sum E_{Loss,Component} + E_{Task}\right) \times EF \tag{3.2}$$

Even though emissions are not directly measurable at a component, its energy demand has a causal influence on total emissions. Each component's energy loss contributes to the overall demand for energy supply, thereby to the amount of diesel burned in the combustion engine finally resulting in GHG emissions. This contribution is shown in Figure with the dark grey arrow being part of the grey hatched input energy. The component's contribution can be represented by a ratio of the component's energy loss to the machine's energy input. The first approach uses the same ratio for the component's contribution to machine level GHG emissions leading to:

$$\frac{E_{Loss,Component}}{E_{In}} = \frac{GHG_{Component}}{GHG_{Machine}} \tag{3.3}$$

Hence the machine's energy input is used as Activity Data for machine level emissions calculation, in this first approach the energy loss of a component is taken as its Activity Data for calculating its use-phase GHG emissions ($GHG_{Component}$), as shown in the following expression:

$$GHG_{Component} = E_{Loss,Component} \times EF \tag{3.4}$$

Therefore, the dark grey arrow in Figure depicts the Activity Data of the hydraulic pump.

This aligns conceptually with the delta approach described in the Product Environmental Footprint (PEF): Additional consumption compared to a reference is assigned to a product. Here, the losses represent the deviation from an ideal, loss-free component. An alternative would be to consider the total energy throughput of a component. However, this would include energy that is passed on to downstream components, leading to double-counting in the GHG balance of the machine. The sum of component emissions would then exceed the measurable tailpipe emissions of the machine. Therefore, the use of loss energy is preferable for accurate allocation.

The useful energy ultimately delivered by the machine (E_{Task}) also represents Activity Data and can be used to assign the remaining GHG emissions to the task (GHG_{Task}) like driving or working:

$$GHG_{Task} = E_{Task} \times EF \tag{3.5}$$

This approach pushes the idea to not only optimize components, rather holistically including the machine's task. The total machine emissions ($GHG_{Machine}$) can thus be partitioned into task and component emissions:

$$GHG_{Machine} = GHG_{Task} + \sum GHG_{Component} \tag{3.6}$$

Consequently, this enables a causal allocation of GHG emissions to components (e. g. the hydraulic pump) and functional processes (e. g. driving or working tasks).

In the second step, the same approach is applied to the electric variant of the reference vehicle. Importantly, the hydraulic pump's operating points are assumed to remain unchanged across both drivetrain configurations. Therefore, the energy loss (and thus the Activity Data) of the hydraulic pump is considered identical in both systems (see dark grey arrow in Figure).

However, just as with machine level, the electric system emits no operational GHG emissions, which could be measured like tailpipe emissions, the only relevant emissions stem from the provision of electricity. Accordingly, no emissions from operation could be assigned to a component and therefore only the Emission Factor for energy provision applies, just as on machine level.

Finally, the proposed answer by the first approach to the research question is Equation 3.4:

The GHG emissions of a component can be calculated by multiplying the average loss energy of the component over a usage profile for a reference lifetime as Activity Data with the respective Emission Factor for operation and energy provision for the energy carrier.

3.2. Validation

During test drives with the diesel-powered wheel loader, pressure, swivel angle, and shaft speed of the hydraulic pump were recorded across work cycles from the same usage profile as defined earlier. Using the pump's characteristic maps, volumetric and hydromechanical losses were determined. Losses from the integrated charge pump are disregarded. The power loss is integrated over each cycle and divided by the duration of the cycle to calculate the pump's mean power loss during the respective cycle. For each cycle in the usage profile this is then weighted by its time share to obtain the mean power loss for the specific usage profile. Multiplied by the assumed reference lifetime of 10.000 h, this yields for this specific example a total energy loss of 27.1 MWh ($E_{Loss,Pump}$) which is according to the first approach equivalent to the Activity Data of the pump.

Special consideration is required for overrun (regenerative) operation, e. g. during coasting. In such cases, no diesel is injected [21], hence no GHG emissions occur, although energy is dissipated. Concluding, only load operation is relevant for accounting emissions. To exclude this, fuel flow measurements are used to identify periods with 0 l/h diesel consumption, which indicate coasting. Only the losses during load operation are included in the Activity Data.

The same Activity Data are applied for both the diesel and electric configurations, as operating points of the hydraulic pump are assumed identical. Use-phase GHG emissions of the pump are then calculated for the diesel-powered wheel loader by values from Table 1 in Equation 3.7.

$$\mathrm{GHG}_{\mathrm{Pump,Diesel}} = \mathrm{E}_{\mathrm{Loss,Pump}} \times \left(\mathrm{EF}_{\mathrm{Operation,Diesel}} + \mathrm{EF}_{\mathrm{Provision,Diesel}}\right) \quad (3.7)$$

Table 1: Values for parameters in Equation 3.7

$\mathbf{E_{Loss,Pump}}$	$\mathbf{EF_{Operation,Diesel}}$	$\mathbf{EF_{Provision,Diesel}}$
27.1 MWh	0.271 kg CO_2e/kWh [16]	0.081 kg CO_2e/kWh [16]

For calculating use-phase GHG emissions of the hydraulic pump in the electric wheel loader Equation 3.8 is used with values from Table 2.

$$\mathrm{GHG}_{\mathrm{Pump,Electric}} = \mathrm{E}_{\mathrm{Loss,Pump}} \times \mathrm{EF}_{\mathrm{Provision,Electric}} \quad (3.8)$$

Table 2: Values for parameters in Equation 3.8

$\mathbf{E_{Loss,Pump}}$	$\mathbf{EF_{Provision,Electric}}$
27.1 MWh	0.427 kg CO_2e/kWh [17]

Figure depicts the resulting GHG emissions assigned to the hydraulic pump in each configuration.

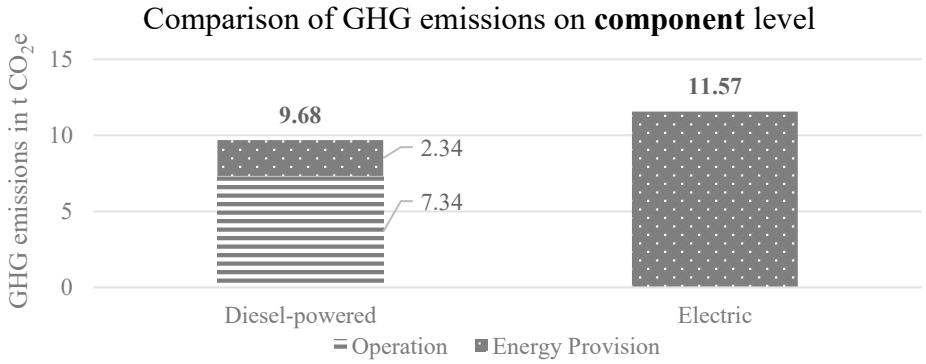


Figure 4: Comparison of hydraulic pump's use-phase GHG emissions for operation and energy provision with different energy carriers

Both bars show the GHG emissions assigned to the hydraulic pump. The left bar shows the hydraulic pump's emissions in the diesel-powered wheel loader, grey-white striped for operation and white dotted on grey background for energy provision. The right bar shows only emissions assigned to the hydraulic pump during energy provision, as no operation emissions occur in the electric wheel loader which could be allocated. However, the total use-phase emissions assigned to the hydraulic pump in the electric wheel loader are higher.

3.3. Discussion and Conclusion

The result contradicts the machine level finding that the electric variant emits fewer use-phase GHGs. To analyse this discrepancy, the ratio of GHG emissions from Equations 3.7 and 3.8 is formed:

$$\frac{\mathrm{GHG_{Electric}}}{\mathrm{GHG_{Diesel}}} = \underbrace{\frac{\mathrm{Activity\ Data_{Electric}}}{\mathrm{Activity\ Data_{Diesel}}}}_{\text{Ratio of Activity Data}} \times \underbrace{\frac{\mathrm{EF_{Provision.Electric}}}{\mathrm{EF_{Operation,Diesel}} + \mathrm{EF_{Provision,Diesel}}}}_{\text{Ratio of Emission Factors}} \tag{3.9}$$

This concludes that, independently of machine or component level, the ratio of GHG emissions from different energy carriers is equal to the ratio of Activity Data and the ratio of Emission Factors.

At machine level, as found in section 2.2, the electric variant requires 49 % less input energy, equalling a ratio of Activity Data between diesel and electric of 0.49. Though the ratio of Emission Factors, with values previously used, concludes that electricity has a 21 % higher total Emission Factor. Consequently, for the electric drivetrain in comparison to the diesel-powered less energy is required in total, but per unit of energy provided more GHG emissions are emitted. However, the higher efficiency outweighs the worse Emission Factor, leading to a 41 % reduction in total use-phase GHG emissions.

At component level, the Activity Data are assumed identical for both configurations. The ratio of Emission Factors still differs. Therefore, the ratio of Activity Data is not able to outweigh the ratio of Emission Factors like it did on machine level. This leads to higher assigned emissions for the hydraulic pump in the electric wheel loader, despite the machine emitting less overall. This inconsistency arises because the share of the hydraulic pump's energy loss relative to the machines total input energy ($\mathrm{E_{Loss,Component}}/\mathrm{E_{In}}$) increases due to higher primary energy supply efficiency and therefore less input energy needed:

- In diesel-powered wheel loader: 4 % of total input energy
- In electric wheel loader: 8.2 % of total input energy

Not only the share of the hydraulic pump's energy loss relative to the machines total input energy increases but according to Equation 3.3 a larger fraction of the machine's emissions is allocated to the hydraulic pump in the electric case. This leads to higher absolute component's emissions, despite the absolute machine's emissions being lower. The total GHG emissions of the component appear to grow solely because of being accounted for a larger fraction of the machine's emissions.

This reveals a key weakness of the first approach: It does not maintain proportional consistency across energy carriers. While suitable for comparing different components with everything else staying the same, therefore changing only Activity Data due to efficiency differences of the components, it leads to qualitatively misleading results when introducing other changes like comparing across energy carriers.

The first approach offers an answer to the research question and enables to allocate use-phase GHG emissions to a hydraulic component with minimalistic data input. As soon as more than one energy carrier is of relevance, it fails since misleading conclusions concerning impacts of different energy carriers can be made on component level. A second approach is required to resolve this issue and ensure component level emission assessments align with machine level trends, also with different energy carriers.

4. SECOND APPROACH

4.1. Methodology

In the first approach, the ratio of Activity Data between the two energy carriers remained constant, while only the ratio of Emission Factors changed. This simplification neglects a crucial interdependency: a change in energy carriers and Emission Factors inherently implies a change in the primary energy supply, which in turn alters the overall system efficiency and thus the machine's total energy input. By keeping the component's Activity Data constant between energy carriers, these upstream effects were not reflected. Consequently, the resulting component emissions did not scale consistently with the machine level emissions when switching energy carriers. Since components do not emit measurable GHGs themselves, their calculated GHG impact is always linked to the chosen energy carrier via the Emission Factor. When comparing different energy carriers, not only the Emission Factor but also the primary energy supply system changes. The first approach captured only the former, dismissing the latter. This led to only comparing Emission Factors when comparing component's emissions. Therefore, in the second approach, the Activity Data are extended to incorporate the influence of the primary energy supply efficiency.

While it would be most straightforward to base energy carrier comparisons directly on machine level energy inputs, this is often not feasible in practice. Component manufacturers typically lack access to complete machine level energy data and instead only possess information about the operating points and energy losses of their components. Furthermore, a methodology relying solely on machine level data would not support the practical need for component-specific GHG reporting. Therefore, the proposed method focuses on extending component level data in a way that remains applicable even without full system information.

Energy flows in the machine for each energy carrier configuration are shown in Figure .

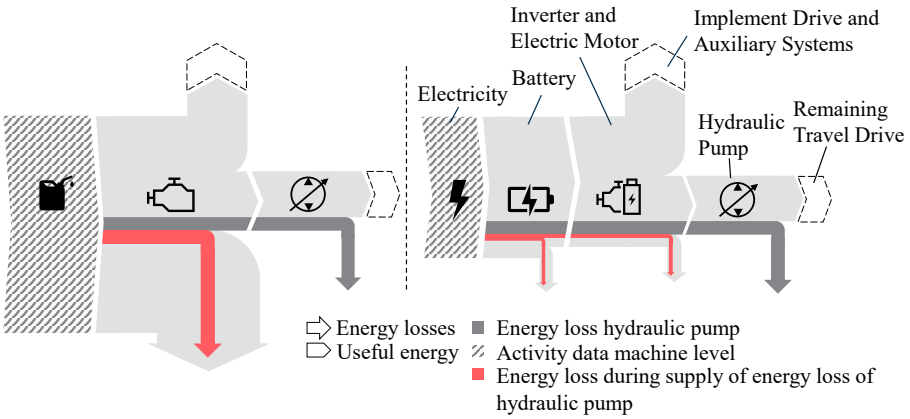


Figure 5: Sankey diagrams for electric and diesel-powered wheel loader

The dark grey arrow represents the hydraulic pump's energy losses, as used in the first approach. The first approach failed to account for the different efficiencies of the primary energy supply, leading to inconsistent results when comparing diesel and electric configurations. In the second approach, the Activity Data are therefore extended to include the energy required to supply the pump's losses, represented by the red arrows. These arrows

indicate the additional energy that must be provided – depending on the primary energy supply's efficiency – to compensate for the component's energy losses.

The adjusted Activity Data for the hydraulic pump are calculated by expanding the measured loss energy by the inverse of the primary energy supply efficiency ($\eta_{\text{PrimaryEnergySupply}}$), as expressed in Equation 4.1:

$$\text{Activity Data}_{\text{Component}} = \frac{E_{\text{Loss,Component}}}{\eta_{\text{PrimaryEnergySupply}}} \tag{4.1}$$

By incorporating these upstream losses, the Activity Data now reflect the full energetic impact of the component within its respective drivetrain, now including the 40 % efficiency of the diesel engine versus 81 % for the electric drivetrain.

The second approach extends the Activity Data in comparison to the answer of the first approach to the research question as follows:

Activity Data for components must include not only the energy losses occurring at the component itself, but additionally the energy losses at the primary energy supply when supplying these losses. To translate Activity Data into use-phase GHG emissions Emission Factors for energy provision and operation must be used.

4.2. Validation

In the following the same assumptions for efficiencies of the primary energy supply system are used as before. Component's Activity Data are now expanded according to Equation 4.1. This leads to 67.75 MWh as Activity Data in the diesel configuration and 33.46 MWh in the electric configuration. As a result, the ratio of Activity Data in Equation 3.9 on component level between the two energy carriers is 0.49, matching the ratio used at the machine level. The share of the pump's Activity Data relative to the machine's total energy input (Activity Data on machine level) becomes identical across energy carriers: 10 % for both configurations. The resulting emissions are calculated as in the first approach with the same emission factors but with the adjusted Activity Data and are shown in the center bar chart of Figure 6. Consequently, with Equation 3.3 the emission share attributed to the hydraulic pump aligns proportionally with the overall emissions of the reference machine. This resolves the mismatch observed in the first approach.

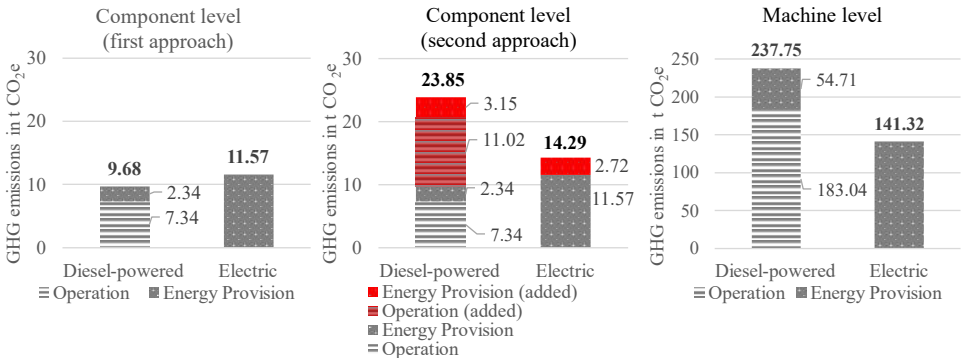


Figure 6: Comparison of use-phase GHG emissions of different approaches on component level and machine level

In Figure 6, the left bar chart shows the result of the first approach on component level for comparison. The center bar chart shows the result of the second approach on component level. The red-hatched section equals emissions linked to the red arrows in Figure 5. The red arrows represent additional energy from losses in the primary energy supply to supply the pump's energy losses, which are now included in the Activity Data of the hydraulic pump. The right bar chart depicts the total emissions of the machine as shown before in Figure 2.

Coming from the left bar chart the red hatched sections in the center bar chart add to GHG emissions of the hydraulic pump exactly so that the ratio of its emissions in diesel and electric drivetrain matches the ratio on machine level comparison. With the second approach, the pump in the electric drivetrain is assigned approximately 41 % fewer GHG emissions, consistent with the machine level analysis.

4.3. Discussion and Conclusion

In the first approach when the drivetrain becomes more efficient, total machine emissions decrease, yet the relative share of individual components in those emissions rises. In other words, a component appears "responsible" for a larger portion of a smaller total. Resulting in higher absolute emissions in comparisons, even though machine level comparisons show reduced emissions.

The second approach demonstrates that the consistency between component level and machine level GHG assessments critically depends on capturing all dependencies, including primary energy supply losses. By extending the Activity Data with the inverse of the primary energy supply efficiency, the second approach establishes a link between the component and the primary energy supply system. A change in energy carrier (for instance, from diesel to electricity) not only modifies the Emission Factor but also alters the efficiency of the primary energy supply system. These two effects are interdependent: the Emission Factor determines the environmental impact per energy unit, while the primary energy supply efficiency dictates how much energy must be provided to sustain component operation. Neglecting this link leads to incomplete or misleading comparisons. Using the second approach, the resulting emissions attributed to components scale proportionally with the overall machine behavior, thereby resolving the inconsistency observed in the first approach, allowing for a holistic and physically consistent comparison of different drivetrain configurations.

The second approach offers a practical compromise for real-world implementation. In industrial practice, complete machine data are often unavailable to component manufacturers. The proposed approach allows them to approximate realistic emissions by expanding known component level Activity Data with a representative efficiency factor of the primary energy supply. This ensures a balance between methodological rigor and practical applicability, enabling consistent use-phase emission allocation for reporting purposes.

5. CONCLUSION AND OUTLOOK

This study developed and validated a methodology for calculating the use-phase greenhouse gas (GHG) emissions of hydraulic components in non-road mobile machinery (NRMM). The work addresses a key gap in current sustainability assessment frameworks: while machine level GHG emissions can be quantified using established standards, no consistent procedure exists to attribute these emissions to individual hydraulic components, whose emissions cannot be measured.

Using a wheel loader as a reference machine, the machine level emission calculation was first established by combining Activity Data, defined as the total supplied energy, with Emission Factors for energy provision and operation. On this basis, two allocation approaches were derived and validated for a hydraulic pump in the travel drive, addressing the central research question:

How can use-phase GHG emissions of hydraulic components be determined, while allowing consistent comparisons across different energy carriers?

The first approach, based solely on the component's loss energy as Activity Data, provides a clear causal link between component efficiency and total machine energy demand. As long as only one energy carrier is relevant, it allows to report GHG emissions of hydraulic components with minimal effort. Alternative energy carriers are rising in importance though, creating the need for comparisons. This is where the first approach leads to inconsistent results, because it neglects the impact of the primary energy supply efficiency only taking the difference in Emissions Factors into account. Although the overall machine emissions decreased for the electric configuration, the pump appeared to contribute higher emissions due to the constant Activity Data definition. Thus, the first approach failed to maintain proportional consistency between component and machine level analysis.

The second approach resolves this inconsistency by extending the Activity Data to include losses in the primary energy supply system required to supply the component's losses. This adjustment ensures that both dependencies, the Emission Factor and the primary energy supply efficiency, are consistently reflected in the component level emission assessment. The approach thereby aligns component and machine results, enabling a physically consistent and holistic comparison between drivetrain configurations such as diesel and electric systems. It bridges the gap between machine level energy-based emission accounting and component level data availability, offering a robust foundation for sustainable design and emission transparency in fluid power systems.

Further concluding from the developed method, efficiency improvements, for example through optimized component operating points, directly translate into lower use-phase emissions within the operational boundary. Similarly, improvements in energy provision, such as the use of renewable electricity, further reduce emissions in the energy provision boundary, leading to two main levers to reduce use-phase emissions:

- Efficiency improvements
- Usage of "green" energy

The model neglects potential optimization of the hydraulic pump's operating points when comparing different primary energy supplies. Building on this, system optimization offers high potential for efficiency improvements. While this work mainly addressed reporting requirements for components, future research should aim to expand the method beyond energy carrier to system comparisons.

To extend the demonstrated method to other components in other machines measurement data, from real-world testing with a suitable machine-level operating profile, is needed to calculate the needed loss data. However, real-world testing introduces non-repeatable conditions, such as varying soil properties, weather, and operator behavior. To ensure reproducibility and comparability and to address high efforts in collecting measurement data for the high variance of NRMM a shift from real-world testing to a simulation-based approach will be required in future work.

In the long term, the methodology presented with the second approach can form the foundation for a standardized, modular, and simulation-supported emission-accounting framework for fluid power systems. By providing a consistent way to attribute use-phase emissions to individual hydraulic components, it contributes to greater transparency, supports the development of sustainable hydraulic designs, and thereby advances the decarbonization of NRMM and fluid power applications.

6. REFERENCES

[1] L. Schmidt and A. H. Hansen, “Estimating the Global Emission Impact of Mobile & Industrial Hydraulic Machinery.” 2024.

[2] Bundesministerium für Arbeit und Soziales, “Den nachhaltigen Wandel beschleunigen: Aktuelle Entwicklungen um den Green Deal der EU-Kommission,” EU Green Deal. Accessed: Aug. 07, 2025. [Online]. Available: https://www.csr-in-deutschland.de/DE/CSR-Allgemein/CSR-Politik/CSR-in-der-EU/EU-Green-Deal/eu-green-deal.html

[3] D. Jepsen, T. Zimmermann, F. Hauschke, S. Schunkert, J. Siewert, and G. Smolka, “Carbon Footprint und Treibhausgasneutralität von Produkten,” Umweltbundesamt, Dessau-Roßlau, Jan. 2025.

[4] World Resources Institute and World Business Council for Sustainable Development, *Product Life Cycle Accounting and Reporting Standard*. 2011.

[5] DIN Deutsches Institut für Normung e. V., *DIN EN ISO 14067:2019-02, Greenhouse gases – Carbon footprint of products – Requirements and guidelines for quantification*, 2019.

[6] Empfehlung (EU) 2021/2279, *Empfehlung (EU) 2021/2279 der Kommission vom 15. Dezember 2021 zur Anwendung der Methoden für die Berechnung des Umweltfußabdrucks zur Messung und Offenlegung der Umweltleistung von Produkten und Organisationen entlang ihres Lebenswegs*. 2021.

[7] H. Hottenroth, B. Joa, and M. Schmidt, “Carbon Footprints für Produkte,” Institute for Industrial Ecology, Pforzheim, 2013.

[8] DIN Deutsches Institut für Normung e. V., *DIN EN ISO 14083:2024-06, Greenhouse gases – Quantification and reporting of greenhouse gas emissions arising from transport chain operations*, 2024.

[9] M. Geimer, *Mobile Working Machines*. Warrendale: SAE International, 2020.

[10] Volvo Construction Equipment, “LCA/Carbon Footprint Principles,” Product Environmental Documentation. Accessed: Oct. 05, 2025. [Online]. Available: https://www.volvoce.com/global/en/products-and-services/environmental-declarations/

[11] Q. Yuan, “Real World Duty Cycle Development Method for Non-road Mobile Machinery (NRMM),” *SAE Int. J. Commer. Veh.*, vol. 09, no. 2, pp. 306–312, Sept. 2016, doi: 10.4271/2016-01-8118.

[12] Q. Dejaegere and S. Verhelst, “Renewable Alternatives for Fossil Fuels in Non-Road Mobile Machinery: A Multicriteria Analysis,” presented at the 16th International Conference on Engines & Vehicles, Capri, Italy, Aug. 2023, pp. 2023-24–0086. doi: 10.4271/2023-24-0086.

[13] T. A. Minav, J. E. Heikkinen, and M. Pietola, “Direct driven hydraulic drive for new powertrain topologies for non-road mobile machinery,” *Electr. Power Syst. Res.*, vol. 152, pp. 390–400, Nov. 2017, doi: 10.1016/j.epsr.2017.08.003.

[14] J. M. Ratzinger, S. Buchberger, and H. Eichlseder, “Electrified powertrains for wheel-driven non-road mobile machinery,” *Automot. Engine Technol.*, vol. 6, no. 1–2, pp. 1–13, June 2021, doi: 10.1007/s41104-020-00072-z.

[15] N. Froehlich and P. Pütz, “Treibhausgasemissionen im Transportsektor,” Umweltbundesamt, 2024.

[16] R. Anthes and M. Borschel, “Environmental Methodology and Data Update 2025, ISO 14083 Version 4.0.”

[17] P. Icha and T. Lauf, “Entwicklung der spezifischen Treibhausgas-Emissionen des deutschen Strommix in den Jahren 1990 - 2024,” Umweltbundesamt, 2025. doi: 10.60810/OPENUMWELT-7844.

[18] A. U. Khan and L. Huang, “Toward Zero Emission Construction: A Comparative Life Cycle Impact Assessment of Diesel, Hybrid, and Electric Excavators,” *Energies*, vol. 16, no. 16, p. 6025, Aug. 2023, doi: 10.3390/en16166025.

[19] A. Lajunen, P. Sainio, L. Laurila, J. Pippuri-Mäkeläinen, and K. Tammi, “Overview of Powertrain Electrification and Future Scenarios for Non-Road Mobile Machinery,” *Energies*, vol. 11, no. 5, p. 1184, May 2018, doi: 10.3390/en11051184.

[20] G. Bauer and M. Niebergall, *Ölhydraulik: Grundlagen, Bauelemente, Anwendungen*. Wiesbaden: Springer Fachmedien Wiesbaden, 2020. doi: 10.1007/978-3-658-27027-8.

[21] F. Küçükay, *Grundlagen der Fahrzeugtechnik: Antriebe, Getriebe, Energieverbrauch, Bremsen, Fahrdynamik, Fahrkomfort*. Wiesbaden: Springer Fachmedien Wiesbaden, 2022. doi: 10.1007/978-3-658-36727-5.

Biographies



Felix Ruthmann received the master's degree in mechanical engineering from the Karlsruhe Institute of Technology (KIT) in 2024. Since 2025, he has been pursuing a Ph.D. at the Institute of Mobile Machines (Mobima) at KIT in cooperation with Bosch Rexroth AG. His research focuses on calculating greenhouse gas emissions and therefore energy efficiency of hydraulic components and systems in non-road mobile machinery (NRMM).



Prof. Dr.-Ing. Marcus Geimer received the Diploma degree in mechanical engineering from the RWTH Aachen University, Germany, in 1990, and the Ph.D. degree from the Institute of Hydraulics and Pneumatics, today named the Institute for Fluid Power Drives and Systems, RWTH Aachen University. He started his industrial career in 1995 in the field of construction at the company Krupp Berco Bautechnik GmbH, Germany, where he was the Leader of the research group for hydraulic breakers. In 2000, he was with Bucher Hydraulics GmbH, Germany, where he led the construction and customer development for mobile hydraulics. Since 2005, he has been a Full Professor and the Director of the Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), Germany. His research and teaching focuses on automation systems, drive concepts, hydraulics and simulation. In addition, he is Faculty Director of Mechatronics.



Prof. Dr.-Ing. Gerhard Geerling received his Diploma degree in mechanical/ aeronautical engineering from the RWTH Aachen University, Aachen, Germany in 1992. In collaboration with the Airbus Technology Transfer Center in Hamburg-Finkenwerder (THF) he received his Ph.D. from the Technical University of Hamburg (TUHH), where he worked as scientific system project engineer at the Institute of Aircraft Systems Engineering (FST) from 1993 until beginning of 1998. He started his industrial career in 1998 with Brueninghaus Hydromatik GmbH, at this time part of the Mannesmann Rexroth AG, nowadays integrated in the business unit Mobile Solutions (DC-MH) of the Bosch Rexroth AG. He has led various management positions at engineering, technology and innovation management, product and project management as well as marketing, technical sales and project procurement. Today he is the responsible Senior Sustainability Manager of DC-MH. Furthermore, he is an honorary professor at the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, giving a lecture about the planning and design of hydraulic systems.



Ulrich Lenzgeiger received his Diploma degree in mechanical engineering from the Technical University of Munich, Germany in 2006. He started his industrial career at Bosch Rexroth AG in System Development, focusing on system integration and functional development for mobile machinery. From 2019 onwards, he has been a Senior Manager in System and Software Development, responsible for mechatronics, reliability, and vehicle system testing.

A Holistic Approach Towards Energy Efficient Excavators

Heiko Steinbach, Dr. Kim Heybroek

Volvo Construction Equipment Germany GmbH, Max-Planck-Str. 1, 54329 Konz

Volvo Construction Equipment AB, Bolindervägen 5, 635 10 Eskilstuna, Sweden

Email: heiko.steinbach@volvo.com, kim.heybroek@volvo.com

Abstract

In the context of the need to reduce greenhouse gas emissions, the construction equipment industry is experiencing a transformative shift.

For decades, excavators have been designed to deliver high performance while maintaining cost-effectiveness. Hydraulic systems provide an easy way of energy distribution from the prime mover to the individual motion function. Robust components with high energy density have been a key factor. Energy efficiency has often been taken with lower priority.

By introduction of electric drives into general purpose machinery, energy consumption becomes a critical factor, especially in case of using batteries as a primary energy storage. At the same time new ways of energy distribution are provided.

A holistic approach will be demonstrated by utilizing architectural methods evaluating established and emerging technologies to identify innovative solutions that can optimize energy efficiency while considering economic viability.

Through a comprehensive analysis of technological advancements alongside their economic implications and commercial availability, this exercise will provide valuable insights for academia and industry stakeholders by driving the future of energy-efficient construction machinery.

Keywords: Energy efficiency, fuel consumption, carbon footprint, BEV, hydraulic systems

1. INTRODUCTION

This study focuses on motion system concepts for battery-electric medium-size excavators with an operating weight between 15 and 30 t. As a reference, a state-of-the-art diesel-driven 22 t crawler excavator is used. The goal is to identify, structure, and evaluate potential solutions for energy- and cost-efficient motion systems that are suitable for battery-electric operation.

Both academic concepts and industrial motion systems are considered. Using an architectural approach, all relevant system variants are mapped in a structured manner, enabling a systematic comparison in terms of energy efficiency, technical feasibility, and expected costs.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. FUNDAMENTALS ON HYDRAULIC SYSTEMS FOR EXCAVATORS

A comprehensive review of hydraulic system concepts from academia and industry suitable for excavator applications reveals several fundamental system principles. Today's market is still dominated by throttle valve-controlled metering systems, while alternative concepts are emerging.

Valve metering systems remain highly attractive due to their excellent controllability at comparatively low cost. Power distribution is provided in a quite effective way. Energy efficiency is typically considered with lower priority.

In conventional internal combustion engine (ICE) driven machinery, this drawback has been tolerated because the energy density of fossil fuels is high, and fuel costs have been in an acceptable range.

With the rising of battery-electric construction machinery, the need for more energy-efficient motion systems is more urgent than ever. Batteries are costly and have limited energy density. Simply replacing the diesel engine with an electric motor without changing the hydraulic architecture results in limited operating time per charge due to the low overall efficiency of conventional motion systems. Just increasing the battery capacity is constrained both by cost and by the limited installation space in the machines.

At the same time, electric drives open new possibilities. Battery-electric machines can employ multiple electric motors on board, enabling advanced hydraulic and electro-hydraulic architectures that directly support more efficient motion systems.

When energy efficiency is prioritized, it becomes evident that inherent system losses must be minimized. Besides throttling losses at individual control functions, pressure compensation losses between functions play a significant role in hydraulic systems. Energy recuperation is rarely used in current excavator applications, even though a medium-size excavator typically comprises 5 to 10 main working functions, frequently operated simultaneously.

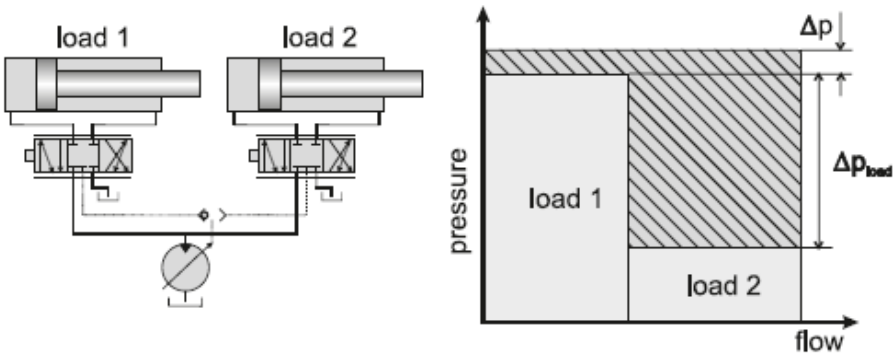


Figure 2.1. Typical single-circuit load-sensing hydraulic system [1]

Higher efficiency can be achieved by using multiple hydraulic circuits supplied by dedicated pumps or by employing more sophisticated architectures that systematically reduce throttling and compensation losses.

2.1. Paths Towards Throttle-Free Hydraulic Control

Two main technological paths towards throttle-free hydraulic control can be identified:

- the Common Pressure Rail (CPR) hydraulics path
- the Displacement Controlled Hydraulics (DCH) path,
  or rather known as Pump Controlled Hydraulics (PCH)

In Common Pressure Rail hydraulics, the system employs a plurality of nearly constant pressure rails, typically supported by hydraulic accumulators. This architecture inherently provides the possibility of energy recovery. A combination of discrete valves and hydraulic transformers is used to control the actuators. For linear actuation, multi-chamber cylinders can be applied to enable advanced flow and pressure management [1]. Hydraulic transformers can appear in both technological paths, but in this study, they are primarily allocated to the Common Pressure Rail approach.

In Displacement Controlled Hydraulics, a linear actuator is directly driven by the displacement of dedicated hydraulic pump unit. The actuator speed is controlled by the pump flow, without interfering throttle elements. Speed can be adjusted via pump rotational speed and/or variable displacement, while the motion direction is determined by the direction of the hydraulic flow.

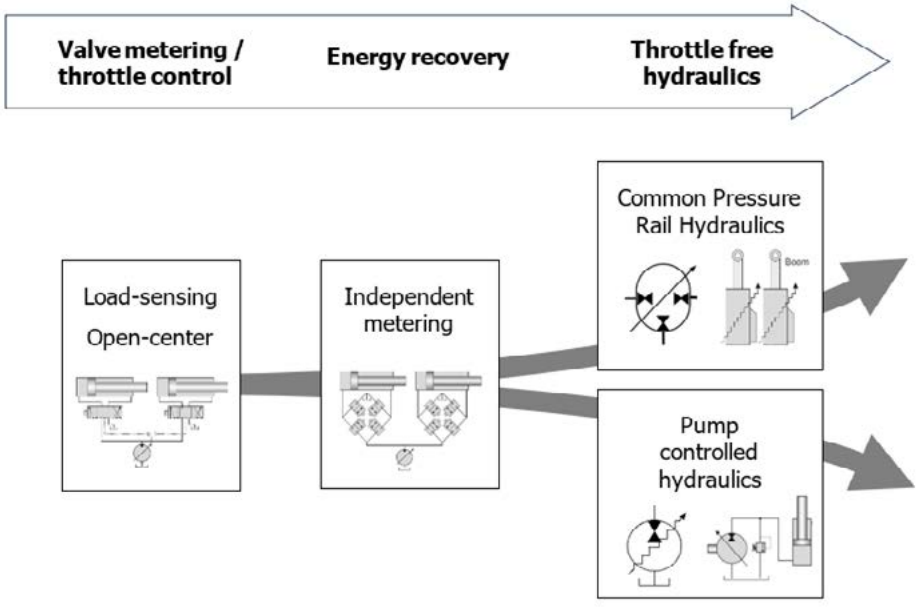


Figure 2.2. Possible paths towards throttle-free hydraulic control

The above-shown basic system principles can be combined in various ways or complemented by additional elements, resulting in hybrid architectures that further optimize efficiency, controllability and cost while the CPR and PCH principles are rather diverging.

3. ARCHITECTURAL APPROACH

The excavators' motion and control system can be clustered into its basic elements. Following the energy flow from the supply side via distribution & control towards the actuation of the working functions.

Those basic elements could be either commercially available components out of the shelf, parts under development or even theoretical ideas from academia.

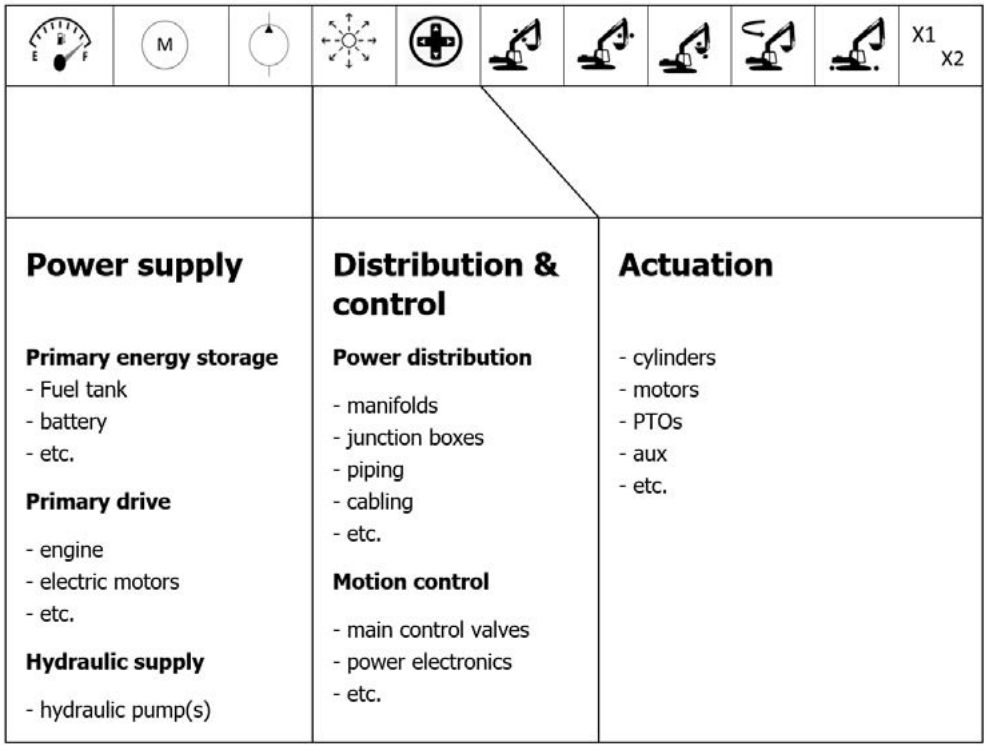


Figure 3.1. Architectural approach

3.1. Concept Creation

The excavator's motion and control system is decomposed into its basic architectural elements along the energy flow: from the energy supply (prime mover and energy storage), through the energy distribution and control subsystems, to the actuation of the individual working functions. These basic elements may be:

- commercially available off-the-shelf components,
- components under industrial development, or
- conceptual solutions proposed in academic research.

This architectural decomposition allows systematic generation and comparison of motion system concepts by combining compatible elements into complete system configurations.

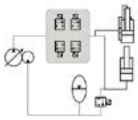
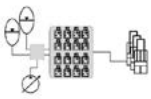
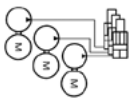
	Throttle Control	**IMVT**	**CPR**	**PCA**	**...**
					
Primary energy storage	**Tank A** xxx litre xxx €	**Tank A** xxx litre xxx €	**Tank B** xxx litre xxx €	**Battery A** xxx kWh xxxxx €	
Primary Drive	**Engine A** xxx kW xxxx €	**Engine A** xxx kW xxxx €	**Engine B** xxx kW xxxx €		
Hydraulic supply	**Pump A** xxx cc/rev xxxx €	**Pump C** xxx cc/rev xxxx €	**Pump C** xxx cc/rev xxxx €		
		Pump D xxx cc/rev xxxx €			
Power distribution	**Hose A** xxx bar xx €	**Hose A** xxx bar xx €	**Hose A** xxx bar xx €	**Cable A** xx mm² xx €	
		Hose B xxx bar xx €	**Manifold A** xx ports xx €	**Junction box A** xx ports xxx €	
Motion control	**MCV A** xx functions xxxx €	**MCV B** xx functions xxxx €	**MCV C** xx functions xxxx €	**Inverter A x3** xxx A xxxx €	
Actuation boom	**Cylinder A** d x D x l xxx €	**Cylinder A** d x D x l xxx €	**Cylinder B** d x D x l xxx €	**Motor A** xxx kW xxxx €	
				Pump E xxx cc/rev xxx €	
Actuation arm	**Cylinder C** d x D x l xxx €	**Cylinder C** d x D x l xxx €	**Cylinder D** d x D x l xxx €	**Motor A** xxx kW xxxx €	
				Pump E xxx cc/rev xxx €	
Actuation bucket	**Cylinder E** d x D x l xxx €	**Cylinder E** d x D x l xxx €	**Cylinder F** d x D x l xxx €	**Motor B** xxx kW xxxx €	
				Pump F xxx cc/rev xxx €	
Actuation swing	**Motor D** xxx cc/rev xxxx €	**Motor D** xxx cc/rev xxxx €	**Motor E** xxx cc/rev xxxx €	**Motor C** xxx kW xxxx €	
...					

Figure 3.2. Concept matrix

3.2. Concept Selection and Evaluation

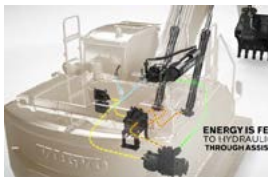
By combining the different architectural sub-components, a broad variety of motion system concepts for medium-size excavators is generated. In parallel, a streamlined power and loss calculation methodology is developed to quantify the potential energy efficiency of each concept under representative duty cycles. The outcome is a comparative assessment of motion system cost versus input energy reduction for the entire excavator. This provides an overall picture of the trade-offs between efficiency, complexity, and economic viability.

Several of the investigated concepts are already commercially available in the construction equipment industry. Others have been analysed in more depth using simulation or experimentally validated on demonstrator machines. Battery-electric excavators in particular act as a strong driver for the adoption of more efficient motion systems, as they directly benefit from reduced energy demand via smaller batteries, lower charging requirements, and improved total cost of ownership.

3.3. Advanced Hydraulic Systems - Commercially Available

As a brief excursus a few examples on already commercially available advanced hydraulic motion systems:

Single function IMVT



Open-Center system combined with independent metering for boom-down energy recuperation.

Commercially available for medium and large excavators.

+ limited complexity
+ cost efficient
+ robust and proven technology

- limited energy efficiency

Full-scale IMVT



Full scale independent metering valve technology

Commercially available for large excavators.

+ significant energy efficiency improvement
+ limited impact on general machine layout

- system complexity

CPR System



Common Pressure Rail system with Digital Valve Actuators and Secondary Controls.

Commercially available for large excavators

+ high energy efficiency improvement
+ supply system downsizing

- complex motion control
- increased system costs

Figure 3.3. Commercially available products with advanced hydraulics

4. HYBRID SYSTEM APPROACH

As electric construction machinery is emerging, the range of possible motion and control systems is expanding. Inspired by industrial robotics, one might consider driving all functions with electric motors or electro-mechanical actuators (EMA). For rotary functions like swing of the upper carriage of an excavator or the traction drives, it comes rather naturally to directly apply those with electric propulsion.

Excavators digging attachment, however, must provide high forces and high speeds while being robust against mechanical shocks and dynamic overload in a rough terrain. Meeting all these requirements simultaneously with purely electro-mechanical actuators becomes challenging [3]. In addition, packaging constraints, component masses, and costs must be carefully considered.

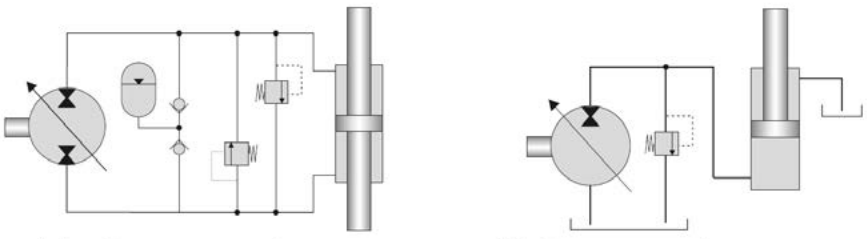
This raises also the question of energy efficiency: Will electro-mechanical solutions automatically outperform hydraulics?

In an EMA-based system, all propulsion and recuperation power flows through the electro-mechanical drive chain, and therefore experiences the associated losses:

- Bearing and mechanical transmission losses
- Electric motor losses
- Power electronics and conversion losses
- Energy storage system (ESS) losses

Considering an excavator working function with high recuperation potential, such as the boom, offers a useful benchmark. Up to 50 % of the required lifting capacity may be due to the dead weight of the structure and its attachment. This justifies a more complex and potentially more expensive motion system, as the benefits in terms of reduced primary energy storage demand and lower energy consumption (and thus total cost of ownership) can outweigh the additional system cost.

In contrast to EMA, Electro-Hydraulic Actuator (EHA) systems allow various different layout concepts that can be tailored to the specific use case [1][5].



(a) *Pump control in a closed circuit arrangement.*

(b) *Pump control in an open circuit arrangement.*

Figure 4.1. Two basic EHA system examples

For example, open- or closed-circuit layouts can be chosen (figure 4.1). Depending on the application, either plunger-, synchronous- or differential cylinders may be the favored choice. Whereas differential cylinders dominate in construction machinery. With the differential flow one of the challenges for EHA systems needs to be considered. Those cannot be simply driven by conventional pumps or pump-motors without further valve technology and storage devices. This will add cost and complexity on one hand, but also provides opportunities to increase the system's efficiency. If a pressurized reservoir in form of a hydraulic accumulator is built-in, this could offer the chance for short term energy buffering or counter-balancing a load.

For an excavator's boom dead weight compensation, a power-split concept (as illustrated schematically in Figure 4.2) can be highly advantageous in terms of energy efficiency. The static load component can be compensated in the hydraulic domain using hydraulic accumulators, while the electric drive primarily provides speed and directional control. Consequently, the power throughput in the electric subsystem is reduced, lowering conversion losses. A significant share of the energy flow remains within the hydraulic domain, and overall efficiency of the specific motion function of up to 80 % can be achieved.

Optimizing and partially decoupling a single motion function not only reduces the power consumption of that particular function. The remaining (hydraulic) motion system also benefits through reduced supply and compensation losses, thereby improving the efficiency of the entire machine significantly.

Considering the above-mentioned aspects, a hybrid system approach for an electric driven excavator could be drawn as shown in figure 4.2.

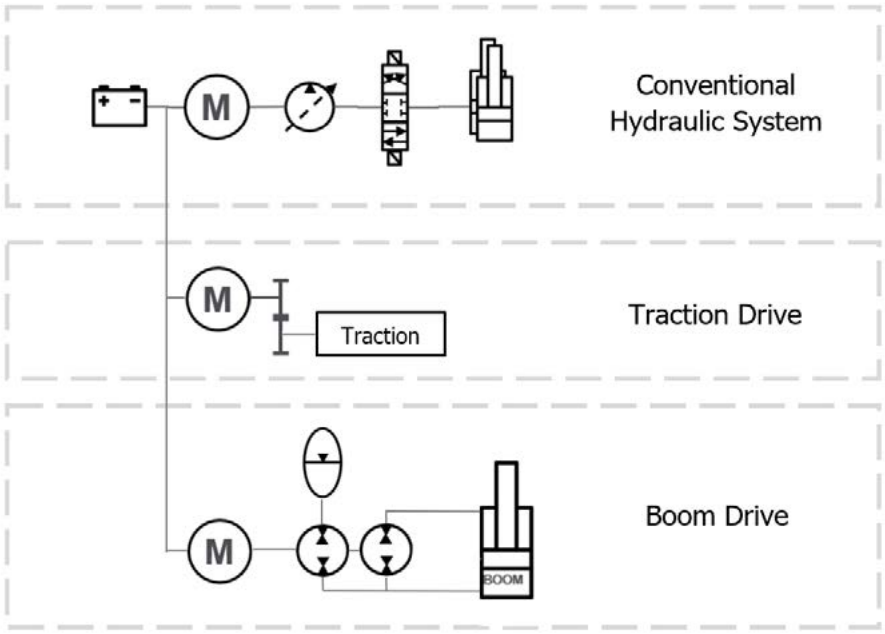


Figure 4.2. Hybrid motion system approach

A hybrid system, as illustrated above, emerged as a promising conceptual result of the architectural approach. The corresponding analysis indicates both a significant improvement in energy efficiency and a considerable economic potential.

The system was simulated and subsequently implemented in a demonstrator machine. As shown in figure 4.3, a substantial reduction of the total input power of approximately 27% was achieved in the dedicated load cycle, although other applications may show different figures. The energy portion of the boom function was reduced by about 10%, which lies within the expected range, while the energy efficiency of the remaining functions improved by approximately 17% in parallel. This can be explained since in conventional hydraulic systems, the boom drive constitutes the most dominant function and typically causes considerable compensation losses for the other functions used in parallel, such as arm, bucket, and swing. By decoupling the boom function, the compensation losses of the remaining functions are significantly reduced.

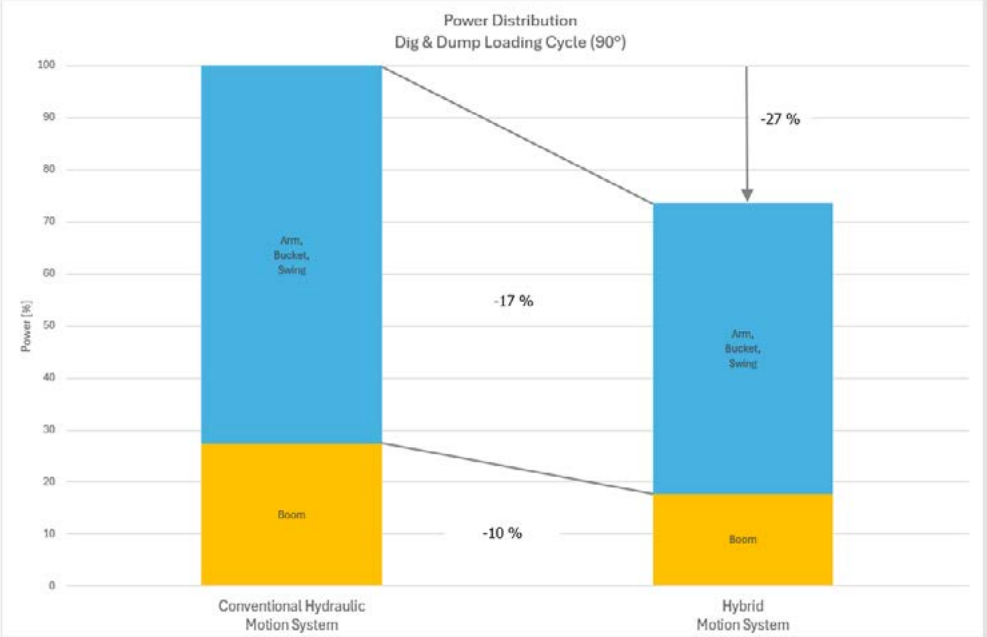


Figure 4.3. Comparison of the power distribution of a dig & dump loading cycle

5. CONCLUSION AND OUTLOOK

The study has been carried out by a cross-functional, international project team within Volvo Construction Equipment. The objective was to derive a motion system technology roadmap for construction machinery, with a focus on excavators and their transition towards electrification.

Technologies and concepts from both academia and industry were evaluated using an architectural approach. Several solutions have been implemented on demonstrator machines or selected as candidates for series development. The analysis confirms that there is no universal, one-fits-all solution that simultaneously delivers maximum performance at minimum cost. In the hydraulic domain in particular, a wide variety of technical solutions exist, suited to different applications and boundary conditions.

As one example a hydraulic hybrid system approach has been showcased in this study to demonstrate that it could be advantageous to optimize only one single machine function by a rather complex system. Nevertheless, there was a substantial benefit for the rest of the hydraulic system as compensation and supply losses are significantly reduced.

However, by systematically applying architectural methods and smartly combining appropriate elements, viable pathways towards significantly higher energy efficiency can be defined. These pathways support the development of economically attractive, energy efficient construction machinery and provide a structured foundation for future research and industrial implementation.

6. REFERENCES

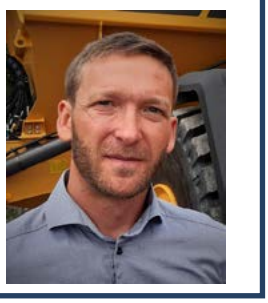
[1] K. Heybroek, ‘On Energy Efficient Mobile Hydraulic Systems with Focus on Linear Actuation’, PhD Thesis, Linköping, 2017
[2] M. Vukovic, ‘Hydraulic Hybrid Systems for Excavators’, PhD Thesis, Aachen 2016
[3] D. Peitsmeyer, ‘Decentralized Hydraulic-Electric Linear Axis‘, Technical article, Bucher Hydraulics GmbH, Klettgau 2020
[4] Volvo CE press release, ‘Pioneering electro-hydraulic solution significantly improving fuel efficiency in construction equipment’ Volvo CE 2020
[5] S. Ketelsen et al., ‘Classification and Review of Pump-Controlled Differential Cylinder Drives’, Paper, Aalborg University, 2019

Nomenclature

IMVT	Independent Metering Valve Technology
PCH	Pump Controlled Hydraulics
CPR	Common (or constant) Pressure Rail hydraulic system
DCH	Displacement Controlled Hydraulics
EMA	Electro-Mechanical Actuator
EHA	Electro-Hydrostatic Actuator
ICE	Internal Combustion Engine
ESS	Energy Storage System

Biographies





Heiko Steinbach received his diploma degree in Mechanical Engineering from Fachhochschule Trier in 2005, with focus in hydraulics as well as commercial vehicle and off-road applications.

He has more than 20 years of experience in the hydraulic domain, including several years in the automotive industry, where he focused on electro-hydraulic transmission controls.

Since 2011, he has held various roles in excavator development at Volvo Construction Equipment. He is currently working as System Owner for Motion Systems.

Kim Heybroek received his MSc degree in Mechanical Engineering from Linköping University, Sweden in 2006. In 2008, he joined Volvo Construction Equipment in Eskilstuna. In 2017, he received his PhD degree at the Department of Fluid Power and Mechatronic Systems at LiU. He is currently working as a network and research area leader in the field of hydraulics.

Innovative self-sufficient clamping system for machine tools

Johannes Gattinger, Marc Summer, Markus Grossauer, Markus Schwimmer

WEBER-HYDRAULIK GmbH, johannes.gattinger@weber-hydraulik.com

Abstract

In the course of optimizing milling processes, automatic workpiece clamping plays an essential role for a high degree of automation. Hydraulic workpiece clamping offers additional advantages, such as good integrability due to the high power density and the generation of variable clamping forces. The aim of this work is to develop a hydraulic, retrofittable, universal pressure supply for automated clamping fixtures. This clamping system should work without hydraulic and electrical interfaces. An integrated pressure accumulator in the closed system can be charged by the spindle of the machine tool via a separable coupling. This provides hydraulic energy for building up and maintaining the clamping force. Other functions, such as reducing the clamping force and opening or closing the clamping jaws, can be performed either manually by the operator or by actuating the corresponding valves with the machine spindle. The clamping force can be monitored wirelessly via Bluetooth using the integrated pressure sensor.

Keywords: hydraulic clamping, retrofit, automation, closed hydraulic system, interface-free clamping system.

1. INTRODUCTION

When optimizing milling processes, a high degree of automation is usually aimed for, and automatic workpiece clamping plays an important role here [1]. In addition to the good automation capability of the clamping process, hydraulic workpiece clamping offers additional advantages such as good integrability due to the high power density and the generation of variable clamping forces [2]. Apart from enabling automation through hydraulic clamping systems, further general advantages are also expected with manual workpiece clamping. While mechanical clamping fixtures are tightened manually with a suitable tool and defined torque, clamping with hydraulic clamping fixtures can simply take place by switching a valve. This could increase productivity through faster part change processes in order to achieve a high Overall Equipment Effectiveness (OEE) [3, 4, 5]. Another positive effect on the OEE is that quality can be increased through repeatable clamping force, which can also be monitored and documented using integrated pressure sensors [6].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

State-of-the-art technology for the pressure supply of hydraulic clamping devices are clamping units integrated in the machine tool. If pallet changing systems are used, appropriate interfaces must be provided for automatic coupling of the hydraulic and, if necessary, electrical connections. Today, many milling machines are equipped with hydraulic clamping units as standard, but retrofitting existing machines is expensive and therefore rarely done.
The aim of this work is to develop a universal pressure supply for hydraulic clamping devices. In order to ensure optimal retrofitting capability and robustness, this clamping system should not require any hydraulic and electrical interfaces. The main focus here is on the question of whether it is possible to drive the pump of the clamping system via the machine spindle using a separable coupling, requiring as little time as possible.

2. CONCEPT AND DESIGN

The basic concept of the clamping system presented here is based on the idea of generating, accumulating and supplying the hydraulic energy directly at the clamping devices in the clamping tower. This enables hydraulic clamping to be carried out directly on the pallet, eliminating the need for an interface to the machine. Each pallet therefore has a self-sufficient clamping unit. All power transmission to the unit is to take place via the machine spindle. This primarily means the drive of the hydraulic pump and also any valve switching operations. This allows the main objective of avoiding hydraulic and electrical interfaces to be achieved. In addition, this design allows the clamping system to be extremely compact and cost-effective, since no additional electric motor is required to drive the pump. The machine spindle drives the hydraulic pump of the clamping system by rotating. A separable coupling is used for this purpose, with one hub acting as a spindle tool that connects to a corresponding hub in the clamping system. The coupling is designed to eliminate the need for precise angular alignment between the machine spindle and the clamping system hub. Even if the machine spindle could, in principle, couple at a defined rotational angle, this approach is avoided because it would increase complexity and error susceptibility. More importantly, it allows for a functionality where the exact position of the clamping system's coupling hub is not always known. It should be possible for the pump to be operated independently of the machine tool, for example, to enable the operator to reload hydraulic energy when changing parts in the pallet station. An additional coupling hub is designed for this purpose, which can be inserted into a cordless screwdriver. After manual operation of the pump by the worker using a cordless screwdriver, the angular position of the coupling hub of the clamping system is undefined. In order to make safe coupling possible for the machine tool, an approach is pursued in which, on the one hand, the highest possible probability is achieved that the coupling parts slide into each other regardless of their angular position and do not collide and block on the end face when they are pushed together axially and, in addition, axial compensation is made possible during coupling so that the coupling parts can move out of the way when they are pushed together. A curved-tooth gear coupling from KTR is selected for this purpose, in which the teeth are sharpened on the end faces, making it less likely that flat end faces will collide. In addition, the sleeve of this special coupling is fixed on the pump-side hub, but can move axially on the hub by 4.5 mm. An O-ring is integrated into the sleeve/hub combination as a resilient element to press the sleeve into its initial position and allow elastic displacement (Figure 2.1). To engage, the spindle must now rotate slowly as it moves axially towards the coupling of the pump. Due to the

continuous rotation of the hub on the spindle with the possibility of elastic displacement of the coupling sleeve on the pump-side hub, the coupling parts can slip into each other. The pump-side hub is also rotated from the point of engagement and the spindle can move to the axial end position. As soon as the coupling process is complete, the spindle can drive the pump at maximum pump speed in order to generate the hydraulic power.

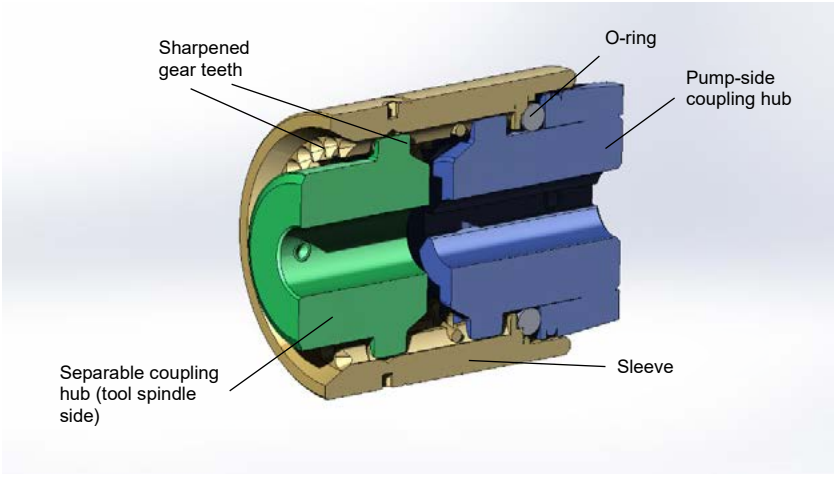


Figure 2.1. Separable coupling

In the unlikely event that the hub does not slide into the sleeve because something is jammed or the teeth actually meet exactly with the end faces and do not slide into each other due to the rotation, the machine has a force monitoring device on the spindle, which prevents a crash by switching off in the latter instance.

A hydraulic accumulator is integrated to enable the clamping jaws to be operated outside the machine during workpiece changes. This accumulator is loaded when the pump is driven by the tool spindle and must be dimensioned so that it is possible to carry out several clamping operations without having to reload in between. This system design (Figure 2.2) with accumulator also makes it possible to maintain the clamping pressure reliably over a long period of time. The accumulator size and preload pressure must be selected so that the maximum clamping pressure to be achieved can be safely maintained over a long period of time, despite internal droplet leakage. In addition, there must be sufficient oil volume in the accumulator to be able to carry out several clamping operations. However, a compromise can be made here in that the full clamping pressure does not have to be reached when clamping the workpiece. When clamping, it is sufficient if the workpiece is held securely in position. The final clamping pressure can then be built up in the machine by the machine spindle before machining begins. On the other hand, the accumulator would have to be disproportionately large. The accumulator design is calculated using the formula for the polytropic change of state of nitrogen with n=1.4 [7]. This results in a calculated available oil volume of 370 ml for the 1 l accumulator with 70 bar gas pre-charge between the maximum pressure of 250 bar and the assumed minimum clamping pressure of 100 bar to be achieved during clamping of the workpiece.

Since the required clamping force can vary greatly depending on the application and component and is often not known precisely [8], but is sometimes simply clamped with the maximum available clamping force, the clamping system is designed so that a standard maximum clamping pressure of 240 bar can be achieved. In order to reliably provide the required maximum clamping pressure over a long period of time, a higher pressure of 250 bar must be maintained in the accumulator. At the same time, however, the maximum pressure limits of the clamping device must not be exceeded, which is why the clamping pressure is regulated by pressure reducing valves. This has the additional advantage that the clamping pressure required for the respective application can be easily adjusted.

The clamping system is also equipped with valves for reducing the clamping force for fine machining. The circuit is designed so that it is possible to switch between high and low clamping pressure without having to build up pressure again via the pump. It is possible to switch from high pressure to low pressure and vice versa. To ensure that this can be done without leakage, a combination of a pressure reducing valve and a pressure relief valve is selected for the low-pressure supply, as shown in Figure 2.2. The switchover is performed by a mechanically operated 4-way, 2 position valve, which can be switched manually or with a movement of the machine spindle (Figure 2.3). The clamping system is designed so that several clamping jaws can generally be operated on the clamping tower. In the application example, there are 4 clamping jaws, each of which is controlled individually via manually operated 3-way, 2-positions valves. This allows the operator to easily clamp and release when changing parts by simply turning the valve switch.

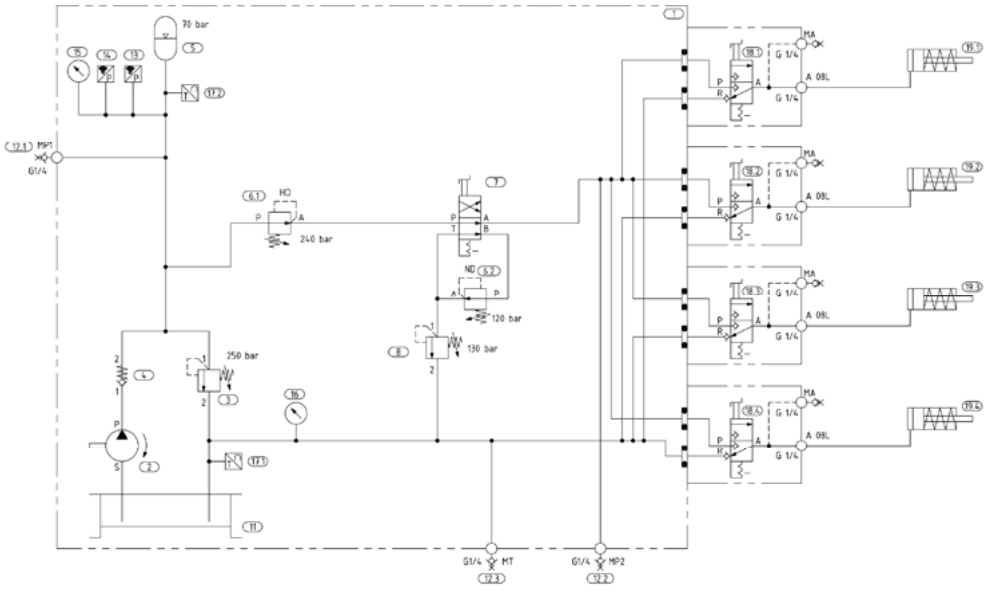


Figure 2.2. Hydraulic circuit diagram of clamping system

Single-acting hydraulic vices with an effective area of 25 cm² are used as clamping jaws. This results in a maximum clamping force of 60 kN at a maximum pressure of 240 bar. The maximum clamping stroke is 4.5 mm, but is additionally limited mechanically to 2 mm in

order to prevent the operator's body parts from becoming trapped during clamping as far as possible. This results in a required oil volume of 5 cm³ per vice, respectively per clamping operation. This means that approximately 74 clamping operations are possible within the limits of 250 bar (fully charged accumulator) and 100 bar (selected lower clamping pressure limit during workpiece insertion).

An external gear pump with a displacement of 2.1 cm³/rev is selected as the hydraulic pump. At 3000 rpm, this allows a calculated reloading time of 0.2 s, assuming that each of the four vices has been relaxed and clamped once. With the selected pump, speeds of up to 6000 rpm would even be possible, resulting in a calculated reloading time of 0.1 s.

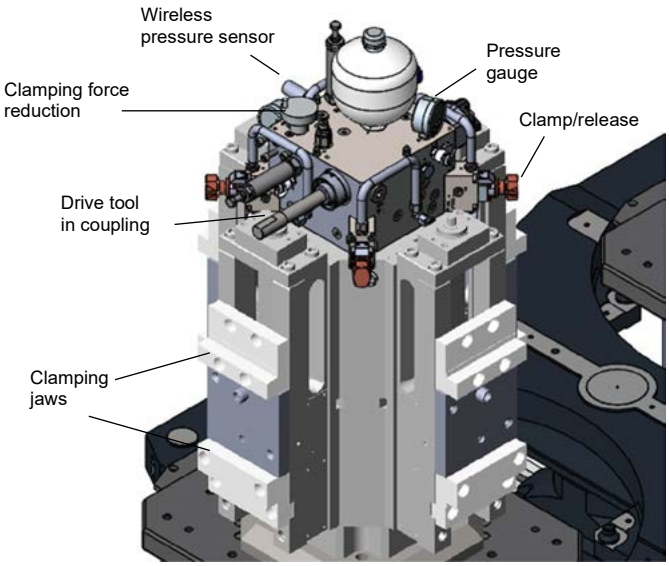


Figure 2.3. Clamping system on pallet tower

The oil tank is flanged to the bottom of the hydraulic manifold and seals hermetically with the manifold. This is necessary to protect the oil from moisture and dirt ingress from chips and cooling lubricant. The tank has a nominal volume of 2.4 l and is filled with 0.8 l of oil. The required exchange volume is 0.6 l. This results in an approximate pressure of 0.73 bar absolute in the tank for an isothermal change of state at maximum oil extraction (accumulator fully charged), which is still acceptable for proper suction by the pump.

Since reliable fixation of the workpiece is relevant to safety, pressure monitoring must be implemented during the machining process [9]. In order to keep complexity low and avoid interfaces, the focus is on wireless pressure monitoring. This also ensures seamless pressure monitoring over a longer period of time in the pallet station. One option for wireless clamping force transmission, as investigated by Denkena, would be to integrate strain gauges on the clamping jaws and transmit the data wirelessly via a CAN master on the pallet to the machine control system via an industrial PC [10]. However, such a solution requires a relatively large amount of electrical energy at the clamping tower. Since no electrical interfaces are desired here, a solution using battery-powered pressure sensors with a wireless

interface (Bluetooth) is implemented. Two relatively new wireless sensors available on the market, the PEW-1000 from WIKA and the M5600 from TE Connectivity, are being tested.

A pressure gauge is integrated so that the operator can easily check the available pressure at any time when changing workpieces. In addition, temperature sensors, another pressure gauge on the tank, and additional measuring connections have been provided for validation purposes.

3. RESULTS

When testing the coupling process, it becomes apparent that the strategy of allowing the coupling sleeve to deflect via the O-ring does not work. The stiffness is too high and the resulting travel too low, so that when the front sides of the coupling teeth meet, the machine's sensitively adjusted force monitoring system is triggered. This means that the coupling process must be carried out very slowly. On the other hand, it can be seen that the basic approach of allowing the coupling sleeve to deflect while the spindle rotates toward the coupling is effective. This is more important than the geometrically adapted front faces of the teeth, as these do not completely prevent a collision. It is therefore assumed that the special geometry of the coupling is not necessary if sufficient deflection is provided. With the current design and the cautious coupling process, coupling takes approx. 14 s. A measurement of the accumulator charging process shows that charging from 160 to maximum pressure at 1000 rpm takes approx. 3.5 s (Figure 3.1). This means that the time required for coupling significantly outweighs this and should be improved.

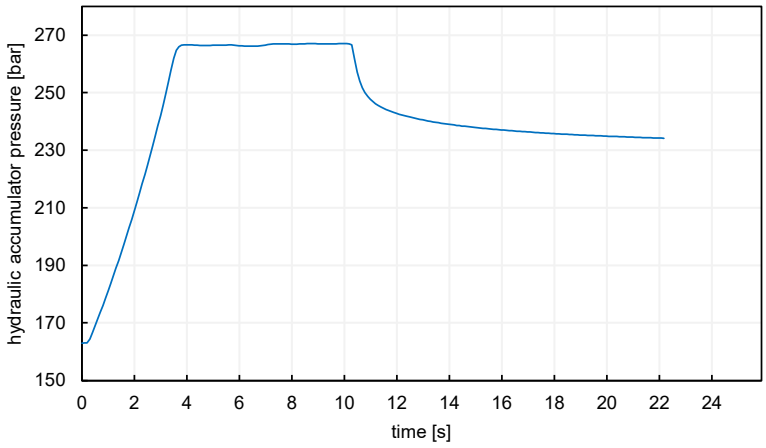


Figure 3.1. Time-controlled charging over 10 s at 1000 rpm

It is also noticeable that the pressure drops rapidly by approx. 30 bar at the end of the charging process. This is attributed to the hysteresis of the pressure relief valve. This effect can probably be avoided if the pressure connection of the pressure relief valve is not connected between the check valve and the accumulator, but between the check valve and the pump. It would also be helpful if the maximum pressure or setting pressure of the relieve valve were not so close to the maximum clampinig pressure. This would also give more

reserve capacity in terms of the stored oil volume above the maximum clampinig pressure, which would mean that it might not be necessary to reload after every insertion process. Currently, only 2.3 clamping procedures between 250 bar and 240 bar storage pressure are calculated to be possible. In reality, even a single cycle of release and clamping results in a pressure drop from 240 bar to approx. 225 bar, as all channels are briefly connected to each other when the 3-way, 2-position poppet valve is switched, resulting in additional oil loss to the tank. If the required clamping pressures are more precisely defined, the pre-charge pressure of the accumulator can also be set more advantageously, so that even at higher pressures a better ratio between the amount of oil available and the pressure difference is possible.

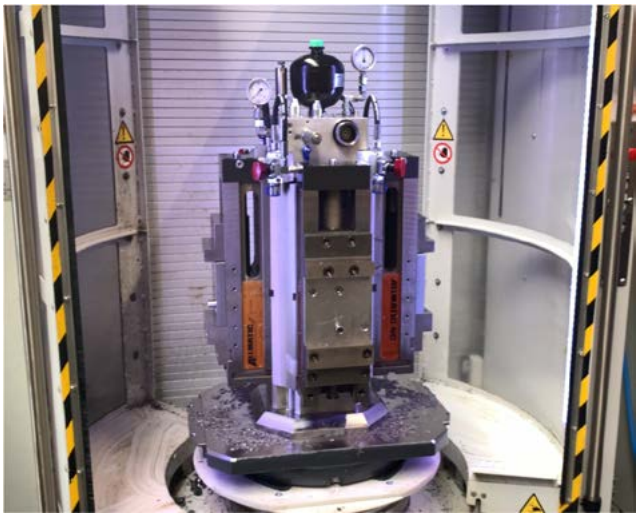


Figure 3.2. Clamping tower with installed clamping system

During validation, the following experiences were recorded with regard to time savings and operational convenience compared to mechanical clamping. In terms of time savings, an improvement of approx. 5 to 10 seconds per workpiece change was observed with a standard 4-sided clamping tower, but this alone does not lead to a profitable amortization period. From an ergonomic point of view, the hydraulic clamping device does lead to an improvement in the process. This effect is sometimes enhanced if the devices are located outside the comfortable working area when changing workpieces, or if the workers are below average in terms of size or strength, which in turn leads to further time savings.

During charging, it can be observed that negative pressure builds up in the tank, but not to the expected extent. When driving the pump, the pressure in the tank quickly drops to approx. 0.85 bar absolute and then does not drop further while the pump continues to fill the accumulator. Therefore, a leak in the tank is suspected. The tank is a welded steel tank that seals radially to the flange via an O-ring on a collar on the manifold, which is the most likely location for a leak. Since the tank is located below the manifold and thus within the clamping tower, and therefore in a fairly protected area, the risk of contamination and moisture entering the hydraulic system is relatively low.

The general suitability of the Bluetooth pressure sensors was demonstrated. Pressure recordings were possible using the manufacturers' application software, and in some cases measurement data could also be exported as CSV files. Measurement and transmission intervals vary and can be parameterized within certain limits. According to the data sheet, the sensors can also communicate with application-specific controllers via standard protocols, but this was not tested.

4. SUMMARY AND CONCLUSION

In summary, it can be stated that the principal functionality of the intended concept is in place. There is further potential for optimization, particularly with regard to the drive shaft coupling. This can be improved in terms of design to enable faster engagement of the tool spindle and thus reduce the loading time, which is directly included in the machining time per pallet. Furthermore, the integration of the wireless sensor for clamping pressure monitoring into the machine control system remains to be completed. This would also enable a pressure-controlled loading process, which would optimize the loading time to the necessary duration compared to time-controlled loading. The root cause of the tank leak needs to be identified and rectified. Alternatively, a second accumulator with a very low gas pre-charge could be utilized as a tank. It is recommended that a further optimization be made by switching to hydraulically vented pressure reducing valves. The pressure reducing valves used are atmospherically vented, which poses a general risk of contamination in this operating environment.

5. REFERENCES

[1] A. W. H. Trummer, “Flexible Werkstückspanntechnik,” in Vorrichtungen der Produktionstechnik, Wiesbaden, 1994, p. 179–202.

[2] J.-M. Kraus, “Hydraulik und Pneumatik bleiben wichtiger Bestandteil spanender Werkzeugmaschinen,” MaschinenMarkt, 2007.

[3] S. Nakajima, Introduction to TPM: Total Productive Maintenance, Portland: Productivity Press, 1988.

[4] R. Hedman, M. Subramaniyan and P. Almström, “Analysis of critical factors for automatic measurement of OEE,” Procedia CIRP, vol. 57, pp. 128-133, 2016.

[5] C. Costa, F. J. G. Silva and R. P. M. Ronny M. Gouveia, “Development of hydraulic clamping tools for the machining of complex shape mechanical components,” Procedia Manufacturing, vol. 17, pp. 563-570, 2018.

[6] H.-C. Mohring, D. Biermann, F. Bleicher, S. Melkote and G. Kappmeyer, “Fixtures and workpiece clamping systems in machining,” CIRP Annals - Manufacturing Technology, vol. 74, pp. 945-969, 2025.

[7] W. Bauer, Hydropneumatische Federungssysteme, Berlin: Springer, pp. 17-22, 2008.

[8] Q. Feng, W. Maier, S. Braun and H.-C. Möhring, “Dynamic modeling of the workpiece-fixture contact behavior for intelligent fixture design,” Procedia CIRP, vol. 119, pp. 58-65, 2023.

[9] Berufsgenossenschaft Holz und Metall, Ed.“Maschinen der Zerspanung,” DGUV Information 209-066, p. 67, 2012.

[10] B. Denkena, D. Dahlmann and J. Kiesner, “Sensor integration for a hydraulic clamping system,” Procedia Technology, vol. 15, pp. 465-473, 2014.

Biographies



Johannes Gattinger completed his master's degree in mechanical engineering at the Technical University of Munich in 2013 and subsequently earned his doctorate there in 2018 at the Chair of Medical Technology. Since then, he has been working as a project leader and technical product owner in the field of hydraulic systems and manifolds at Weber-Hydraulik GmbH in Wörth an der Isar.

Digital Architecture Design and Implementation for Performance Optimization of Electromechanical-Hydraulic Systems: A Case Study on Injection Molding Machines

Zhiqiang He[1], Fugang Zhai[1], Zhuofan Yang[1], Tianran Yang[2], Jing Yao[1,*]

[1]*School of Mechanical Engineering, Yanshan University, Qinhuangdao 066000 China*

[2]*Beijing Research Institute of Automation for Machinery Industry Co., Ltd. 100120 Beijing*

**Corresponding author e-mail: jyao@ysu.edu.cn*

Abstract.

Performance degradation under temperature rise conditions is a persistent challenge in the motion clamping system of injection molding machines, where unmeasurable variations in oil viscosity led to pump efficiency loss and positioning accuracy deterioration. To address this issue, this study proposes a platform-assisted approach for system performance modelling and control optimization. A digital platform integrating data acquisition, simulation analysis, and data review is developed to enable unified analysis of measured operating data and model predictions. Using the temperature-dependent viscosity effect as a representative case, a data-driven inverse identification method is implemented through platform-based parameter mapping and implicit indicator extraction, allowing an engineering-oriented redefinition of leakage characteristics in the pump flow model and explicit incorporation of the viscosity-temperature relationship into system modelling. Based on the established model, a pump flow compensation channel is introduced into the original position-based PID control architecture to provide feedforward correction for temperature-induced flow degradation. Experimental results under multiple operating conditions demonstrate that the proposed method significantly improves positioning accuracy and control consistency. The study provides an effective modelling and control optimization framework for performance degradation driven by unmeasurable latent variables in electromechanical-hydraulic systems.

Keywords. Electromechanical-hydraulic systems; Digital technology architecture; Digital platform; Performance optimization; Injection molding machine

1. INTRODUCTION

In recent years, digital twins and digital prototypes have attracted increasing attention in both industry and academia, with digital technologies emerging as a key driver for overcoming existing limitations and enabling transformative upgrades [1]. In equipment

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

manufacturing in particular, digital technologies—leveraging production data as a central link—have introduced new models and methods across all stages of production, significantly advancing the industry toward higher efficiency, intelligence, and sustainability [2, 3].

Electromechanical-hydraulic (EMH) systems are the core driving units of industrial equipment, with their performance directly determining functional value across the entire life cycle. FIGURE 1.1 shows the key stages from development to decommissioning, among which the performance tuning stage plays a pivotal role in optimizing the preliminary prototype into a high-performance product. However, the multidisciplinary coupling of EMH systems introduces substantial challenges in both forward design and reverse debugging of industrial equipment [4].



Figure1.1. Equipment Lifecycle and Performance Debugging Phase

1.1. Industry Status and State of the Art

First, the dynamic responses of multidisciplinary components are inconsistent, and complex parameter matching makes it difficult to accurately predict operating performance at the design stage [5]. Meanwhile, nonlinear factors such as friction and inertial loads increase uncertainties in debugging, compromising system stability and responsiveness [6]. Moreover, conventional debugging relies heavily on manual experience [7], with limited data integration, poor documentation, and weak knowledge reuse, which constrain efficiency and global optimization [8]. Therefore, introducing digital technologies, virtual debugging, and intelligent methods is essential to optimize the process, shorten the cycle, and ensure accurate and efficient EMH system debugging [9].

Currently, research on digitalization and platform construction is emerging across equipment design, debugging, manufacturing, and operation and maintenance [10–13]. Within debugging, studies focus on assembly debugging, which verifies functional integrity, and performance debugging, which optimizes operational performance. In various industries, digital technologies have partially addressed key challenges, including variable operating conditions [14], conflicts between development cycle and quality [15], high costs [16], equipment discreteness [17], complex and error-prone design processes [18], and strongly time-varying performance indicators [19]. However, for complex EMH equipment, application cases remain limited, and substantive support is lacking.

This study focuses on the EMH system of an injection molding machine (IMM) and addresses the demand for motion accuracy optimization of the motion clamping system (MCS) by proposing a performance tuning–oriented digital system architecture. The proposed architecture consists of a physical layer (PL), a simulation layer (SL), a data relay layer (DRL), and a visualization layer (VL), and establishes a data visualization workflow for system diagnosis and optimization based on high-fidelity multidisciplinary models. The PL corresponds to the actual IMM system. The DRL enables real-time acquisition and updating of multiple key operating variables using a limited number of sensors. The VL serves as the human-machine interface, providing the display of diagnosable variables and supporting performance analysis. The SL implements virtual sensing through multidisciplinary simulation to characterize both dynamic and static system behaviours, conduct coupled multidisciplinary analyses, and virtually estimate implicit parameters that are difficult to measure directly. In this way, the proposed architecture reveals both explicit and implicit factors affecting system performance, thereby improving system tunability and motion accuracy.

2. OVERALL DESIGN OF SYSTEM ARCHITECTURE

2.1. Overview of the MCS of the IMM

The research object—MCS of an IMM—is a core unit of plastic forming. It must travel about 1 m within 1 s, maintaining millimeter-level positioning while withstanding high-speed motion of mold components weighing several tons. The system operates via motor-driven hydraulic pumps that actuate the MCS. System configuration and motion accuracy are shown in FIGURE 2.1. Scatter plots indicate that actuator motion raises hydraulic oil temperature and positioning errors. High inertia, rapid response, and strict precision requirements pose severe challenges to conventional commissioning methods.

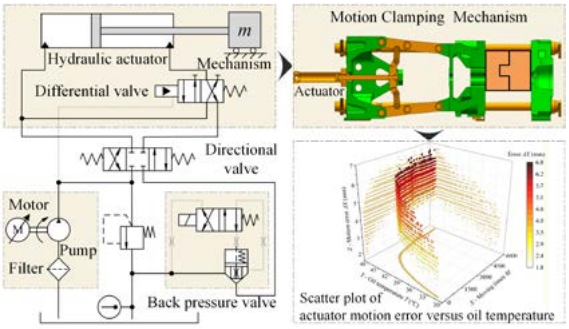


Figure 2.1. The EMH System of IMM and Its Existing Problems

2.2. Requirement Analysis

Based on the overview and current challenges of the MCS, the logical relationships of the requirements in the context of digital upgrading are illustrated in FIGURE2.2 and can be summarized as follows:

(1) High-precision simulation models: Digitalization requires high-fidelity models that accurately map physical entities. Incorporating external, inertial, and frictional forces enables faithful reproduction of the platen's dynamics. Such models allow analysis of operational deviations, and support improvements in motion accuracy and stability.

(2) Feature data support: The MCS requires a data acquisition system to collect and process key state data, enabling the simulation model to accurately depict equipment operation.

(3) Visualization platform: The platform integrates simulation models with the data acquisition system, providing intuitive display of real-time states, trends, and implicit variables to enhance system transparency, while offering interactive interfaces for functional expansion and optimization analysis.

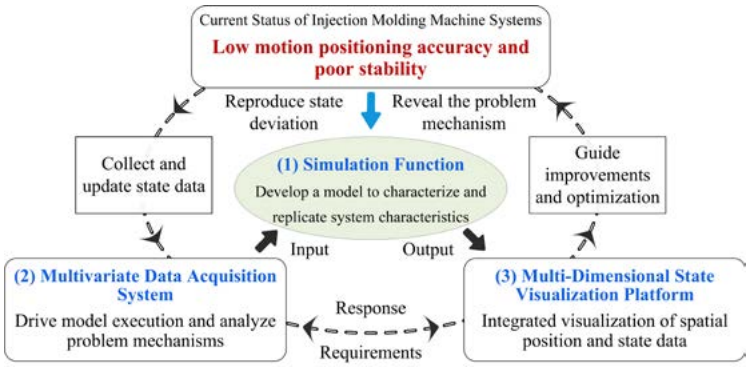


Figure2.2. Digital Technology Application Requirements for the MCS of the IMM

2.3. Architectural Design

Based on the aforementioned logical relationships of requirements, a digital application scheme is designed, encompassing both functional hierarchy design and technology framework design.

(1) Functional Hierarchy Design

This study follows an application-driven and incremental digitalization paradigm, rather than pursuing a fully instantiated, end-to-end global digital twin in a single step. Based on the current state and requirements of the MCS, the digital upgrade should adopt a four-level architecture: physical entity, system model, multi-source data acquisition system, and visualization platform. The following sections define each level and its objectives.

① PL: The MCS provides geometric and state data, forming the dynamic basis of the digital architecture. This layer includes components and sensors for data acquisition.

② DRL: This layer collects and manages multi-source data between the physical entity and digital space, serving as the core medium for their interaction. It includes transmission systems, protocols, and databases to enable efficient data transfer, storage, and management.

③ SL: Integrating the physical system's geometric and state models with its operational logic, this layer creates a virtual representation of the entity. Using data from the DRL as constraints, it replicates EMH system operation, serving as the core of digital applications.

④ VL: Serving as the interface between the physical entity and simulation model, this layer manages data acquisition, simulation execution, and comparative analysis of physical and simulated states. By identifying optimization parameters, it enhances the application of high-precision SL models, providing intuitive support for performance analysis and optimization.

(2) Technical Architecture Design

The above four functional levels establish a digital application framework for the MCS, centered on a high-precision simulation model. On this basis, it is necessary to further develop operational logic and implementation standards to form a complete digital technology architecture, which can guide system application and deployment. This technology architecture is shown in FIGURE2.3.

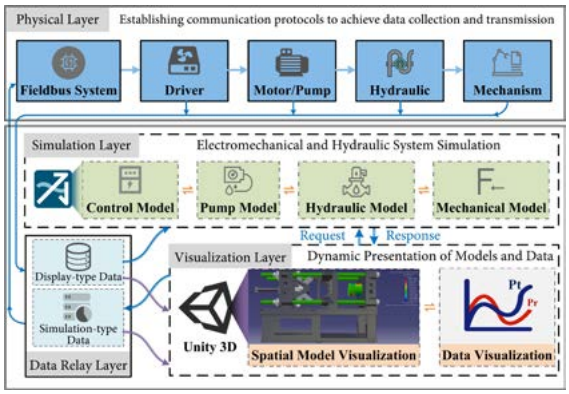


Figure 2.3. Digital Application Technology Architecture of the MCS

Based on this technology framework, the system operates as follows: During the operation of the MCS, the VL monitors system performance and collects state data for computation in the SL and for visual display. The acquired data and simulation data are synchronously stored in the DRL, supporting dynamic comparison and subsequent analysis on the VL. By linking all layers through data, the framework drives the high-precision simulation model to reproduce motion deviations in the MCS. The state variables in the model are abstractly represented on the VL, and through explicit comparison of observable states and implicit calculation of latent states, the system's operational mechanisms are revealed, providing guidance for optimizing motion positioning accuracy in the MCS.

3. DIGITAL PLATFORM DEVELOPMENT

This section builds a performance optimization digital platform for the MCS based on the proposed digital application technology architecture. By integrating multidisciplinary simulation, data flow processing, and visualization technologies, the platform enables the digital upgrade of the debugging process.

3.1. SL Construction

Multidisciplinary co-simulation is essential for the study of EMH systems. By characterizing the interactions among multiple physical fields, it reveals the system's dynamic behaviour and coupling effects, providing theoretical support for optimization. In this study, a hierarchical modelling strategy is employed to construct the SL for the MCS.

(1) Mechanical Structure Model

As the actuator of the system, the MCS is required to achieve precise positioning, rapid dynamic response, and continuous coordination across multiple motion stages during operation. However, latent factors such as component deformation, coupled inertial effects,

and friction exert a significant influence on its dynamic performance. To address these issues, this study proposes a symmetric recursive method based on Newton–Euler dynamics and introduces a deformation allocation strategy grounded in the ratio of component stiffness coefficients. By exploiting measurable displacement variables, unmeasurable parameters, including coupled inertia and structural deformation, are inferred, enabling the construction of explicit dynamic models for sub-stages. This framework facilitates virtual sensing and thereby enhances the observability of the motion process.

In addition, an equivalent friction modelling approach is adopted to provide a unified representation of friction effects arising at multiple positions and governed by different friction mechanisms. A position–pressure dual-parameter identification and switching logic is further developed, enabling the establishment of a dynamic model that spans the entire operational process of the mechanism. Experimental results demonstrate that the accuracy of each stage-specific sub-model exceeds 90%, while the overall model accuracy remains stably above 97% under continuous multi-cycle operating conditions, thereby achieving high-fidelity dynamic simulation of the MCS throughout its full process. Detailed procedures and analyses of the related study can be found in Ref. [6].

(2) Hydraulic-Electro-control Model

The electrohydraulic control subsystem is characterized by numerous components and complex parameter coupling. During the research process, model development is often constrained by practical factors such as industrial confidentiality arising from market competition, non-standardized designs, and the difficulty of directly measuring fluid parameter characteristics, which together pose significant challenges to accurate modelling. To address these issues, this study integrates multiple approaches: characteristic curves relating rotational speed to volumetric efficiency and torque based on performance and efficiency samples for the pump; component disassembly, inspection, and dimensional measurement of spool geometry and spring stiffness for the directional control valve; and black-box principle-based validation for the back-pressure valve. By fully accounting for multidisciplinary coupling, as well as nonlinear effects such as temperature, viscosity, and friction, electrohydraulic control subsystem models of the MCS are established.

(3) Integrated Model

Since the MCS couple mechanical, hydraulic, and electrical domains, this study develops an integrated EMH model based on existing subsystems. By coupling these models, a unified multi-domain dynamic simulation platform is established. Here, the electro-control model issues motion commands to the hydraulic model, which simulates pressure and flow and drives the mechanical structure. As the actuator, the mechanical structure feeds back displacement, velocity, and acceleration, enabling the electro-control system to adjust parameters accordingly. The approach, model, and results are shown in FIGURE3.1.

3.2. DRL Construction

Based on the proposed technology architecture, this section elaborates on the construction of the DRL from three aspects: data acquisition, data storage, and data-driven applications.

(1) Data Collection

Data acquisition should be based on the EMH mechanism model of the MCS, in order to identify key parameters and thereby determine the types of data required. These data encompass three aspects: mechanical, hydraulic, and electro-control, as detailed below:

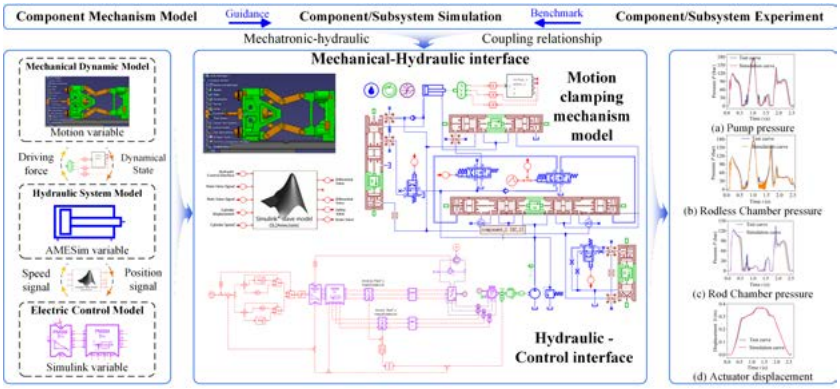


Figure3.1. EMH Combined Simulation Model of the MCS

(1) Mechanical part: Focus on monitoring the displacement-acceleration mapping relationship between the input and output ends of the MCS, as well as the deformation of tie bars, in order to calculate the clamping force and analyse the effects of inertia, friction, and deformation on the system.

(2) Hydraulic part: Monitor the inlet and outlet pressures and flow rates of each hydraulic component to evaluate the effects of hydraulic losses, cavitation, and pressure build-up processes on system performance. Meanwhile, a combined pressure–temperature sensor is installed at the return port of the hydraulic oil tank to enable real-time monitoring of the oil temperature, thereby revealing the intrinsic relationship between temperature rise and performance degradation.

(3) Electro-control part: Monitor motor speed, torque, and solenoid valve enabling signals to analyse the impact of signal delays on system performance.

Considering the diversity of data types, the dispersion of measurement points, and the stringent requirements on transmission cycle and sensitivity, this study employs a fieldbus motion control unit based on the EtherCAT communication protocol as the master-slave transmission device, enabling efficient acquisition and integration of multi-source data. In addition, the Modbus TCP protocol is used to facilitate data and command interactions between the fieldbus system and the host computer platform. Ultimately, a multi-source data acquisition and interaction system is established, in which the fieldbus system serves as the relay, with EtherCAT and Modbus TCP protocols operating collaboratively, as illustrated in FIGURE 3.2.

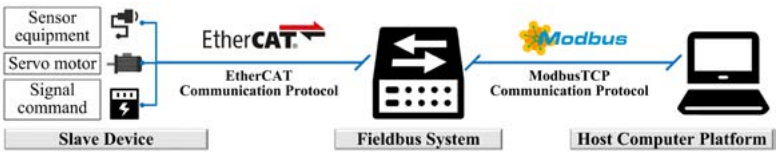


Figure 3.2. Multi-Source Data Acquisition and Interaction System

The real-time performance of data updates is a key indicator of system performance. In this study, the system communication delay was measured to be approximately 50 ms. Compared with the 2.3 s cycle of continuous motion tests and the production operation cycles exceeding 10 s, this delay has a negligible impact on the performance of the digital platform, thereby ensuring the reliability of real-time data interaction.

(2) Data Storage

Efficient data storage and management directly affect the operational efficiency of the digital platform. In this study, system data are categorized into display data and simulation data, and hierarchical storage is implemented using MySQL, with Navicat employed for management.

① Visualization Data: This includes system test and simulation data, encompassing system state parameters such as actuator displacement, clamping force, and component pressures and flow rates, and is primarily used for monitoring and result visualization.

② Simulation Data: This involves the modelling parameters of individual components and the system simulation model, such as spool diameter and damping, which serve as inputs for model computation and simulation analysis.

To enable convenient "one-click simulation" functionality and support comparative analysis of multiple data types, this study integrates a MySQL database communication interface within the Python simulation scripts, allowing for automatic storage and visualization of simulation results. To address the comparative debugging requirements of step-type components, a component parameter modification window was added (see Section 4), supporting parameter adjustment and simulation analysis for components of different specifications, thereby enhancing the flexibility and specificity of performance tuning.

(3) Data Drive

The key to achieving digital-assisted debugging of the MCS lies in the platform's ability to call and drive multi-source data and simulation models, transforming simulation results into visualized optimization guidance and gradually realizing virtual debugging. As shown in Figure 3.3, this study connects the multidisciplinary simulation software AMESim via a software interface, utilizing Python scripts to drive simulation execution while interacting with the database. The specific process is as follows:

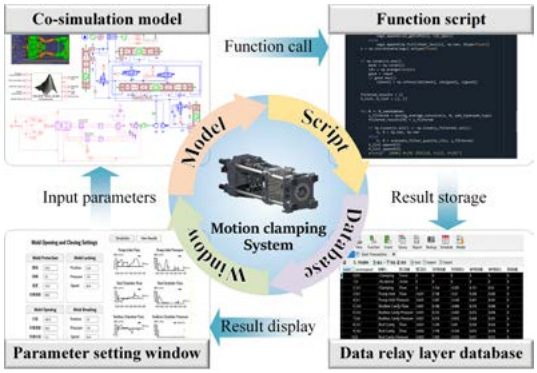


Figure 3.3. Platform-Driven Integrated Simulation Framework Logic

① Parameter Modification: Users adjust component parameters via the platform interface, and the modified data are written to the database through data communication.

② Simulation Execution: The script calls the AMESim simulation model, retrieves the updated parameters from the database, and drives the simulation computation.

③ Data Storage: During simulation, the model dynamically references parameter variables from the database, and upon completion, the results are stored in the database.

④ Visualization Display: The platform retrieves the simulation results and presents them intuitively through the Unity interface, enabling full-process visual interaction for parameter adjustment, simulation execution, and result review.

It is worth noting that the proposed DRL is designed to be protocol-agnostic. Given the widespread use of EtherCAT and Modbus TCP in industrial environments, these protocols are adopted in this study, while the relay layer abstracts raw data into unified data objects. This design enables higher-level industrial communication standards, such as OPC UA, to be integrated as gateway or middleware solutions without affecting the proposed modelling and control optimization approach.

3.3. VL Construction

The VL, serving as the carrier for both the SL and the DRL, not only supports the MCS and integrated data management but also acts as the interface for the digital platform's functional output. This section describes the construction of the VL from three aspects: functional positioning, architecture, and operational logic.

(1) Platform Functionality Positioning

Within the context of achieving virtual equipment commissioning, the platform's functional positioning encompasses three core functions:

① Status Monitoring: Utilizing a 3D model of the IMM in combination with data communication technology, the platform synchronously mirrors the physical system, dynamically presenting the operational process and spatial states of the MCS.

② Virtual Commissioning: Leveraging the simulation model and measurable parameters of the EMH system, the platform maps unmeasurable variables to potential system performance, thereby enhancing transparency in system debugging.

③ Data Management: The platform provides efficient and intuitive data query and analysis capabilities, supports historical data retrieval, and integrates comparative analysis of experimental and simulation data to reveal underlying patterns in system parameter variations, offering a foundation for optimization decisions.

(2) Platform Architecture

Based on the functional positioning, the DRL includes display data and simulation data. Display data are connected to the data acquisition system through experimental data communication, enabling data integration and storage. Simulation data rely on software interfaces to establish communication between the SL and scripts, supporting co-simulation of the MCS, storing computation results, and facilitating comparison between experimental and simulation data.

The status monitoring function maps physical entity data to the experimental data layer via data communication, ensuring synchronous updates between the VL and the PL and enabling intuitive monitoring of system states. The virtual simulation function allows parameter adjustments through the platform and visualizes simulation results, providing support for system optimization. The data management function integrates experimental and simulation data, facilitating data comparison, trend analysis, and optimization decision-making on the platform, thereby enhancing the tunability and optimization capability. The development environment and tools used in this study are listed in TABLE3.1.

Table3.1. Digital Platform System Development Environment and Tools

Item	Describe
Operating System	Window10
Programming Language	C#, Python2.7.15
Open-Source Tool	Numpy, Matplotlib
Database	My SQL8.0, Navicat
Visualization Tool	Unity3D, 3Ds Max, Xchart

4. PLATFORM-ASSISTED MODELING AND CONTROL OPTIMIZATION

This section presents the functional modules of the platform developed based on the aforementioned architecture and, taking the MCS performance degradation induced by oil temperature rise shown in FIGURE 2.1 as a case study, verifies the effectiveness of the proposed technical architecture in system diagnosis and performance optimization.

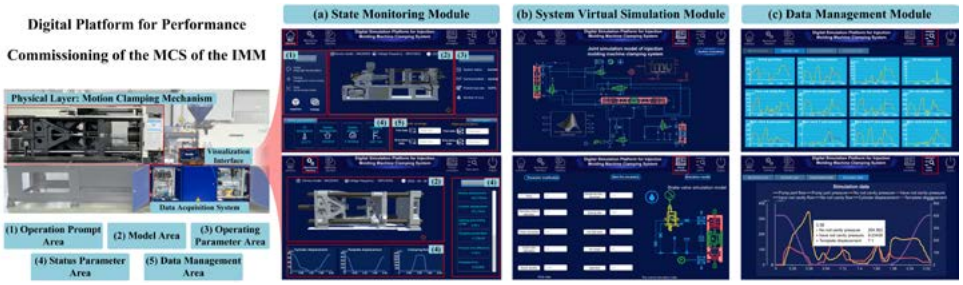


Figure 4.1. Digital Platform Interface of the MCS of IMM

4.1. Platform Functions

The platform takes the MCS as the research object and, through multi-level modelling as well as parameter interaction and analysis, realizes visual representation of system operating states, online evaluation of key parameters, and closed-loop support for the design and commissioning processes, thereby providing a unified tool environment for performance degradation diagnosis and validation of optimization strategies. The overall interface of the platform is shown in FIGURE 4.1.

(a) *State Monitoring and Visualization Function*

The platform integrates state monitoring functions at the system, mechanical, and hydraulic levels. Through the coordinated visualization of three-dimensional models and parameter panels, the operating state of the equipment is presented in an intuitive manner. During system operation, key parameters such as oil temperature, platen displacement, system pressure, and clamping force are updated in real time and synchronously displayed with the motion process. At the mechanical level, the platform directly reflects state variations of the MCS, while at the hydraulic level, dynamic schematic diagrams illustrate hydraulic circuit switching and actuator actions. This enables comprehensive perception of the multi-physical states of the EMH system, providing an intuitive basis for operating condition assessment and anomaly identification.

(b) *System- and Component-Level Virtual Simulation Function*

The platform adopts a hierarchical virtual simulation architecture, enabling parametric simulation and result analysis at both the system and component levels. Based on the system

model, users can adjust key component parameters and trigger corresponding simulation calculations, with the simulation results fed back to the platform in real time for performance evaluation. This functionality establishes a rapid mapping between variations in component parameters and the dynamic response of the system, providing an efficient virtual validation approach for structural design, parameter matching, and performance tuning.

(c) *Data Management and Comparative Analysis Function*

The platform incorporates a unified data management module that enables centralized storage, retrieval, and comparative analysis of experimental and simulation data. Users can access historical data by operating cycle and perform synchronized comparative visualization of multiple categories of system parameters to reveal performance differences under varying operating conditions or parameter configurations. By supporting visualization of multiple physical quantities across multiple coordinate axes, the platform significantly enhances data interpretation efficiency, providing robust data support for system performance degradation analysis and optimization decision-making.

4.2. Optimization Support Based on the Digital Platform

Conventional performance tuning of the MCS in IMM primarily relies on an offline trial-and-error process driven by human experience. This process typically takes directly measurable explicit performance indicators, such as cylinder displacement error and system pressure response, as references, and compensates for performance degradation caused by changes in leakage characteristics, evolution of friction behaviour, and deterioration of actuator dynamic response during long-term operation by repeatedly adjusting parameters including control loop gains, pressure thresholds, or servo pump speed limits. However, this tuning approach is essentially an experience-based correction method relying on outcome feedback, and its effectiveness is highly dependent on the subjective judgment of the operator. In particular, for key influencing factors that are difficult to measure online, their effects can only be indirectly reflected through system-level performance degradation, making it challenging to distinguish whether observed performance variations originate from improper control parameter settings or from the evolution of component intrinsic characteristics under varying operating conditions. As a result, a clear causal mapping between parameter adjustments and system responses is lacking. Owing to these limitations, traditional manual tuning is often characterized by long commissioning cycles and poor reproducibility of parameter settings, and it provides limited support for subsequent mechanism analysis, model refinement, and performance optimization.

This study takes the empirically observed phenomenon of system performance degradation induced by variations in hydraulic oil temperature as the starting point. By combining theoretical derivation and analysis with the mapping capability of a digital platform, a quantitative relationship is established between the viscosity-temperature characteristics of the hydraulic oil, the leakage behaviour of the gear pump, and the corresponding flow equations. On this basis, explicit characterization of pump efficiency variations and the development of a corresponding compensation model are achieved.

As the core driving unit of a servo pump-controlled system, the hydraulic pump plays a decisive role in overall system performance, and its internal leakage is one of the primary factors leading to performance degradation, which is commonly characterized by the pump volumetric efficiency. Owing to the compact structure of gear pumps, the dynamic microscopic clearances and the associated leakage paths that vary with operating conditions

are difficult to measure directly using conventional methods. Therefore, this study adopts an experimental data–driven parameter inverse identification approach to systematically test and model the efficiency characteristics of the pump under different oil temperature, pressure, and rotational speed conditions. Specifically, based on the developed digital platform, multi-condition experimental data and the corresponding simulation results are managed in a unified manner and visualized synchronously. By aligning and analysing key variables such as pump outlet flow rate, pressure, and rotational speed, and by minimizing the discrepancy between experimental measurements and model outputs, the key parameters related to internal leakage in the flow rate equation are inversely identified. In this way, a quantitative mapping between pump performance indices and theoretical model parameters is established. Consequently, the flow rate equation of the engineering-oriented internal gear pump can be redefined as:

$$Q_{\mathrm{a}} = v_{\mathrm{c}} n_{\mathrm{p}} \cdot \left[1 - \frac{C_{\mathrm{leak}} \cdot \Delta p}{\mu(T) \cdot v_{\mathrm{c}} n_{\mathrm{p}}} \right] \tag{4.1}$$

where, Q_{a} denotes the actual output flow rate, v_{c} is the constant displacement of the pump, n_{p} represents the actual rotational speed of the pump, Δp is the pressure difference between the high-pressure and low-pressure chambers, μ(T) denotes the dynamic viscosity of the hydraulic oil as a function of temperature, and C_{leak} is the fitted leakage coefficient.

The volumetric efficiency of the pump can then be defined as:

$$\eta_{\mathrm{v}} = \frac{Q_{\mathrm{a}}}{v_{\mathrm{c}} \cdot n_{\mathrm{p}}} = 1 - \frac{C_{\mathrm{leak}} \cdot \Delta p}{\mu(T) \cdot v_{\mathrm{c}} n_{\mathrm{p}}} \tag{4.2}$$

Equation (4.2) indicates that the pump's flow delivery capability is primarily influenced by the combined effects of rotational speed n_{p}, hydraulic oil viscosity μ(T), and pressure differential Δp. Among these factors, the temperature-dependent evolution of oil viscosity is difficult to measure directly and represents a key source of uncertainty affecting the accuracy of pump volumetric efficiency estimation. To address this issue, the digital platform was employed to uniformly collect and manage system operating data for the hydraulic oil used (ISO VG 46) under varying temperature conditions. A parametric model of the relationship between oil viscosity and temperature was then developed, yielding an exponential viscosity-temperature empirical correlation, as shown in FIGURE 4.2 When this viscosity-temperature model is explicitly embedded in the controller, the current oil viscosity can be continuously estimated based on real-time readings from a temperature sensor, and, in combination with Equation (4.2), the pump volumetric efficiency η_{v} can be calculated and compensated online.

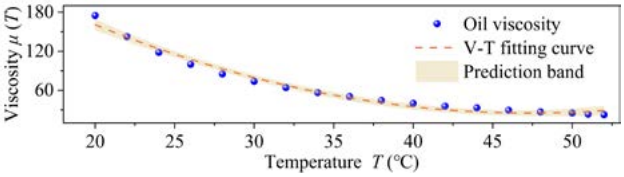


Figure 4.2. Viscosity-temperature fitting curve of hydraulic oil

4.3. Platform Performance

By integrating the oil viscosity–temperature model into the controller, this study introduces a pump flow compensation channel based on the viscosity-temperature model on top of the

original system's position-based PID controller, aiming to mitigate system positioning accuracy degradation caused by oil temperature rise. The corresponding framework is illustrated in FIGURE 4.3.

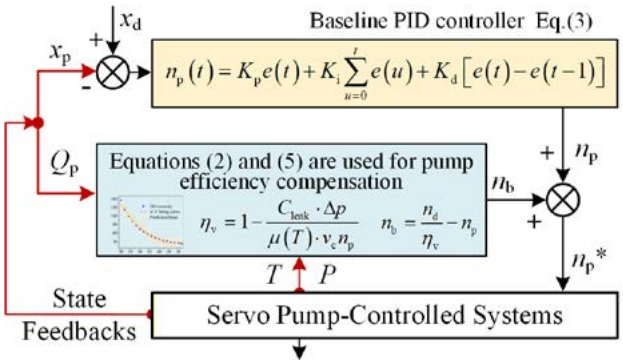


Figure 4.3. Pump Flow Compensation Strategy Based on the Platform-Fitted V-T Model

PID control is a classical method for servo pump-controlled systems. In the original system, position-tracking errors are constructed based on discretely sampled position signals, and the control law can be expressed as:

$$n(t) = K_p e(t) + K_i \sum_{u=0}^{t} e(u) + K_d \left[e(t) - e(t-1) \right] \tag{4.3}$$

where, e(t) denotes the tracking error, K_p, K_i, and K_d are the controller tuning parameters, and n(t) represents the controller output. The tuning parameters are set based on engineering practice and remain fixed.

Using the volumetric efficiency η_v estimated from the viscosity–temperature model, the compensation channel adjusts the pump command to offset the effects of leakage. Specifically, when the baseline PID controller requests a desired displacement x_d (or n_d), the compensation channel proportionally amplifies the speed command so that the actual flow Q_p closely matches the ideal demand $Q_d = n_d \cdot v_c$. The corresponding compensation flow Q_b can then be obtained as the reciprocal of the pump efficiency:

$$Q_b = Q_p^* - Q_p = \frac{Q_d}{\eta_v} - Q_p = \frac{Q_d \cdot \mu(T) \cdot v_c n_p}{\mu(T) \cdot v_c n_p - C_{leak} \cdot \Delta p} - Q_p \tag{4.4}$$

where Q_p* is the compensated flow, such that $Q_p = \eta_v \cdot Q_p^* = Q_d$.

Equivalently, the speed command of the servo motor can be scaled by the reciprocal of the efficiency to obtain the compensated speed n_b as:

$$n_b = n_p^* - n_p = \frac{n_d}{\eta_v} - n_p \tag{4.5}$$

where n_p* denotes the compensated rotational speed.

As illustrated in FIGURE 4.3, this channel functions as a parallel feedforward path, taking the baseline output n_p along with the current temperature T and pressure differential Δp as inputs to the compensation module, which outputs the gain-corrected command n_b. When η_v<1, n_b increases the pump speed appropriately to compensate for the flow deficit.

4.4. Performance Evaluation

To validate performance, multiple representative operating conditions were designed, under each of which 4,000 motion cycles were executed and data were collected for more than 4 hours in total. This design sufficiently captures the slow-varying characteristics of oil temperature during long-term operation, thereby verifying the general applicability and superiority of the platform and the proposed method under different conditions. The operating conditions were defined by varying actuator stroke and velocity, including a typical application condition with a 378 mm stroke, a short-stroke condition of 290 mm, a long-stroke condition of 406 mm, and a high-speed motion condition under the same long-stroke configuration. The detailed parameters of each condition are listed in TABLE 4.1, where the stroke and speed were set as percentages of the maximum hydraulic actuator stroke specified in the test standard [20]. Under the same temperature range, the tuning results obtained for different operating conditions are shown in FIGURE 4.4.

Table4.1. Detailed Experimental Operating Conditions

Scene	Distance (mm)	Velocity percentage	Purpose
1	378	80%	Frequent
2	190	70%	Short distance
3	406	70%	Long distance
4	406	99%	Speediness

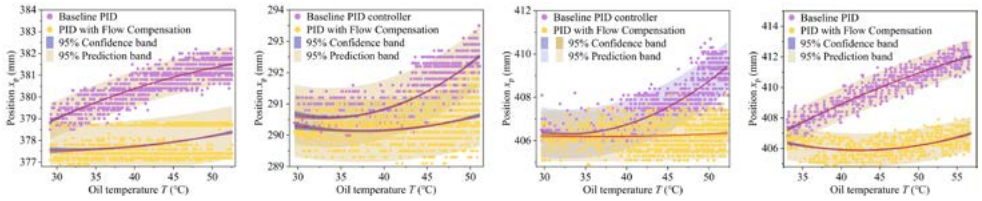


Figure4.4. Comparison of optimization effects for MCS. (a) Common operating condition. (b) Short-stroke condition. (c) Long-stroke condition. (d) High-speed condition

A cost function is constructed with the objective of minimizing the displacement tracking error to evaluate the optimization performance, and its magnitude of the real system under long-term operating conditions. The formulation is given as follows:

$$J = \sum_{k=1}^{N} \frac{\left(x_{\mathrm{p}} - x_{\mathrm{d}}\right)^2}{N} \tag{4.6}$$

where x_{p} denotes the actual terminal displacement of the system in the k-th experiment, and x_{d} represents the desired terminal displacement.

Table 4.2. The minimum cost values before and after optimization

Type	Common Operating	Short-Stroke	Long-Stroke	High-speed
J_{b}	7.61	2.32	3.75	18.33
J_{c}	0.45	0.37	0.41	0.44
R	94.1%	84.1%	89.1%	97.6%

With the minimization of the cost function J as the optimization objective, the minimum cost values corresponding to all tested configurations obtained from the displacement

tracking error data are summarized in TABLE4.2. Compared with the cost function of the baseline system J_b, the optimized cost function J_c achieved using the digital platform is reduced by 94.1%, 84.1%, 89.1%, and 97.6%, respectively, demonstrating a significant improvement in system tracking accuracy.

5. CONCLUSION

This study addresses the issue of performance degradation in the MCS of IMM under temperature rise conditions by proposing a digital platform–supported approach for performance modelling and control optimization. By developing a platform that integrates data acquisition, simulation analysis, and data review, unified analysis of system operating data and model predictions is achieved. Focusing on the reduction of pump efficiency and deterioration of system positioning accuracy caused by temperature-dependent oil viscosity, the approach employs the digital platform for parameter mapping, inverse identification, and implicit indicator extraction to realize an engineering-oriented redefinition of pump leakage characteristics and explicitly incorporates the viscosity-temperature model into system modelling. On this basis, a pump flow compensation channel is introduced into the original position-based PID control structure to implement feedforward correction for the effects of oil temperature variation. Experimental results demonstrate that the proposed method significantly improves system positioning accuracy under multiple operating conditions, providing an effective interactive control optimization strategy for performance degradation driven by unmeasurable latent variables in EMH systems. Moreover, the proposed digital platform architecture and modelling methodology are designed with strong engineering scalability and system-level compatibility, enabling seamless alignment with both existing and emerging industrial digitalization standards. This provides a solid foundation for future integration with standardized communication frameworks, digital twin architectures, and collaborative simulation environments.

ACKNOWLEDGEMENTS

This work was supported in part by National Science and Technology Major Project under Grant No.12124778012.

REFERENCES

[1] T. Böttjer, D. Tola, F. Kakavandi, et. al., ‘A review of unit level digital twin applications in the manufacturing industry’, CIRP J. Manuf. Sci. Technol., Vol. 45: pp. 162-189, Oct ,2023. DOI: 10.1016/j.cirpj.2023.06.011.

[2] N. Wang, J. Wan, Z. Ma, et. al., ‘How digital platform capabilities improve sustainable innovation performance of firms: The mediating role of open innovation’, J. Bus. Res., Vol. 167: pp. 114080, 2023. DOI: 10.1016/j.jbusres.2023.114080.

[3] J. Mohd, H. Abid, S. Rajiv, ‘Digital twin applications toward industry 4.0: a review’, Cognitive Robotics, Vol. 3: pp. 71-92, 2023. DOI: 10.1016/j.cogr.2023.04.003.

[4] C. Qiu, S. Zhou, Z. Liu, et. al., ‘Digital assembly technology based on augmented reality and digital twins: a review’, Virtual Reality & Intelligent Hardware, Vol. 1: pp. 597-610, 2019. DOI: 10.1016/j.vrih.2019.10.002.

[5] R. Jin, L. Li, X. Liang, et. al., ‘Energy-efficient design of the powertrain for mechanical-electro-hydraulic equipment via configuring multidimensional

controllable variables', Renew. Sust. Energ. Rev., Vol. 201: pp. 114511, 2024. DOI: 10.1016/j.rser.2024.114511.
[6] Z. He, F. Zhai, Y. Zhang, et. al., 'Research on dynamics model of the motion clamping mechanism in injection molding machines addressing inertia and deformation', J. Manuf. Process., Vol. 119: pp. 987-1004, 2024. DOI: 10.1016/j.jmapro.2024.04.027.
[7] Y. Ran, Q. He, W. Huang., et. al., 'Analysis of the coupling mechanism of the dynamic response and mechanical-thermal deformation in mechanical seals', Tribol. Int., Vol. 192: pp. 109257, 2024. DOI: 10.1016/j.triboint.2024.109257.
[8] M. Tang, Y. Wu, J. Wang, 'An intelligent decision system for debugging engine fuel regulators', Expert Syst. Appl., Vol. 265: pp. 125987, 2025. DOI: 10.1016/j.eswa.2024.125987.
[9] S. Ha, Y. Park, J. Kim, et. al., 'Research trends of digital platforms: A survey of the literature from 2018 to 2021', Telecommun. Policy, Vol. 47: pp. 102543, 2023. DOI: 10.1016/j.telpol.2023.102543.
[10] J. Leng, J. Guo, J. Xie, et. al., 'Review of manufacturing system design in the interplay of Industry 4.0 and Industry 5.0 (Part I): Design thinking and modeling methods', J. Manuf. Syst., Vol. 76: pp. 158-187, 2024. DOI: 10.1016/j.jmsy.2024.07.012.
[11] N. Striffler, T. Voigt, 'Concepts and trends of virtual commissioning–A comprehensive review', J. Manuf. Syst., Vol. 71: pp. 664-680, 2023. DOI: 10.1016/j.jmsy.2023.10.013.
[12] H. Hajar, Q. Qu, S. Mojtaba, et. al., 'Digital twin empowering manufacturing paradigms: lean, agile! just-in-time (jit), flexible, resilience, sustainable', 10th Int. Conf. on Infor. Tech. and Quan. Mana., Vol. 221: pp. 1258-1267, Oxfordshire, 2023. DOI: 10.1016/j.procs.2023.08.114.
[13] R. Dinter, B. Tekinerdogan, C. Catal, 'Reference architecture for digital twin-based predictive maintenance systems', Comput. Ind. Eng., Vol. 177: pp. 109099, 2023. DOI: 10.1016/j.cie.2023.109099.
[14] X. Zhu, J. Chang, Q. Qin, et. al., 'Research and application of intelligent debugging platform for emu braking system', 2017 Chinese Auto. Cong., pp. 23-28, Jinan, 2018. DOI: 10.1109/CAC.2017.8243200.
[15] J. Puggaard, E. Ribeiro, A. Bilberg, C. Bro, 'Design and development of automation equipment based on digital twins and virtual commissioning', 54th CIRP Conf. on Manu. Sys., Vol. 104: pp. 1167-1172, virtual conf., 2021. DOI: 10.1016/j.procir.2021.11.196.
[16] E. Martin, S. Christian, P. Josef, et. al., 'A novel automated NFC interoperability test and debug system', 2019 Euro. Conf. on Smart Objects, Sys. and Tech., pp. 27-34, Magdeburg, 2019. Web: https://ieeexplore.ieee.org/abstract/document/8835961.
[17] H. Liu, O. Chernova, 'For intelligent debugging management of offshore oil engineering with big data', J. Intell. Fuzzy Syst., Vol. 40: pp. 6091-6101, 2021. DOI: 10.3233/JIFS-189448.
[18] A. Ivannikov. 'Test set generation for computer simulation design debugging of control digital systems with finite alphabet of functions' 2020 Int. Conf. on Ind. Eng. Appl. and Manu., pp. 1-5, Sochi, 2020. DOI: 10.1109/ICIEAM48468.2020.9112024.
[19] Y. Zhai, Q. Yang, X. Ji, et. al., 'Key technologies of intelligent debugging platform for the whole process of automation system', 2021 IOP Conf. Series: Earth and Envir. Sci., Vol. 701: pp. 012064, Nanjing, 2021. DOI: 10.1088/1755-1315/701/1/012064.
[20] General Specifications and Inspection Methods of Rubber and Plastics Injection Molding Machines, Chinese Standard GB/T 25156-2020, 2021. Web: https://www.chinesestandard.net/PDF.aspx/GBT25156-2020.

BIOGRAPHIES



Zhiqiang He received his B.Eng. degree in Mechanical Engineering from Yanshan University, Qinhuangdao, China, in 2019, and is currently pursuing a Ph.D. degree in Mechanical Design and Theory at the School of Mechanical Engineering, Yanshan University. His research interests focus on the control and optimization of electromechanical-hydraulic systems in industrial equipment, and their digital transformation.



Fugang Zhai received the B.Eng. degree in Mechanical Engineering from Yanshan University, Qinhuangdao, China, in 2004, and the Ph.D. degree in Mechatronic Engineering from the School of Mechanical Engineering, Yanshan University, in 2012.

Since 2012, he has been with the School of Mechanical Engineering at Yanshan University, where he serves as a professor, doctoral supervisor, and head of the Department of Mechanical Design. His research interests include system dynamics, transmission, and control of electromechanical-hydraulic systems in industrial equipment, and digital twin technology.



Zhuofan Yang received his B.Eng. degree in Mechanical Engineering from Yanshan University, Qinhuangdao, China, in 2023, and is currently pursuing a M.Eng. degree in Mechanical Design and Theory at Yanshan University. His main research interest is the digital control of electromechanical-hydraulic systems.



Tianran Yang received the B.Eng. degree in Mechanical Engineering from Central South University, Changsha, China, in 2014, and his M.S. degree in Mechanical Engineering from the University of Florida, Gainesville, US, in 2019.

Since 2021, he has been with the Beijing Research Institute of Automation for Machinery Industry Co., Ltd., where he serves as a Research and Development Engineer. His main research interest is the optimization of hydraulic systems.



Jing Yao received the B.Eng. degree in Mechatronic Engineering from Yanshan University, Qinhuangdao, China, in 2001, and her Ph.D. degree in Mechatronic Engineering from the School of Mechanical Engineering, Yanshan University, in 2009.

Since 2009, she has been with the School of Mechanical Engineering at Yanshan University, where she serves as a professor, doctoral supervisor, and dean of the School of Mechanical Engineering. Her research interests include the transmission and control of the flow power in industrial equipment, as well as the research on their digitalization and intelligence.

Downsizing of hydraulic power units by accurate prediction of disperse flows

Dr.-Ing. Mark-Patrick Mühlhausen*

Dr.-Ing. Markus Perschall*

M.Sc. Alexander Huber*

**Bosch Rexroth AG, Zum Eisengießer 1, 97816 Lohr am Main*

Corresponding author: mark-patrick.muehlhausen@boschrexroth.de

Abstract.

Downsizing of hydraulic power units means installing the same hydraulic power within a smaller footprint or increasing hydraulic power within the same footprint.

To achieve this, knowledge and influencing the condition of the hydraulic fluid and the tank flow are crucial. The condition of the hydraulic fluid depends primarily on dispersed contaminations by air bubbles, water droplets, and dirt particles. All of these should be separated in the reservoir, which can be significantly improved through suitable flow guidance. The resulting improved fluid condition contributes to significant resource savings over the entire product life cycle, thus generating tangible economic and technological advantages.

The difficulties in implementation lie, on the practical side, mainly in the strong influence of the system's operating conditions, such as pressure and flow rate curves, and on the modeling side, due to the complex properties of hydraulic fluids, which also vary across the entire spectrum of fluid classes. Both challenges often prevent generally valid statements and therefore require a multitude of CFD simulations for various systems and even the same systems under different load situations.

We combine the solution to the resulting high demands on the model quality and the performance of the numerical method in bTrack, our in-house solver for dispersed flows. In this presentation, we provide an overview of the physical and numerical fundamentals of bTrack. We start with simple equations of motion and extend these continuously with algorithms capturing more complex phenomena like collision, breakup and coalescence. We also present and discuss coupling strategies between the disperse and continuous phase implemented in bTrack. The performance and efficiency of the models are presented and discussed using various application examples for bubble, particle, and droplet flows in tanks.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

bTrack has thus been an essential component of the myCro Downsizing toolbox from Bosch Rexroth for more than a decade.

Keywords. downsizing, air bubbles, mycro, cfd, simulation

1. INTRODUCTION

The hydraulic industry faces increasing pressure to reduce the footprint and weight of hydraulic power units while maintaining or increasing performance. This downsizing trend is driven by demands for improved energy efficiency, reduced material costs, and enhanced system integration in mobile and industrial applications. A critical component is the hydraulic reservoir, which traditionally has been sized conservatively to ensure adequate fluid conditioning, heat dissipation, and contaminant settling. Significant reservoir size reduction is only feasible if fluid quality can be maintained despite reduced residence times and settling volumes. Hydraulic fluid condition is strongly determined by three types of dispersed contaminations: entrained air bubbles, emulsified water droplets, and solid dirt particles. Air enters through turbulent mixing at free surfaces, and suction-side leakage; water through condensation and seal leakage; and particles through external ingress and internal wear.

These dispersed contaminants cause numerous detrimental effects on hydraulic system performance and durability. Entrained air reduces the effective bulk modulus, leading to spongy system response, reduced positioning accuracy, and increased noise and vibration as well as oxidation and Diesel effect. Water contamination accelerates fluid degradation through hydrolysis, promotes corrosion of metallic components, reduces lubrication effectiveness, and catalyzes oxidation processes. Solid particles cause abrasive wear on precision surfaces, leading to internal leakage, reduced efficiency, and component failure.

The extent of air and water contamination strongly depends on both fluid properties and operating conditions. Air solubility in hydraulic fluids follows Henry's law, making gas absorption and release pressure-dependent phenomena. This can be illustrated by a carbonated mineral water bottle: when sealed under pressure, carbon dioxide remains dissolved; upon opening, pressure drops and gas bubbles form and rise. Similarly, hydraulic fluid returning from high-pressure zones to the atmospheric reservoir experiences pressure reduction, causing dissolved air to nucleate into bubbles requiring separation. Water solubility and emulsion stability are likewise influenced by temperature, pressure, and fluid additive packages, with mineral oils, synthetic esters, and water-glycol fluids exhibiting drastically different behaviors.

Successful reservoir downsizing through improved contamination control offers substantial lifecycle benefits. Reduced reservoir volume directly translates to lower fluid inventory, decreased system weight, and smaller footprints – critical advantages in mobile hydraulics where space and weight constraints are severe. Improved fluid quality extends component life by reducing wear rates, increase maintenance intervals by slowing fluid degradation, and enhances system reliability through consistent performance. Economically, these improvements reduce both initial investment and total cost of ownership. Environmental benefits include reduced fluid consumption and lower energy requirements from improved system efficiency. Furthermore, high fluid quality enables tighter manufacturing tolerances

in hydraulic components, permitting higher operating pressures and power densities – reinforcing downsizing objectives.

Effective downsizing requires accurate prediction of fluid flow patterns and contaminant separation efficiency within reservoirs. The reservoir must facilitate air bubble and solid particle separation while providing adequate deaeration surface, appropriate flow velocities to prevent particle resuspension, and sufficient residence time for settling. Computational Fluid Dynamics (CFD) enables optimization of reservoir designs, flow guidance structures, baffle arrangements, and inlet/outlet configurations without costly prototyping. However, reservoir flows present significant modeling challenges. The flow is inherently three-dimensional and highly unsteady, driven by turbulent structures and time-varying boundary conditions reflecting realistic operation. During typical operation, volume flow rates entering the reservoir vary by orders of magnitude depending on actuator movements, valve switching, and pump displacement changes. These transient conditions continuously alter the flow field, affecting particle trajectories and separation efficiency in ways not captured by steady-state simulations. Additionally, the wide range of hydraulic fluid types – from mineral oils to water-based fluids – each with distinct density, viscosity, surface tension, and foaming characteristics, prevents universal design rules.

Several numerical approaches exist for modeling dispersed multiphase flows. The Euler-Euler approach treats both phases as interpenetrating continua, solving conservation equations for each with interphase exchange terms. This method is computationally efficient at high volume fractions but requires closure models for interfacial forces and cannot easily track individual particle characteristics. The Euler-Lagrange approach treats the continuous phase as a continuum while tracking individual particles or bubbles as discrete entities moving according to Newton's equations. This method naturally accommodates polydisperse size distributions, enables direct implementation of collision, breakup, and coalescence models, and provides detailed particle trajectory statistics. For dilute to moderately dense dispersed flows – typical of hydraulic reservoirs where volume fractions generally remain below 10% – Euler-Lagrange offers optimal balance between physical fidelity and computational efficiency.

We employ a three-dimensional CFD approach combining Eulerian description of the continuous phase with Lagrangian tracking of dispersed contaminants. This framework captures complex, transient three-dimensional reservoir flow structures while resolving trajectories of individual air bubbles, water droplets, and solid particles. Lagrangian tracking naturally accounts for size-dependent separation, enables physically sound particle-particle and particle-wall interaction models, and facilitates bubble coalescence and breakup submodels. Two-way coupling ensures dispersed contaminants appropriately modify continuous phase momentum and turbulence fields, essential for accurate prediction at higher concentrations. This methodology is implemented in our in-house solver bTrack, developed and validated specifically for dispersed flows in hydraulic systems. bTrack is well matured and has been used in the analysis and optimization of hydraulic power units of Bosch Rexroth since more than a decade. Since the



Figure 1: mycro logo

beginning, it has been an essential part of the myCro Downsizing toolbox [1] and is today a further growing key factor of its success contributing to both customer-specific ETO applications as well as Bosch Rexroth Standard HPU solutions.

2. METHODOLOGY

The continuous fluid phase is described by the incompressible Navier-Stokes equations for mass and momentum conservation:

$$\nabla \cdot \mathbf{u} = 0$$

$$\frac{\partial}{\partial t}\rho \boldsymbol{u} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}\boldsymbol{u}) = -\boldsymbol{\nabla} \mathrm{p} + \boldsymbol{\nabla} \cdot \left(2\mu_{eff}\boldsymbol{S}\right) + \rho\, \boldsymbol{g} + \boldsymbol{S}_{b,mom}$$

with $\boldsymbol{u}$ the fluid velocity, p the pressure on the Euler grid, $\boldsymbol{S} = \frac{1}{2}[\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T]$ the deformation rate tensor. The momentum source term $\boldsymbol{S}_{b,mom}$ are caused from the bubbles and it is nonzero only in case of two-way coupling. The density is the pure liquid density $\rho = \rho_l$ and assumed constant, throughout. The viscosity is the liquid viscosity $\mu_{eff} = \mu_l$ in case of a laminar flow. These equations are solved on a 3D domain with Cartesian coordinates. Turbulence is modeled using RANS or LES approaches.

The bubble velocity is computed by solving an ordinary differential equation for the linear momentum considering different forces on the bubble. Generally, these forces may be divided into fluid dynamic forces (e.g. drag, pressure, virtual mass, and Saffman forces) and field forces (gravity, electrostatic forces, etc.). The new bubble position is then obtained in each Lagrange time step by integrating the bubble velocity so that the equations read

$$\frac{d\boldsymbol{x}_b}{dt} = \boldsymbol{u}_b$$

$$m_b \frac{d\boldsymbol{u}_b}{dt} = \boldsymbol{F}_D + \boldsymbol{F}_L + \boldsymbol{F}_{vm} + \boldsymbol{F}_p + \boldsymbol{F}_g + \boldsymbol{F}_{coll}$$

where $\boldsymbol{x}_b$ is the bubble position vector, m_b the bubble mass and $\boldsymbol{u}_b$ the bubble velocity vector. The forces on the right-hand side are drag force, lift force, virtual mass force, pressure gradient force, and gravity, respectively. For solid particles, it may be required to account for their rotation by additionally solving an equation of the angular momentum. Due to the extremely small mass of bubbles and their behavior during collisions, this is typically not considered.

The drag force results from the flow around the particle and can be expressed with the following equation

$$F_D = \frac{3}{4} m_b \frac{\rho_l}{\rho_b} \frac{C_D}{D_b} (\boldsymbol{u} - \boldsymbol{u}_b)|\boldsymbol{u} - \boldsymbol{u}_b|$$

where C_D is the empirical drag coefficient, and D_b is the bubble diameter. Because the drag force is one of the most important forces for disperse flows the right choice of the drag coefficient is crucial. There are several correlations proposed in literature, e.g. [2].

The lift force is caused by a shear flow around the bubble. It acts perpendicularly to the direction of the flow and can be expressed as proposed by [3].

$$F_L = C_L m_b \frac{\rho_l}{\rho_b}[(\boldsymbol{u} - \boldsymbol{u}_b) \times (\nabla \times \boldsymbol{u})]$$

where the lift force coefficient depends on the deformation of the bubble or droplet and can be expressed by the correlation of [4].

The virtual mass force accounts for the additional work that the bubble performs to accelerate or decelerate the surrounding fluid and can be written

$$\boldsymbol{F}_{vm} = C_{vm} m_b \frac{\rho_l}{\rho_b}\left(\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} - \frac{d\boldsymbol{u}_b}{dt}\right)$$

like proposed by [5]. The virtual mass force coefficient is typically modeled constant.

The sum of the pressure force and the gravity force can be expressed as [6]

$$\boldsymbol{F}_p + \boldsymbol{F}_g = m_b \frac{\rho_l}{\rho_b}\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + m_b \boldsymbol{g}\left(1 - \frac{\rho_l}{\rho_b}\right).$$

The collision force can be incorporated using hard-sphere or soft-sphere collision paradigm. All given equations are relying on choosing the respective coefficients carefully. Most of the coefficients from literature assumes ideal isolated lab conditions or deal with water or purified fluids. Due to the strong influence of the fluid properties and the deviation to hydraulic fluids an arbitrary error in prediction is introduced. To avoid these, the used relevant coefficients are based on experiments with hydraulic fluids.

Numerical Framework

The bTrack solver is an in-house tool designed for disperse multiphase flows in hydraulic systems, operating standalone or two-way coupled with ANSYS Fluent using first and second-order time integration. The solver is fully parallelized using MPI and written in Python and C, leading to both great flexibility and high performance. The software architecture is modular and designed to intercept and influence the operation at all relevant points during simulation to maximize flexibility during model development. Setup, monitoring and output files follow common open-source standards. Setup creation is simplified via a simple UI reducing input errors. The output format is also chosen to work seamlessly with Paraview. bTrack can be fully controlled via a command line interface, which makes it very suitable for parameter studies and automation.

In case of two-way coupling interpolation schemes from disperse to continuous phase and vice-versa are required. Besides robust and well validated schemes for interpolation of volume fraction to the Cartesian grid and undisturbed liquid velocity to each Lagrangian particle we also developed and validated interpolation schemes reducing computational effort.

Modelling turbulence in multiphase flows can be computational expensive due to the interplay between the turbulent continuous flow and the disperse phase, because it could demand a fully two-way coupled approach. To limit the effort while retaining the accuracy carefully modelling the effects is important. bTrack incorporates models for bubble induced turbulence to modify turbulent characteristics of the continuous phase and an advanced random walk model to take turbulent fluctuations into account for bubble motion calculation. bTrack relies on local fluid and flow properties by default, so that multiphase

flows with spatial and/or temporal varying properties like density, temperature, viscosity, etc. can also be simulated. Based on local flow properties bTrack can simulate bubble coalescence, deformation and breakup. As a foundation of coalescence and collision an efficient detection kernel based on Sweep-and-prune algorithm is implemented. An overview about the features can be found here [7, 8].

3. RESULTS

Within this chapter we present several reservoir flows where the capabilities of bTrack can be shown. Figure 2 shows a typical reservoir flow. The flow inside is highly transient caused by two aspects: boundary conditions and flow instabilities. The typical load cycle of an application leads to a time dependent flow and pressure profile at each inlet or outlet of the reservoir. Unsteadiness and magnitude of flow rate and pressure lead then to flow instabilities inside the reservoir. Our experience has shown that capturing these effects is crucial in many applications. mycro optimized reservoirs are intended to work even with large oil volume changes, so called working or pendulum volume, inside the reservoir. To capture all relevant effects a multiphase simulation is often necessary. In figure 2 bTrack is used to calculate bubble motion and behaviour of different sizes within the oil volume, while the oil and air phase and their interface are simulated via a Volume-of-Fluid approach. bTrack can not only output all raw data but can compute many relevant characteristics and statistics during simulation. This saves disk space and speeds up post processing.

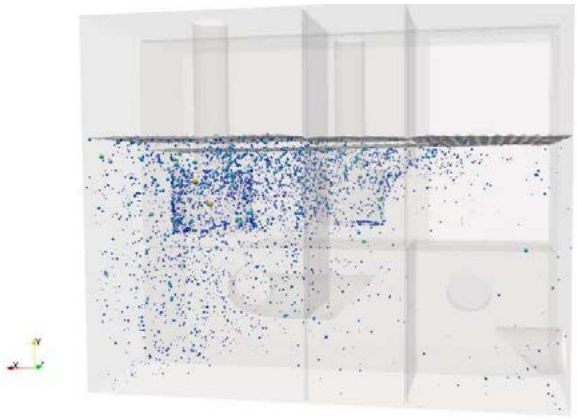


Figure 2: Transient multiphase flow

In figure 3 a bTrack water droplet simulation is shown. The droplets enter the reservoir through one of the pipes mounted from the top. The flow in this example is from the top to the bottom. Of main interest here is the accumulation of water droplets leading to puddle depending on the different vertical reservoir sections. Due to the Lagrangian approach of bTrack the motion and position of each droplet can be calculated in detail. Using statistical methods, the locations and extensions of puddles can be predicted. These predictions were in good agreement with experimental

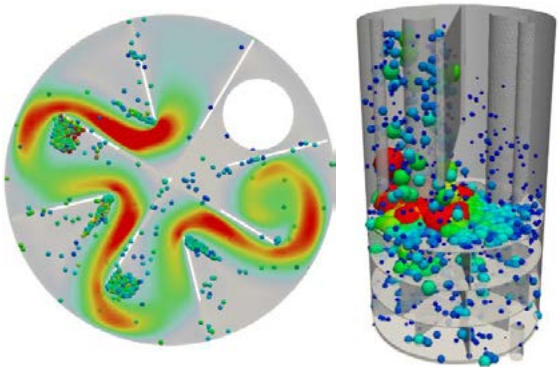


Figure 3: bTrack droplet simulation

investigations. The knowledge of the motion and position of disperse pollutions, like droplets, bubbles or particles are used in application dependent reservoir design.

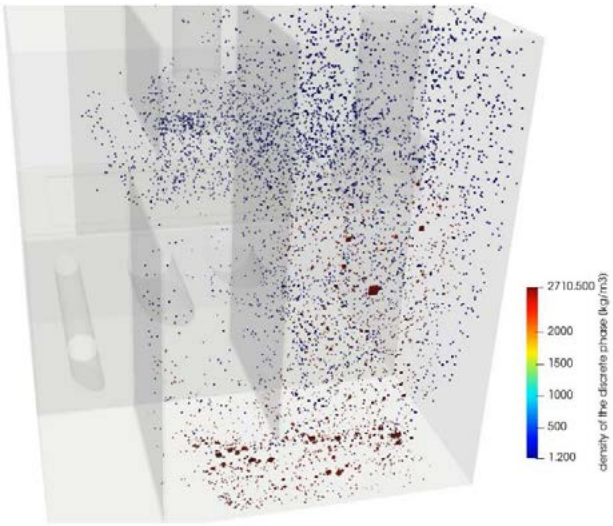


Figure 4: bTrack calculates multiple disperse phases

Figure 4 demonstrates how bTrack can handle multiple disperse phases at the same time, which enables for multi objective optimization of contaminant separation. In the given application metal particles in red are settling at the bottom, while air bubbles in blue rise to the oil surface. Both phases are influencing the velocity of continuous phase, the oil flow. This fact impacts the settling and rising times and are important for reservoir design and sizing. Each disperse phase can be described with their own properties, their own physical models and model constants. This allows for very accurate predictions and saves time.

Although the separation and settling takes place in the reservoir the flow inside is heavily depending on the application. One of these applications is shown in figure 5. If necessary bTrack is not limited to flows inside the reservoir. Disperse flows in suction and return lines can also be regarded in the same simulation, because disperse phase models, e.g. for bubble size, are taking local flow properties like pressure, temperature, density, volume fraction, etc. and numerical parameters like CFL number into account. This means that one description of a disperse phase can be valid for a wide range of flow conditions.

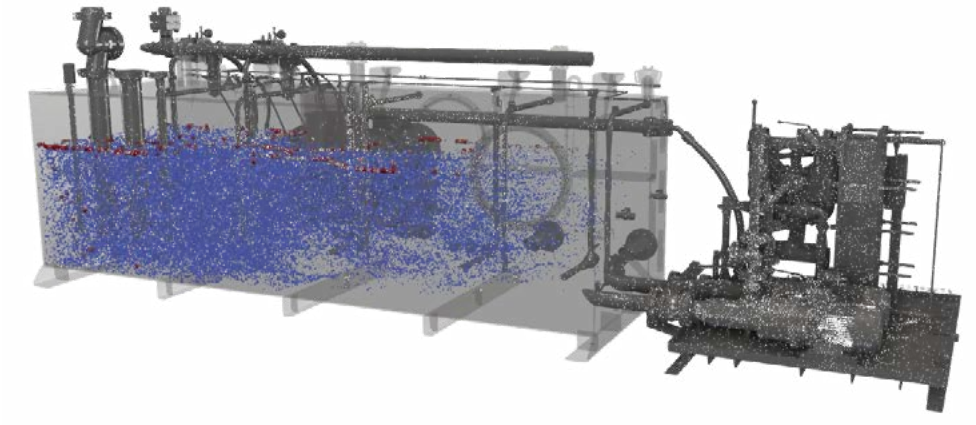


Figure 5: bTrack can capture flow in return and suction lines

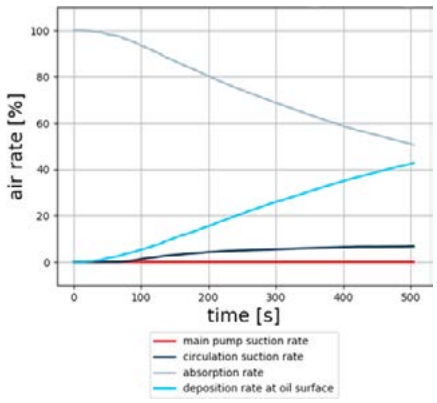


Figure 6: Air rate statistics

Detailed statistics like the air rates shown in figure 6 are calculated by default in bTrack and used for competitive design studies as well as in the analysis how robust a design is under changing fluid, ambient or operating conditions. These statistics are done for test bubble size spectra derived from several inhouse air release experiments for different hydraulic fluids. Application specific bubble size spectra are the foundation of reliable prediction.

Mycro reservoirs must fulfill highest standards and pass bTrack simulations with most conservative bubble size spectra.

4. CONCLUSION AND OUTLOOK

In this work we presented the capabilities of our three-dimensional in-house solver bTrack for simulation of disperse multiphase flows. bTrack is used since more than one decade in analysis and optimization of hydraulic power units and is integral part of mycro, the downsizing toolbox of Bosch Rexroth. The huge amount of implemented models, the modular architecture and robust and performant numeric schemes makes it suitable for a wide range of applications. Experimental campaigns to derive new models, integrate new fluids and improve numerics are continuously carried out, to improve accuracy and reduce computational effort.

5. REFERENCES

[1] “Bosch Rexroth,” 21 01 2026. [Online]. Available: https://www.boschrexroth.com/de/de/produkte/industrielle-loesungen/industriehydraulik/aggregate/mycro/.

[2] A. K. I. Z. I. S. T. Tomiyama, “Drag coefficients of single bubbles under normal and micro gravity conditions,” *JSME Int. J. Ser.,* vol. 41, no. 2, p. 472–479, 1998.

[3] P. Saffman, “The lift on a small sphere in a slow shear flow,” *J. Fluid Mech.,* vol. 22, no. 2, pp. 385-400, 1965.

[4] A. T. H. Z. I. H. S. Tomiyama, “Transverse migration of single bubbles in simple shear flows,” *Chem. Eng. Sci.,* vol. 57, no. 11, p. 1849–1858, 2002.

[5] C. Crowe, Multiphase Flow Handbook. Mechanical and Aerospace Engineering Series, CRC Press, 2005.

[6] S. B. D. S. M. G. M. Laín, “Modelling hydrodynamics and turbulence in a bubble column using the Euler-Lagrange procedure,” *Int. J. Multiph. Flow,* vol. 28, no. 8, p. 1381–1407, 2002.

[7] X. &. M. M.-P. &. F. J. Yang, “Interpolation methods for two-way coupled Euler-Lagrange simulation of finite-size bubbles,” *Chemical Engineering Science,* vol. 238, 2021.

[8] X. &. M. M.-P. &. F. J. Yang, “Efficient simulation of bubble dispersion and resulting interaction,” *Experimental and Computational Multiphase Flow,* vol. 3, pp. 1-19, 2020.

Biographies

Mark-Patrick Mühlhausen received his Diploma degree in mechanical engineering from Karlsruhe Institute of Technology in 2009 and Ph.D. in Mechanical Engineering from Karlsruhe Institute of Technology in 2012, respectively. He is currently working at Bosch Rexroth in the Engineering Domain Virtual Engineering and Verification Centre. He focusses on model and tool development.

Sensor-Efficient Health Monitoring for Brownfield Industrial Hydraulic Systems

Reiner Stammberger[1], Arne Platzbecker[1], Joachim Nagelmann[1],
Maximilian Schmidt[2], Ahmad Al-Issa[2], Tobias Schulze[2], Jürgen Weber[2]

[1]*SMS group GmbH, Germany,* reiner.stammberger@sms-group.com, arne.platzbecker@sms-group.com, joachim.nagelmann@sms-group.com

[2]*Dresden University of Technology, Germany,* maximilian.schmidt2@mailbox.tu-dresden.de, ahmad.al-issa@mailbox.tu-dresden.de, tobias.schulze2@tu-dresden.de, fluidtronik@mailbox.tu-dresden.de

Abstract.

SMS group GmbH presents a hierarchical health tree condition monitoring system specifically engineered for brownfield industrial hydraulic installations, where brownfield installations refers to the process of upgrading, expanding, or re-purposing existing facilities, for example by integrating online monitoring technologies or modern control systems. Based on SMS group's long-term OEM and service experience in the metal industry, this paper presents a framework to encode expert knowledge into an automated monitoring architecture that avoids the extensive labelled datasets required by many machine-learning approaches. The framework aggregates sensor data and operational parameters into a single, transparent Health Index, providing intuitive, real-time feedback for both maintenance and production crews. A novel, aggressive multiplication logic ensures that any component degradation is immediately reflected at the system level, preventing the failure-masking effects of conventional aggregation schemes.

The solution was developed and validated on a 1970s-era rolling mill, designed by SMS group GmbH. The implementation required only a minimal set of strategically placed sensors and a few hours of planned downtime for installation. The results confirm robust fault detection under realistic industrial conditions, thus leading to a significant reduction in unplanned downtime and a significant resource efficiency improvement. Furthermore, the system is able to quantify gradual degradation trends - a practical foundation for future predictive maintenance programs in challenging brownfield environments.

Keywords. Condition Monitoring, Brownfield equipment, Industrial Hydraulics, Health Index, Expert Knowledge, Predictive Maintenance.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Complex servo-hydraulic systems have been installed widely in rolling mills since the early 1970s to control the rolling force respectively the roll gap among other functions. With the demand for higher product quality and output, an increasing number of hydraulic functions were introduced into rolling mills, many of which were retrofitted into existing plants. This longevity, however, creates a fundamental tension: while the core machinery endures for decades, its components—comprised of countless moving parts—are inherently subject to continuous wear.

To increase overall equipment effectiveness (OEE) and product quality, a well-maintained hydraulic system is key. The above-mentioned brownfield hydraulic systems, which have grown organically over decades, present a formidable challenge for modernization. They are complex, hydraulically coupled networks of components from various eras, operating under immense pressure to maintain non-stop production.

This research builds upon an SMS group GmbH patent titled 'Method for monitoring the condition of a hydraulic system of a metal forming plant and condition-monitoring device' [1]. The scientific and technical realization of this patented concept was driven by a unique trilateral collaboration: SMS group GmbH, as the original equipment manufacturer (OEM) with decades of service experience, providing the foundational equipment know-how and an operator of a brownfield SMS plant contributing the critical perspective on real-world customer needs. This combined industrial expertise was then scientifically processed by the Chair of Fluid-Mechatronic Systems (TU Dresden), allowing to develop the methodology and implementation described in chapter 4 and following.

The central question this work addresses is how to bridge the gap between sophisticated maintenance experience and the need for digital, systematic monitoring in an environment where resources are severely constrained. Our answer is rooted in the formalization of decades of hands-on experience from maintenance and commissioning experts within the SMS group GmbH. This deep, practical knowledge is the essential ingredient to translate the complex, often tacit, operational experience from a customer's production environment into a systematic, digital assessment. Prevailing condition monitoring paradigms often prove unsuitable for this context. Approaches rooted in machine learning and artificial intelligence frequently require vast amounts of labelled fault data. This prerequisite fundamentally clashes with the reality of economically viable brownfield systems for two reasons: first, fault states are by necessity rare, making such data inherently scarce; second, their root causes are often obscured by complex system-level interactions that defy simple generalization. This data scarcity and ambiguity often result in “black box” models that are met with distrust by experienced maintenance crews [2]. While valuable for algorithmic development, solutions demonstrated on laboratory test rigs under controlled conditions often fail to translate to the operational volatility of a real factory floor.

Indeed, sophisticated approaches like hierarchical health index frameworks have been successfully implemented in data-rich sectors like semiconductor manufacturing [3]. However, these successes are typically enabled by extensive sensor networks and dedicated data science teams capable of managing complex statistical aggregation methods, such as the Analytic Hierarchy Process (AHP). This creates an economic and logistical barrier that is insurmountable for most legacy brownfield plants. Furthermore, these methods often

involve complex weighting debates and can mask critical single point degradations through averaging.

As the SMS group GmbH, we are proud to introduce an elegant approach to implement condition monitoring for existing hydraulic systems: a hierarchical framework designed specifically for resource-constrained, expert-driven environments in brownfield applications of the SMS group GmbH. With this method, we are able to elevate operation of old plants to a new level of availability. For the rolling mill, a reliable operation of the hydraulic system is of absolute importance. Therefore, this paper describes the approach to assessing “Operational Production Health” rather than monitoring single component wear. An 80% worn pump, for example, may still provide 100% production health if it reliably meets the system’s current operational demands. Our methodology enables the direct encoding of decades of maintenance expertise into intuitive, transparent operational boundaries. Through a principle of aggressive bottom-up multiplication, any deviation from these expert-defined boundaries at the lowest level immediately impacts the overall system health score. This deliberate choice prevents the masking effects common in weighted aggregation schemes [3], where a single critical failure can be statistically obscured by a majority of healthy components. By ensuring that one bad item "pulls down" the entire branch, we allow the operator to view the tree in a folded state without fear of missing a hidden fault, thereby reducing cognitive overload and allowing for focused troubleshooting only when necessary.

For a more detailed evaluation of the machine condition, the system can also provide further data on individual component wear. This is important for maintenance planning and will be further developed and described in future publications.

This work presents the design, implementation, and results of this framework on a 1970s-era rolling mill hydraulic system. We demonstrate that with the strategic addition of only a few sensors, a comprehensive, scalable, and trustworthy condition monitoring system can be built. The primary contributions are:

- A methodology for capturing and digitalizing expert knowledge without programming.
- A hierarchical aggregation model based on multiplication that ensures transparency and prevents failure masking.
- A clear distinction between component wear and production health, tailored for brownfield systems.
- A practical, module-by-module implementation strategy that minimizes risk and investment.

The following sections will detail the state of the art, present the methodology, describe its implementation and results, and discuss the implications of this approach for the sustainable operation of aging industrial assets.

2. STATE OF THE ART AND RELATED WORK

The field of condition monitoring is vast, but its application to brownfield hydraulic systems require navigating a specific set of challenges. Following established taxonomies, monitoring approaches can be broadly categorized as model-based, signal-based (including

data-driven methods), and knowledge-based. The existing literature can be reviewed through this lens to highlight the specific gap our work addresses.

2.1. Data-Driven and Model-Based Approaches

Condition monitoring approaches are often categorized as either model-based, requiring deep system knowledge, or data-driven, relying on historical process data [4]. Data-driven condition monitoring, particularly using machine learning, represents the most active area of research. The typical workflow involves detecting anomalies, diagnosing their location, and predicting the severity of the damage [5]. However, the success of these methods is critically dependent on the availability of large, comprehensive, and accurately labelled datasets. This prerequisite is a major obstacle in the context of brownfield industrial plants, where historical fault data is sparse or non-existent [6]. Researchers have explicitly acknowledged this "lack of data problem" as a fundamental hurdle, showing that classifiers trained only on simple, single-site damage data struggle to reliably identify more complex, multiple-site fault scenarios [7]. Much of the research is conducted using public benchmark datasets generated on laboratory test rigs, such as the widely used hydraulic system dataset from Helwig et al. [8]. The reviewed study predominantly utilizes machine learning and signal processing techniques. Although effective for pattern recognition, these approaches lack a physical basis, rendering them highly sensitive to noisy or anomalous measurements that may not reflect actual failures or performance degradation.

While invaluable for comparing algorithms, these datasets cannot capture the environmental noise and unique failure modes of a specific industrial installation. More complex "Big Data" frameworks [9], while powerful, assume a level of data velocity and IT infrastructure that is simply not present in the target environment.

Model-based approaches, such as those using Unscented Kalman Filters, offer an alternative by comparing sensor readings to the output of a mathematical system model. They can achieve excellent results in detecting and isolating faults with minimal sensor sets in real industrial machines [10]. However, their primary drawback is the significant upfront effort required to develop an accurate physical model, which is prohibitively complex for an organically grown system with components from different manufacturers and decades of undocumented modifications. Our work positions itself as a hybrid, knowledge-based approach, using data but relying on expert-defined rules rather than statistical models.

2.2. Hierarchical Health Indices

The concept of aggregating multiple sensor inputs into a single, hierarchical health index is an established practice for managing complex assets. Lee et al. developed a comprehensive Equipment Health Index (EHI) framework for the semiconductor industry, which successfully enables real-time monitoring of thousands of machines. It uses a hierarchical equipment structure which relies on statistical process control charts at the leaf level and uses AHP for weighted aggregation. This introduces a layer of statistical complexity and necessitates expert consensus on weighting factors, a process our methodology deliberately avoids.

Similarly, Yuan et al. proposed a Fuzzy Analytic Hierarchy Process (FAHP) to evaluate the health of gas insulated switchgear in power plants [11]. This method also leverages a

hierarchical structure and incorporates expert knowledge to define the hierarchy and pairwise comparison matrices. While this validates the use of expert input, it again depends on a complex, weighted aggregation logic that can obscure the impact of a single critical failure and is less intuitive for maintenance personnel without a background in fuzzy logic.

2.3. The Research Gap and Our Contribution

While the literature provides powerful tools for component-specific monitoring and sophisticated frameworks for data-rich environments, a critical gap remains. There is no established methodology that unifies decentralized expert knowledge into a single, transparent system health indicator for brownfield hydraulic systems under severe resource constraints. Existing hierarchical methods rely on complex statistical or weighted aggregations that are ill-suited for production environments where a single failure can halt operations and where trust in the system is paramount.

Furthermore, many data-driven approaches implicitly shift a significant burden onto domain experts: the selection and engineering of relevant features. This is a time-consuming task that experts are often unwilling or unable to perform, creating a bottleneck. Moreover, statistical models are notoriously brittle; their performance degrades when the system's behaviour changes due to modifications or new operational patterns, a common occurrence in brownfield plants, leading to false alarms and a loss of trust.

This paper closes this gap by presenting a framework that is:

1. **Expert-Driven at its Core:** It directly translates operational experience into deterministic rules.
2. **Transparent by Design:** It uses multiplication, avoiding opaque weighting schemes.
3. **Resource-Efficient:** It is built for minimal sensor additions and leverages existing infrastructure.
4. **Pragmatic:** It is scalable and deployable module-by-module without disrupting production.

By doing so, we provide a reproducible template for monitoring the vast number of aging, critical systems that will not be replaced but must be made smarter.

3. METHODOLOGY: THE HEALTH TREE FRAMEWORK

The health tree framework is engineered to overcome the specific challenges of monitoring brownfield industrial systems. The developed methodology rests on three foundational pillars: a deterministic bottom-up aggregation logic that prevents failure masking, a novel interface for capturing expert knowledge without programming, and a hierarchical structure that mirrors the physical plant, eliminating abstract complexity, as shown in Figure 1.

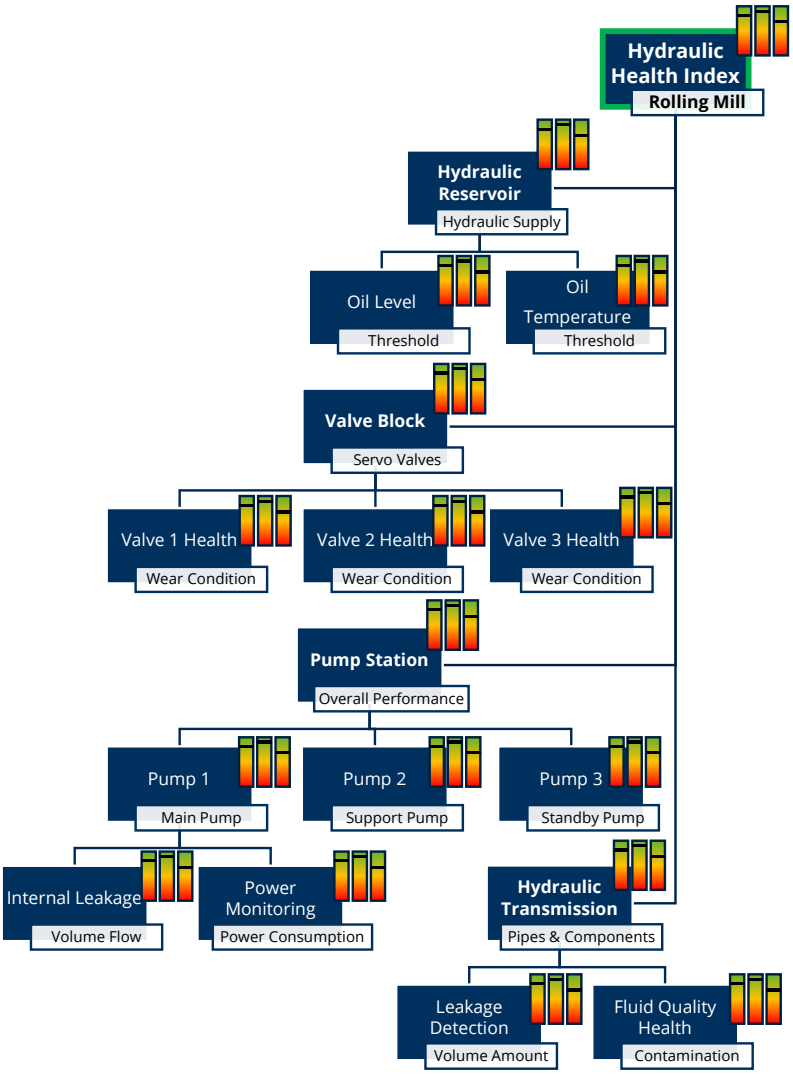


Figure 1 The health tree Framework for the Rolling Mill Hydraulic System (excerpt). This figure shows the root node (Overall System Health Index), branching into major subsystems (Pump Station, etc.), which in turn branch into individual component health indices

3.1. Hierarchical Structure Without Weighting Debates

The framework's main structure, shown in Figure 1, is not an abstract model but a direct digital representation of the physical plant architecture. This hierarchical tree is immediately intuitive to plant personnel who work with the physical system daily.

Level 1 (Leaves): Represent individual sensor inputs or calculated values, each with its expert-defined health function.

Level 2 (Branches): Represent physical components or subsystems, such as a complete pump station or cooling circuit.

Level 3 (Root): Represents the entire hydraulic installation. Its health, the System Health Index (SHI), is the product of all its underlying branches.

Crucially, this structure contains no weighting coefficients. This design choice deliberately eliminates the subjective, political, and often endless debates about whether the pump station is "more important" than the cooling system. The philosophy is absolute: if any single component's condition poses a risk to production, the overall system health must reflect that risk. This deterministic structure ensures that every SHI value is fully traceable. So, any Level 1 harm will be fully recognizable at Level 3 – the SHI.

3.2. Expert Knowledge Capture Without Programming

The core innovation that makes this framework practical is the method for defining the health functions f(s). The challenge was to create a digital representation of the nuanced, often unwritten rules that an expert with decades of experience uses to judge a system's health. To achieve this without requiring any programming skills, the solution was to replace traditional coding with a visual design paradigm. Instead of requiring engineers to write code, we developed an intuitive graphical user interface (GUI) where experts can directly "draw" their experience. The process is simple and powerful:

Graphical Threshold Definition: An expert sketches the acceptable operating region for a parameter directly onto a chart displaying its ranges, comparing it with live or historical data.

Real-Time Visualization: As the expert draws the curve, the resulting mathematical transformation and its effect on the health index are immediately visualized.

Decentralized and Autonomous Work: This tool empowers specialists to work independently. The pump expert defines pump leakage boundaries based on decades of observation, while the cooling systems expert defines temperature limits derived from past summer emergencies.

One-Click Code Generation: Once satisfied, a single click generates the complete, production-ready code for that health function.

This methodology mitigates the persistent challenges of compartmentalized expertise and complex system integration. It provides a structured process to transform valuable, undocumented operational knowledge into a scalable and digital asset, independent of the maintenance staff's programming capabilities.

3.3. Bottom-Up Health Index Through Multiplication

The framework employs a radical bottom-up construction with an multiplication logic at every aggregation step. The process is executed in three stages:

Leaf Node Assessment: At the lowest level of the tree, an individual measurement channel is transformed into a normalized Health Index (HI). All raw sensor data is collected and stored in an on-premise database. A leaf node does not always represent a raw sensor signal but rather a meaningful diagnostic parameter s, which is often the result of pre-processing:

> **Direct Sensor Values:** The parameter s is a direct, unprocessed signal, such as oil temperature.
>
> **Calculated Values:** Simple conversions, such as converting a tank height sensor reading (mm) into a tank volume (Liters).
>
> **Synthesized Parameters:** Complex metrics derived from multiple inputs, such as the heat exchanger efficiency calculated from temperatures and flow rates.
>
> **State-Dependent Metrics:** Values evaluated only when the system is in a specific state, such as assessing a valve characteristic during quasi-stationary periods.

The subsequent transformation of this parameter s into a normalized Health Index (HI) between 0 (complete failure) and 1 (perfect health) is performed by the expert-calibrated function $f(s)$.

Branch Node Aggregation: The health indices of all channels belonging to a single component or subsystem (a "branch") are aggregated into a single subsystem HI. Instead of averaging, we use the multiplication operator:

$$\boldsymbol{HI_{subsystem} = \Pi\,(HI_1, HI_2, \dots, HI_n)} \qquad \text{Equation 3-1}$$

This approach is deliberately chosen to compound the negative effect when multiple issues arise within the same subsystem. It ensures that the severity of the alert escalates when multiple concurrent faults are present, reflecting a more critical state. A pump unit with two moderately degraded parameters is considered less healthy than a unit with only one, a nuance that a min ( ) operator would miss.

While multiplication is the standard for assessing production readiness, it is important to note that alternative aggregations (e.g., weighted averages for specific sub-goals) are possible and can be configured in coordination with the plant operator.

System-Level Aggregation: Finally, the health indices of all subsystems are aggregated into the single, overall System Health Index (SHI), again through multiplication:

$$\boldsymbol{SHI_{system} = \Pi\, HI_{subsystem}} \qquad \text{Equation 3-2}$$

This consistent use of multiplication acts as a logical AND operator at every level. This approach is crucial for production assurance, ensuring any degradation at any level is immediately and unambiguously reflected in the final score.

4. IMPLEMENTATION STRATEGY

The successful modernization of an operational brownfield system hinges on the pragmatism of its implementation. Deploying a new monitoring framework in a live industrial environment, without interrupting production and on a constrained budget, requires a strategy that prioritizes immediate value, minimizes risk, and builds trust with the existing maintenance crews. Our approach was founded on two core principles: first, maximizing the diagnostic potential of existing infrastructure, and second, deploying the system in self-contained, value-driven modules. This stands in contrast to all-in-one solutions or model-based approaches which can require extensive upfront system modelling before any results are delivered [10].

4.1. Leveraging Existing Infrastructure

Our primary objective was to achieve maximum insight with minimal hardware intervention. This principle of using a small, targeted number of sensors is based on several strategic considerations:

- **Suitability for Brownfield Plants:** It allows for the maximum utilization of signals that are already available.
- **Rapid Deployment:** A minimally invasive approach enables installation and commissioning without production downtime.
- **Reduced Complexity:** Focusing on a few key indicators with clear diagnostic value is more effective than managing a deluge of data points.
- **Ensuring generalization:** The methodology is designed to be applicable across multiple plants and system configurations.

An audit of the plant's existing instrumentation revealed about 50 signals already installed for basic control but unutilized for condition monitoring. These included pressure transmitters, temperature sensors, cooling circuit on/off-signals, and position feedback sensors. While this provided a strong foundation, four new sensor types were added to close key diagnostic gaps:

1. Ultrasonic flow meters were installed on each pump leakage drain pipe. This enables a direct comparison of in-production pump performance against datasheet limits, a measurement typically only possible on a test rig [12].

2. Multiple temperature sensors were mounted on the main cooling unit to calculate the thermal power being extracted, providing direct insight into the thermal health of the system.

3. A particle counter was integrated into the main return line for continuous tracking of oil contamination, a validated method in industrial hydraulics [13].

4. A high-resolution oil level sensor was installed inside the reservoir to precisely track the system's total oil volume, making it a highly sensitive tool for detecting small, persistent external leaks.

This philosophy of using a small number of strategically placed sensors is supported by other research where system-level properties have been successfully modelled using common control signals [14]. The additional hardware investment for this upgrade was in the low

single-digit percentage range of the original plant investment, a stark contrast to the costs of production stops and emergency maintenance that can accumulate over a year.

4.2. Module-by-Module Deployment

To manage risk and foster adoption, the health tree was rolled out incrementally, module by module. The first module implemented was external leakage monitoring. Using data from the new high-resolution level sensor, it identified a small leak that had previously gone unnoticed, immediately proving the system's value and building trust. Following this, the cooling system health module was activated, using existing temperature sensors and a mass flow calibration from a non-persistent flow meter to track heat exchanger efficiency. Subsequently, modules for pump and valve health were deployed, providing continuous performance data.

A practical method for assessing pump condition involves comparing the measured drain flow with predefined baseline values derived from theoretical nominal flow calculations. The optimal conditions for monitoring the case drain flow are those under which the flow reaches its maximum, specifically:

- Dead head (nearly zero delivered flow rate)
- Constant nominal speed
- Constant nominal pressure

The case drain flow of axial-piston variable displacement pumps is a key indicator for evaluating internal leakage and overall pump efficiency. Monitoring this flow allows for early detection of wear-related degradation, particularly when manufacturer-provided leakage data is unavailable.

The servo valves play an important role within the gap and force control system. Three indicators have thus far been evaluated as health indices for each servo valve within the valve block and cylinders: null bias shift (NBS), signal deviation, and rated flow rate.

1. Null bias shift (NBS): This parameter reflects the tendency of a valve to drift from its neutral position as a result of variations in temperature, supply pressure, or internal leakage occurring in downstream components. In such situations, correction of the electrical current or adjustment of the mechanical null screw is required to re-establish the precise neutral point. Any deviation in the null position introduces errors in closed-loop systems, particularly in positional servomechanisms or pressure control applications, and is especially critical in zero-lapped valves.
2. Signal deviation: In electrohydraulic cylinders controlled by servo valves, control accuracy is determined by the dynamic relationship between the command signal (setpoint) and the feedback signal (measured response). Signal deviation is therefore defined as the difference between these two signals.
3. Rated flow rate: Assessing the actual flow characteristics is essential to compare the rated flow specified in the manufacturer's datasheet with the experimentally measured flow rate. Two degradation trends may be observed. An increase in the rated flow over time indicates spool wear, which enlarges the effective cross-sectional area of the spool ports. Conversely, a decrease in the rated flow signifies increased internal leakage in downstream components.

5. RESULTS FROM SEVEN MONTHS OPERATION

The ultimate measure of any industrial monitoring system is its tangible impact on plant availability, maintenance efficiency, and operational stability. Over a seven-month period of continuous operation, the health tree framework has transitioned from a pilot project to an indispensable tool for the maintenance team. Its success can be measured across four distinct areas: a direct comparison of system response, the technical detection of incipient failures, the cultural transformation of maintenance practices, and the resulting economic benefits.

5.1. Comparative Baseline and System Response

A comparison of system behaviour over two consecutive 7-month intervals provides the clearest evidence of the framework's impact. In the period before live monitoring, retrospective analysis of historical data shows that the System Health Index (SHI) dropped to as low as 8% during fault events. This extremely low score reflects the late discovery of issues where degradation was allowed to progress to a catastrophic state before manual detection.

In the subsequent 7 months with the active health tree, the SHI never fell below 70%. This stark difference is not due to a lack of faults, but rather the system's ability to trigger alarms early - typically when health drops from 100% to the 85-90% range. This enables proactive decisions long before a critical state is reached, effectively "flattening the curve" of degradation.

Notably, the number of fault-initiating events was of a similar magnitude in both periods; the crucial difference lies in the system's ability to make degradation visible in real-time.

Crucially, this multiplication logic does not lead to a permanently low system score (e.g., three 80% healthy pumps resulting in ~50% system health), as might be feared in a purely wear-based assessment. This is because our "Health Index" measures production readiness, not just physical wear. A pump with 80% remaining physical life can still be assigned a hight Health Index like 98% if its performance is fully sufficient to ensure production targets. This distinction allows the system to remain sensitive to critical deviations without generating constant, low-level alarms for functional but aging components.

For this comparison, the historical baseline was retrospectively calculated using the identical leaf functions $f(s)$ as the live system.

5.2. Detected Anomalies and Prevented Failures

The system’s ability to provide early warnings was validated on multiple occasions:

Leakage Prevention:

The early detection of an external leak, identified by the “Leakage Health” module prevented a potential emergency shutdown. More importantly, it averted product quality issues that arise from hydraulic oil contaminating the cooling water or emulsion—a subtle degradation that, without monitoring, is often discovered days later during quality control checks.

Heat Exchanger Performance:

As ambient temperatures rose, combined with maintenance work on the plant’s water management system, the “Cooling Health” index showed a progressive decline. This

visualized the decreasing efficiency of the main heat exchanger due to warmer inlet cooling water, allowing the team to proactively adjust production schedules.

Valve Health Verification:

The in-situ determination of valve characteristics via the "Valve Health" module showed a strong correlation with a manual valve measurement performed during a machine shutdown. This confirmed the system can support the definition of maintenance intervals and potentially justify component replacement without requiring external characteristic measurements.

5.3. Transformation of Maintenance Culture

Perhaps the most significant impact of the health tree framework transcends technical metrics. It has transformed the daily workflow and mindset of the maintenance crew.

Team members report a marked reduction in stress associated with unexpected failures. The system typically provides a 4- to 12-hour advance warning of developing issues, shifting the team from a reactive, firefighting mode to a proactive, planning mode. This newfound predictability allows for better coordination with production planning and provides maintenance managers with clear, data-backed justifications for necessary interventions.

One senior technician summarized the change perfectly: "For 20 years, I knew something was wrong from the sound, but I could never explain it to management. Now the health score system shows exactly what I hear, with numbers they understand."

The system has not replaced human expertise; it has amplified it.

5.4. Economic Impact

The economic benefits became apparent through a fundamental shift in operational efficiency:

The primary value is realized through avoided ad-hoc emergency deployments, the ability to conduct planned interventions, and the significant reduction of risks to both product quality and overall plant availability. While concrete monetary figures vary depending on the production context, the swift transition from reactive to proactive maintenance has generated a clear and positive financial impact.

As plant personnel universally acknowledge, the true value lies in the unquantifiable benefits of risk mitigation and enhanced stability for the entire production process.

6. 6. DISCUSSION

The health tree's effectiveness stems from a series of deliberate methodological choices that directly address the operational, cultural, and economic realities of aging factories.

6.1. Why Aggressive Multiplication Works

The framework's consistent use of multiplication for aggregation is a deliberate departure from conventional methods. It avoids the failure-masking flaw of weighted averages while also capturing the compounding risk of multiple concurrent faults—a nuance lost when only considering the single worst value (a min() operator). A focused on system optimization might tolerate averaging, but this fundamentally fails for production assurance.

Production in a rolling mill is often binary—either the product meets strict quality specifications, or it does not. The health tree's multiplication acts as a deterministic logical AND operator at every level, ensuring that any degradation threatening production readiness is immediately and unambiguously reflected in the final score.

While this multiplication deliberately maps production readiness, alternative aggregations (e.g., weighted averages for specific sub-goals) are possible and can be configured on a project-specific basis with the plant operator. However, transparent multiplication remains the standard for ensuring overall production security.

6.2. The Role of Localization: Awareness before Pinpointing

A key aspect of this resource-efficient strategy is the focus on awareness over precise localization. The primary goal is to make developing problems visible to the expert team early. Maintenance crews with years of plant-specific knowledge are highly effective at pinpointing the physical source of a fault once they are alerted to its existence. Providing a general but robust early warning, therefore, delivers the highest value by leveraging this existing human expertise. A more granular, sensor-intensive localization is technically feasible but economically complex and is reserved for future expansion stages.

6.3. Scalability Through Methodology, Not Standardization

The framework is designed to be scalable across different facilities, not through rigid standardization of thresholds, but through the replication of its core methodology. Each new plant implements the same proven approach:

1. **Map the physical plant architecture** to a logical tree structure.
2. **Audit available sensors and identify** critical diagnostic **gaps**.
3. **Implement a data pre-processing layer** to convert raw signals into meaningful diagnostic parameters.
4. **Empower local domain experts** to define operational boundaries independently for each leaf node's diagnostic parameter.
5. **Aggregate health indices** from the bottom up using the deterministic multiplication logic.
6. **Visualize** the results through intuitive, hierarchical dashboards.

This approach allows the specific health functions and thresholds to remain entirely plant-specific, respecting the autonomy and empowering the expertise that exists at each individual site.

6.4. Limitations and Future Development

The current implementation, while successful, also illuminates clear paths for future enhancement. The framework's modularity is one of its greatest strengths. One limitation is that the operational boundaries are static. Future versions could implement dynamic thresholds that adapt to the specifications of the product currently being manufactured. Furthermore, a key future development is the implementation of a prognostic layer at the leaf node level. This would involve extrapolating the observed degradation trend to provide

an estimated operational time window before a critical threshold is breached, analogous to a ‘range indicator’ in a vehicle.

Finally, the current focus is explicitly on production health rather than direct component wear. While this was a deliberate choice, the framework is fully compatible with integrating more advanced wear monitoring methods. For instance, methods for tracking pump performance degradation through case drain flow analysis [12] represent a prime candidate for a future “leaf” in the health tree, allowing for better prediction of the remaining useful life of key components.

7. CONCLUSION

This paper presented a novel health tree framework for the condition monitoring of brownfield industrial hydraulic systems. By combining a hierarchical structure, an aggressive multiplication-based aggregation logic, and an intuitive graphical interface for capturing expert knowledge, the system successfully bridges the gap between decades of hands-on experience and the need for digital oversight. The implementation in a live rolling mill demonstrated that a resource-efficient approach can yield a significant and rapid return on investment. The framework not only prevented unplanned downtime but also fostered a proactive maintenance culture. This work provides a practical and scalable methodology for extending the operational life and reliability of critical aging assets.

8. REFERENCES

[1] Stammberger R, ‘Method for monitoring the condition of a hydraulic system of a metal forming plant and condition-monitoring device’, U.S. Patent No. US11415156B2 2019.

[2] Isermann, R. Fault-Diagnosis Systems; Springer: Berlin/Heidelberg, Germany, 2006; ISBN 978-3-540-24112-6.

[3] Lee, C-Y., Dong, Z-H., ‘Hierarchical Equipment Health Index Framework’, 2019. https://doi.org/10.1109/TSM.2019.2925362

[4] Biggio, L., Kastanis, I., ‘Prognostics and Health Management of Industrial Assets: Current Progress and Road Ahead’, 2020. https://doi.org/10.3389/frai.2020.578613

[5] Worden, K., Manson, G., ‘The Application of Machine Learning to Structural Health Monitoring’, 2007. https://doi.org/10.1098/rsta.2006.1938

[6] Ahanga, M., Charter, T., Abbasi, M., Khadivi, M., Ogunfowora, O., Najjaran, H. ‘Intelligent Condition Monitoring of Industrial Plants: An Overview of Methodologies and Uncertainty Management Strategies’, 2025 https://doi.org/10.48550/arXiv.2401.10266

[7] Barthorpe, R. J., ‘On Model- and Data-based Approaches to Structural Health Monitoring’, PhD-Thesis, 2010.

[8] Helwig, N., Pignanelli, E., Schütze, A., ‘Condition Monitoring of a Complex Hydraulic System using Multivariate Statistics’, 2015. https://doi.org/10.1109/I2MTC.2015.7151267

[9] Kumar, A., Shankar, R., Thakur, L. S., ‘A big data driven sustainable manufacturing framework for condition-based maintenance prediction’, 2018. https://doi.org/10.1016/j.jocs.2017.06.006

[10] Razavi, B., de Silva, C.W., ‘Condition Monitoring in a Hydraulic System of an Industrial Machine Using Unscented Kalman Filter’, 2010. https://doi.org/10.1142/S0219878910002191

[11] Yuan, C., Sun, Y., et al., ‘Health Status Evaluation of GIS Equipment Based on FAHP’, 2022. https://doi.org/10.1109/ICPRE55555.2022.9960463

[12] Al-Issa, A., Schulze, T., Weber, J., ‘CONDITION MONITORING THE PERFORMANCE DEGRADATION OF AXIAL PISTON PUMPS BASED ON CASE DRAIN FLOW EVALUATION AND PRINCIPAL COMPONENT ANALYSIS’, 2025. https://doi.org/10.1115/DETC2025-169401

[13] Jeon, H-G., Kim, J-K., Na, S-J., Kim, M-S., Hong, S-H., ‘Application of Condition Monitoring for Hydraulic Oil Using Tuning Fork Sensor: A Study Case on Hydraulic System of Earth Moving Machinery’, 2022. https://doi.org/10.3390/ma15217657

[14] Al-Issa, A., Schulze, T., Weber, J., ‘Rapid Thermal Energy Modeling and Analysis of Complex Industrial Hydraulic Systems’, 2024. https://doi.org/10.13052/ijfp1439-9776.2614

Biographies



Reiner Stammberger received his diploma in hydraulic engineering from TU Dresden in 1989 and joined SMS subsequently. Having engineered countless hydraulic systems for various steel mill applications all over the world, he switched to the fluid service department in 2007. As senior project manager and engineer he is responsible for modernisations and development projects with regard to fluid systems. In this position he can draw from centuries of experience in the steel and hydraulic industry.



Arne Platzbecker received the B.Sc. and M.Sc. degrees in mechanical engineering from RWTH Aachen University, Germany in 2020 and 2022 respectively. In 2022 he joined SMS group in Mönchengladbach for a two-year trainee program with focus towards fluid engineering. In his current position he supports customers to improve their fluid systems in the aftermarket and is responsible for the hydraulic test benches from SMS group around the world..



Joachim Nagelmann has been working at SMS in the IIoT and SmartSensor Solutions division since 2022 and has many years of experience in industrial automation as well as industrial data processing. In his current research project, he is investigating hydraulic systems using modern IIoT technologies to advance data-driven industrial optimizations.



Maximilian Schmidt received the Diplom-Ingenieur degree in mechanical engineering from TU Dresden, Germany, in 2017, and worked at thyssenkrupp Carbon Components GmbH until 2022, developing resin transfer moulding production systems and digitalization solutions for automotive manufacturing. Since 2022, he has been a Research Associate at the Chair of Fluid-Mechatronic Systems (Fluidtronics) at TU Dresden, where his research focuses on developing physics-based condition monitoring frameworks for industrial hydraulic systems, bridging domain expertise with modern information technology for predictive maintenance applications.



Ahmad Al-Issa received the B.Sc. and M.Sc. degrees in mechanical engineering from AL-Mustansiriya University, Baghdad, Iraq in 2009 and 2012, respectively. This was followed by approximately 2 years as a university lecturer and then a 6-year industrial phase as a hydraulic engineer at General Company for Grain Processing, Ministry of Trade, Baghdad, Iraq. Currently, he is a research assistant and pursuing his PhD degree at Chair of Fluid-Mechatronic Systems, TU Dresden, Germany. His research focuses on thermal analysis and energy condition monitoring of complex industrial hydraulic systems.



Tobias Schulze is working as a research associate and pursuing his Ph.D. at the Chair of Fluid-Mechatronic Systems (Fluidtronics) at TU Dresden, Germany. He serves as the team leader for the Fluid-Mechatronic Systems Group, with his research focusing on stationary hydraulic systems, specifically hydraulic deep drawing presses and electrohydraulic compact drives.



Jürgen Weber has been appointed in 1st March 2010 as a University Professor and the Chair of Fluid-Mechatronic Systems as well as the Director of the Institute of Fluid Power at TU Dresden, and took on the leadership of Institute of Mechatronic Engineering in 2018. He finished his doctorate in 1991 and was an active Senior Engineer at the former Chair of Hydraulics and Pneumatics until 1997. This was followed by a 13-year industrial phase. Besides his occupation as the Head of the Department Hydraulics and Design Manager for Mobile and Tracked Excavators, starting in 2002, he took on responsibility for the hydraulics in construction machinery at CNH Worldwide. From 2006 onwards, he was the Global Head of Architecture for hydraulic drive and control systems, system integration and advance development CNH construction machinery. Furthermore, Jürgen has been head of the Consulting Board for HYDAC, Sulzbach/Saar, for 10 years, still being a member, and was also appointed as a member of the Supervisory Board of Musashi Europe GmbH. He is a fellow and now chair of the Global Fluid Power Society. The membership of 5G Lab Germany at TU Dresden is a further indicator for more than 15 years of experience in management and coordination of research alliances as well as the activities as surveyor, PhD supervisor, over 300 publications, technical books and lecture notes.

Energy-efficient handling through passive vacuum generation and biomimetic functional principles

Kuolt, Harald[1]*; Opolka, Anna Marie[2] (*https://orcid.org/0009-0000-2236-6140*); Kappel, Peter[3] (*https://orcid.org/0000-0001-6973-8618*); Dröder, Klaus[2] (*https://orcid.org/0000-0002-6424-4384*); Speck, Thomas[4] (*https://orcid.org/0000-0002-6357-1180*); Poppinga, Simon[3] (*https://orcid.org/0000-0001-5341-9188*)

[1]*J. Schmalz GmbH, Johannes-Schmalz-Str. 1, 72293 Glatten, Germany*

[2]*Technische Universität Braunschweig, Institute of Machine Tools and Production Technology (IWF), Langer Kamp 19b, 38106 Braunschweig, Germany*

[3]*Technical University of Darmstadt, Botanical Garden, Schnittspahnstraße 2, 64287 Darmstadt, Germany*

[4]*University of Freiburg, Plant Biomechanics Group, Botanic Garden, Schänzlestr. 1, 79104 Freiburg im Breisgau, Germany*

* Corresponding author: Tel.: +49 7443 2403-462; E-mail: harald.kuolt@schmalz.de

Harald Kuolt, Anna Marie Opolka and Peter Kappel contributed equally.

Abstract.

Automated handling is an established process within the context of industrial and automated production, usually carried out by robotic systems with gripper systems. Conventional suction grippers rely on compressed-air-based vacuum generation – with a significant associated energy consumption and required infrastructure. A possible alternative is the usage of passive vacuum-based handling technologies, which offer the opportunity to entirely avoid the previously necessary expensive infrastructure. However, within the current state of development, passive systems inherently lack pressure regulation capabilities, which introduces the risk of vacuum leakage and therefore a reduced gripping reliability. The present research therefore aims to develop a new type of energy-efficient, reliable and location-independent handling technology based on passive vacuum generation and the development of a framework for identifying and avoiding impending sealing failures. The approach is biologically inspired and designed to operate entirely without the active supply of external energy for vacuum generation. On the basis of these developments, opportunities for significant potential savings in terms of energy use, CO_2 emissions and operating costs could be enabled.

Keywords. Vacuum handling, energy efficiency, biomimetics, Mediterranean medicinal leech, *Hirudo verbana*.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Vacuum handling technologies are widely used in industrial automation due to gentle handling, high cycle rates and flexibility across object geometries [1]. As deployments scales and task diversity increases, the efficiency and infrastructural footprint of vacuum generation become critical constraints, especially in mobile and decentralised settings. The rapid growth of e-commerce and warehouse automation has significantly increased the demand for fast, flexible handling systems. Automated storage and retrieval systems and autonomous picking stations frequently rely on vacuum gripping to manage packaging materials, cartons and irregularly shaped goods [2, 3].

In conventional industrial automation, the pressure difference is commonly generated pneumatically via compressed air due to the inherent simplicity and low component costs. On mobile and decentralised platforms, however, compressed air becomes impractical: the energy density of compressed air is low, on-board air tanks are not viable for extended duty cycles, even with periodic recharging. The use of hoses to connect to the compressed air supply constrains the workspace and safety envelope. Electrical vacuum generators such as pumps or blowers require continuous electrical power and increase system complexity and noise [1, 4]. These constraints motivate the passive generation of pressure differences with the goal of creating and maintaining the pressure difference during the handling cycle without an external power supply. However, passive systems lack active pressure regulation. As a result, leakage events can erode the holding force and compromise process reliability [5, 6]. For market acceptance, passive vacuum handling must not only generate the pressure differences efficiently but also ensure high process safety. This requires minimising leakage through an innovative gripper design as well as in situ detection and avoidance of leakage events.

This paper addresses this challenge via three complementary strategies, which are developed in the following chapters. Section 2 surveys passive generation of the pressure difference and safety measures for vacuum maintenance. Biomimetic compensation concepts for enhancing the sealing and compliance are presented in section 3. Section 4 develops a process-safety framework for the detection of imminent leakage-induced failures. Section 5 concludes with an overview of the addressed challenges, future potentials and an outlook.

2. ACTIVE AND PASSIVE SUCTION GRIPPERS

2.1. General Operation Principles for Suction grasping

The operating principle of suction grippers is based on a pressure difference between the ambient pressure and the internal volume of the suction cup, which is positioned close to the surface of the handled object. When the gripper and the object are pressed against each other, the suction cup is sealed against the ambient pressure, enabling a frictional connection and the creation of a pressure difference [7]. Since handling is conventionally carried out under ambient conditions, a negative pressure must be generated inside the suction cup [8, 9].

The achievable holding force F for a given task can be defined by [8]

$$F=\Delta p \cdot A_{eff} \cdot z \cdot n \cdot \eta \cdot \frac{1}{S} \qquad (2.1)$$

where $\Delta p=p_{amb}-p_i$ describes the difference from the environmental and the internal pressure, A_{eff} the effective sealed area of the gripper and z the number of grippers. The deformation coefficient n, the system efficiency rate η and the safety factor S can be used to take gripper characteristic and process-induced uncertainties into consideration. In principle, this pressure difference can be generated using two different principles, which are described as active and passive principles in the following [10].

2.2. Active Generation of Pressure Differences

The industrially established principle as the usage of "active" grippers. Here, the pressure difference is created (active suction cup [11]) by either an electric or pneumatic vacuum generator [10, 12]. These vacuum generators reduce the gas mass within the suction cup's internal volume, thereby decreasing the pressure within the suction cup. The internal volume of the suction cup can be reduced during the process of evacuation, depending on the design of the suction cup. In the case of the suction cup being designed as a suction plate, the internal volume remains constant. This indicates that the internal volume of the suction cup can be described as $V_1 \geq V_2$ and the absolute pressure as $p_1>p_2$ [7].

2.3. Passive Generation of Pressure Differences

Another possibility is the generation of the pressure difference within a closed system, thereby ensuring a constant mass within the suction cup (passive suction cup [11], see Figure 2.1 (left)). In the "passive" applications, the pressure difference is established through the application of a force resulting in an increase in the internal volume. With this applied force and the deformation of the gripper, the internal volume of the suction cup V is enlarged and the condition $V_1<V_2$ applies. According to the ideal gas law, this results in a decrease in the absolute pressure p within a closed, isothermal system. This pressure difference generation process, which merely results from the deformation of the gripper, is conventionally described as passive and therefore cannot compensate for pressure increases in the suction cup caused by inadequate sealing. "Passive" specifically implies that the required force F is applied by the handling device, thereby eliminating the need for an independently powered vacuum system

.

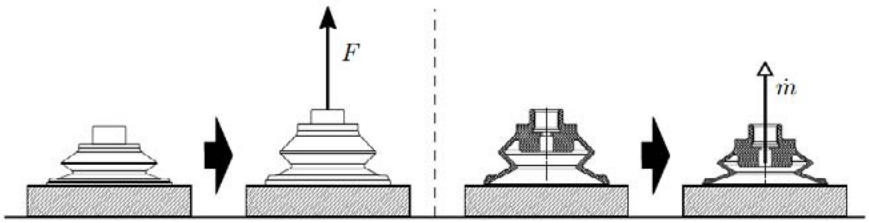


Figure 2.1. Passive vacuum suction cup – closed system (left)
and active suction cup – open system (right) [7]

In contrast to active generation of pressure difference, which relies on a continuous energy supply from pumps or compressors, passive generation of pressure difference relies exclusively on the volume change within a previously defined volume as a result of a discrete external force. For instance, the design of a passive suction gripper may take form of a clamping pivot and a slide damper (Figure 2.2). During operation, a discrete external load presses the cup onto the surface, reducing the chamber volume and expelling air. The elastic lip seals the interface, thereby creating a pressure differential that results in the cup adhering to the surface. The release process is actuated by opening the slide damper, thereby allowing the chamber to vent [8]. However, leakage cannot be actively compensated, so the holding force decays if the seal is disturbed. Consequently, the energy required to generate the vacuum is significantly lower than for an actively operated vacuum generator.

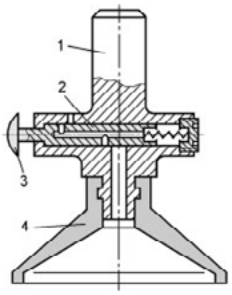


Figure 2.2. Principle of an exemplary passive vacuum design [8]
(1) clamping pivot and carrier, (2) manually operated slide damper,
(3) service button, (4) disk sucker

2.4. Use-Case: E-Commerce and Resulting Challenges

To utilise this high degree of energy efficiency, the systems have to be qualified for industrial usage. Especially the required reliability of the grasping forces is currently not achievable for all surfaces. In e-commerce fulfilment, heterogeneous items, uneven or film-covered surfaces, contamination and short cycle times increase the risk of leakage-induced detachment. Therefore, three practical challenges are focused:

(i) System-level adaptation of the pressure difference,

(ii) Suction-cup design for sealing robustness and

(iii) Leakage-aware handling strategies.

Currently, passive systems struggle with very short cycles and the majority of approaches are only capable of functioning in a leak-free manner when operating within an ideal environment. In practice, the likelihood of a small leakage in the seal of the passive suction cups during handling cannot be ruled out, raising the risk of a loss of the handled component. Existing systems have not yet been designed to compensate for this impending leakage and there is no established process setup for a proactive reduction of handling parameters such as speed or acceleration, nor a move to a safe position. Other promising approaches involve a flexible membrane that can easily adapt to the surface of a handled, which can generate sufficient pressure differences for a handling process through a small stroke to increase the volume beneath the membrane [13]. Another focus lays on concepts that minimise leakage. Soft sealing lips that adapt optimally to the respective workpieces are of key importance for this [6]. Initial concepts have been developed and implemented to combine extremely soft sealing lips with harder suction cups [1].

In the following chapters, the different possibilities (i) - (iii) are investigated.

3. POSSIBLE SAFETY SYSTEMS FOR ACHIEVING RELIABLE HANDLING WITH PASSIVE GRIPPERS

In order to achieve the highest possible process reliability when using passive vacuum gripping systems, there are generally two possible development goals: First, the development of suction grippers with optimal sealing properties should be considered. These grippers should be designed to achieve optimal sealing with the handled components, thereby allowing the lowest possible leakage airflow into the evacuated suction gripper. If a leakage airflow eliminates the created pressure difference, the gripping force is also removed, resulting in the handled object being dropped. Secondly, an additional safety system can be employed to prevent the component from falling off in the event of high leakage, thus ensuring the continuity of the current running process without damaging the component.

The following approaches are therefore being further considered as safety systems:

- Redundant vacuum generator
  - Active, e.g. using a small electric pump
  - Passive, e.g. using a second passive suction cup or a matrix of suction cups that can compensate for the failure of a suction cup
- Additional storage volume
  - Rigid, e.g. an additional pneumatic cylinder as storage
  - Flexible, e.g. a bellows
- Additional safety element
  - Mechanical, e.g. a clamp
  - Electric field
  - Magnetic (if the workpiece is ferromagnetic)
  - Adhesive

Within this approach, this could only be achieved via a redundant pump that intervenes in an emergency. However, this would then no longer be a purely passive solution. Two implemented approaches (Figure 3.1 and Figure 3.2) can therefore only represent the starting point for targeted further developments for a leakage-robust passive vacuum gripper design.

The redundant system in Figure 3.1 consists of a high flow suction cup and an adhesive tape. For a passive holding solution, the adhesive tape mounted on the gripper will grasp the workpiece. If the holding force is not high enough or if a failure occurs, the vacuum gripper can be activated to hold the workpiece like in a standard vacuum handling application.

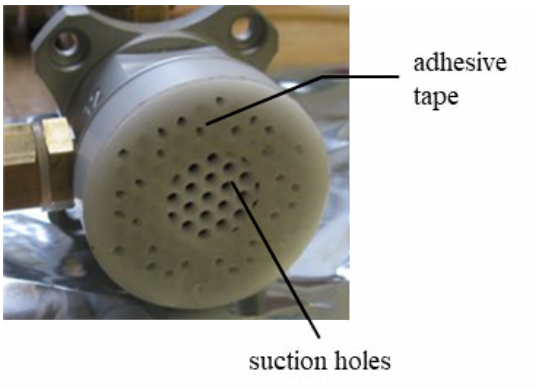


Figure 3.1. Redundant system with an additional adhesive tape that can hold the component in case of vacuum system failure

A first functional prototype shown in Figure 3.2 was equipped with a membrane, a pressure mechanism, a venting valve, a vacuum sensor, an additional pump (top of Figure 3.2) and a mechanical connection to common robot flanges. If leakage occurs during passive handling, this can be recognized by the vacuum sensor and the additional pump can be activated. Alternatively, there could be a solution with an additional vacuum realized by a pre-loaded air cylinder or a bellow.



Figure 3.2. Redundant system with an additional pump (top), a system with a compressed air cylinder as a pre-loaded accumulator (bottom left) or with a flexible bellows (bottom right)

4. LEAKAGE COMPENSATION BASED ON BIOMIMETIC PRINCIPLES

In biology, multiple reversible attachment mechanisms based on suction have evolved, providing engineering with a wealth of role models for vacuum-based, flow-free, energy-efficient suction devices. Most organisms utilising suction are aquatic, especially leeches are of particular interest for biomimetic approaches due to their semi-aquatic nature and ability to attach underwater as well as on land. The hematophagous Mediterranean medicinal leech (*Hirudo verbena*) evolved a remarkable mechanism to hold on the body surface of its hosts (mammals and fish), which could inspire the solution to a classical engineering problem: how to attach securely and reversibly to a wide range of substrates without the assistance of an external vacuum source. The leech accomplishes this through two muscular suction cups located at the anterior and posterior ends of its body [14], (Figure 4.1). These two suckers enable the leech to attach and move on surfaces underwater and on land that are smooth, rough, or even porous [15, 16].



Figure 4.1. Mediterranean medicinal leech *Hirudo verbana* with its anterior (A) and posterior (B) suckers. White arrow: Grasping-like deformation of the posterior sucker.

The leech alternately attaches and detaches both suckers for locomotion (inchworm-like gait or vermiform crawling), while the anterior sucker additionally provides a stable fixation point for the jaws during feeding, i.e. blood intake [14] as seen in Figure 4.1A. For attachment, both suckers employ the same physical principle – generation of negative pressure by musculoelastic sucker deformation – yet differ in their functional design and redundancy. The posterior suction disc is exclusively used for locomotion and attachment, while the anterior sucker fulfils multiple roles in locomotion and feeding. The procedure of surface contact formation and attachment from the inside out of the suction discs is equal in both types. This process dramatically reduces the dead volume and increases the contact area with the object's surface when compared to the classical suction cup designs commonly used in industrial applications. Even though the “normal” suction performance of both suckers is already impressive, the anterior sucker can additionally actively adjust the pressure difference between the sucker chamber and the surrounding medium (water or air) via the mouth orifice in the case of an approaching leakage. Thereby, the anterior sucker reaches slightly higher attachment force maxima than the larger posterior sucker in pull-off tests, while the posterior sucker sets the pressure difference and needs to adjust to such events solely through sucker deformation and cavity sealing. [15] The latter is demonstrated in Figure 4.1B in which the posterior sucker is attached to a flexible surface and pulled, leading to a particular pronounced, almost grasping-like deformation of the sucker to generate an increased pressure difference and counter an imminent leakage. The reliable performance of both suckers demonstrates that strong, flow-free suction can be achieved without an active pump, making the leech an ideal biomimetic role model for passive, vacuum-based suction grippers.

4.1. Functional Basis of the Leech Suction System

The anterior sucker of *H. verbana* is formed by its first four body segments and the posterior sucker by its seven terminal body segments. Both suckers contain strong musculature organised in circular, radial and other orientations, as well as connective and glandular tissue. When the leech attaches with the posterior sucker, the sucker flattens and spreads over the surface, thereby increasing its contact area. The rim forms an airtight seal, while the sucker musculature enlarges the cavity volume, reducing the internal pressure. No valves or ducts are involved: all pressure differentials arise from the mechanical interplay of elastic deformation and sealing. This mechanism is inherently reversible; releasing the sucker simply involves local muscular contraction at the rim to admit air or water flows into the cavity. [15]

A defining feature of both *H. verbena* suckers is their distributed actuation: most of the sucker contact surface is controlled by fine musculature and capable of fine deformation. This allows continuous adjustment of local contact pressure and shape, providing adaptive and consistent surface contact even on rough, uneven or porous substrates. Although capillary bridging is necessary for maximum attachment forces, the attachment remains suction-dominated across many substrate conditions (even those characterised by porosity) as the leech is capable of closing surfaces cavities and holes with its highly deformable suckers. [15] The ability to maintain suction on non-airtight substrates illustrates the robustness of the system and highlights the leech's highly adaptable attachment strategy. When confronted with objects that do not offer sufficient surface area for suction (e.g. hairy surfaces or mesh structures), the leech's ability to locally control and deform its suckers provides secondary attachment mechanisms, such as clamping, spacing and locking at object edges or into object spaces [15]. The muscles and connective tissue in the suckers that cause these unique deformation and attachment properties are anatomically derived from the tissue layers in the leech's basic body segments. The way in which the tissues within the suckers are anatomically and functionally organised has not yet been investigated in detail. However, a complete understanding of the morphology and anatomy is a prerequisite to encoding the functional principle and abstract them for engineering.

4.2. Design Implications for Passive Suction Grippers

Both suckers, but especially the posterior one demonstrates the possibility of a stable suction through shape change and sealing alone. From this follow three critical engineering requirements for passive suction cups: 1) negative pressure generation by structural deformation, 2) minimisation of leakage and 3) delay of peel or detachment events. These are realised biologically through geometry, compliance and muscular control.

The first important aspect is negative pressure generation by structural deformation. Instead of state-of-the-art suction cups, that get deformed and attached to a surface when their dead volume is evacuated by an applied vacuum (Figure 2.1), a biomimetic passive sucker should be capable of deforming itself to generate a negative pressure difference, preceded by an initial surface contact that is as large as possible, with minimal dead volume.

The second aspect is leakage control: The rim of the leech sucker is highly compliant and structured with radial furrows. These grooves act as fluid reservoirs and distribute mucus

uniformly across the contact area, forming a tortuous path for air leakage. Since an industrial requirement are residual-free handling devices, a biomimetic optimised suction gripper cannot rely on applied liquids for maximal sealing and attachment forces. Possible alternatives could be the use of materials with different stiffness and micro-patterned rims to increase the contact area and to provide a comparable sealing function without residues.

The third aspect is detachment control: In mechanical terms, failure of suction typically initiates as a peel at the sucker edge. The leech mitigates this for the posterior suction disc by a joint-like neck constriction connecting sucker and body, which allows pitch–roll flexibility and maintains uniform compression along the rim. Additionally, the secondary attachment mechanisms could inspire suckers with mechanisms that mechanically counteract the detachment of the sealing lip, prevent it altogether, or even lead to automatic re-gripping or jamming of the formerly attached object in the event of suction failure. This could be achieved through structural and material-immanent functions in the suction cup.

In sum, *H. verbana* provides a rigorous template for flow-free, purely passive suction and leakage control. The posterior sucker demonstrates that a pump-less vacuum can be generated and stabilised through geometry and sealing alone. In contrast, the anterior sucker highlights the importance of a pump as an additional fail-safe mechanism where feasible. With these elements, passive suction grippers can achieve the leech's hallmark combination of adaptability, robustness and passive attachment.

5. INCREASING PROCESS SAFETY

Passive vacuum handling eliminates the need for compressed-air infrastructure; however, it removes the ability to actively re-establish or regulate vacuum during operation. As a result, leakage caused by surface topography or off-axis loading can lead to a seal deformation and weakening and result in the detachment of the handled object. The object loss following the sealing detachment can occur within milliseconds. The early detection of the sealing deformation is therefore crucial for the purpose of either adapting the robot motion or ensuring the safe placement of the object. Therefore, the development of an integrated approach represents a potential solution for enabling passive vacuum handling technologies in industrial automation in order to establish a sufficient process safety.

Process safety in this context can be understood as the probability of completing a pick-and-place operation cycle without an occurring object loss. To prevent leakage-induced object loss, the critical moment of the imminent object detachment at the gripper-object interface has to be identified and analysed. Leakage-induces object loss can occur from a variety of sources. These can be grouped into following [3]:

(i) Influences from the gripper characteristics, e.g. seal degradation

(ii) Influences from the object properties such as surface roughness, load, mass eccentricity or surface contamination

(iii) Process-induced influences, e.g. shear load spikes during acceleration or jerk as well as off-axis loading during the pick process and transport phase.

In order to react to an imminent object loss, it is necessary for the detachment to be detected in advance. This results in the need of a sufficient holding force level to be able to ensure the object is being attached to the gripper for the required time between the detection and potential reaction strategy.

5.1. Empirical Characterisation and Derived Optimisation Potential

In the following, the behaviour of the sealing detachment and resulting leakage is investigated. Especially the sealing itself is focussed, therefore the influences from the gripper (i) and object characteristics (iii) are reduced as much as possible. For this purpose, systematic static pull-off experiments were performed with a 7-axis robot on a smooth, non-porous surface using a standard bellows suction cup (FG 62 NBR-55 N018 from Schmalz) functioning as a passive gripper to initially derive the main process-induced effects (iii). For these experiments, an in-cup pressure sensor was integrated into the suction cup setup to obtain the internal suction cup pressure $p_i(t)$. Furthermore, a 6-DoF force/torque transducer located at the robot flange captured the occurring load, while an ultrasonic distance sensor measured the end-effector extension along the surface normal (Figure 5.1).

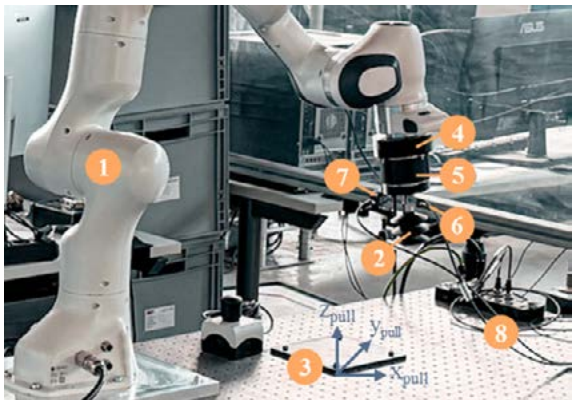


Figure 5.1. Experimental setup
(1) robot arm, (2) bellow suction cup functioning as a passive gripper,
(3) PLA plate, (4) flange adapter, (5) force transducer, (6) vacuum sensor,
(7) ultrasonic sensor, (8) IO-link master

Within the experimental investigation, the approach depth, approach angle, process speed and acceleration for linear motion were varied by the factorial experiments. The process depth is defined as the distance from the ultrasonic sensor to the pull-off surface with a maximum possible compression d_{max} with approx. -20 mm for the depth value compared to the initial undeformed state of the suction cup. The approach depth influences the deformation and sealing and the loads to the sealing lip as well as the handling object. The velocity and acceleration were varied as a percentage of the maximum robot speed and acceleration to investigate the influences on the sealing and release behaviour through occurring dynamic effects, e.g. viscoelastic effects. The variation of the rotational angle

between the gripper verticals and the surface normal was varied within the range of $w \in \{-20°, 0°, 20°\}$ to examine its impact on the sealing behaviour and force distribution.

The investigated response variables were the measured internal pressure difference, the pressing force applied to the sealing to the surface as well as the pull-off force. The experiments were performed with 20 process variable combinations, each with $n=50$ trials (see Table 5.1). For each parameter combination, the maximum level of pressure difference and the minimum and maximum vertical force values were measured and averaged across the repetitions. To unify the sampling rate, the lower-rate vacuum data was anti-alias filtered and resampled to 150 Hz to match the force sampling rate. Due to the data acquisition of the vacuum and ultrasonic sensor via IO-Link-master and TCP/IP, a delay in the data stream compared to the USB-based force/torque data may occur. To avoid this distortion in the temporal overlap of the various data sources, an initial pull-off movement was performed before each trial to align the data streams by an offset correction based on the first prominent peak during the start-up motion, compression and pressure establishment.

The initialisation phase is characterised by a contactless interaction between the suction cup and the pull-off surface. The robot approached the surface orthogonally without any load besides the end effector setup consisting of gripper as well as cables and without cup deformation or pressure change ($F_Z=p_i(t)=0$) (Figure 5.2 (a)). The behaviour of the vacuum gripper during the pull-off experiments could then be segmented into three phases (Figure 5.2 (b)-(d)):

- Sealing phase,
- Leakage onset,
- Contact rupture.

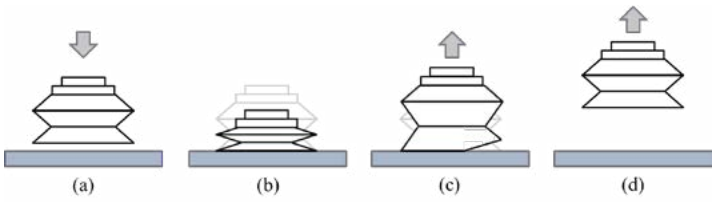


Figure 5.2. Phases of the pull-off experiments
(a) contactless phase, (b) sealing phase, (c) leakage onset due to local occurring sealing detachments, (d) contact rupture leading to complete sealing detachment

During the approach to the defined approach depth, the normal force -F_Z occurs, leading to a sealing lip deformation, thereby sealing the gripper to the pull-off surface. When a positive pull force is applied, a pressure difference in the gripper leads to an adhesion of the seal to the surface until the collapse of the holding force $+F_Z$ at contact break. After release, the gripper and kinematic structure exhibit a short elastic ring-down (Figure 5.3). These phase markers provide consistent anchors for feature computation.

Table 5.1. Section of the performed factorial pull-off experiments

No.	v_{robot}	a_{robot}	w [°]	Δp_{max_avg} [mbar]	$F_{Z_min_avg}$ [N]	$F_{Z_max_avg}$ [N]	$\sigma(p_{max_avg})$ [mbar]	$\sigma(F_{Z_max_avg})$ [N]
1	50%	50%	0	451.72	-39.65	57.64	16.91	4.52
2	100%	50%	0	475.21	-38.17	59.42	22.17	4.27
3	50%	100%	0	477.18	-39.37	56.13	24.8	7.64
4	100%	100%	0	495.65	-39.61	56.71	19.94	9.41
5	50%	100%	20	432.93	-37.58	66.92	9.41	4.1
6	100%	50%	20	454.20	-43.82	72.24	1.78	1.58
7	100%	100%	20	440.81	-44.31	69.58	3.12	1.14
8	50%	50%	-20	393.20	-57.69	63.51	14.15	2.57
9	100%	100%	-20	375.98	-58.35	58.46	24.82	5.07
10	50%	100%	-20	391.83	-58.79	64.06	11.01	1.93

Over the 50 repetitions for each parameter set, the standard deviation remained comparatively small in relation to their respective means, thereby signifying a high degree of repeatability. Varying the initial movement as well as the pull-off speed and acceleration showed negligible effects in the tested static conditions due to short distances and therefore short acceleration distances, though dynamic trajectories are expected to introduce viscoelastic and inertial effects.

The impact of a varied approach angle was demonstrated and indicated, that non-orthogonal approaches reduced maximum pressure difference by 15-25 % due to asymmetric sealing. An orthogonal approach has been demonstrated to achieve the highest average pressure difference values, while average pull-off forces F_{Z,max_avg} show a slightly higher level at w=20°. This discrepancy can be attributed to the fact that the achievable holding force is also dependent by an angle-induced pre-compression, which is described with the deformation coefficient n. A tilted approach has been shown to increase local lip compression and bending stiffness. It has been observed that the strongest average contact force occurs at a negative approach angle, indicating a higher pre-load which influences the initial sealing behaviour while achieving a lower pressure difference. Therefore, the selection of approach angle should be dependent on the specific characteristics of the surface and the task. Research in future will focus on the quantification of the trade-off between angle-induced precompression and maximum cavity pressure difference with the aim of deriving task-adaptive approach strategies.

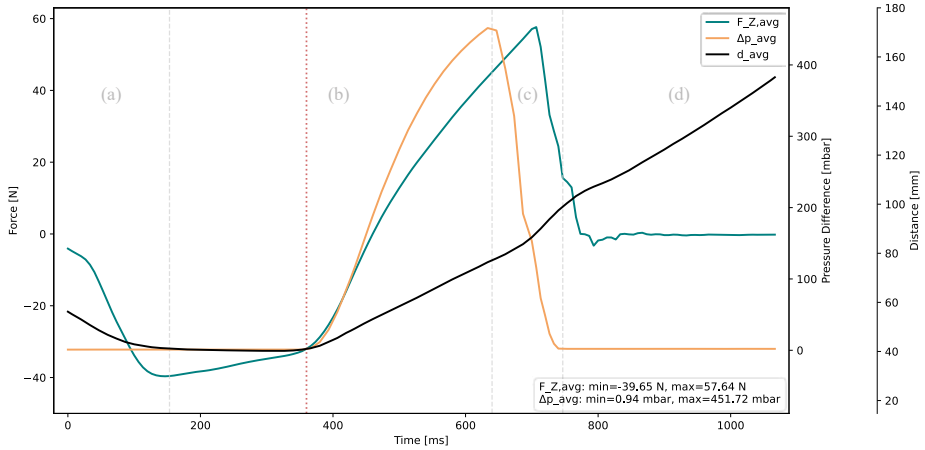


Figure 5.3. Plot of average force in Z-direction, pressure difference and distance values divided into the phases of the pull-off experiment (grey dashed lines) pull-off experiment no. 1, pull-off at t_{pull}=360ms (red dotted line)

6. CONCLUSION AND OUTLOOK

The overall goal of the presented publication was the investigation of energy efficient handling systems with a focus on passive suction grasping. However, even though the usage of a passive suction cup is a very energy efficient solution to decrease energy consumption, process safety topics such as system failure and the possible loss of the workpiece must be considered. As a result, the three different aspects of achieving reliable grasping with passive grippers were investigated, derived from a use-case of a usage in short-cycle e-commerce handling tasks. Next to systemic safety concepts, biomimetic-inspired suction designs were focussed. Additionally, redundant safety strategies and mechanisms were investigated to enable a wide range of possible applications in automated processes.

Overall, the findings highlighted the possibilities and possible further research avenues for passive suction grippers to achieve industrially acceptable reliability. As a next step, the highlighted research avenues should be analysed further and combined into prototypical functional demonstrators, to empirically investigate possible process limitations. Additionally, modelling and simulation approaches could significantly accelerate development processes for passive vacuum handling systems and should be researched alongside physical demonstrators. Overall, the passive handling technology can be described as an emerging technology with a high potential for various use-cases. If a high degree of reliability can be achieved, significantly enhanced efficiencies for automated handling systems are opened up for various industries. Additionally, elements of these approaches could also be transferred to conventional vacuum handling technologies, opening up further potential for increased energy efficiency.

ACKNOWLEDGEMENTS

The authors acknowledge funding by the German Federal Ministry for Economic Affairs and Energy (BMWE) under the project number: 03EN4072 (“BiPas - Passive vacuum-based handling via biomimetic functional principles”).

7. REFERENCES

[1] Fiedler, M. and Schaaf, W., Novel Engineering and Product Solutions towards Digitalization and Sustainability in Vacuum Handling Automation in 14th International Fluid Power Conference, Dresden, Germany, pp. 888–900, 2024.

[2] Boysen, N. and Koster, R. de, 50 years of warehousing research—An operations research perspective, European Journal of Operational Research, vol. 320, no. 3, pp. 449–464, 2025.

[3] Gabriel, F., Fahning, M., Meiners, J., Dietrich, F., and Dröder, K., Modeling of vacuum grippers for the design of energy efficient vacuum-based handling processes, Prod. Eng. Res. Devel., vol. 14, 5-6, pp. 545–554, 2020.

[4] Kuolt, H., Vacuum generation for handling technology: mobile, autonomous and energy efficient, 2018.

[5] Kuolt, H., Moosavi, A., Tautenhahn, R., Weber, J., Gabriel, F., and Pierri, E. et al., Increase of energy efficiency in vacuum handling systems based on biomimetic principles in Volume 3 - Conference, pp. 17–26, Oct. 2020.

[6] Kuolt, H., Kampowski, T., Poppinga, S., Speck, T., Tautenhahn, R., and Moosavi, A. et al., Biomimetic Suction Cups for Energy-Efficient Industrial Applications in Zukunftstechnologien für den multifunktionalen Leichtbau, Technologies for economic and functional lightweight design, pp. 182–188, 2021.

[7] Straub, D., Methode zur technischen Auslegung von Vakuumgreifsystemen mit einer Mindesthaltedauer auf Basis fluidischer Untersuchungen. Dissertation: Stuttgart : Fraunhofer Verlag, 2020.

[8] Monkman, G. J., Hesse, S., Steinmann, R., and Schunk, H., Robot Grippers: Wiley, 2006.

[9] Seegräber, L., Greifsysteme für Montage: Handhabung und Industrieroboter ; Grundlagen - Erfahrungen - Einsatzbeispiele. Ehningen bei Böblingen: Expert-Verl., 1993.

[10] Tomokazu, T., Kikuchi, S., Suzuki, M., and Aoyagi, S., Vacuum gripper imitated octopus sucker-effect of liquid membrane for absorption- in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, pp. 2929–2936, 2015.

[11] Chen, W., Zhao, S., and Chow, S. L., Grippers and End-Effectors in Handbook of Manufacturing Engineering and Technology, pp. 2035–2070, 2015.

[12] Murrenhoff, H. and Reinertz, O., Fundamentals of Fluid Power: Part 2: Pneumatics, 2nd ed. Aachen: Shaker Verlag, 2016.

[13] Caruso, F., Shea, H., and Cacucciolo, V., Electroadhesion Suction Cups, Advanced materials (Deerfield Beach, Fla.), vol. 37, no. 27, e2420231, 2025.

[14] Kampowski, T., Eberhard, L., Gallenmüller, F., Speck, T., and Poppinga, S., Functional morphology of suction discs and attachment performance of the Mediterranean medicinal leech (*Hirudo verbana* Carena), Journal of the Royal Society, Interface, vol. 13, no. 117, 2016.

[15] Kampowski, T., Thiemann, L.-L., Kürner, L., Speck, T., and Poppinga, S., Exploring the attachment of the Mediterranean medicinal leech (*Hirudo verbana*) to porous substrates, Journal of the Royal Society, Interface, vol. 17, no. 168, p. 20200300, 2020.

[16] Kampowski, T., Schuler, B., Speck, T., and Poppinga, S., The effects of substrate porosity, mechanical substrate properties and loading conditions on the attachment performance of the Mediterranean medicinal leech (*Hirudo verbana)*, Journal of the Royal Society, Interface, vol. 19, no. 188, p. 20220068, 2022.

Development and Experimental Validation of a Novel Origami-Inspired Pneumatic Actuator Concept with Progressive Force-Stroke Characteristics

Philipp Hüpper, Olivier Reinertz, Katharina Schmitz

RWTH Aachen University, Institute for Fluid Power Drives and Systems, Aachen, Germany, philipp.huepper@ifas.rwth-aachen.de

Abstract.

Pneumatic actuators currently face limitations in achieving a progressive correlation between force and stroke under isobaric conditions. This characteristic bears the potential to significantly improve applications such as vibration isolation, which benefits from minimal stiffness and low eigenfrequency, and automation, where reduced air consumption can be achieved, e.g., in clamping devices. Conventional actuator designs struggle to provide these features because they require the active area to increase with the stroke.

This paper deals with a novel origami-inspired actuator concept, which offers a promising solution by increasing the active area progressively over its stroke. The primary challenge lies in the structure's design. From a mathematical point of view, origami actuators with rigid facets represent shape-variable polyhedrons, which are isochoric despite a possible form change, making them unsuitable as pneumatic actuators. However, engineering approaches demonstrate that slight elongations of selected edges during form change can enable significant volume changes. The goal is to overcome technical challenges by developing an airtight origami structure that transmits force along its facets and edges while permitting minor facet deviations.

A working prototype is designed, manufactured and experimentally investigated on a force-stroke test rig. Measurement results validate the feasibility of the concept and the achievement of the targeted properties.

Keywords. origami, pneumatic, actuators, design, progressive force actuators

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Origami is the art of paper folding. Although origami was traditionally used for artistic purposes, it found technical applications in the 20th century [1]. Examples include space-efficient, folded structures for rocket transportation [2]; medical products, such as stents [3]; and foldable concrete structures designed to achieve high local stiffness [4]. Complex patterns are created by introducing mountain and valley folds into a sheet of paper [5]. This results in rigid facets connected by movable creases. Points where these creases meet are referred to as vertices [5].

In a new approach, Reinertz demonstrated that origami can be used for designing a pneumatic actuator chamber with a progressive force-stroke characteristic. By inverting the principle of a bellow actuator, this approach utilizes rigid facets to transmit force while allowing deformation along the creases. Currently, no commercially available pneumatic actuators offer a progressive stroke characteristic, though such actuators would be highly beneficial for suspension systems due to their low eigenfrequency and could contribute to energy savings in automation. [6]

This paper will turn this new approach, as shown in Figure 1.1.a), into a working prototype. Manufacturing difficulties will be solved by tailoring the mechanical properties of Reinertz's proposed geometry. An experimental investigation will demonstrate the desired progressive force-stroke characteristic of the actuator, which could result in a low eigenfrequency.

Sadeghi et al. already employed origami techniques to develop a quasi-zero stiffness pneumatic actuator [7]. The proof-of-concept prototype revealed both the desired progressive force-stroke characteristic and low eigenfrequency [8]. However, this design lacks practicality because the end facets of the used Miura-Ori origami change significantly in size during actuation, as shown in Figure 1.1.b). This is not beneficial for a pneumatic actuator because the required sliding seals between the facets and adjacent end faces induce significant amounts of leakage and friction. Additionally, the point of force application moves relative to the end faces.

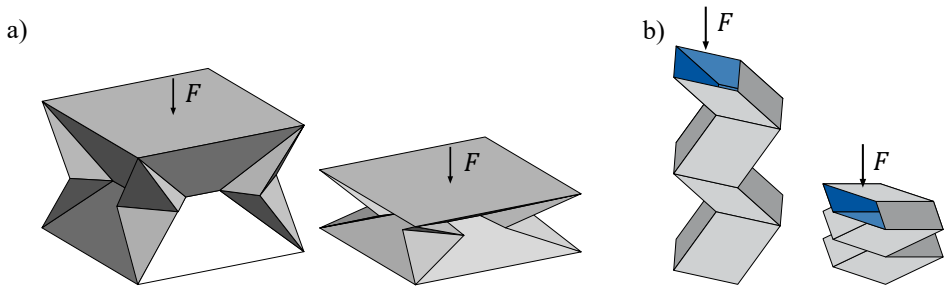


Figure 1.1. a) Reinertz's proposed geometry [6]; b) origami structure of Sadeghi et al. [7]

Building upon Reinertz's approach, the goal is to create a polyhedron whose facets do not change during deformation [6]. From a mathematical perspective, however, it is impossible to alter the volume of a deformable closed polyhedron without changing its facet shapes [9], making such structures infeasible as pneumatic actuators. Nevertheless, Reinertz's design

creates a change in volume through slight deviations from a mathematically exact polyhedron, achieved by the elasticity of a few creases [6].

The scope of this paper is to first optimize the proposed geometry for manufacturing a working prototype. After analyzing critical points, the origami pattern will be optimized numerically to minimize facet deformation. The manufactured prototype will finally be tested on a testbed to characterize the pneumatic properties of the actuator, focusing on the force stroke correlation under isobaric conditions.

2. ANALYSIS OF CRITICAL POINTS

To further improve feasibility, the proposed geometry, shown in Figure 2.1.a), is slightly modified. For the realization of technical origami structures, vertices are critical points since multiple creases converge there. The complexity of sealing and force transmission increases as the number of meeting creases increases. As shown in Figure 2.1.b), one way to reduce this complexity is by lowering the maximum number of creases that meet at each vertex. The size of these new facets can be tailored so that the pneumatic properties of the actuators remain comparable.

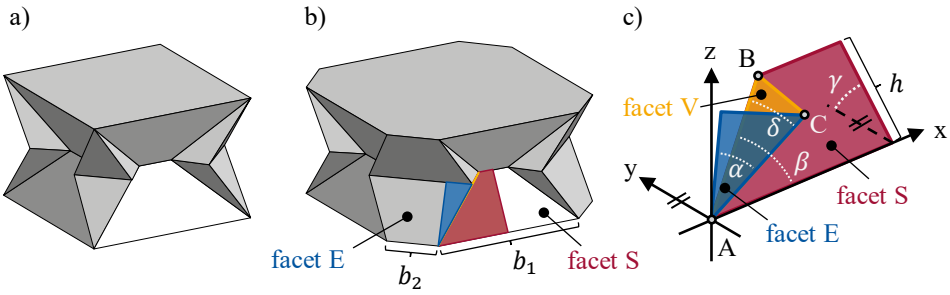


Figure 2.1. a) Reinertz's proposed geometry [6]; b) modified geometry; c) one vertex of the origami-pattern

For analytical purposes, the structure can be simplified to a single vertex, as illustrated in Figure 2.1.c). This reduction is possible because the entire actuator structure can be formed by reflecting this vertex multiple times. In this configuration, the facets are labeled E, S, and V with corresponding angles α, β, and δ. The actuator stroke is defined by angle γ. For clarity in further discussions, key points A, B, and C are also marked.

In a classic origami pattern, angles α, β, and δ are related by Equation 2.1 and do not change. This relation is not effective for a functional actuator. As Reinertz demonstrated, an actuator can be created by allowing a small elongation of the crease $\overline{BC}$ [6]. Therefore, the angle δ must change during actuation.

$$\alpha + \beta + \delta = 90° \tag{2.1}$$

The goal is to minimize stress on facet V and its connecting creases, as well as to simplify manufacturing, by keeping changes in δ as small as possible. The alteration in δ over a full actuator stroke is defined as $\Delta\delta$ by Equation 2.2:

$$\Delta\delta = \delta_{\max} - \delta_{\min} \tag{2.2}$$

Here, δ_{max} is the maximum value of δ during the actuator stroke, and δ_{min} is the corresponding minimum value. Minimizing $\Delta\delta$ requires numerical optimization of the angles α, β, and δ. This optimization process will be discussed in detail in the following section.

3. NUMERICAL OPTIMIZATION OF THE ORIGAMI PATTERN

To minimize $\Delta\delta$, a geometric analysis is required to describe the angle δ as a function of the other geometric parameters and the actuator stroke. The distances between points A, B, and C ($\overline{AB}$, $\overline{AC}$ and $\overline{BC}$) are defined as shown in Equations 3.1 to 3.3:

$$\overline{AB} = h \cdot \sqrt{\left(\frac{1}{\tan\beta}\right)^2 + \cos^2\gamma + \sin^2\gamma} \tag{3.1}$$

$$\overline{AC} = h \cdot \sqrt{\frac{1}{2}(\cos\gamma + \tan\alpha)^2 + \frac{1}{2}(\cos\gamma - \tan\alpha)^2 + \sin^2\gamma} \tag{3.2}$$

$$\overline{BC} = h \cdot \sqrt{\left(\frac{1}{\tan\beta} - \frac{\sqrt{2}}{2}(\cos\gamma + \tan\alpha)\right)^2 + \left(\cos\gamma - \frac{\sqrt{2}}{2}(\cos\gamma - \tan\alpha)\right)^2} \tag{3.3}$$

Using the law of cosines, the angle δ can be calculated by Equation 3.4:

$$\delta(\alpha,\beta,\gamma) = \arccos\left(\frac{\overline{AB}^2 + \overline{AC}^2 - \overline{BC}^2}{2\cdot\overline{AB}\cdot\overline{AC}}\right) \tag{3.4}$$

$\delta(\alpha,\beta,\gamma)$ is independent of the height h, as it is eliminated. By choosing values for α and β, δ can be determined depending on γ. It can be shown that δ follows a trajectory similar to Figure 3.1.c). As γ approaches zero, $\frac{\partial\delta}{\partial\gamma}$ also approaches zero. Depending on the values of α and β, δ reaches a minimum before always increasing as γ approaches 90°, which represents the maximal possible stroke.

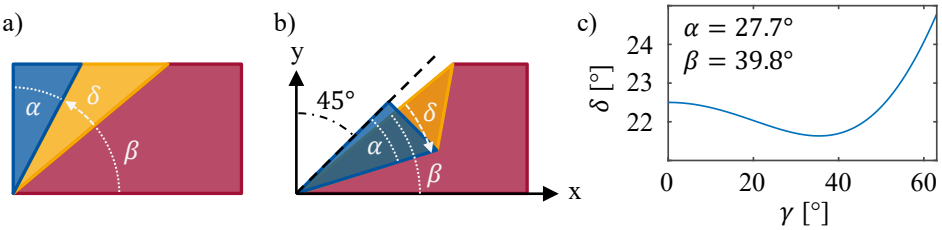


Figure 3.1. Exemplary course of $\boldsymbol{\delta}$ over the stroke $\boldsymbol{\gamma}$

To maximize the stroke while minimizing changes in δ, γ could theoretically be set near zero or around the minimum point of δ. However, setting γ close to zero is not technically feasible because it would cause actuator components to intersect. Therefore, γ is defined as $\gamma = [21°, 45.8°]$. This range includes the minimum point of δ and allows for a maximal stroke that is twice as large as the minimal stroke.

To ensure that the structure closely resembles origami, the relationship given by Equation 2.1 is taken into consideration. This relationship guarantees that the origami pattern can be folded from a flat sheet of paper, as illustrated in Figure 3.1.a). The pattern

can be folded according to Figure 2.1.c). When the pattern is folded flat ($\gamma = 0°$), Figure 2.1.c) becomes Figure 3.1.b) when viewed from above. Then, α, β, and δ have the relationship given by Equation 3.5:

$$\alpha + \beta - 45° = \delta \tag{3.5}$$

Equating Equation 2.1 and Equation 3.5 yields $\delta = 22.5°$. Inserting this into Equation 2.1 yields the condition between α and β given by Equation 3.6:

$$\alpha + \beta = 67.5° \tag{3.6}$$

The maximum and minimum values of δ during the stroke $\gamma = [21°, 45.8°]$ are determined using Equation 3.4 with values of α and β that satisfy the condition in Equation 3.6. Then, the change in δ, $\Delta\delta$, is calculated using Equation 2.2.

With $\alpha = 27.70°$ and $\beta = 39.80°$, the change in δ, $\Delta\delta$, is minimal and less than 0.5°. Finally, δ is redefined as $\delta_{\text{new}} = 21.81°$. This value is the mean between $\delta_{\max}$ and $\delta_{\min}$. Although this contradicts the relationship of Equation 3.5, δ_{new} ensures that compression and elongation of the crease $\overline{BC}$ are balanced.

Subsequent work will proceed using these optimized parameter values.

4. PROTOTYPE DESIGN AND MANUFACTURING

Building on the previously optimized geometry, this section describes the process of translating the design into a functional prototype and addresses the associated technical challenges.

One of the main challenges in transferring an origami-inspired structure to a technical application is accounting for material thickness. While paper's negligible thickness allows for ideal folding, technical structures require facets with significant thickness to enable force transmission. This consideration has already influenced the lower limit of γ. It also allows Tachi's approach to be applied, in which origami pattern creases serve as rotational axes and material is strategically removed from areas where intersections would otherwise occur [10].

This actuator uses hinges, as used by Shao et al. [11], in creases that transmit compressive forces for several reasons. They are durable and cannot be wedged between facets under compression. They also avoid brittleness over time, which could occur with flexible materials. However, since traditional hinges are not airtight, an air pouch must be inserted inside, as in Sadeghi et al.'s method [8]. The relative movement of the hinges must be considered, as it could damage the air pouch.

Since facets E and S are the largest facets of the actuator, they bear most of the force that needs to be transmitted to the end facets of the actuator. These compressive forces need to be transmitted via the creases, so these facets are connected via hinges. This results in numerous precisely aligned hinge axes across the entire actuator, necessitating tight production tolerances.

During small strokes, facet V becomes wedged between facets E and S. In order to accommodate the material thickness of facet V without any material intersection, the material from facets E and S must be moved out of their respective planes. This adjustment does not affect pneumatic performance. The force output remains unchanged, as long as the

hinge axes remain correctly positioned. However, this modification results in less favorable internal load path within facets E and S, so these parts require reinforcement with support structures. Support structures on the inside of the actuator should be avoided to protect the air pouch.

These requirements, including hinged connections with precise positioning and small tolerances, out-of-plane material displacement, and reinforcement, result in highly complex components for facets E and S. Consequently, additive manufacturing is chosen for prototyping these parts. To be cost-efficient, FDM (fused deposition modeling) printing with PLA (polylactic acid) is used. However, the anisotropic mechanical properties resulting from print orientation must be considered, as demonstrated by Abdul Sani [12]. In general, to prevent hinge failure, the hinge axes should be perpendicular to the print layers. In case of this actuator, this orientation would expose parts of the components to perpendicular force transmission through the print layers, which could result in component failure. As a compromise, the components are printed at an angle (see Figure 4.1.). The wall thickness is set at $t = 4\,\text{mm}$, and steel rods with a diameter of $d = 2\,\text{mm}$ are used as hinge axes. Given these choices in materials and manufacturing methods, maximum internal pressure is limited to $p_A = 1\,\text{bar}_{rel}$. Support structures were iteratively improved under these conditions using an FEM (finite element method) simulation for facets E and S. The resulting parts for facet E and S are shown in Figure 4.1.. Figure 4.3. shows the parts assembled The dimensions corresponding to Figure 2.1. of the parts are $b_1 = 170\,\text{mm}$ and $b_2 = h = 50\,\text{mm}$.

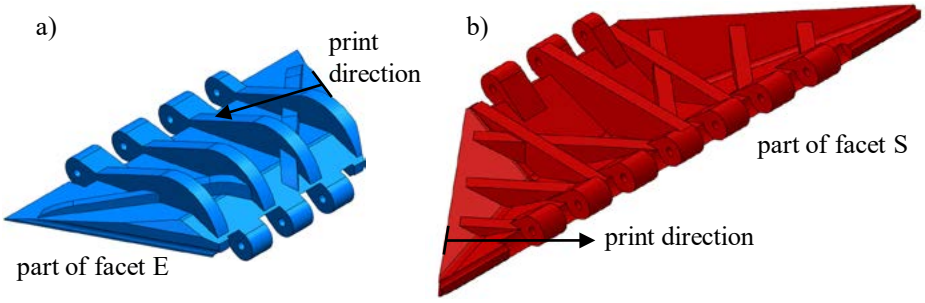


Figure 4.1. Parts of facet E and S, the shown print direction is perpendicular to print layers

A combination of steel and textile is used for both the hinges of facet V and the facet itself for several reasons. The fabric serves as a hinge, requiring significantly less space than a conventional mechanical hinge. These flexible fabric hinges are also well-suited for transmitting tensile forces, which are present along the outward-facing creases of facet V. A 1.5 mm sheet metal panel forms the main body of facet V, ensuring that it remains rigid and does not bulge under pressure. However, as it cannot accommodate for changes in δ, it is sandwiched between two layers of fabric. As the actuator operates, these fabric layers not only create the hinges but also compensate for variations in δ by stretching alongside the metal.

The sheet metal is held in place solely by stitching together the two surrounding fabric layers. In principle, combining the textile with an elastomer to form a composite material could make both facet V and its hinges airtight. Deleo et al. [13] and Zhakypov et al. [14] used similar approaches. However, since an air pouch is already used to achieve airtightness

due to the non-sealing hinges on facets E and S, this additional step is unnecessary. Linen is selected as the textile because it allows controlled stretching, can efficiently transmit tensile forces, and remains thin enough for compact integration. The resulting structure of facet V is illustrated in Figure 4.2..

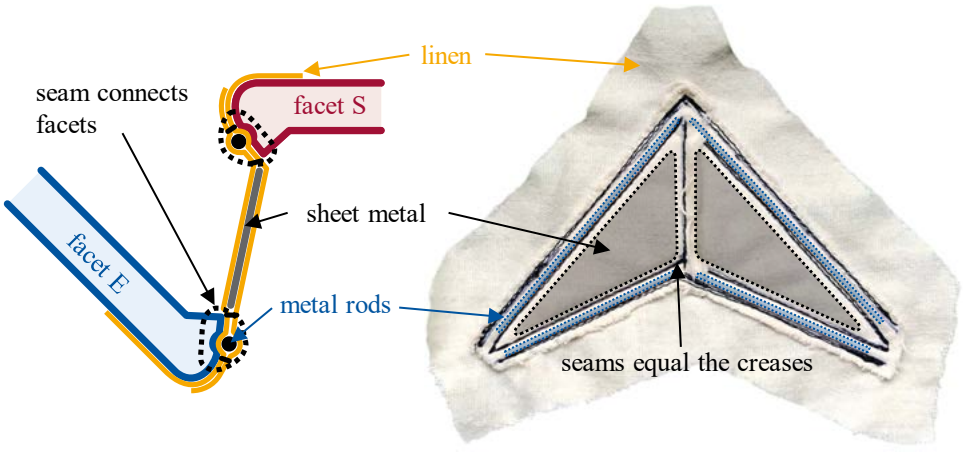


Figure 4.2. Design of facet V; Cross-section through facet E, V and S (left) and assembled facet V (right)

This design provides sufficient space for the air pouch while allowing facet V to fit between facets E and S throughout actuator movement. Metal rods, also enclosed within the two fabric layers of facet V, are stitched to facets E and S to connect facet V securely at both ends, as shown in Figure 4.2..

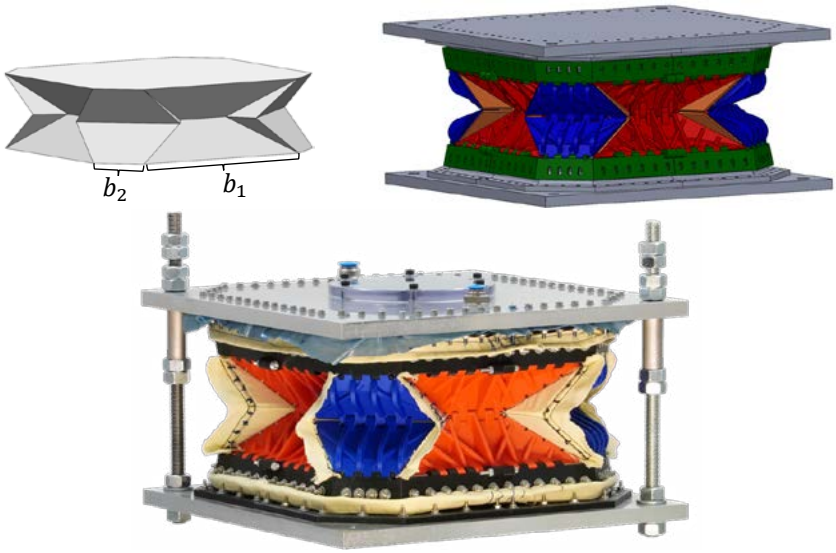


Figure 4.3. top: Origami-pattern and CAD-model of the prototype (screws are not shown); bottom: prototype of the actuator

Based on Sadeghi's approach, the actuator is sealed using a custom-made, tube-shaped air pouch made of 0.1 mm thick, low-density polyethylene film [8]. To protect the air pouch, all interior surfaces are kept flat with rounded corners and small gaps between parts. Additionally, a textile sleeve separates the moving origami elements from direct contact with the air pouch to prevent frictional wear or pinching, even when gaps open around moving joints. The air pouch's circumference exceeds what is needed. This ensures that the air pouch does not stretch but rather forms harmless folds of excess material.

Finally, the top and bottom of the actuator are closed off with 13 mm aluminum plates. These plates provide mounting points for mechanical and pneumatic connections. The air pouch is sandwiched between the printed origami parts and the aluminum plates. Silicone sealant ensures airtightness. Figure 4.3. illustrates the final prototype.

Like Tachi's approach, the rotational axes of the actuator correspond to the origami creases [10].

5. FORCE APPROXIMATION

Before conducting experimental investigations, the force-stroke characteristic of the developed actuator is first approximated mathematically. This preliminary analysis supports both the design of the testbed and the validation of subsequent measurement results.

To simplify the calculations, the origami structure is reduced to a two-dimensional bar model, as shown in Figure 5.1.. In this approximation, only facets E and S are considered. The contribution from facet V is neglected because its effect on the total output force is relatively small. The resulting pressure force of each facet F_{res} is concentrated at its centroid. Hereby z_s is the distance between the centroid and the hinge.

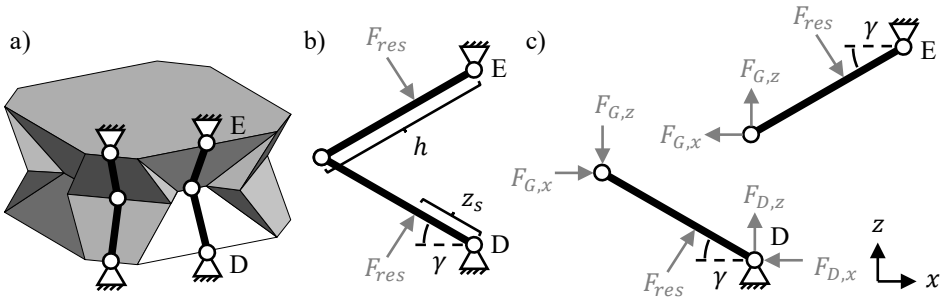


Figure 5.1. simplified mechanism for force approximation

The force $F_{G,z}$ can be determined for the lower and upper beams in Figure 5.1.c) if moment equilibrium is assumed at points D and E. Equating the two moment equilibria, it follows that $F_{G,z} = 0$. The force $F_{D,z}$ can be calculated using force equilibrium for the lower beam, as shown by Equation 5.1:

$$\Sigma F_z = 0 \Rightarrow F_{D,z} = -F_{res} \cdot \cos\gamma + F_{G,z} = F_{res} \cdot \left(\frac{z_s}{\cos\gamma \cdot h} - \cos\gamma\right) \tag{5.1}$$

The overall force generated by the actuator can then be expressed as shown in Equation 5.2:

$$F_{actuator} = 4 \cdot F_{D,z-facet\,S} + 4 \cdot F_{D,z-facet\,E} + F_{end\,face} \tag{5.2}$$

The force, $F_{end\ face}$, is the internal pressure multiplied by the area of the end face. The approximation of the actuator's output force will be used to validate the experimental results.

6. EXPERIMENTAL INVESTIGATION

To validate the proposed concept of an origami-inspired pneumatic actuator with a progressive force-stroke characteristic, this section describes the experimental procedures used to evaluate the previously designed prototype.

The primary focus of the experiments is to determine the force-stroke relationship of the actuator at constant internal pressures. To achieve this, a testbed is set up that can maintain a constant internal pressure, allow for variable stroke adjustments, and accurately measure the resulting force. Due to limitations of the available testing equipment, measurements can only be taken up to a maximum load of $5{,}000\ \mathrm{N}$. This restricts the maximum internal pressure p_A during testing.

During each test series, the internal pressure of the actuator is set to an initial value of $0.05\ \mathrm{bar}_{rel}$, which remains constant while the stroke is varied. The stroke is repeatedly decreased and increased three times before the internal pressure is raised by another increment of $0.05\ \mathrm{bar}$. This process continues until the actuator's output force reaches $5{,}000\ \mathrm{N}$. All measurements were recorded at a sampling rate of $1{,}000\ \mathrm{Hz}$ for high temporal resolution.

7. ACTUATOR CHARACTERIZATION

This section presents and analyzes the results obtained from experimental testing of the novel origami-inspired pneumatic actuator.

The previously described measurements were performed using the presented prototype. The maximum pressure is limited to $0.7\ \mathrm{bar}_{rel}$ as the output force would otherwise exceed $5{,}000\ \mathrm{N}$. For this reason, the maximum pressure for which the actuator is designed cannot be tested. To visualize the results, force is plotted against actuator stroke. The stroke x can be calculated from the angle γ. The angle $\gamma = [21°, 45.8°]$ corresponds to a stroke $x = [35.8\ \mathrm{mm},\ 71.6\ \mathrm{mm}]$ with the proposed dimensions of the prototype. To eliminate the influence of the end stops, x is set within the range of $x = [36.5\ \mathrm{mm},\ 72.3\ \mathrm{mm}]$. The lower end stop is defined by the actuator parts. The upper end stop can be set freely. Since the lower limit of x increased to eliminate the influence of the lower end stop, the upper limit is also raised to achieve the intended total stroke. In Figure 7.1. the force-stroke characteristic is exemplified for an internal pressure of $p_A = 0.3\ \mathrm{bar}_{rel}$. To reduce measurement noise each data point is averaged over the previous 100 samples. This is acceptable because the stroke changes relatively slowly.

The resulting plot shows a hysteresis between increasing and decreasing stroke. The hinge friction could be the main cause of this effect. When retracting, frictional forces act in the same direction as the measured force. When extending, they oppose it. Consequently, measured force values are lower during extension than during retraction at any given stroke position.

Figure 7.1. also displays the approximated force curve for $p_A = 0.3\,\mathrm{bar}_{rel}$ based on the previous calculations from section 5. Both the magnitude and trend of measured values closely match these theoretical predictions.

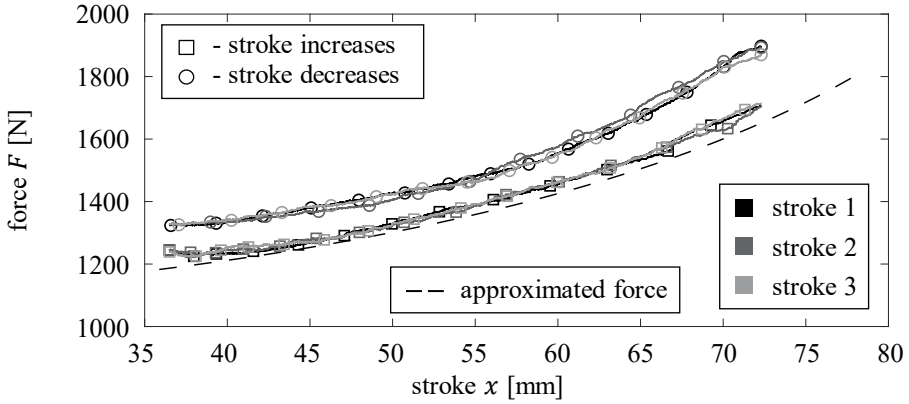


Figure 7.1. Actuator output force F plotted against actuator stroke x, with internal pressure $p_A = 0.3\,\mathrm{bar}_{rel}$ and approximated force according to section 5

To further analyze performance and minimize the effect of pressure fluctuations during testing, each measured force value was divided by its corresponding pressure reading. The resulting "active area" metric A_{active} is proportional to the output force, yet independent of the pressure. These active area results for an internal pressure of $p_A = 0.3\,\mathrm{bar}_{rel}$ are shown in Figure 7.2..

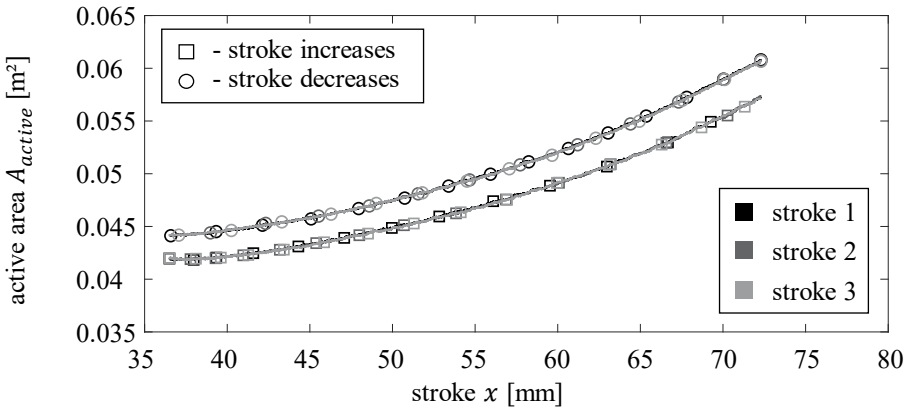


Figure 7.2. Active area A_{active} plotted against actuator stroke x, with internal pressure $p_A = 0.3\,\mathrm{bar}_{rel}$

The three stroke cycles now align more closely with each other. This indicates that the previously observed fluctuations were largely due to minor variations in internal pressure rather than inconsistent actuator behavior.

The active area serves as a robust measure for comparing different pressure levels independently of the internal pressure. Figure 7.3. presents the active area of the actuator at different internal pressures. The internal pressure, p_A, ranges from $0.05\ \mathrm{bar}_{rel}$ to $0.7\ \mathrm{bar}_{rel}$ and describes the internal pressure of the actuator.

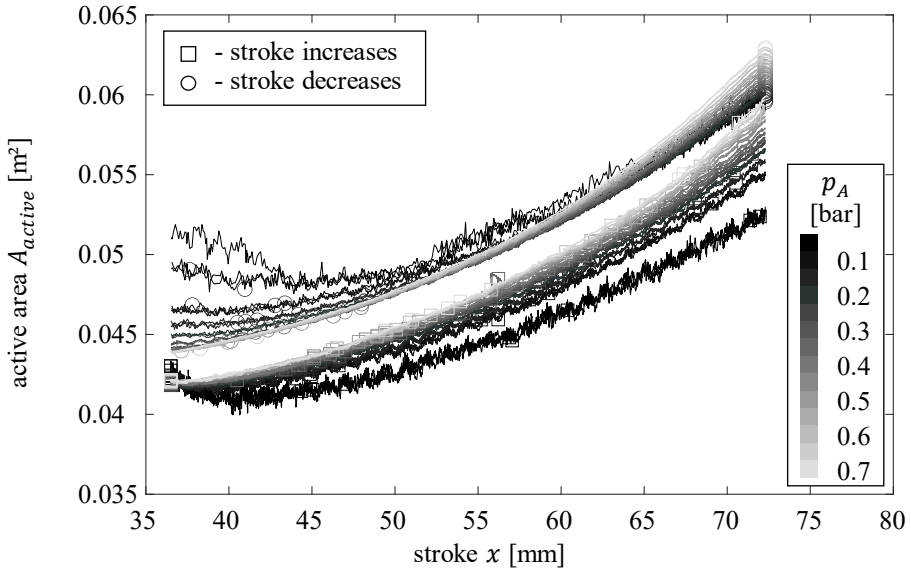


Figure 7.3. Active area A_{active} plotted against actuator stroke x, with different internal pressures p_A

At lower pressures, greater deviations are observed. These are likely because friction has a proportionally larger effect at low pressure. Friction could increase at low strokes, as the support structures of Facet E move closer together and rub against each other. At higher pressures the curves converge consistently. For strokes above approximately $x = 60\ \mathrm{mm}$, some divergence appears at higher pressures. This is likely due to stretching or deformation of parts under load, which results in minor changes in γ. These changes result in greater variations in output force at extended positions.

Overall, these results demonstrate that the actuator exhibits repeatable behavior and delivers a progressive force-stroke characteristic consistent with prior calculations.

8. CONCLUSION AND DISCUSSION

In this work, a novel origami-inspired pneumatic actuator concept featuring a progressive force-stroke characteristic was developed. An initial analysis of the geometric relationships served as a basis for the design and prototype implementation with minimized stretching of the facets.

The design uses 3D-printing. A textile-metal composite is used to achieve the necessary deformations in the actuator's facets. This approach allows the complex origami-geometry to be realized in a technical structure.

Experimental validation shows that the prototype works as intended. The measured force-stroke curves match the mathematically predicted progressive behavior. Results are consistent across multiple tests, showing that the actuator's performance is reproducible.

These findings confirm that it is possible to build an origami-based pneumatic actuator with a progressive force-stroke characteristic, by integrating flexible materials at key points.

Despite these achievements, high manufacturing effort remains a limitation due to complex components that require significant assembly effort. Future work can further develop this concept. Various origami-patterns could be explored to tailor pneumatic properties further based on usage requirements. Material selection and production methods can also be improved. Reducing manufacturing complexity and employing alternative materials for enhanced mechanical performance could facilitate the industrial use of such actuators. Potential applications include suspension systems or automation technology. In these fields, actuators with tailored force-stroke characteristics offer significant advantages. A simpler design could miniaturize the actuator, qualifying it for broader applications.

Overall, this study demonstrates both the feasibility and potential of origami-inspired pneumatic actuators for advanced engineering applications.

9. REFERENCES

[1] R. J. Lang, 'The complete book of origami', Dover Publications Inc., New York, 1988. ISBN: 978-0-486-25837-9.

[2] S. Dixit, R. Hyde, A. Weisberg, J. Early, M. Rushford, J. Britten, 'Development of Large Aperture, Light-Weight Fresnel Lenses for Gossamer Space Telescopes', 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, American Institute of Aeronautics and Astronautics, Reston, Virigina, 2002. https://doi.org/10.2514/MSDM02.

[3] K. Kuribayashi, K. Tsuchiya, Z. You, D. Tomus, M. Umemoto, T. Ito, M. Sasaki, 'Self-deployable origami stent grafts as a biomedical application of Ni-rich TiNi shape memory alloy foil', Materials Science and Engineering: A, pp. 131–137, 2006. https://doi.org/10.1016/j.msea.2005.12.016.

[4] S. Bhat, K. Heins, C. Bertram, O. Reinertz, K. Schmitz, T. Gries, 'Ein Ausblick auf das Zusammenspiel von Faltprozess und Drapierbarkeit der Textilbewehrung', Beton und Stahlbetonbau, pp. 925–935, Ernst & Sohn GmbH, 2024. https://doi.org/10.1002/best.202400062.

[5] L. M. Fonseca, G. V. Rodrigues, M. A. Savi, 'An overview of the mechanical description of origami-inspired systems and structures', International Journal of Mechanical Sciences, 2022. https://doi.org/10.1016/j.ijmecsci.2022.107316.

[6] O. Reinertz, 'Static Properties Based Classification of Pneumatic Actuators - Potential of Progressive Effective Area Over the Stroke', International Conference Fluid Power 2025: Conference Proceedings, pp. 111–130, University of Maribor Press, Maribor, 2025. https://doi.org/10.18690/um.fs.7.2025.

[7] S. Sadeghi, S. Li, 'Harnessing the Quasi-Zero Stiffness From Fluidic Origami for Low Frequency Vibration Isolation', Volume 2: Modeling, Simulation and Control of

Adaptive Systems; Integrated System Design and Implementation; Structural Health Monitoring, American Society of Mechanical Engineers, 2017. https://doi.org/10.1115/SMASIS2017-3754.

[8] S. Sadeghi, S. Li, 'Fluidic origami cellular structure with asymmetric quasi-zero stiffness for low-frequency vibration isolation', Smart Mater. Struct., 2019. https://doi.org/10.1088/1361-665X/ab143c.

[9] I. K. Sabitov, 'On the problem of invariance of the volume of a flexible polyhedron', Russ. Math. Surv., pp. 451-452, 1995. https://doi.org/10.1070/RM1995v050n02ABEH002095.

[10] T. Tachi, 'Rigid-Foldable Thick Origami', Origami 5, pp. 253–263, A K Peters/CRC Press, 2016. https://doi.org/10.1201/b10971.

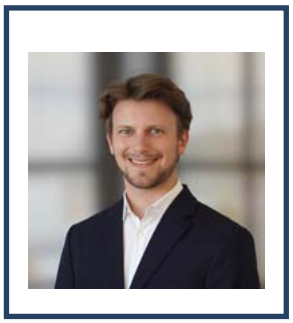
[11] Z. Shao, W. Zhao, Z. Zuo, J. Li, I.-M. Chen, 'An Origami-Inspired Negative Pressure Folding Actuator Coupling Hardness with Softness', Actuators, p. 35, 2023. https://doi.org/10.3390/act12010035.

[12] A. S. Abdul Sani, M. N. Osman Zahid, M. R. Mohamad Yasin, S. Z. Ismail, M. Z. Mohd Zawawi, A. R. Abdul Manaf, S. N. Mohd Saffe, R. Abd Aziz, F. Mohd Turan, 'Enabling Industry 4.0 through Advances in Manufacturing and Materials', Springer Nature Singapore, Singapore, 2022. ISBN: 978-981-19-2889-5.

[13] A. A. Deleo, J. O'Neil, H. Yasuda, M. Salviato, J. Yang, 'Origami-based deployable structures made of carbon fiber reinforced polymer composites', Composites Science and Technology, 2020. https://doi.org/10.1016/j.compscitech.2020.108060.

[14] Z. Zhakypov, F. Heremans, A. Billard, J. Paik, 'An Origami-Inspired Reconfigurable Suction Gripper for Picking Objects With Variable Shape and Size', IEEE Robot. Autom. Lett., pp. 2894–2901, 2018. https://doi.org/10.1109/LRA.2018.2847403.

Biographies



Philipp Hüpper received the bachelor's degree in mechanical engineering in 2022 and the master's degree in product development in 2025 from the RWTH Aachen University. He is currently working as a Research Associate at the Institute for Fluid Power Drives and Systems, Faculty of Mechanical Engineering, RWTH Aachen University. His research areas include Gas- and Micro-Systems with a special focus on high pressure hydrogen systems.



Prof. Olivier Reinertz received his diploma and his doctoral degree in mechanical engineering from RWTH Aachen University, Germany. He is currently Scientific Director at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. His research focuses on the model-based analysis of fluid power components and systems and the derivation and validation of innovative strategies for efficiency and performance optimization with an emphasis on compressed air and gas-powered systems.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

AAS-Orchestrated Production – End-to-End Digital Integration from Machine to Visualization

Micha Koebele[1], Jan Edler[2], Philipp Niemann-Stryczek[3], Heiko Baum[1]

[1] *FLUIDON GmbH, micha.koebele@fluidon.com, heiko.baum@fluidon.com*
[2] *Emerson Aventics GmbH, jan.edler@emerson.com*
[3] *IQstruct Engineering GmbH, philipp.stryczek@iqs-e.de*

Abstract.

This paper demonstrates how the Asset Administration Shell (AAS) can be used as an integration and orchestration layer for automated, product-specific processing of real machine data. Using a ballpoint pen assembly machine, standardized AAS mechanisms enable cloud-based visualization, energy analysis, and CO_2 accounting. The approach highlights product-related sustainability assessment in fluid power systems and illustrates the AAS as a key enabler for interoperable and scalable digital engineering workflows.

Keywords. Fluid 4.0, Asset Administration Shell (AAS), AAS-orchestrated data processing, Digital Twin, Fluid Power Systems, Energy and CO_2 analysis, Sustainability assessment, Industry 4.0

1. INTRODUCTION

The increasing demands for transparency, energy efficiency, and sustainability in industrial production require new, integrated digital concepts for information and data processing. In fluid-based production systems in particular, there has been a lack of standardized approaches for consistently utilizing real machine data for analysis, visualization, and product-related sustainability assessments.
This article presents an end-to-end solution orchestrated entirely by the Asset Administration Shell (AAS) [1], which is demonstrated using a ballpoint pen assembly machine as an example. The aim is to present the end-to-end flow of information from the real machine via cloud-based services to product-related CO_2 accounting and visualization in a comprehensible manner.
To provide context for the application example, the following chapter first introduces the machine used.

2. MACHINE AND SYSTEM OVERVIEW

The left-hand side of **Figure 2.1** shows the machine used to demonstrate the AAS-orchestrated data flow. The machine is an *Industry 4.0* [2] demonstrator from the *Fluid 4.0*

[3] research project. The machine, built by project partner *Emerson Aventics* [4], produces ready-to-use ballpoint pens from four basic components: the housing, refill, spring, and cap. The actual assembly process is carried out by valve-controlled pneumatic grippers and rotary actuators, while electromechanical linear axes transfer the components from the transfer points of the storage magazines to the respective assembly positions. After successful assembly, the machine assigns a unique serial number to each manufactured ballpoint pen, which is printed on a Fluid 4.0 product sticker and attached to the pen.
This serial number then allows the owner to access the product-related production data for their ballpoint pen using any AAS-enabled application.

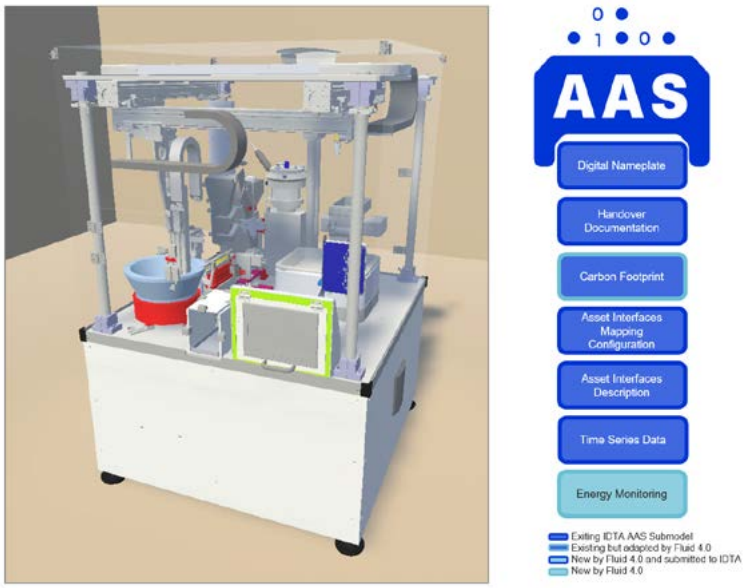


Figure 2.1 : Ballpoint pen assembly machine and AAS submodels used

During the assembly process, the machine collects the data listed below and continuously transmits it to the *MQTT broker* [5] of the project partner *Fluidon* [6]:

- pneumatic air consumption,
- electrical energy consumption,
- valve signals
- Positions of the linear axes
- serial number of the ballpoint pen being produced.

This production data forms the basis for AAS-orchestrated data processing.
The right-hand side of Figure 2.1 shows an overview of the AAS submodels used in the application example. The legend distinguishes between submodels that have already been standardized by the *IDTA* [7] and those that have been newly developed or expanded as part of the Fluid 4.0 research project.

3. AAS-ORCHESTRATED DATA PROCESSING WORKFLOW

Data processing is carried out via a data processing workflow created in the SaaS development environment *Fluidon Cube* [8]. For the sample application, the processing of production data is deliberately implemented in small steps in order to present the individual interactions with the AAS submodels involved in a didactically comprehensible manner. **Figure 3.1** shows the sequence of process steps with which the project partners implemented AAS-orchestrated data processing.

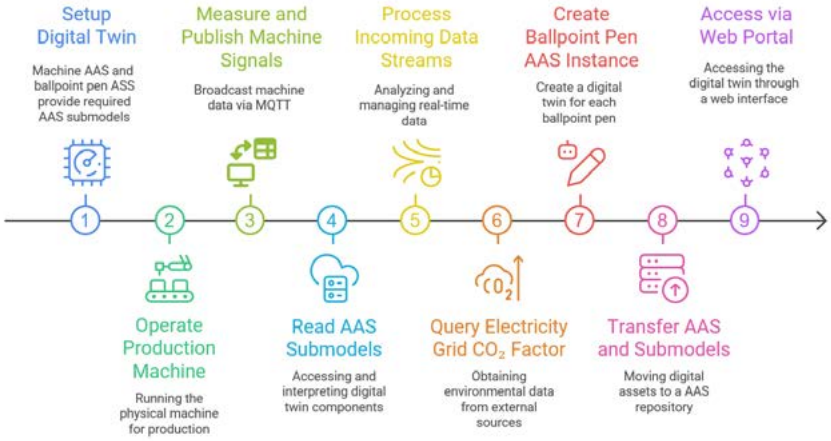


Figure 3.1 : Overview of the process steps

3.1. Preparatory work

In the first process step, an instance AAS of the machine and a type AAS of the ballpoint pen were first created (see Figure 3.1-1, *"Setup Digital Twin"*). The following process step (see Figure 3.1-2, *"Operate Production Machine"*) is representative of the preparatory work on the machine. Pneumatic pressure and flow sensors, a system for recording the electrical power consumption of the linear axes, and a label printer were retrofitted. The third process step (see Figure 3.1-3, *"Measure and Publish Machine Signals"*) represents the integration of all necessary signals into the control system's data acquisition, their processing, and their transmission to the MQTT broker.

3.2. The data processing workflow

The fourth process step (see Figure 3.1-4, *"Read AAS Submodels"*) symbolizes the start of the actual AAS-orchestrated data processing. After the workflow has retrieved the machine's AAS submodels and connected to the MQTT broker, signal processing takes place (see Figure 3.1-5, *"Process Incoming Data Streams"*). Following the query of the current CO_2 emission factor (see Figure 3.1-6, *"Query Electricity Grid CO_2Factor"*) and calculating the CO_2 parameters, an instance AAS of the ballpoint pen just produced is created in the seventh process step (see Figure 3.1-7, *"Create Ballpoint Pen AAS Instance"*).

3.3. Provision and use of data

In the eighth process step, the generated instance AAS and the associated submodels are transferred to any AAS repository server (see Figure 3.1-8, *"Transfer AAS and Submodels"*).

The information contained therein or linked to it is then available for interoperable access, for example via a web portal or other AAS-enabled applications (see Figure 3.1-9, *"Access via Web Portal"*).

4. PARTICIPATING ASSET ADMINISTRATION SHELLS

At the heart of AAS-orchestrated data processing are the instance AAS of the machine and the type AAS of the ballpoint pen. As already presented in the right-hand section of the image at Figure 2.1, both standardized AAS submodels and those developed jointly with other Fluid 4.0 project partners were used to create them.

4.1. Instance AAS of the ballpoint pen assembly machine

With the aim of graphically illustrating the structural design of an AAS, **Figure 4.1** shows a screenshot of the instance AAS of the machine. In the context of the application example, the instance AAS can be accessed via the AAS repository server of the project partner *IQstruct Engineering* [9]. Access to the machine's AAS instance is controlled by IQstruct Engineering via an access rights management system based on preliminary work from *Eclipse BaSyx* [10] and prototypically implemented as part of the Fluid 4.0 research project.

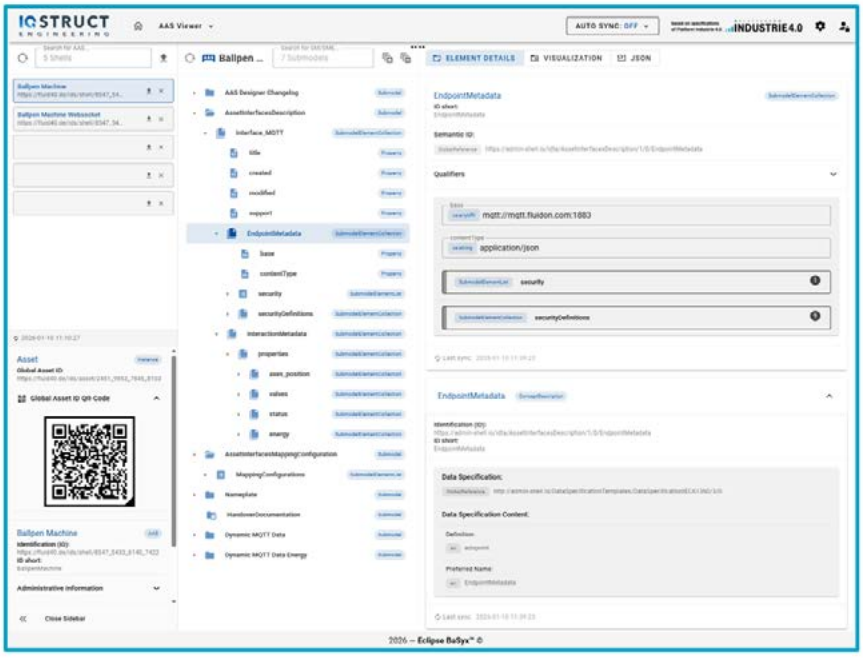


Figure 4.1 : Instance AAS of the ballpoint pen assembly machine

The middle section of the image at Figure 4.1 shows all the submodels contained in the AAS. The two AAS submodels *Asset Interface Description* (AID) [11] and *Asset Interface Mapping Configuration* (AIMC) [12] are used for the application example.
The right-hand side of the image shows elements from the *Submodel Element Collection* (see[1, p. 56]) *"EndPointMetadata"* of the AID submodel (see[11, p. 16]). All information

required for the automatic connection of the data evaluation workflow can be read from the elements in this collection. In the application example, the MQTT broker of the project partner Fluidon is entered here.
The AIMC submodel is used for the formal description of the mapping of technical machine interfaces to semantically defined AAS properties. It enables a clear decoupling between the physical or IT data source (e.g., fieldbus, MQTT, OPC UA) and the semantic structure of the administration shell. This allows production data to be automatically collected, interpreted, and further processed by the data processing workflow without project-specific adjustments. The submodel thus forms the central basis for AAS-orchestrated, scalable data processing.

4.2. Type AAS of the ballpoint pen

Figure 4.2 shows a screenshot of the type AAS of the ballpoint pen. The submodels required for the application example are listed in the left-hand part of the image at Figure 4.2. In addition to the *Nameplate* submodel [13], these are the *Handover Documentation* [14] and *Time Series Data*[15] submodels, as well as the *Carbon Footprint* [16] submodel, which was expanded as part of Fluid 4.0.



Figure 4.2 : Type AAS of the ballpoint pen

To demonstrate interoperability, the Type AAS in the application example is provided via Fluidon's AAS repository server and not via IQstruct Engineering's AAS repository server. A direct comparison of the screenshots from Figure 4.1 and Figure 4.2 shows that the two AAS are also managed on two different AAS repository server implementations. IQstruct Engineering uses the Eclipse *BaSyx Java V2* [17], while Fluidon uses the *Eclipse AASX Server* [18]. This is to illustrate that it is irrelevant for AAS-orchestrated data processing where and with which server implementation the participating AAS are hosted. Interoperable access to information is possible at any time via the standardized AAS mechanisms.
The following chapter describes how the two administration shells and their submodels are used in AAS-orchestrated data processing.

5. AAS-ORCHESTRATED DATA PROCESSING

The AAS-orchestrated data processing workflow is explained using the sequence diagram shown in **Figure 5.1**. Four actors are involved in the workflow:

- *Operator*, who interacts with the data processing workflow
- *Fluidon Cube*, which hosts the workflow
- *Web APIs*, representing the web services involved
- *Real Machine*, as the process data source

It is clear that the actors *Operator* and *Real Machine* are only actively involved in the process at the beginning and end. The actual data processing, on the other hand, is fully automated, as the distributed systems involved can be accessed via the standardized, interoperable access mechanisms of Asset Administration Shell.

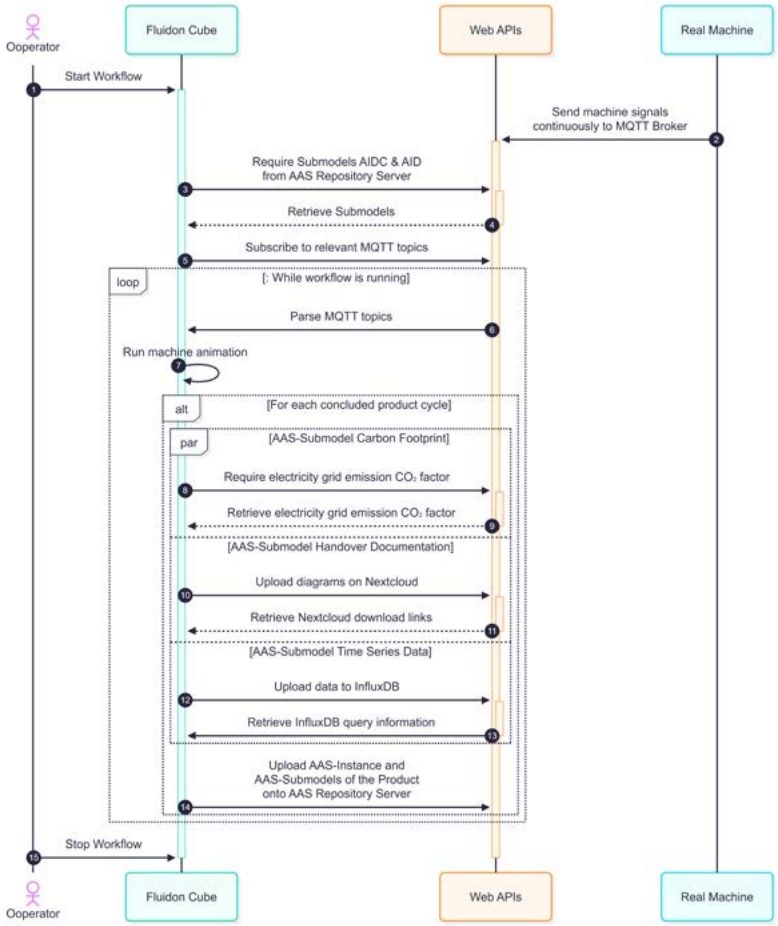


Figure 5.1 : Sequence diagram of AAS-orchestrated data processing

5.1. Start of AAS-orchestrated data processing

In the first step (see Figure 5.1-1), the operator starts the data processing workflow. It is irrelevant whether the machine is already in production at this point or – as shown in the figure (see Figure 5.1-2) – only starts the production process after the workflow has been started.

5.2. Data transfer and data preprocessing

The data processing workflow consists of several tasks, the respective functional scope of which is tailored to the process steps described in Figure 5.1. Each task performs a maximum of one AAS-supported interaction with external services in order to keep the data processing sequence comprehensible for the user of the application example. In the example, the workflow has 8 processor cores at its disposal, so that data processing takes place in parallel tasks. **Figure 5.2** shows the view of the data processing workflow in the *Fluidon Cube* development environment [8].

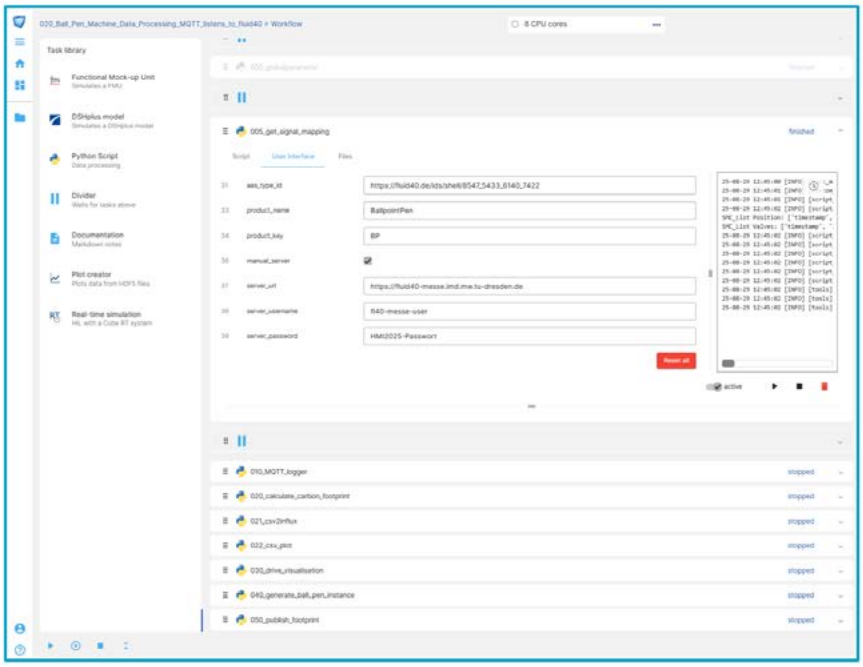


Figure 5.2 : Fluidon Cube data processing workflow [8]

The individual data processing steps were implemented using Python scripts, specifically using freely accessible libraries from the BaSyx environment [10], such as *AAS Python SDK* [19] or *aas-core-python* [20], as well as AAS-specific libraries [21], which were developed as part of Fluid 4.0 and will be handed over to the BaSyx community after the project is completed. The data-processing Python scripts can be viewed by the user on request and can serve as a template for using the BaSyx libraries in their own applications.
The user specifies under *31 aas_type_id* which machine the workflow is connected to and defines the name of the product under *33 product_name*. In addition, they specify on which AAS repository server the generated product instances are to be stored. In the application example, a repository server of the project partner *TU Dresden* [22] is used for this purpose. At the beginning of the process flow (see Figure 5.1-3), the workflow first authenticates itself with the AAS repository server and receives the AAS submodels AID and AIMC from it (see Figure 5.1-4). After successfully connecting to the MQTT broker (see Figure 5.1- 5), the workflow subscribes to the corresponding topics and starts continuous data recording.

The workflow receives the current production data (see Figure 5.1-6) and stores it temporarily.

5.3. Visualization of production data

Some of the data is processed numerically and forwarded to a 3D visualization of the assembly machine in the workflow dashboard (see Figure 5.1-7). **Figure 5.3** shows this dashboard, which integrates online graphics to display the temporal progression of selected process signals in addition to the 3D visualization.

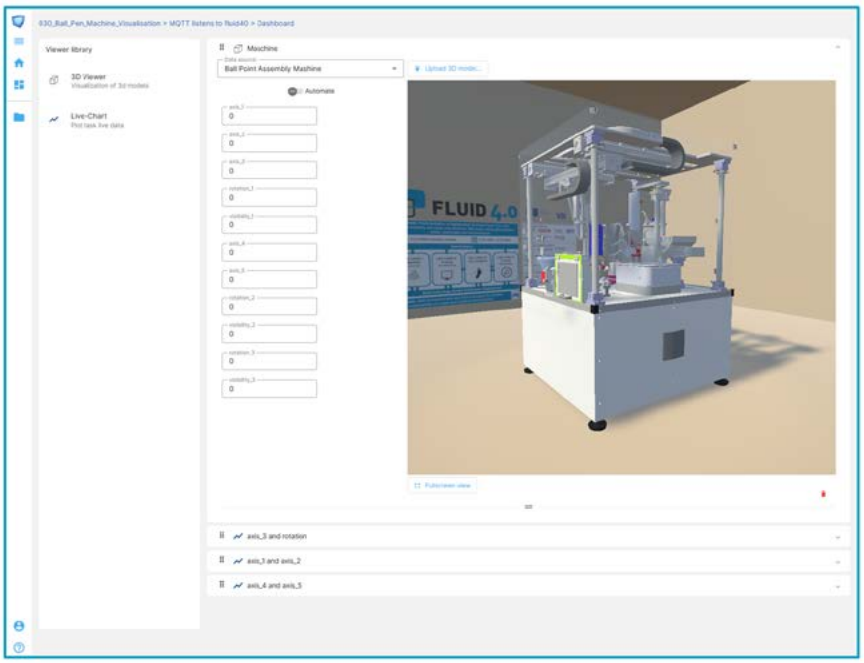


Figure 5.3 : Visualization of the ballpoint pen assembly machine

The visualization can be displayed:

- **live,** parallel to the ongoing production process, or
- **retrospectively** based on the stored time series data.

This creates a digital twin of the machine that visualizes the current or historical machine status regardless of location.

5.4. Energy analysis, CO_2 accounting, and documentation

As soon as the machine reports that the production of a ballpoint pen has been completed, energy analysis, CO_2 accounting, and product-related documentation are initiated in parallel tasks. Based on the recorded production data, the energy consumption per production cycle is determined, distinguishing between

- pneumatic energy (amount of compressed air used) and
- electrical energy (movements of the axes).

This differentiation is particularly important for fluid technical systems, as pneumatic energy is often insufficiently taken into account in classic CO_2 balancing. In the application example, this circumstance is taken into account by using a *conversion factor* that converts the consumed pneumatic energy into an electrical power equivalent for its generation. **Figure 5.4** shows the corresponding input field for this conversion factor in the workflow task. The conversion factor is multiplied by the balanced pneumatic energy; the resulting electrical equivalent is then added to the balanced electrical energy.

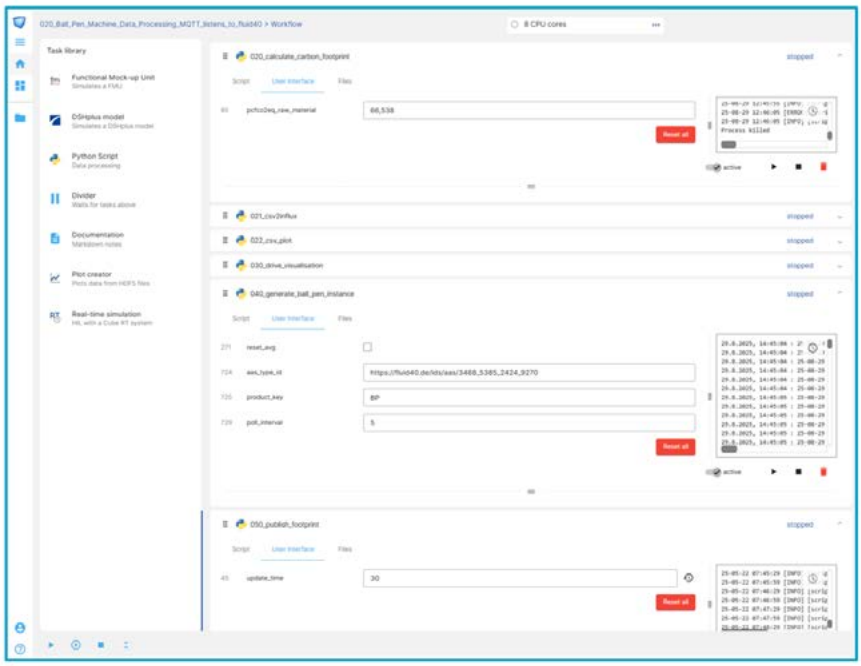


Figure 5.4 : Workflow detail for CO2 accounting

To determine the CO_2 footprint, the CO_2 emission factor of the power grid valid at the respective time of production is automatically queried via external servers (see Figure 5.1-8 and -9). Multiplying this emission factor by the determined electrical energy results in an individual, product-related CO_2 value for each ballpoint pen. This is stored in the *Carbon Footprint* submodel of a newly created product instance AAS for the ballpoint pen.
At the same time, another task generates diagrams (see Figure 5.1-9) that show, for example, the axis positions of the handling system or the development of pneumatic and electrical energy consumption over time. The diagrams are stored on a Nextcloud drive belonging to the project partner Fluidon; the download links generated (see Figure 5.1-11) are stored in the *Handover Documentation* submodel of the product instance AAS.

In addition, the production data is stored long-term in a time series database (InfluxDB) (see Figure 5.1-12). The associated database queries (see Figure 5.1-13) are stored in the *Time Series Data* submodel of the product instance AAS. An AAS-based interface thus enables targeted access to historical production data – for example, based on the serial number or a time period – which supports further analysis even long after production has been completed.

5.5. Uploading the instance AAS of the ballpoint pen

Once the data analysis and submodel creation are complete, a product-related instance AAS for the ballpoint pen is generated in the final process step (see Figure 5.1-14). To do this, the type AAS presented in Figure 4.2 is loaded and instantiated with the results determined in the workflow, creating a concrete product instance AAS. Finally, this is transferred to the previously specified AAS repository server.

6. USER ACCESS TO THE BALLPOINT PEN AAS

The owner of a ballpoint pen can use a web application to retrieve the corresponding product instance AAS by entering the serial number printed on the product. This access provides both product-related sustainability indicators such as the carbon footprint and process-related visualizations and documents. **Figure 6.1** shows an example of user access via the Fluid 4.0 research project website. The image shows an entry from the *Carbon Footprint* submodel.

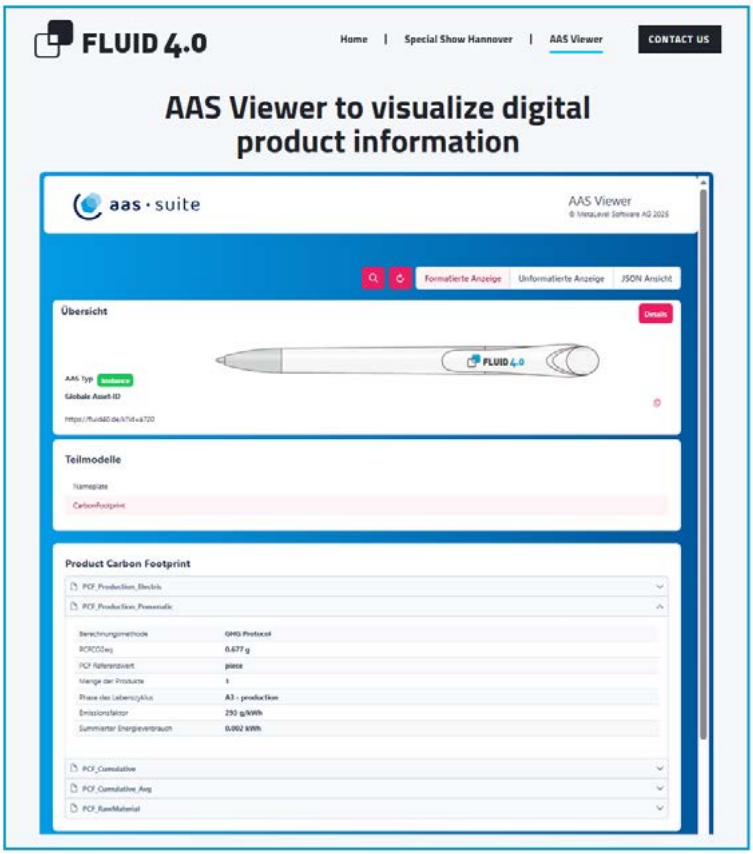


Figure 6.1 : User access to the ballpoint pen AAS

The functionality was implemented using a browser plugin provided by the associated project partner *Metalevel* [23]. This plugin can be integrated into any website with minimal

effort and is connected to any AAS repository server on the backend side by Metalevel. **Figure 6.2** illustrates user access using a sequence diagram. The *client application* represents the browser plugin in the user's web application and the backend at Metalevel.

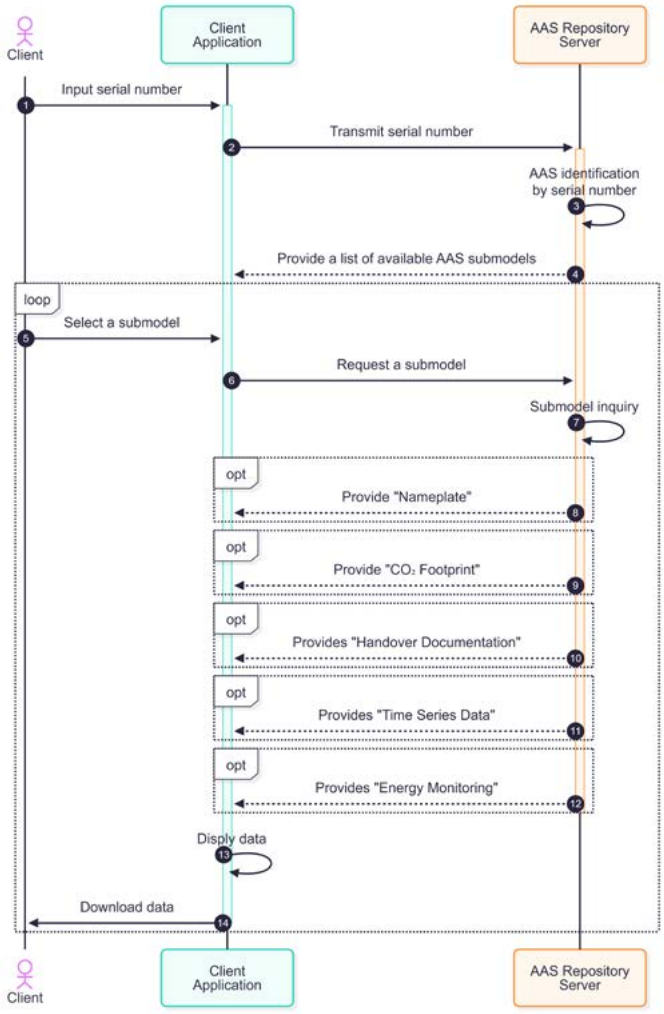


Figure 6.2 : Sequence diagram of user access

The serial number entered by the user (see Figure 6.2-1) is transmitted to the *client application*, which then queries the AAS repository server for the AAS associated with the serial number (see Figure 6.2-2). The AAS repository server determines the appropriate product instance AAS (see Figure 6.2-3) and transmits the metadata required to access the contained submodels to the *client application* (see Figure 6.2-4).
The user selects the relevant submodel in the browser plugin (see Figure 6.2-5), which is then requested by the *client application* from the AAS repository server (see Figure 6.2-6). The AAS repository server determines the requested submodel and transfers its contents to

the browser plugin (see Figure 6.2-7 to 11), where they are visualized (see Figure 6.2-12). The user can then access and download the displayed data. **Figure 6.3** shows an example of a results graph depicting the axis positions of the handling system during the production of a ballpoint pen.

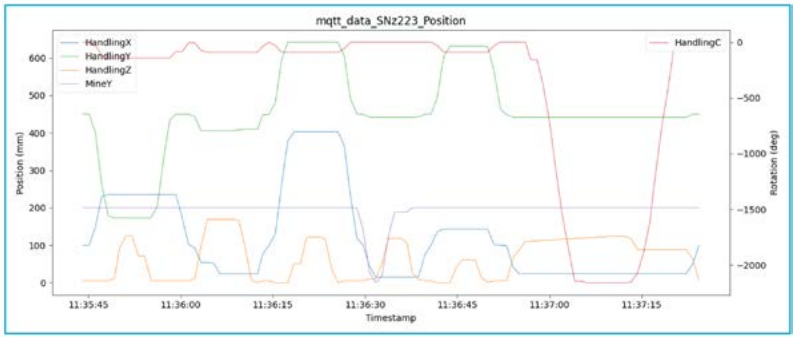


Figure 6.3 : Result diagram from the handover documentation

Furthermore, access to the product AAS is not limited to a single application. Thanks to the standardized interfaces and semantic descriptions of the Asset Administration Shell, the stored information can also be used by any AAS-enabled application. These include, for example, mobile applications for product information, analysis tools for quality or sustainability assessments, and engineering or simulation tools that access historical production data.

Another such use case is access to the time series data referenced in the *Time Series Data* submodel. Authorized users can use the database queries stored there to access specific historical production data, for example, to analyze process deviations, track individual batches, or perform further statistical evaluations.

The separation between data storage, semantic description, and access via standardized AAS mechanisms thus enables flexible and future-proof use of product data throughout the entire lifecycle. New applications or services can be connected to the product AAS without adapting the existing data processing, allowing the AAS to function as a central integration interface for product-related information.

7. SUMMARY

This article uses an industrial demonstrator to show how the Asset Administration Shell can be used as a consistent integration and orchestration layer to automatically collect, process, and provide real machine data for specific products. Using the example of a ballpoint pen assembly machine, it was demonstrated how production data can be linked to cloud-based services via standardized AAS mechanisms and used for visualization, energy analysis, and CO_2 accounting.

The solution presented illustrates in particular the interaction between machine instance AAS and product type or product instance AAS, as well as the role of standardized and project-specific extended AAS submodels for modular data processing. Through AAS-orchestrated workflow execution, central processing steps – from data acquisition and analysis to documentation – are fully automated and decoupled from specific IT or server implementations.

A particular focus was placed on the product-related sustainability assessment of fluid technology systems. The separate accounting of pneumatic and electrical energy and its consolidation into a product-specific carbon footprint shows how AAS-based data structures can be used for transparent and reproducible sustainability assessment.
Finally, it was presented how the generated product instance AAS can be used by different AAS-enabled applications via standardized interfaces. User access via a web application is only one example of a use case. Overall, the example underscores the potential of the Asset Administration Shell as a central basis for interoperable, scalable, and sustainable digital engineering and production processes, especially in the context of fluid power production systems.

8. ACKNOWLEDGEMENT

This work has been funded by the European Union (EU) and supported by the German Federal Ministry for Economic Affairs and Energy (BMWE) in the Fluid4.0 project – Revolutionizing Fluid Engineering: Implementation of digitalization for fluid power 4.0 in the cross-industry and cross-manufacturer data space using asset administration shells, submodels, and demonstrators (project numbers 13IK039E, 13IK039D, and 13IK039H). The views and opinions expressed are solely those of the authors and do not necessarily reflect the views of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.



9. REFERENCES

[1] IDTA, 'Specification of the Asset Administration Shell - Part 1: Metamodel'. Industrial Digital Twin Association, 2025. [Online]. Available: https://industrialdigitaltwin.org/en/content-hub/aasspecifications/specification-of-the-asset-administration-shell-part-1-metamodel-idta-number-01001

[2] 'Industrie 4.0', *Wikipedia*. Dec. 17, 2025. Accessed: Jan. 10, 2026. [Online]. Available: https://de.wikipedia.org/w/index.php?title=Industrie_4.0&oldid=262510167

[3] 'Joint Research Project Fluid 4.0 – Revolutionizing Fluid Engineering: Implementation of digitalization for fluid power 4.0 in the cross-industry and cross-manufacturer data space using asset administration shells, submodels and demonstrators.' Accessed: Jan. 03, 2026. [Online]. Available: https://fluid40.de/

[4] 'AVENTICS | Pneumatische Automatisierungslösungen'. Accessed: Jan. 10, 2026. [Online]. Available: https://discreteautomation.emerson.com/de/brands/aventics

[5] 'MQTT', *Wikipedia*. Dec. 17, 2025. Accessed: Jan. 10, 2026. [Online]. Available: https://de.wikipedia.org/w/index.php?title=MQTT&oldid=262501931

[6] 'FLUIDON | Ihr Experte für Fluidtechnik und Mechatronik', Mainpage. Accessed: Jan. 03, 2026. [Online]. Available: https://fluidon.com/

[7] ‘IDTA – Der Standard für den Digitalen Zwilling’, IDTA. Accessed: Jan. 03, 2026. [Online]. Available: https://industrialdigitaltwin.org/

[8] ‘FLUIDON | Cube’. Accessed: Jan. 03, 2026. [Online]. Available: https://cube-lab.app/

[9] ‘Home < IQstruct Engineering’, IQstruct Engineering. Accessed: Jan. 10, 2026. [Online]. Available: https://iqs-e.de/

[10] ‘BaSyx Wiki — Eclipse BaSyx™’. Accessed: Jan. 11, 2026. [Online]. Available: https://wiki.basyx.org/en/latest/index.html

[11] IDTA, ‘AAS-SM - Asset Interfaces Description’. Industrial Digital Twin Association, Frankfurt am Main, Germany, Jan. 2024. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2024/01/IDTA-02017-1-0_Submodel_Asset-Interfaces-Description.pdf

[12] IDTA, ‘AAS-SM - Asset Interfaces Mapping Configuration’. Industrial Digital Twin Association, Frankfurt am Main, Germany, Jan. 2024. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2024/06/IDTA-02027-1-0_Submodel_AssetInterfacesMappingConfiguration.pdf

[13] IDTA, ‘AAS-SM - Digital Nameplate for Industrial Equipment’. Frankfurt am Main, Germany, Oct. 2025. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2025/10/IDTA-02006-3-0-1_Submodel_Digital-Nameplate.pdf

[14] IDTA, ‘AAS-SM - Handover Documentation’. Frankfurt am Main, Germany, Mar. 2023. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2023/03/IDTA-02004-1-2_Submodel_Handover-Documentation.pdf

[15] IDTA, ‘AAS-SM - Time Series Data’. Frankfurt am Main, Germany, Mar. 2023. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2023/03/IDTA-02008-1-1_Submodel_TimeSeriesData.pdf

[16] IDTA, ‘AAS-SM - Carbon Footprint’. Frankfurt am Main, Germany, Mar. 2025. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2025/03/IDTA-02023_Submodel_CarbonFootprint.pdf

[17] ‘eclipse-basyx/basyx-java-server-sdk’. Accessed: Jan. 27, 2026. [Online]. Available: https://github.com/eclipse-basyx/basyx-java-server-sdk

[18] ‘eclipse-aaspe/server: C# based server for AASX packages’. Accessed: Jan. 27, 2026. [Online]. Available: https://github.com/eclipse-aaspe/server

[19] *eclipse-basyx/basyx-python-sdk*. (Jan. 10, 2026). Python. Eclipse BaSyx™. Accessed: Jan. 11, 2026. [Online]. Available: https://github.com/eclipse-basyx/basyx-python-sdk

[20] ‘aas-core-works’, GitHub. Accessed: Jan. 05, 2026. [Online]. Available: https://github.com/aas-core-works

[21] ‘Fluid4.0’, GitHub. Accessed: Jan. 13, 2026. [Online]. Available: https://github.com/fluid40

[22] work T. ServiceCenterStudium, ‘Technische Universität Dresden’, TU Dresden. Accessed: Jan. 11, 2026. [Online]. Available: https://tu-dresden.de?set_language=de

[23] M.-L. S. AG, ‘Meta-Level Software AG’, Meta-Level Website. Accessed: Jan. 13, 2026. [Online]. Available: https://meta-level.de

Biographies



Micha Koebele received his M.Sc. degree from RWTH Aachen University in 2021. Since 2024, he has been working at FLUIDON GmbH as a development engineer, focusing on dynamic system simulation, model-based engineering, and automated simulation workflows for hydraulic and pneumatic systems. His work particularly addresses the integration of simulation models with digital twin concepts and Asset Administration Shell–based data structures in the context of Industry 4.0.



Heiko Baum studied mechanical engineering at RWTH Aachen University and completed his doctorate in 2001. Since 2001, he has been the managing director of FLUIDON GmbH, where he focuses on model-based engineering, digital twins, and automated simulation workflows for hydraulic and pneumatic systems. His long-standing experience in the simulation and development of fluid power systems bridges industrial applications and research projects in the context of Industry 4.0 and the Asset Administration Shell.



Jan Edler studied mechatronics at the University of Magdeburg and graduated with a Dipl.-Ing. After serving as the construction leader for all customer-specific pneumatic cylinders for Aventics GmbH, he began introducing the Emerson Intelligent Automation Software stack into global factories. His vast knowledge of pneumatic construction, data generation, and analytics converges with his close collaboration with universities and the digital twin initiatives driving the AAS.



Philipp Niemann-Stryczek has studied Computer Science at Trier University of Applied Science after graduating as a State-certified Technician, specialized in Automation Technologies at Aachen Technical College. Before founding and becoming Managing Director of IQstruct Engineering GmbH in 2018 he has been commissioning automobile production lines and developed distributed server backend systems. His passion creating digital twins is widely used within AAS related software and virtual commissioning solutions.

CFD-Based Optimization of a Supersonic Vacuum Ejector Using Combined 2D–3D Simulations and Latin Hypercube Sampling

Cong Liu*, Lennard Günther, Jürgen Weber

Chair of Fluid-Mechatronic Systems (Fluidtronics) – TUD Dresden University of Technology, Helmholtzstraße 7a, 01069 Dresden, Germany

E-Mail: cong.liu@tu-dresden.de

* *Corresponding Author*

Abstract.

This study proposes a combined two-dimensional (2D) and three-dimensional (3D) CFD-based optimization workflow with Latin Hypercube Sampling (LHS) for vacuum ejector nozzles operating in the supersonic flow regime. The reliability of the numerical framework is first established through detailed validation against experimental measurements under both open- and closed-suction operating conditions. The results demonstrate that the adopted CFD setup, boundary conditions, and the RNG k-ε turbulence model can accurately reproduce the measured ejector performance. The numerical simulations further provide detailed insights into the internal supersonic flow structures, including oblique shocks, shock reflections, and downstream shock trains.

Building on this validated CFD framework, the combined 2D–3D optimization workflow is applied, in which a 2D axisymmetric model is coupled with LHS to perform large-scale parametric screening of ejector geometries with enhanced suction performance. The most promising configurations identified from the 2D–LHS optimization are subsequently reconstructed and verified using full 3D simulations. The close agreement in predicted performance improvement between the 2D optimization results and the corresponding 3D validation confirms the reliability of the proposed workflow. Owing to its reduced computational cost and validated predictive accuracy, the methodology provides a practical and efficient design tool for the development of energy-efficient vacuum ejectors and is well suited for industrial applications.

Keywords. Vacuum ejector, Supersonic flow, CFD simulation, Turbulence modeling, RNG k-ε model, Axisymmetric model, Latin Hypercube Sampling (LHS), Optimization

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Vacuum ejectors are widely used for generating negative pressure in industrial automation and material handling, particularly in systems for gripping, holding, and transporting components.

To better understand and quantify the influence of ejector geometry on overall performance, Computational Fluid Dynamics (CFD) has been extensively applied. Several studies have reported that 3D CFD simulations can accurately reproduce the flow phenomena observed in Schlieren visualization experiments, confirming the capability of numerical models to capture shock-wave and pseudo-shock structures within ejector nozzles. Zhu and Jiang [1] demonstrated that experimentally observed shock trains can be accurately predicted by 3D simulations, while Chen et al. [2] confirmed close agreement between numerical reconstruction of the first shock wave and experimental data. Similar investigations on Laval nozzles showed that Large-Eddy Simulation (LES) achieves excellent consistency with Schlieren results, capturing multi-shock pseudo-shock systems and shock–boundary layer interactions [3].

Although 3D CFD simulations can provide highly accurate representations of ejector flow fields, their considerable computational expense makes them impractical for large-scale optimization studies and industrial applications. Consequently, an increasing number of publications have focused on two-dimensional (2D) CFD analyses as a more efficient alternative. Gagan et al. [4] and Schieder [5] demonstrated that 2D axisymmetric simulations can approximate 3D results with reasonable accuracy for most ejector configurations. More recently, Little et al. [6] compared 2D CFD predictions with Schlieren and shadowgraph visualization results for a large-scale air ejector and reported strong agreement between simulation and experimental observations under on-design operating conditions, confirming that 2D CFD can reliably capture key flow features and performance trends. These findings indicate that, while 3D modeling remains indispensable for resolving complex flow interactions, 2D CFD provides a validated and computationally efficient approach for ejector design and performance evaluation.

CFD-based optimization studies have identified several geometric parameters that strongly influence ejector performance. Wang et al. [7] found that the divergent angle and throat length of the primary nozzle are critical to the entrainment ratio, while Ruangtrakoon et al. [8] showed that the throat diameter and area ratio affect both entrainment capability and back pressure. Dong et al. [9] highlighted an optimal range of mixing chamber length for peak performance, and Wang et al. [10] demonstrated that 2D CFD can effectively guide design optimization before 3D validation. Building on this, Yann et al. [11] optimized three key geometries—the constant-pressure mixing chamber length, throat diameter, and constant-area mixing chamber length—under varying nozzle configurations, showing that the coupling between these dimensions and nozzle geometry plays a decisive role in shock structure and energy efficiency.

Building upon these previous efforts, the present study proposes an integrated methodology that combines 2D and 3D CFD simulations with Latin Hypercube Sampling (LHS) for efficient and accurate ejector optimization. The LHS method ensures uniform coverage of the parameter space with far fewer samples compared to conventional random sampling [12]. In this framework, the ejector geometry is parameterized so that each key geometric

parameter can be systematically varied within a predefined range. Using the LHS method, multiple combinations of these structural parameters are generated, resulting in a large set of sample ejector configurations that comprehensively cover the multidimensional design space. The 2D model, established through appropriate geometric simplification, is applied to simulate each sample configuration under identical boundary conditions, allowing rapid evaluation of suction performance. The most promising configurations obtained from the 2D simulations are subsequently reconstructed as full 3D models under identical computational and boundary conditions, to quantify the performance improvement achieved through optimization. This combined CFD–LHS framework enables systematic and accelerated optimization of ejector geometry, providing both high computational efficiency and reliable performance enhancement.

2. MEASUREMENT

The vacuum ejector investigated in this study is the nozzle module of a compact industrial ejector supplied by a leading manufacturer of vacuum-handling technology. A custom-designed adapter was used to mount the ejector on the experimental test bench while preserving the original inlet, suction, and outlet flow paths and ensuring airtight connections.

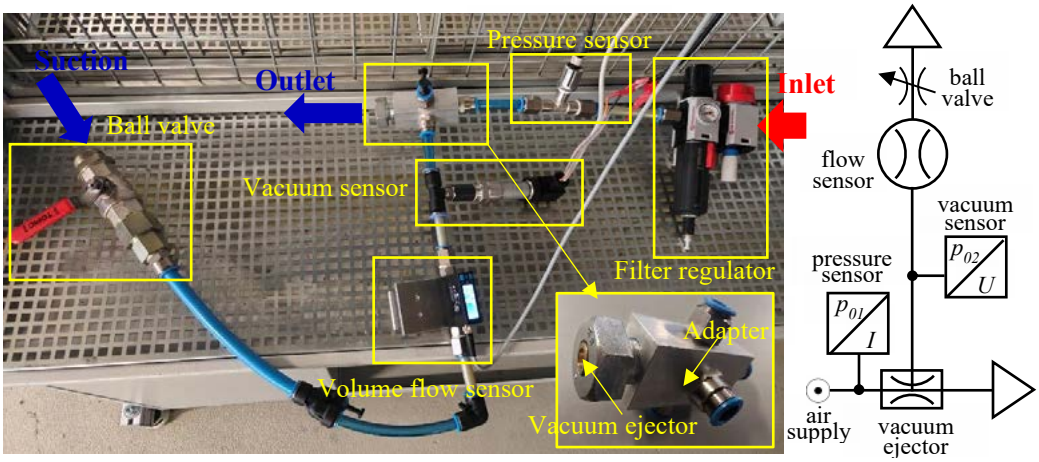


Figure 2.1 Experimental setup for vacuum ejector performance measurement with corresponding schematic diagram.

As shown in Figure 2.1, compressed air from the supply line passes through a filter–regulator unit and a pressure sensor before entering the motive nozzle as the primary flow. The inlet pressure p_{01} is measured upstream of the ejector, while the suction pressure p_{02} is recorded at the suction port using a vacuum sensor. The entrained air flow rate is monitored by a volume-flow sensor installed in the suction line.

A ball valve positioned between the flow sensor and the atmosphere allows adjustment of the suction-port boundary condition. With the valve open, ambient air is entrained into the ejector and the suction pressure can be continuously varied to represent different loading

conditions (**open case**). When the valve is fully closed, the suction port is sealed and no air is entrained, corresponding to zero suction flow (**closed case**). This setup enables systematic characterization of ejector performance over a wide range of suction pressures and operating conditions. The experimental results are presented in Section 4 and compared with numerical simulations to evaluate the predictive accuracy of the CFD model.

3. CFD MODELING

Following the experimental measurements presented in the previous chapter, this section focuses on the numerical modeling of the ejector system. The modeling procedure begins with the governing conservation equations and the applied turbulence modeling approach, which together form the theoretical basis of the simulation framework. Subsequently, the 3D computational model is introduced, followed by a grid-independence study to ensure solution accuracy. To enhance computational efficiency, a 2D axisymmetric model is then developed as a simplified representation of the ejector geometry.

3.1. Fundamentals

3.1.1. Governing equations

The governing equations of the steady compressible flow inside the ejector are expressed in conservative form as follows. The mass conservation equation is given by

$$\frac{\partial(\rho u_i)}{\partial x_i} = 0 \ , \tag{3.1}$$

where ρ denotes the local fluid density, u_i the velocity component in the x_i direction.

The momentum conservation equation is written as

$$\frac{\partial(\rho u_i u_j)}{\partial x_j} = \rho f_i - \frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j}\underbrace{\left[\mu\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right) - \frac{2}{3}\mu\delta_{ij}\frac{\partial u_k}{\partial x_k}\right]}_{\tau_{ij}} \ , \tag{3.2}$$

where p represents the static pressure, f_i the body-force component per unit mass, and μ the dynamic viscosity of the working fluid; subscripts i, j and l denote Cartesian coordinates following the Einstein summation convention; δ_{ij} is the Kronecker delta and τ_{ij} the viscous stress tensor.

The energy conservation equation is expressed as

$$\frac{\partial}{\partial x_j}[u_i(\rho E + p)] = -\frac{\partial}{\partial x_i}\left(\sum_j h_j J_j\right) + S_h \ , \tag{3.3}$$

where E denotes the total energy per unit mass, h_j the specific enthalpy of the j^{th} species, $J_{j,i}$ the diffusive mass flux of species j in the x_i direction, and the S_h the volumetric energy source term accounting for additional heat transfer or dissipation effects.

3.1.2. Turbulence model

Accurate turbulence modeling is fundamental to the numerical prediction of high–Mach-number flows, making the selection of an appropriate turbulence model a critical component of the computational setup. Several studies have examined the applicability of different turbulence models for ejector flow simulations. Bartosiewicz et al. [11] conducted a comparative assessment of six turbulence closures for supersonic ejectors and reported that the RNG k-ε and SST k-ω models yielded the best agreement with experimental pressure distributions. Zhu et al. [13] performed a geometry-optimization study involving three turbulence models and found that the RNG k - ε formulation provided the closest correlation with the measured entrainment ratios. In a subsequent investigation, Zhu and Jiang [1] combined Schlieren visualization with three-dimensional CFD analysis and confirmed that the RNG k-ε model reproduced the shock-wave wavelength and mixing characteristics more accurately than other models. Collectively, these studies demonstrate that the RNG k-ε model delivers reliable and consistent performance in predicting the flow features of single-phase supersonic ejectors. Therefore, this turbulence model is adopted in the present work to describe the turbulent flow field within the vacuum ejector.

For compressible turbulent flows, the instantaneous velocity field is decomposed into mean and fluctuating components using Favre (density-weighted) averaging. Accordingly, the velocity components are expressed as [14]

$$u_i = \tilde{u}_i + u_i'', \quad \tilde{u}_i = \frac{\overline{\rho u_i}}{\bar{\rho}}, \quad \overline{\rho u_i''} = 0 \ , \tag{3.4}$$

where $\tilde{u}_i$ denotes the Favre-averaged velocity and u_i'' represents the corresponding density-weighted fluctuation.

Applying the Favre-averaging procedure to the compressible Navier–Stokes equations yields the Favre-averaged momentum equations, in which additional terms arise due to the turbulent velocity fluctuations. The resulting Favre-averaged momentum equation can be written as:

$$\frac{\partial(\bar{\rho}\tilde{u}_i\tilde{u}_j)}{\partial x_j} = \bar{\rho}\tilde{f}_i - \frac{\partial \bar{p}}{\partial x_i} + \frac{\partial \bar{\tau}_{ij}}{\partial x_j} + \frac{\partial}{\partial x_j}\underbrace{\left(-\bar{\rho}\widetilde{u_i''u_j''}\right)}_{R_{ij}} \tag{3.5}$$

In Eq. (3.5), the term $-\bar{\rho}\widetilde{u_i''u_j''}$ represents the Reynolds stress tensor, which accounts for the momentum transport induced by turbulent fluctuations. Since this term introduces additional unknowns, the system of equations requires closure through an appropriate turbulence model.

In the present study, the Boussinesq eddy-viscosity hypothesis is adopted to relate the Reynolds stresses to the mean strain-rate tensor. Under this assumption, the Reynolds stress tensor is defined as

$$R_{ij} = -\bar{\rho}\widetilde{u_i''u_j''} = \mu_t\left(\frac{\partial \tilde{u}_i}{\partial x_j} + \frac{\partial \tilde{u}_j}{\partial x_i}\right) - \frac{2}{3}\bar{\rho}k\delta_{ij} \ , \tag{3.6}$$

where μ_t is the turbulent viscosity and k the turbulent kinetic energy.

Based on the Reynolds-stress closure in Eq. (3.6), the production of turbulent kinetic energy due to mean velocity gradients, G_k, is defined as the contraction of the Reynolds stresses with the mean velocity gradients and is written as

$$G_k = R_{ij}\frac{\partial \tilde{u}_i}{\partial x_j} \ . \tag{3.6b}$$

To complete the turbulence closure, transport equations for the turbulent kinetic energy k and its dissipation rate ε are introduced. In this work, the Renormalization Group (RNG) k-ε model is employed, as it has been shown to provide reliable predictions for high-speed compressible flows with strong shock–turbulence interactions.

The steady Favre-averaged transport equation for the turbulent kinetic energy k, including compressibility effects, is given by

$$\frac{\partial(\bar{\rho}\tilde{u}_i k)}{\partial x_i} = \frac{\partial}{\partial x_j}\left(\bar{\mu}\frac{\partial k}{\partial x_j} + \frac{\mu_t}{\sigma_k}\frac{\partial k}{\partial x_j}\right) + G_k + G_b - \bar{\rho}\varepsilon - Y_M + S_k \ , \tag{3.7}$$

where G_k denotes represents the production of turbulent kinetic energy due to mean velocity gradients, G_b denotes the production due to buoyancy effects, and S_k is an optional user-defined source term. The term $Y_M = 2\bar{\rho}\epsilon M_t^2$ accounts for the additional dilatational dissipation arising from compressibility, following the model proposed by Sarkar (1991). The turbulent Mach number is defined as $M_t = \sqrt{2k}/a$, where a is the local speed of sound.

The corresponding transport equation for the dissipation rate ε is written as

$$\frac{\partial(\bar{\rho}\tilde{u}_i \varepsilon)}{\partial x_i} = \frac{\partial}{\partial x_j}\left(\bar{\mu}\frac{\partial \varepsilon}{\partial x_j} + \frac{\mu_t}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial x_j}\right) + C_{1\varepsilon}\frac{\varepsilon}{k}(G_k + C_{3\varepsilon}G_b) - C_{2\varepsilon}\bar{\rho}\frac{\varepsilon^2}{k} - R_\varepsilon + S_\varepsilon \tag{3.8}$$

In this equation, $C_{1\varepsilon}$ and $C_{2\varepsilon}$ are model constants, G_k and G_b retain their definitions from the k-equation, and S_ε denotes a user-defined source term. The additional term R_ε is specific to the RNG formulation and represents a correction accounting for the interaction between turbulence dissipation and mean strain rate, which enhances the model's performance in rapidly strained and high-shear flows. All empirical constants adopt their default values as implemented in ANSYS Fluent 2024R2.

$$\sigma_k = 1.0,\ \sigma_\varepsilon = 1.3,\ C_{1\varepsilon} = 1.42,\ C_{2\varepsilon} = 1.68,\ C_\mu = 0.0845$$

3.1.3. Advanced numerical settings for supersonic simulation

Supersonic CFD simulations are prone to numerical instabilities caused by strong shock interactions and steep flow gradients. To ensure stable and convergent solutions, the High-Speed Numerics (**HSN**) scheme and the built-in Solution Steering strategy of the density-based implicit solver in ANSYS Fluent were employed. These approaches enhance numerical robustness in shock-dominated flows and improve convergence behavior [15].

3.2. 3D modeling

To ensure the accuracy of the numerical results, the 3D simulation model was developed based on the internal configuration of the vacuum ejector assembled with the adapter, as illustrated in Figure 3.1. The annular gap volume, originally present in the compact ejector design, was also reproduced in the model to accurately represent the real operating condition, in which suction occurs from a single direction toward the ejector.

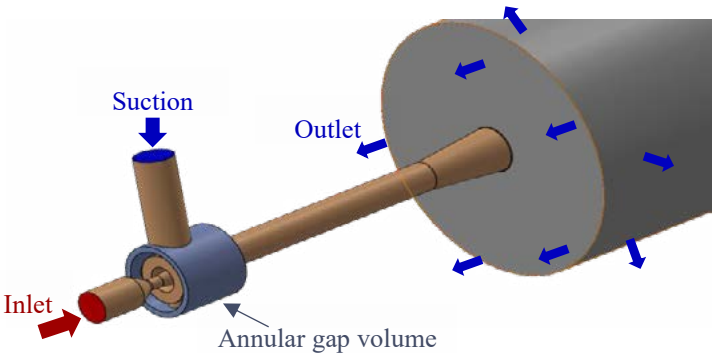


Figure 3.1 3D vacuum ejector model with annular gap volume

The computational domain was discretized using a poly-hexcore mesh generated in Fluent Meshing, as illustrated in Figure 3.2. This hybrid configuration combines structured hex-core regions with polyhedral transition zones, providing a good balance between mesh quality, numerical accuracy, and computational efficiency [16].

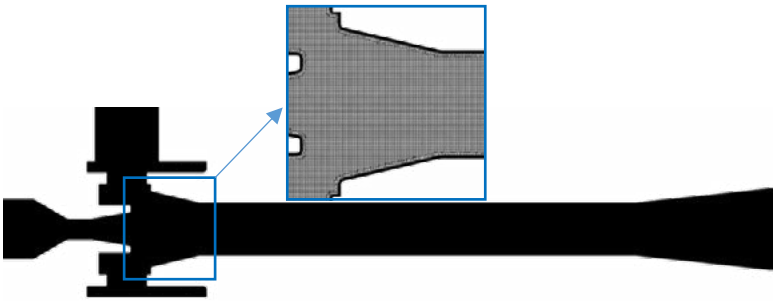


Figure 3.2 3D computational mesh of the vacuum ejector and detailed view

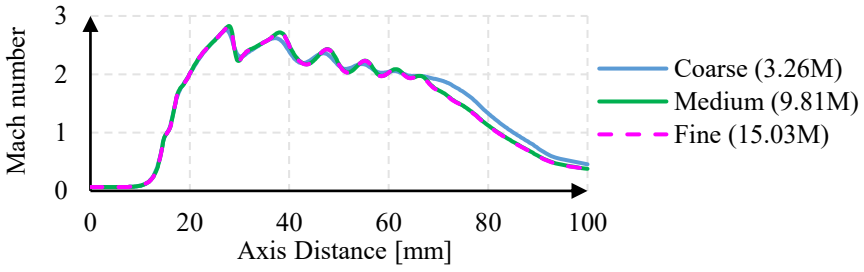


Figure 3.3 Comparison of Mach number distributions along the ejector centerline for different mesh resolutions

A grid independence study with three mesh resolutions was conducted to assess the influence of mesh density on simulation accuracy (Figure 3.3). The coarse, medium, and fine meshes contained approximately 3.26, 9.81, and 15.03 million cells, respectively. While the coarse mesh fails to adequately resolve the shock structures along the ejector centerline, the medium and fine meshes yield nearly identical Mach number distributions. Therefore, the medium mesh was selected for all subsequent simulations as an optimal compromise between numerical accuracy and computational cost.

3.3. Model reduction from 3D to 2D axisymmetric

Although three-dimensional CFD simulations yield highly accurate predictions of the ejector flow field, each case requires approximately five hours to complete on the TU Dresden high-performance computing (HPC) cluster (256 cores, Intel® Xeon® Platinum 8470). Consequently, such simulations are impractical for extensive parametric studies involving multiple geometric variations.

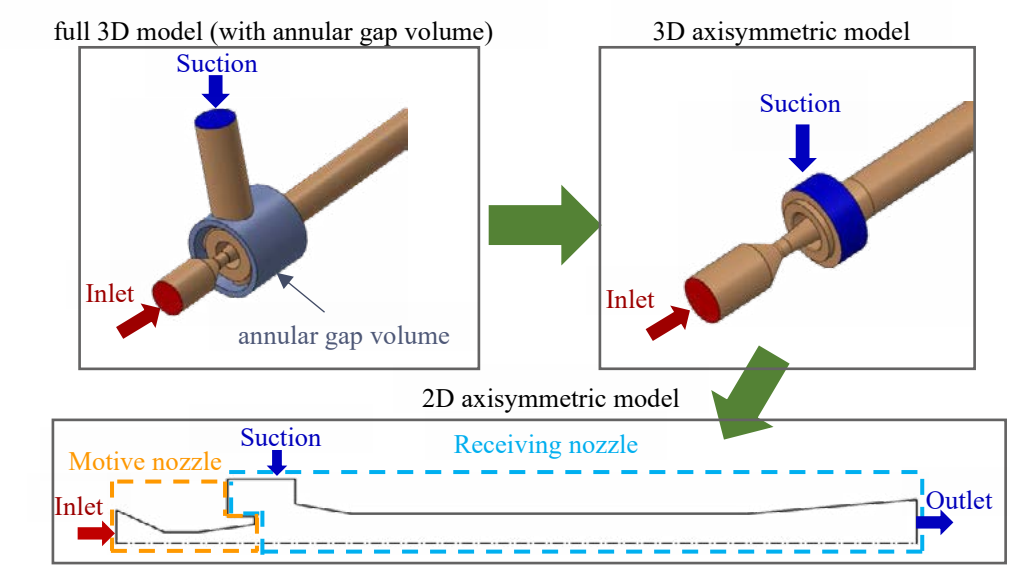


Figure 3.4 Transition from the 3D model to the simplified 2D axisymmetric model

To enhance computational efficiency, the model was progressively simplified, as illustrated in Figure 3.4. In the first step, the 3D ejector with the annular gap volume was reduced to a rotationally symmetric configuration, which preserves the main internal flow structures while omitting external domains. In the second step, this geometry was further simplified to a 2D axisymmetric model for parametric analysis.

In this process, it is important to note that the 2D axisymmetric simulations in ANSYS Fluent are solved in cylindrical coordinates, where all flow variables are expressed as functions of the radial and axial directions. Although the azimuthal (swirl) velocity component is neglected, previous validation work [5] has demonstrated that 2D axisymmetric formulations can still accurately capture the essential supersonic jet

characteristics, including shock-wave structures and entrainment behavior. Therefore, 2D simulations provide a reliable and computationally efficient approach for modeling supersonic ejector flows, particularly during the preliminary optimization phase.

3.4. Boundary condition and numerical setup

To ensure that the numerical model accurately represents the experimental operating conditions, the boundary conditions and numerical settings of the three-dimensional (3D) CFD simulations are summarized in Table 1. Two operating modes were investigated: the open case, in which the inlet pressure was kept constant while the suction pressure was varied to represent different loading conditions, and the closed case, in which the suction port was sealed and defined as a wall boundary while the inlet pressure was varied.

Table 1 Overview of boundary conditions and simulation parameters

		Open case		Closed case	
		Type	Value	Type	Value
Boundary conditions	Inlet	Pressure inlet	4.8 bar	Pressure inlet	2.3…6 bar
	Suction	Pressure inlet	-210…-810 mbar	Wall	–
	Outlet	Pressure outlet	Atmosphere	Pressure outlet	Atmosphere
Numerical setup	Solver	Density-based			
	Formulation	Implicit			
	Turbulence model	RNG k-ε			
	Wall treatment	Standard wall function			
	Material	Air			

4. RESULTS AND DISCUSSION

The boundary conditions and numerical settings summarized in Table 1 were implemented to analyse the flow characteristics and performance of the vacuum ejector under both open- and closed-suction operating conditions. In this section, the predicted internal flow structures obtained from different CFD models are first compared, followed by a quantitative validation against experimental measurements. Based on these results, the applicability of simplified axisymmetric models is discussed.

4.1. Flow-field comparison of different CFD models

Figure 4.1 compares the Mach-number contours obtained from the full 3D model, the 3D axisymmetric model, and the 2D axisymmetric model under identical boundary conditions. All three simulations predict the formation of the first oblique shock at nearly the same axial location, with comparable inclination and strength. This close agreement indicates that the generation of the initial shock is predominantly governed by the motive-nozzle geometry and can be robustly captured by both three-dimensional and axisymmetric formulations.

Downstream of the first shock, noticeable differences emerge between the full 3D model and the two axisymmetric configurations. In the full 3D simulation, the suction flow enters the ejector through a geometrically asymmetric inlet connected to the annular gap volume, resulting in localized three-dimensional flow interactions that influence the subsequent development of the supersonic jet. In contrast, the axisymmetric models employ an idealized annular suction boundary, which leads to a more circumferentially uniform inflow and, consequently, a more symmetric Mach-number distribution within the ejector.

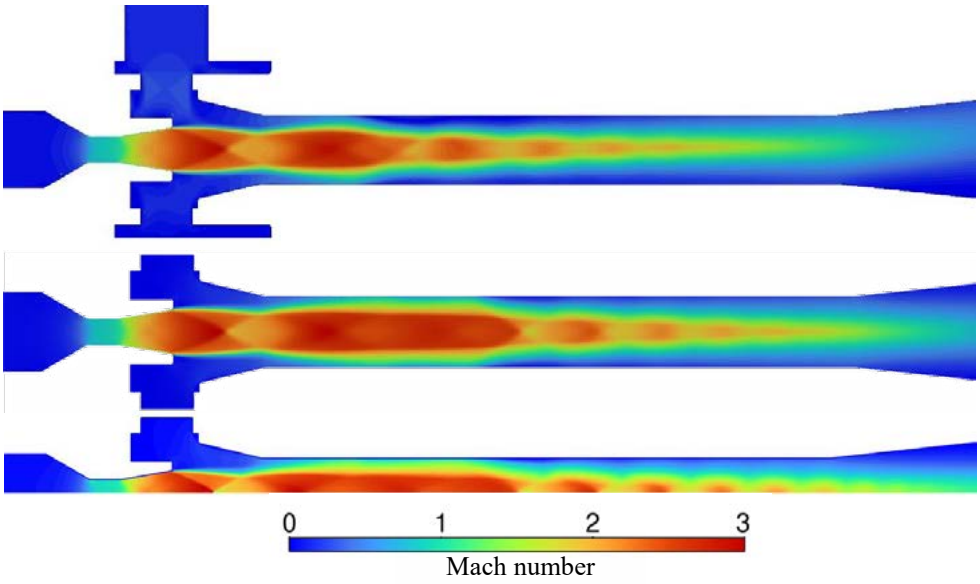


Figure 4.1 Mach-number contours of the full 3D (top), 3D axisymmetric (middle), and 2D axisymmetric (bottom) CFD simulations

When comparing the two axisymmetric simulations, the Mach-number distributions predicted by the 2D and 3D axisymmetric models are found to be qualitatively similar over most of the flow domain. Minor differences are mainly reflected in a smoother and less sharply defined shock train in the 2D solution, which can be attributed to the inherent limitations of the two-dimensional formulation, such as the inability to resolve three-dimensional effects including azimuthal velocity components, asymmetric mixing, and localized shock–turbulence interactions. Nevertheless, the overall expansion–compression pattern, shock spacing, and decay of the supersonic core remain consistent between the two axisymmetric models.

4.2. Quantitative validation against experimental data

The simulation results for both the open- (left) and closed-suction (right) cases within the operating range are presented in Figure 4.2. In this context, the vacuum level is defined as the pressure difference between the ambient atmospheric pressure and the absolute pressure measured near the suction port, i.e., Vacuum $= p_{atm} - p_{abs}$.

For the full 3D model, the predicted suction volumetric flow rates under open-suction conditions show excellent agreement with the experimental measurements. A similarly close correspondence is observed for the vacuum level under closed-suction conditions. This consistency across both operating modes demonstrates the validity of the applied boundary conditions, numerical setup, and turbulence modeling strategy, confirming that the adopted CFD framework can accurately reproduce the ejector performance.

The simplified 3D axisymmetric and 2D axisymmetric models successfully capture the overall performance trends of the ejector but predict slightly higher suction volumetric flow rates under open-suction conditions. As discussed in Section 3.3 and illustrated in Figure 3.4, this deviation arises from the geometric simplification of the suction region. In the axisymmetric configurations, the suction boundary is represented as an idealized annular surface, which effectively enlarges the area available for entrained air compared with the asymmetric suction inlet of the full 3D model. This enlarged entrainment area facilitates a greater inflow of secondary air and consequently leads to an overprediction of the suction flow rate in the open-suction case.

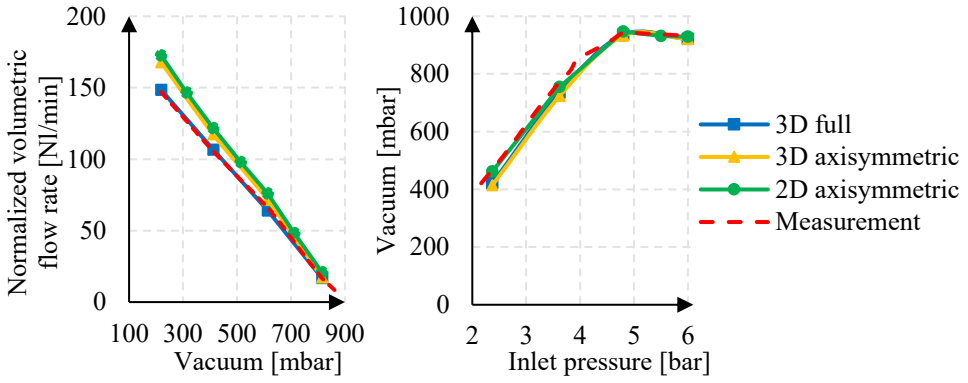


Figure 4.2 Comparison of experimental and simulation results for the open (left) and closed (right) suction cases using different CFD models

Under closed-suction conditions, however, the differences between the three models become negligible. Since the suction port is defined as a wall boundary and no external air is entrained, the internal flow field is no longer influenced by the suction boundary geometry. Instead, the ejector performance is primarily governed by the expansion, shock formation, and recompression processes of the motive jet. As a result, all three modeling approaches yield nearly identical predictions of the vacuum level.

4.3. Discussion and implications for model reduction

Taken together, the qualitative flow-field comparison and the quantitative validation against experimental data demonstrate that the dominant supersonic flow features and overall ejector performance are consistently captured by the axisymmetric formulations. While the full 3D model provides the most detailed representation of the internal flow field, the close agreement between the 2D and 3D axisymmetric results indicates that, once axisymmetry is assumed, the additional dimensional reduction introduces only limited deviations in the predicted flow field.

This finding supports the use of the 2D axisymmetric model for large-scale parametric simulations and preliminary geometry screening, where computational efficiency is essential and the primary interest lies in comparative performance trends rather than in detailed three-dimensional flow structures.

5. EJECTOR OPTIMIZATION WITH LATIN HYPER CUBE SAMPLING

In the present study, the optimization is conducted primarily under open-suction operating conditions, as this case directly reflects the ejector performance relevant to energy efficiency during practical operation. As stated in [17], the geometric parameters of a vacuum ejector strongly interact with one another, and it is therefore more meaningful to vary the key parameters simultaneously rather than individually. The vacuum ejector is parameterized as illustrated in Figure 5.1, and its geometry is defined by several characteristic dimensions that determine the internal flow passages.

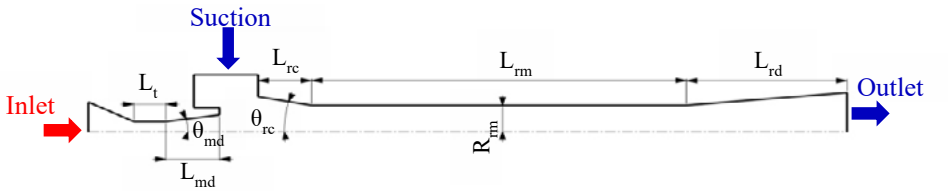


Figure 5.1 Parameterized geometry of the vacuum ejector

Based on previous studies, the throat section of the motive nozzle plays a crucial role in compressing and accelerating the primary flow to Mach number M=1 [18][19]. Variations of the throat radius R_t significantly affect the mass flow rate of the induced secondary stream and, consequently, alter the overall operating range of the ejector. Furthermore, changing R_t would require substantial modifications to both the motive and receiving nozzle geometries, thus reducing the comparability between the optimized and reference geometries. For this reason, R_t was not considered as a design variable in the present optimization. This approach also reflects the practice in commercial compact ejector product lines, where different models are primarily distinguished by their inlet mass-flow capacities rather than by variations in the throat size.

The definitions of the geometric parameters and their respective variation ranges are summarized in Table 2.

Using the LHS method, a total of **1500** sample geometries were generated based on the parameter ranges listed in Table 2. The sample size was selected to ensure sufficient coverage of the high-dimensional parameter space while maintaining a feasible computational cost [20].

Since the primary objective of the present study is the enhancement of energy efficiency, the optimization process was conducted with a clear emphasis on the open-suction operating condition, which accounts for approximately 80% of the overall performance weighting. The closed-suction case was incorporated as a secondary criterion (20%) to ensure that the optimized geometries still maintain an acceptable vacuum level.

In the first selection stage, all geometries exhibiting a higher suction volumetric flow rate than the reference ejector under open-suction conditions were retained as candidate geometries. Because the throat radius R_t was kept constant, the primary mass flow rate at a given operating point remained nearly identical across all simulations, allowing direct comparison of the normalized suction volumetric flow rate without additional correction.

Table 2 Definition and variation ranges of the geometric parameters

Parameter	Description	Variation range
R_t	Throat radius of the motive nozzle	-
L_t	Length of the motive nozzle throat section	-45% ... 750%
L_{md}	Length of diverge section of motive nozzle	-75% ... 150%
θ_{md}	Angle of diverge section of motive nozzle	-85% ... 235%
θ_{rc}	Angle of converged section of receiving nozzle	-85% ... 165%
L_{rc}	Length of converged section of receiving nozzle	-15% ... 470%
R_{rm}	Radius of mixing chamber	-5% ... 25%
L_{rm}	Length of mixing chamber	-35% ... 160%
L_{rd}	Length of diverge section of receiving nozzle	-20% ... 380%

In the second stage, the retained candidate geometries were further evaluated under closed-suction conditions. A weighted performance metric combining the improvements in open- and closed-suction cases was then applied to rank the candidate geometries. Based on this weighted evaluation, two optimized ejector geometries, denoted as **Opti1** and **Opti2**, were selected for detailed comparison.

The simulated performance of the optimized ejector geometries under both open- and closed-suction conditions is presented in Figure 5.2. The relative improvements are evaluated with respect to their corresponding reference configurations: the 2D results are normalized by the baseline 2D axisymmetric model shown in Figure 4.2, while the 3D results are referenced to the baseline full 3D model simulations. This consistent benchmarking strategy ensures that the optimization outcomes are assessed within the same modeling framework and avoids cross-dimensional bias.

In the open-suction case, the 2D axisymmetric and full 3D models exhibit highly consistent improvement trends across the investigated operating range. Both optimized geometries show similar relative gains in suction volumetric flow rate, and the agreement between the 2D-based predictions and their 3D validation results is particularly pronounced. This strong correspondence confirms that the adopted optimization strategy—screening candidate geometries using computationally efficient 2D simulations and subsequently validating the most promising designs with full 3D models—is both reliable and effective. Moreover, it demonstrates that two-dimensional simulations are capable of capturing the dominant performance trends relevant to energy-efficiency enhancement.

Under closed-suction conditions, although the quantitative improvements predicted by the 2D and 3D models differ slightly, the overall trends remain consistent. An important observation from the right-hand plot of Figure 5.2 is that enhanced suction performance under open-suction conditions is accompanied by a reduction in the achievable vacuum level under closed-suction operation. This behavior indicates an inherent trade-off between maximizing energy-efficient suction capacity and attaining high vacuum levels, suggesting that these two objectives cannot be simultaneously optimized within a single ejector geometry.

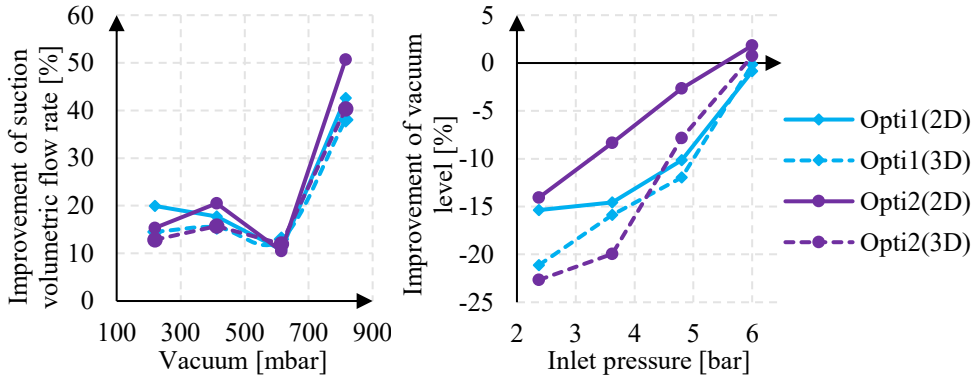


Figure 5.2 Improvement of the optimized ejector geometries under open- (left) and closed-suction (right) conditions

6. CONCLUSION

This study has demonstrated the accuracy and practical feasibility of an integrated CFD-based optimization methodology for vacuum ejector nozzles operating in the supersonic flow regime. By systematically combining experimental measurements with three-dimensional and two-dimensional numerical simulations, the reliability of the adopted CFD setup was established across both open- and closed-suction operating conditions. In particular, the RNG k-ε turbulence model was shown to accurately predict the suction volumetric flow rate in the open-suction case as well as the achievable vacuum level under closed-suction conditions, confirming its suitability for modeling supersonic ejector flows.

A key outcome of this work is the validation of a computationally efficient optimization framework that integrates 2D axisymmetric simulations with Latin Hypercube Sampling

(LHS) for large-scale parameter screening, followed by targeted full 3D model validation. The strong agreement observed between the optimization trends obtained from the 2D simulations and their corresponding 3D validation results demonstrates that the proposed workflow provides a reliable basis for identifying energy-efficient ejector geometries while significantly reducing computational cost. This confirms that simplified axisymmetric models, when properly validated, can be effectively employed in early-stage design and optimization of supersonic flow components.

The optimization results further reveal an inherent trade-off between maximizing suction capacity and achieving high vacuum levels, indicating that these performance objectives cannot be simultaneously optimized within a single ejector configuration. Consequently, the design of high-performance vacuum ejectors requires a balanced consideration of energy efficiency and attainable vacuum level, depending on the intended application.

Future work will extend the present methodology beyond numerical optimization by manufacturing and experimentally testing the optimized geometries to assess their real-world performance. Moreover, the proposed CFD–LHS framework is inherently applicable to a broad class of fluid-dynamic optimization problems in which the geometry can be parameterized and globally optimal configurations are sought. This includes, but is not limited to, ejector designs and other complex flow devices, where robust global exploration of the design space is required under strongly nonlinear flow conditions.

7. REFERENCES

[1] P. Jiang, Z. Zhu, “Experimental and numerical investigation of the effect of shock wave characteristics on the ejector performance,” Energy Conversion and Management, vol. 86, pp. 233–241, 2014.

[2] W. Chen, Y. Hou, J. Zheng, D. Chong, and J. Yan, “Studies on shock wave structures in supersonic ejector and its influence on ejector performance,” Heat Transfer Engineering, vol. 45, no. 7–8, pp. 631–641, 2024. DOI: 10.1080/01457632.2023.2213995

[3] C. Stemmer and S. Hickel, “Large Eddy Simulation of a Pseudo-Shock System within a Laval Nozzle,” Gauss Centre for Supercomputing e.V., Technische Universität München, December 2021. Available: https://www.gauss-centre.eu/results/computational-and-scientific-engineering/large-eddy-simulation-of-a-pseudo-shock-system-within-a-laval-nozzle

[4] B. Gagan, M. Wójcik, W. Nowak, “Comparative study of turbulence models in application to gas ejectors,” Energy Conversion and Management, vol. 77, pp. 137–145, 2014.

[5] M. Schieder, “Characterization and Optimization of Ejectors using Numerical Fluid Dynamics,” Diploma Thesis, TU Wien, Vienna, Austria, Jan. 2023.

[6] A.B. Little, Y. Bartosiewicz, S. Garimella, “Visualization and Validation of Ejector Flow Field With Computational and First-Principles Analysis,” J. Fluids Eng., Vol. 137, No. 5, 2015, Paper 051107. DOI: 10.1115/1.4029534.

[7] L. Wang, J. Yan, C. Wang, X. Li, “Numerical study on optimization of ejector primary nozzle geometries,” International Journal of Refrigeration, vol. 76, pp. 219–229, 2017.

[8] N. Ruangtrakoon, T. Thongtip, S. Aphornratana, T. Sriveerakul, “CFD simulation on the effect of primary nozzle geometries for a steam ejector in refrigeration cycle,” International Journal of Thermal Sciences, vol. 63, pp. 133–145, 2013.
[9] J. Dong, Q. Hu, M. Yu, W. Cui, D. Liang, H. Ma, X. Pan, “Numerical investigation on the influence of mixing chamber length on steam ejector performance,” Applied Thermal Engineering, vol. 174, pp. 115204, 2020.
[10] L. Wang, M. Becker, S. Schneider, “Development of an ejector for passive hydrogen recirculation in PEM fuel cell systems by applying 2D CFD simulation,” International Journal of Hydrogen Energy, vol. 44, no. 32, pp. 17564–17574, 2019.
[11] Y. Bartosiewicz, P. Desevaux, Z. Aidoun, and Y. Mercadier, “CFD–experiments integration in the evaluation of six turbulence models for supersonic ejector modeling,” Proceedings of the International Conference on Ejectors and Applications (ICEA), 2004. https://api.semanticscholar.org/CorpusID:41016682.
[12] E.J. Pebesma, G.B.M. Heuvelink, “Latin hypercube sampling of Gaussian random fields,” Technometrics, Vol. 41, No. 4, pp. 303–312, 1999.
[13] Y. Zhu, W. Cai, C. Wen, Y. Li, Numerical investigation of geometry parameters for design of high performance ejectors, Appl. Therm. Eng. 29 (2009) 898–905.
[14] ANSYS® Fluent Theory Guide, Release 2024 R2, ANSYS, Inc., Canonsburg, PA, USA, 2024.
[15] ANSYS® Fluent User’s Guide, Release 2024 R2, ANSYS, Inc., Canonsburg, PA, USA, 2024.
[16] A. Wróblewski, A. Macek, A. Banasiewicz, and J. Wodecki, “Comparison of Hexcore and Poly-Hexcore computational meshes in the aspect of air flow modeling based on the actual geometry of mining excavations,” IOP Conference Series: Earth and Environmental Science, vol. 1295, 012007, 2024. DOI: 10.1088/1755-1315/1295/1/012007
[17] J. Yan, R. Li, and C. Wang, “Optimization of Three Key Geometries of a Steam Ejector under Varied Primary Nozzle Geometries,” Entropy, vol. 25, no. 1, p. 15, 2023. DOI: 10.3390/e25010015
[18] N. Gebhardt and J. Weber (Eds.), Hydraulik – Fluid-Mechatronik: Grundlagen, Komponenten, Systeme, Messtechnik und virtuelles Engineering, 7th ed., Springer Vieweg, Berlin/Heidelberg, Germany, 2020.
[19] G. N. Abramowitsch, Angewandte Gasdynamik, VEB Verlag Technik, Berlin, 1958.
[20] M. L. Stein, “Large sample properties of simulations using Latin hypercube sampling,” Technometrics, vol. 29, no. 2, pp. 143–151, 1987. DOI: 10.1080/00401706.1987.10488205

8. ACKNOWLEDGEMENT

This research was funded by the Federal Ministry for Economic Affairs and Energy (BMWK), grant number 03EN4087B (EnerVaz: Design and development of innovative pneumatic vacuum generators with improved efficiency and reduced noise emission).

The authors gratefully acknowledge the Center for Information Services and High Performance Computing (ZIH) at Dresden University of Technology (TU Dresden) for providing the computational resources used in this work.

Biographies



Cong Liu received his diploma in Diploma in Mechanical Engineering in the fields of numerical multi-phase flow simulations of cavitation from TU Dresden in 2022. Since 2023 he is working as Research Associate at the Chair of Fluid-Mechatronic Systems (Fluidtronics), Institute of Mechatronic Engineering, Technische Universität Dresden. His research areas include numerical simulation and optimization of pneumatic and hydraulic systems, with a focus on CFD-based analysis and design improvement of flow components in industrial applications.



Lennard Günther received his diploma in Diploma in Mechanical Engineering in the fields of numerical multi-phase flow simulations of cavitation from TU Dresden in 2020. Since 2020 he is working as Research Associate at the Chair of Fluid-Mechatronic Systems (Fluidtronics), Institute of Mechatronic Engineering, Technische Universität Dresden. His research areas include valve development in the field of industrial applications as well as research and numerical investigations of special internal flow phenomena of hydraulic components.



Jürgen Weber has been appointed in 2010 as a University Professor and the Chair of Fluid-Mechatronic Systems as well as the Director of the Institute of Fluid Power at the TUD Dresden University of Technology, Germany, and took on the leadership of Institute of Mechatronic Engineering in 2018. He finished his doctorate in 1991 and was an active Senior Engineer at the former Chair of Hydraulics and Pneumatics until 1997. This was followed by a 13-year industrial phase. Besides his occupation as the Head of the Department Hydraulics and Design Manager for Mobile and Tracked Excavators, starting in 2002, he took on responsibility for the hydraulics in construction machinery at CNH worldwide. From 2006 onwards, he was the Global Head of Architecture for hydraulic drive and control systems, system integration and advance development CNH construction machinery.

Digital fluid technology in the data space: From semantic component search to the development of mobile hydraulic systems using AAS

Paul Lukas Remde[1], Benjamin Beck[1,2], Denis Ritz[2], Jürgen Weber[1,2], Tom Gneuß[3], Nico Braunisch[3], Martin Wollschlaeger[3]

[1] Construction Future Lab gGmbH, benjamin.beck@cflab.de, paul.remde@cflab.de
[2] TU Dresden, Chair of Fluid-Mechatronic Systems, denis.ritz@tu-dresden.de, fluidtronik@mailbox.tu-dresden.de
[3] TU Dresden, Chair of Industrial Communications, tom.gneuss@tu-dresden.de, nico.braunisch@tu-dresden.de, martin.wollschlaeger@tu-dresden.de

Abstract.

The development of hydraulic systems is becoming increasingly complex, especially in mobile machines with a wide range of fluid technology components. The joint research project Fluid 4.0, funded as part of kopa35c and embedded in the objectives of Manufacturing-X, is developing an end-to-end digital approach for the collaborative engineering of fluid power systems. The focus is on the Asset Administration Shell (AAS) as an interoperable digital twin. The article presents solutions for industry-relevant User Stories like “Machine AAS with linkage and physical composition of component AAS”, “Semantic Search within the dataspace to find suitable products” and “Automatic control software parametrization”. Therefore, occurring pain points inside the industry will be highlighted and helpful technologies based on the AAS introduced. This paper provides input to the application of existing and new approaches of the AAS to reduce manual effort and enhance collaborative engineering mainly inside the design process of fluid-mechatronic systems.

Keywords. Fluid 4.0, Digitalisation, Asset Administration Shell, Collaborative Engineering, Dataspace, Submodels, industry-relevant Use Cases

1. INTRODUCTION

The requirements for the development of stationary and mobile machines with hydraulic or pneumatic drive technology are increasing due to targets set by customers and legislation as well as technological developments and product diversity on the market. Particularly in the context of collaborative engineering (across departments and companies), standardised data helps to make efficient development decisions. The Asset Administration Shell (AAS) as a data container with a machine-readable interface and industry-standardised semantics

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

represents a technological option for implementing this data-based decision-making basis. The publication of [1] provides interesting insights about the characteristics and technologies about the AAS with a high number of references to various tools and information about the usage inside the industry and applied technologies. The fluid power industry can already supply many components with digital representation in the form of an AAS. However, the added value only arises when the data is utilised in the development of the machine.

The Fluid 4.0 joint research project is developing solutions for the digitalisation of the following four key areas at the interface of fluid technology components and their use in stationary and mobile machines:

1. Control development of machines
2. Energy monitoring
3. Cross-industry CO2-balancing
4. Circular economy

The content of this publication will take part on the first key area “Control development of machines”, which also includes system layout design. At first in chapter 2, the state of the art and technology will be presented by describing the development of mechatronic systems and interoperable data models. The next chapter follows with the introduction of the main User Stories to support engineering processes based on the AAS. The relevant technologies, that have been used and developed to achieve the desired improvements inside those User Stories will be presented in chapter 4. Chapter 5 shows the usage of those technologies by presenting client applications, applied for the User Stories respectively. Therefore, a simplified example system will be used in addition to a demonstrator, resembling a mobile machine, for further validation. The conference paper will be finished with a conclusion and an outlook in chapter 6.

2. STATE OF THE ART AND TECHNOLOGY

2.1. Development of Mechatronic Systems

As described in VDI Guideline 2206, the development of modern mechatronic systems is subject to a staggered process. Basically, this can be described as a V-model. On the left branch, a system design with rough boundary conditions is repeatedly refined with increasingly fine-grained components until the lowest point of the “V” – the point of greatest detail – is reached. In the development phase – the right branch of the “V” – each hierarchical level is then checked from the lowest level upwards to ensure it fulfils the required function. This involves first testing individual components, subsystems, systems, and finally the entire machine/plant [2]. This so-called macrocycle of the V-Modell is shown as the inner cycle in Figure 2.1.

What also can be seen in the figure is that the inner cycle is layered with more outer cycles, which can be interpreted as different stages of a product, from a (digital) functional mock-up to a (digital) demonstrator, a prototype, a pilot series, and the “final” series product itself. There could be more outer cycles, like for each step of a “model facelift” or development of a new product version release. During the development process, multiple different domains must work together to create a suitable product. Typically, each domain works with its own tools; see [3], which is a major flaw during the design process, because of the limited

possibilities for the exchange of data between the different domains. This problem is often more complex when you look at the whole chain from the different companies that are involved for one product, starting at a “low” level from a TIER-X supplier, which may offer “simple” components like screws or cable connectors, along the path to the TIER II and TIER I suppliers, which offer complex elements and subsystems that the OEM has to integrate into their product.

One common way for exchanging data is PDF files, which often include unstandardized content and are not machine-readable by default. One point at which this circumstance becomes apparent in the development process is the parameterization of a lookup table to describe the often non-linear behaviour of a component during the application software development of a machine. Product data sheets often contain two-dimensional curves describing the control-relevant behaviour of components such as valves and pumps, which provide a good basis for parameterization. However, these curves are in a lot of cases only available as graphical representations of the data. The values for such a lookup table must be recreated manually. This is highly inaccurate and prone to errors. Another problem system developers face is the connection flanges of hydraulic components, which are also represented as a PDF graphic. Those flanges are essential when creating manifolds for complex systems. A remarkable fact is that Automation Studio has a function that has a large database for creating manifolds automatically [4]. Such a high integration of connection interfaces is lacking in many other tools.

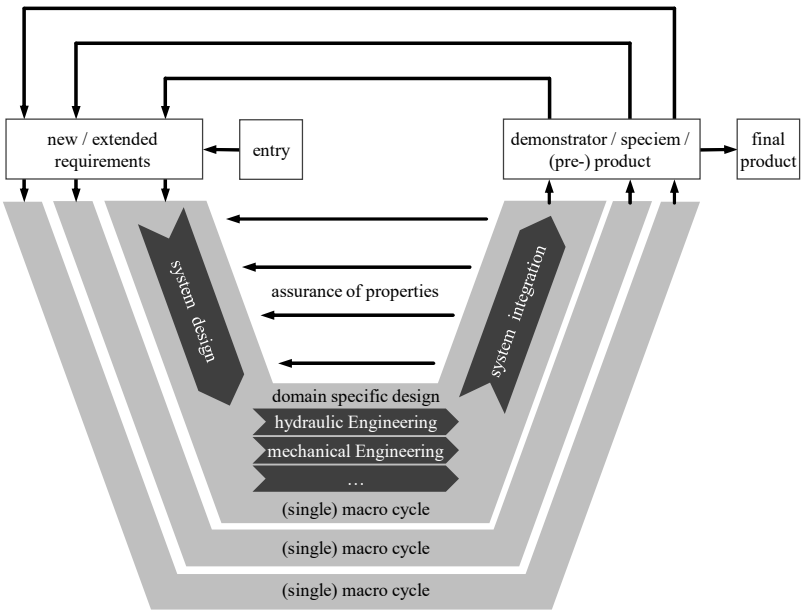


Figure 2.1. Multiple macro-cycles in the V-Model for the product development of a complex mechatronic product according to [2]

When looking at the V-Model again, it is hard to find a program architecture for an application, which is suitable to cover the whole development process and fits all requirements of each system level and domain. To describe a machine with its drive system

and the process, a system simulation tool with lumped parameters can be a good option. The product itself is often represented in a 3D CAD program, which is a geometrical description. Both describe the same product in different ways. But when changing major parameters of the product based on one of those systems, there is just manual feedback and no automatic functions that trigger a redesign. As an example, the systems engineer simulated the lumped model and identified that the system's dynamics are too slow; he decided to use a larger hydraulic cylinder, which offers higher forces. The larger force of the actuator is thus a changed requirement of the frame, which holds the actuator. This may lead to several changes in the mechanical / geometrical design to support the force into the frame of the machine.

It is therefore clear that the development of complex systems such as the hydraulic drive system of machines requires a large amount of data and information that must be exchanged across companies and departments. How can this be done efficiently while avoiding errors?

2.2. Interoperable Data Models

Interoperability can be defined as the ability of two or more systems to exchange and use information seamlessly. [5] In the context of engineering, product and machine data, i.e. layout, interfaces, and connectors, represent a crucial form of information that must be interpreted and applied by a wide range of stakeholders. These include designers and producers, who generate and maintain product data, as well as customers and OEMs, who need to consume the same data in their respective tool environments.

To support this exchange, several standardized formats for engineering data have been developed. Common examples include STEP, AutomationML (AML), and OPC UA, which are reviewed in detail by Ritz [3]. Each of these formats is formally specified and standardized, yet their adoption in practice remains heterogeneous. Not all manufacturers rely on the same data format, and the distribution of data is likewise inconsistent: some producers send engineering files by email after purchase, others make them available for download on websites, and in many cases, product specifications are still provided only as PDFs, which has been discussed above. Furthermore, different formats tend to dominate in different engineering domains, limiting seamless cross-domain data usage [6]. Technologies such as RDF offer higher semantic expressiveness and support reasoning mechanisms, but their adoption in industrial practice has been limited so far [7]. The resulting multitude of formats hinders comparability across components and complicates the composition of systems, a task that OEMs in particular must regularly perform. In contrast, interoperable and vendor-independent data formats have the potential to streamline engineering workflows, enabling consistent data exchange and integration across tools and domains.

Against this background, the AAS has emerged as a particularly promising approach. Developed in context of Plattform Industrie 4.0 and the Industrial Digital Twin Association (IDTA), the AAS provides a standardized digital representation of assets that is semantically enriched and interoperable across domains [8]. Its standardized metamodel, defined in IEC 63278, ensures syntactic interoperability. The AAS organizes asset-specific information into Submodels (SMs), which can be further standardized as Submodel Templates (SMTs) to guarantee the well-defined location of particular data elements. Finally, a standardized application programming interface (API) provides a uniform access

mechanism for reading and utilizing this information across various tools and platforms. Refer to chapter 4.1 for further details on the technical realization of the AAS.

3. USER STORIES

As described in chapter 2, the development process of mobile machines is highly collaborative and data intensive which leads to communication effort and is error prone. During an intensive user story development phase within the Fluid 4.0 project under the support of associated partners the most relevant processes with high potential for data usage and automation were identified. Thereby, three main problems were identified:

1. data on components and the machine are distributed
2. technical properties are rarely semantically standardised and usually not available in machine-readable form
3. there is a diverse tool landscape

Based on this, five user stories have been developed to accommodate this.

User Story I: Machine AAS

Based on the AAS of hydraulic components, the whole drive system of a machine should be described within a machine AAS as well. This would provide a single source of truth for machine-readable information about the system for all developers. The machine AAS should represent the physical and data linkage between the used components and their digital representation. Designers of hydraulic systems should be able to create a layout as the basis for further development steps.

User Story II: Semantic Search

Based on the layout and calculation of the size of the components, the engineers must search for suited products, which is a very manual and time-consuming task today. Hydraulic engineers shall be able to search for fitting products within the dataspace based on standardized technical properties of public information, such as data sheets. This approach has been applied in [9] for the parameterization of simulation models regarding Model Based System Engineering.

User Story III: Tool interfaces and automation

The automated information transfer between different tools shall be possible. Various applications in the engineering process can benefit from standardized data provision and take advantage of automating workflows, which required manual input before. By using freeware and other work-specific software tools, the applied usage of the AAS will be demonstrated and will contribute as an example and motivation for future AAS application integrations.

User Story IV: Simulation

Even though simulation models can be unique and complex, data are required to define the values of the components. Likewise, data from the AAS can be inserted automatically to save time and effort. On top of that, since simulation models often resemble the shape of a hydraulic schematic, a first approach to generate the first instance of a simulation model

automatically can be carried out, which will be presented in additional papers. The applicability of the AAS for the Simulation has been shown in [9], [10].

User Story V: Control development

As basis for virtual commissioning, control relevant parameters of the used components shall be imported automatically into the IDE (integrated development environment). Software developers benefit from fast and safe parametrization of the control units` I/O`s.

Figure 3.1. shows the implementation of these User Stories through five client applications and their interaction via a central AAS – the machine AAS as the single source of truth during the product development. In the focus of this contribution, the clients I, II III and V will be described in detail.

To implement these User Stories with the corresponding Clients a few building blocks and technologies are needed. In [3] a mock-up for a single application was shown in the outlook. This publication focuses on the identified clients, which are described in the following chapter. Basically, there are always three main building blocks:

1. Data modelling
2. Infrastructure and Interfaces
3. Clients and Tools

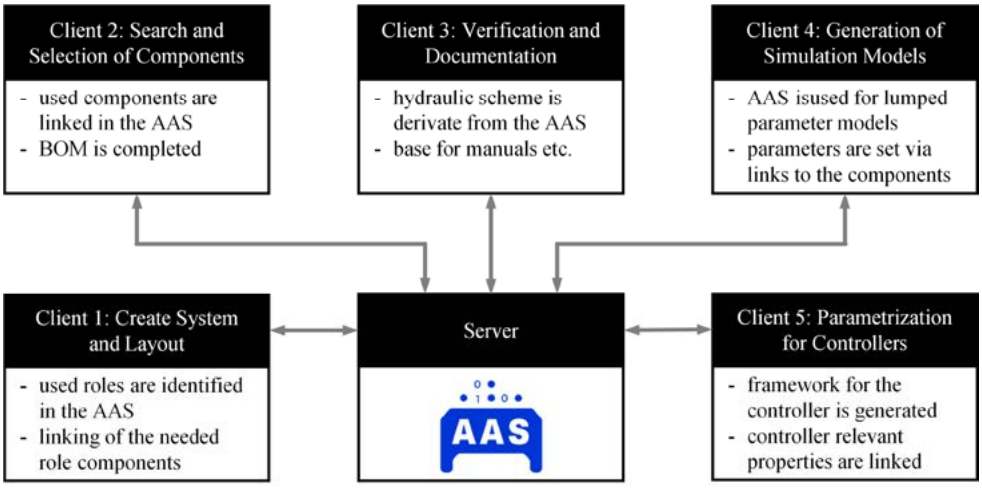


Figure 3.1. Client Applications for the implementation of the defined User Stories

4. INDUSTRY 4.0 TECHNOLOGY

4.1. AAS Infrastructure

The construction of an AAS follows the rules defined in its standardized metamodel [11]. An AAS is composed of SM, each of which contains hierarchically structured SubmodelElements (SMEs). These SMEs can take on various types, such as Properties, Ranges, References, or Collections. The structure and nesting depth of SMEs is highly variable and depends on the specific information that needs to be represented. Different SMEs can be grouped in a SubmodelElementCollection (SMC), which also can include

other SMCs. To provide consistency across use cases, SMTs define a pre-specified structure for SMs and their elements. Such templates are tailored to specific domains or application scenarios, for example the *Digital Nameplate for Industrial Equipment* or the *Generic Frame for Technical Data for Industrial Equipment in Manufacturing* [12]. Refer to chapter 4.3.3 for an in-depth elaboration on the information that these provide. Templates are not developed in isolation but are discussed and refined in dedicated working groups, ensuring cross-industry applicability and acceptance (standardization process).

Two types of AAS infrastructure are distinguished. A Type-1-AAS is exchanged in the form of a serialized file, typically encoded as XML, JSON, or RDF, and often wrapped in an AASX package for convenient bundling [13]. In contrast, a Type-2-AAS is hosted on a server, which provides an application programming interface (API) via HTTP for reading and modifying AAS contents [14]. This includes operations such as changing the value of SMEs, adding or rearranging SMEs, or adding and removing complete AASs and SMs. Beyond basic access, a Type-2-AAS also supports a query language that allows clients to request AASs or SMs from a server that match certain criteria [14]. There exists a vast landscape of open-source AAS applications that implement the AAS metamodel and API specification to varying degrees. Among them, some have just recently included support for the AAS query language.

From a security perspective, the AAS infrastructure already integrates attribute-based access control (ABAC) mechanisms [15]. These mechanisms allow fine-grained regulation of who can access or modify specific AAS contents. However, information signing for ensuring the integrity and authenticity of exchanged data is not yet part of the standard. At the time of writing, there is no publicly available off-the-shelf AAS server application that reliably and completely implements this specification.

For implementing business logic and user stories working with the AAS, the community has created several Software Development Kits (SDKs). The SDKs provide the necessary functions to insert, edit, delete, as well as serialize and deserialize the information inside an AAS. For example, SME can be added to SMs, their values edited, and the result serialized to an exchangeable format like JSON.

4.2. Dataspaces

A dataspace is a decentralized technology for sharing information. Its central focus lies in ensuring data ownership and sovereignty, meaning that data providers retain full control over the conditions under which their data is made available and used [16].

To support interaction among participants, dataspaces typically include centralized services such as a catalogue, which enables the discovery of connector endpoints of other parties, and a wallet, which stores participant certificates. Each participant in a dataspace can act as a data provider, a data consumer, or both, depending on their role in each exchange.

On the provider side, internal rules are defined that govern the conditions under which a consumer can access information. Consumers can issue requests to provider endpoints but must present authentication information; the provider then decides whether the consumer is eligible to receive the requested data. These access rules can go beyond simple eligibility checks and include detailed usage policies, such as restrictions on how often the data may be accessed, how it may be processed, or whether it can be forwarded to third parties. In this

way, policy enforcement mechanisms ensure that data sovereignty is preserved even after access has been granted [17].

The technical realization of these interactions is achieved through dataspace connectors. These connectors must be configured either as providers or consumers, although some implementations support a dual configuration in which a connector can serve both roles. A provider-side connector must be connected to backend resources, such as databases or repositories, to expose the relevant information. At the architectural level, contracting takes place on the control plane, while the actual data exchange occurs on the data plane [18].

Within the Fluid 4.0 key areas and User Stories, each vendor maintains an AAS repository and a SM repository containing AAS (and their linked SMs) for all offered products. At a minimum, each AAS is required to include the SM *Generic Frame for Technical Data for Industrial Equipment in Manufacturing* [12]. One dataspace connector is deployed per vendor, linked to the AAS/SM repositories, and another connector is operated by the customer. The customer issues a query, routed through its connector, to all vendor connectors to retrieve *Technical Data* SM that meet certain criteria – for example, a specific ECLASS number identifying pumps or valves (User Story II). Based on the retrieved SMs, the customer then issues a second query to the AAS repositories (or rather, the dataspace connectors acting on their behalf) to collect all AASs that reference these SMs. The result is a consolidated list of AAS associated with the targeted product category. The whole process and involved components are visualized in Figure 4.1.

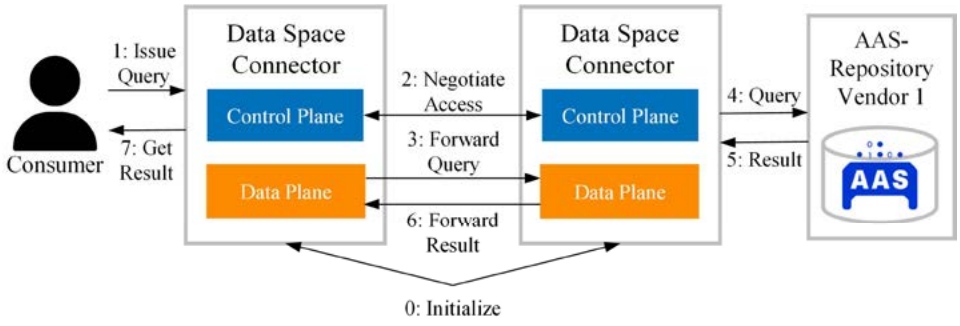


Figure 4.1. Workflow for querying through the Dataspace

4.3. Relevant Submodels for collaborative engineering

SMs are the key elements of an AAS containing domain-specific information in a structured way. Given the well-known location of data according to the SMT, the extraction of information can be carried out automatically, regardless of the manufacturer who populated the information. The diversity in the presentation of product-specific details inside the uniquely designed datasheets of the companies will be bypassed completely.

Up until now, many SMs have been published, e.g., the *Digital Nameplate for Industrial Equipment* [19] or *Handover Documentation* [20]. As the name implies, SMs are mainly designed for specific purposes, like [20], which offers information about the asset's manufacturer and its identification. Beyond that, machine manufacturers can create their own client applications to interact with the machine-readable information of the assets for their individual purposes. In the end, one of the main purposes of the SMTs lies in the

standardized provision of asset-specific information in a machine-readable format. As a result, the engineering process can be designed more efficiently through the new possibilities in automation.

In the process of applying the User Stories by developing the client applications, that has been illustrated in Figure 3.1., required data were missing that couldn't get included in the defined structures of existing SMTs. This includes the information in diagrams, which are only graphically, but not machine-readable provided inside the datasheets of the manufacturers. Those data play a major role inside the control development of User Story V, since software developers need to extract the information manually to handle the control behaviour of hydraulic components accordingly for their unique application. This contributes to the necessity of a SMT to retrieve machine readable information from a graphic in a standardized generic syntax.

Furthermore, to realize the User Stories I, III and IV, a new approach for a SMT has been created to automatically generate a hydraulic schematic. To develop a fluid technological drive system, the hydraulic schematic contains the information about which components are included and how there are connected respectively. The provision of the information in a generic machine-readable syntax of the AAS generates many opportunities for the realization of the User Stories. By applying graphic editors to draw hydraulic schematics, the AAS of the developed system, regarding to Use Case I, can be generated in combination with the graphic information stored inside a SM. This would help to create hydraulic schematics in various data formats for communication and documentation (Use Case III), as well as serving the graphic information as basis for the simulation models of fluid-technological systems (Use Case IV), while interacting with one medium as single source of truth. [9] has been mentioned, that the automatic generation of the structure and topology of a simulation model should be achieved in the future. The content of those new SM approaches will be further discussed in the following chapters 4.3.1 and 4.3.2.

To realise User Story II, existing SMTs need to be applied, like *Generic Frame for Technical Data for Industrial Equipment in Manufacturing* [Data]. This SM contains all necessary information to classify a component and verify its system compatibility by its technical properties. As a result, search algorithms for fitting components can be implemented. A more specific insight of this SMT will be presented in the chapter 4.3.3.

4.3.1. Characteristic Curves

As explained above, the characteristics of hydraulic components and systems are usually characterized by nonlinearities. For this reason, an application was submitted to the IDTA as part of the Fluid 4.0 joint research project to create a new SMT *Characteristic Curves*. The aim is to develop a uniform and universal description of characteristic curves that can be used for the technical properties of a wide variety of products and various applications.

On the one hand, (multidimensional) characteristic curves are often a widely used approach for modelling component behaviour in system simulation. On the other hand, the inverse behaviour of the components is essential for achieving good control behaviour of the respective drive axes. Since the corresponding SM is still under development, no definitive representation can be provided at this point. However, Figure 4.2. Figure 4.2.shows an example of both a characteristic curve mapping $\mathbb{R}^1 \rightarrow \mathbb{R}^1$ and a higher-dimensional

characteristic field representing a mapping of the order $\mathbb{R}^2 \rightarrow \mathbb{R}^1$. This can be extended to any number of dimensions.

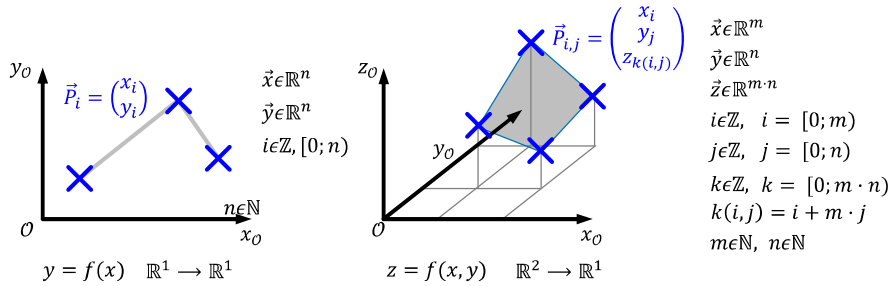


Figure 4.2. Examples for a representation of characteristic curves, according to the planned IDTA SM Characteristic Curves

Generally, it makes sense to create a link to a semantic lexicon, such as ECLASS [21], to build a unique assignment of the contained values to the corresponding variables of the respective components. At the same time, it is planned to store the boundary conditions under which the respective behaviour is mapped, such as the kinematic viscosity υ that was present during a valve measurement.

4.3.2. Model Reconstruction

In the fluid technology industry, the hydraulic schematic is a document of high significance. It contains the integrated components in a visual representation and their connections to one another. It provides the blueprint to achieve the planned movement sequences of the machine via hydraulic components. For engineering tasks, the schematic serves as a fundament for communication, planning and decision making. On top of that, they are also used for maintenance and monitoring of hydraulic systems afterward.

To improve the efficiency of the engineering process, the task occurred to develop a SM, which enables the possibility to reconstruct a hydraulic schematic generically according to User Story I. Schematics are most often available in graphic or software-specific files, especially between employees with different working backgrounds. A standardized template, which includes all necessary parameters of a hydraulic schematic, would offer the opportunity for interaction in a specific work-related environment. This will play a major role in the AAS of hydraulic systems, which will be further discussed in chapter 4.4. Since simulation models and hydraulic schematics are highly related, the information inside the SM can be used for the automated generation of simulation models as well, see User Story IV. This could also be used for the electrical engineering as well with same benefits, ideally using the same BOM as shared data for the systems engineering.

Up until now, the SM consists of two SMCs with the following names: *Model Components* and *Connections*. A *Model Component* is a graphic element inside the schematic, e.g. a pump, valve, or cylinder. It includes specific properties for a unique identification and visual placement inside the schematic. The identification includes the component name, a display designation for tags inside the model, a description for further details, and an element identification to categorize it based on ECLASS [21] standardization. Additionally, a model

component possesses attributes to describe its virtual appearance inside the model. This includes position, size, and orientation, as well as information about the unique positions and IDs of component-specific ports. This approach is visualized in Figure 4.3.

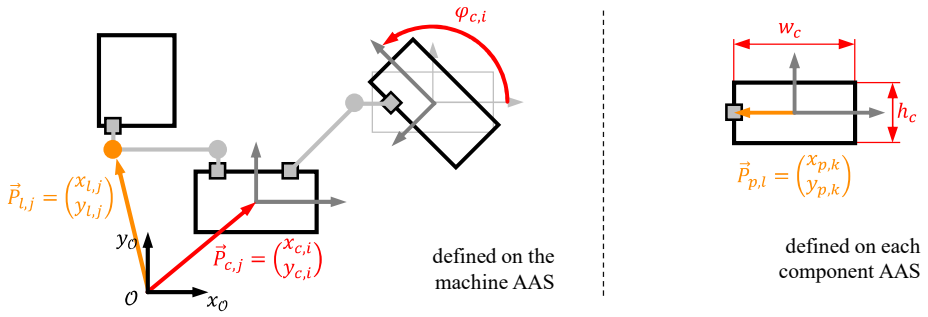


Figure 4.3. Basic approach for Semantic Position Description

Connections is the second SMC inside the SM *Model Reconstruction* and contains the information to describe the linkages between the model components respectively. Inside the SMC is a *Source* and *Destination*, which includes the name of the model component in combination with the specific ID of the port. By determining a source and destination, the direction of flow can additionally be described. To picture the hydraulic schematic more accurately, the SMC *Connections* also include *Points* and *Line*. *Points* determine the coordinates of the key points to describe the path of the line respectively. The coordinates to describe the location of those points share the same reference as the position coordinates of the model components. Additionally, *Line* includes information about the width and shape of the line in first instance.

4.3.3. Generic Frame for Technical Data for Industrial Equipment in Manufacturing

To establish User Story II the SM *Generic Frame for Technical Data for Industrial Equipment in Manufacturing* [12] is used and extended. Besides *Product Classification,* an important property to identify the specific type of hydraulic components, the SM contains most technical data, which used to be found inside the datasheets. From a different point of view, this SM is the provider of machine-readable data, ideally combined with a sematic id based on ECLASS or other dictionaries. From a general view, this SM is the source to retrieve most of the necessary information that is required for automatic product search (User Story II).

This SM includes SMC like *Product Classifications* and *Technical Properties. Product Classifications* associates the asset with product classes in common classifications systems like ECLASS, e.g. an axial piston pump or a hydraulic cylinder. *Technical Properties* includes technical parameters to describe the characteristics of the components respectively. These parameters relate to International Registration Data Identifiers (IRDI) [22], to enable a unique identification of the property. This SM is the basis to implement Use Case II, since the conditions are met to automate the search process for required components, because a model components suitability to be integrated inside the machine can be verified by its product group and technical parameters.

4.4. The AAS for hydraulically-driven machines

Machines with a hydraulic drive system are a composition of hydraulic components according to a defined layout. Even though you could treat the components like pumps or valve as systems themselves, in context of Fluid 4.0, the focus lies on stationary machines, like a hydraulic press, and mobile machines such as wheel loaders or excavators.

According to User Story I, the digital representation of the machine drive system is one fundamental basis for collaborative engineering. From a machine manufacturers perspective data consistency is required from the machine AAS to the component AAS without duplicating data. The physical and logical linkage of component AAS within the machine layout is already part of the SM *Model Reconstruction* in chapter 4.3.2. Besides the idea of a structural reconstruction of hydraulic plans, the SM offers the opportunity for an interaction with the selected system components and their unique AAS by using API. The linkage between specific component AAS and their model-based representation inside the *Model Reconstruction* has yet to be finalized. First approaches are the synergetic application of the SM *Hierarchical Structures enabling Bills of Material* [23], which contains the used components inside the systems and, if available, the reference to their specific AAS.

The combination of both SMs *BOM* and *Model Reconstruction* would enable the possibility to simultaneously provide all required data interactively, since the information have been linked through the *BOM* and an interface for interaction has been provided through the *Model Reconstruction*. A machine AAS, consisting of both SMs, would tremendously support the provision of information, to enable an automated data exchange, the creation and parametrization of simulation models, the efficient exchange of components, and to retrieve component-specific information like the characteristic curves.

5. CLIENTS AND VALIDATION

The User Stories described in chapter 3 will be addressed by the client applications shown in Figure 3.1. This chapter presents an overview of the applications in combination with an assessment of how precisely the requirements have been fulfilled. All client / user applications use the aas-core-works SDKs as basic communication stack with the AAS [24].

5.1. Systems Engineering and Model Reconstruction

The first User Story aims for a digital twin of a hydraulic system that can later be shared and detailed further. Typically, after the objective of the system has been defined and all necessary translational and rotational movement patterns have been detected, the design of a hydraulic system begins with the hydraulic schematic. It includes all necessary actuators for power transmission, e.g., cylinders or hydraulic motors, as well as the required valves and hydraulic lines to control the flow of power, and finally, the providers for hydraulic power, like pumps.

There are various software tools to generate a hydraulic schematic by dragging and dropping the hydraulic components. Usually, there are libraries of commonly used components that can easily be embedded into the own design. Connections between the components via their unique ports are established in the same graphical manner. From that schematic, the initial version of an AAS for the hydraulic system is created. To achieve it, a graph editor program is required to create the library, which contains all the necessary components to draw

hydraulic schematics. In this project, *yEd* [25] has been used, since it is a free tool that can generate various diagrams. Besides the visualization aspect, it is important to embed component parameters inside the library, which can later be retrieved through an XML or GraphML-file. To create the AAS out of the file, a Python program has been developed, which executes the deserialization process by using the aas-core-works SDK [24]. First, the SM *Model Reconstruction* from chapter 4.3.2 will be applied. It contains all the information to recreate the hydraulic schematic as detailed as it was created in *yEd* before. This provides the opportunity for later interactions, like the establishment of references to individual component AAS. In Figure 5.1., a cylinder drive system has been designed inside *yEd* by using the prepared library of elements and later transferred into an AAS by using the implemented Python application.

Figure 5.2. shows an example application of the SM *Model Reconstruction* by recreating the hydraulic schematic inside a vector graphic. To simulate the generation of a hydraulic system, a second Python application has been programmed to identify the parameters inside the SM *Model Reconstruction* and generate a vector graphic (*.svg) out of it. The result is shown on the right side of Figure 5.2. The hydraulic schematic has been created by retrieving information from the AAS. This shows the advantage of an AAS as single source of truth which can be used as harmonization inside multiple tools using different file formats.

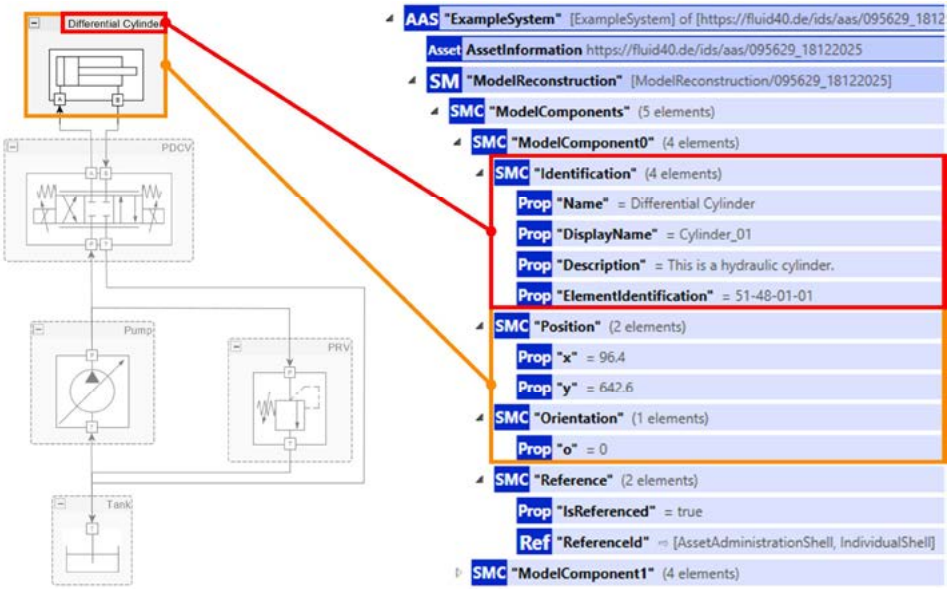


Figure 5.1. Example System of a hydraulic cylinder drive by using yEd (left) and its representation as AAS (right), which was generated automatically by the python script

In chapter 2, the ineffectiveness of the process in finding fitting components for a hydraulic system by manually analysing various datasheets has been highlighted. It contributed to User Story II for a standardized data format, which can be applied for an automized search process. By defining the relevant parameters, the values of various AAS can be analysed and automatically identified as a fitting component. The basic requirement includes the

knowledge of the exact location, where the information will be stored, a fitting identifier, and compliance with the syntactic representation.

To achieve it, the SM *Generic Frame for Technical Data for Industrial Equipment in Manufacturing* will be used to store the individual parameters of the components. By using the IRDI, the technical properties will receive a unique reference for a specific assignment. On one hand, the SM stores the information about the product class of the component by using ECLASS identification. On the other hand, the individual technical parameters for the specific product group will be listed. By using AAS API calls to retrieve the information, the suitability of the component can be assessed through the comparison of the available and required parameters. This process enables automatic search by scanning accessible servers of the component vendors, including AAS of the traded components. Even though the process works as anticipated, difficulties occurred by scanning the server for suitable components. By repetitively requesting the AAS (including *Technical Data*) from the server and checking whether the component matches the desired parameters, the time for computing increased noticeably with a higher number of components. To reduce the client-side computing effort and increase the efficiency of this process, the AAS Query Language [14] has been applied. It assists the search process tremendously, since the server pre-filters the results instead of the client application.

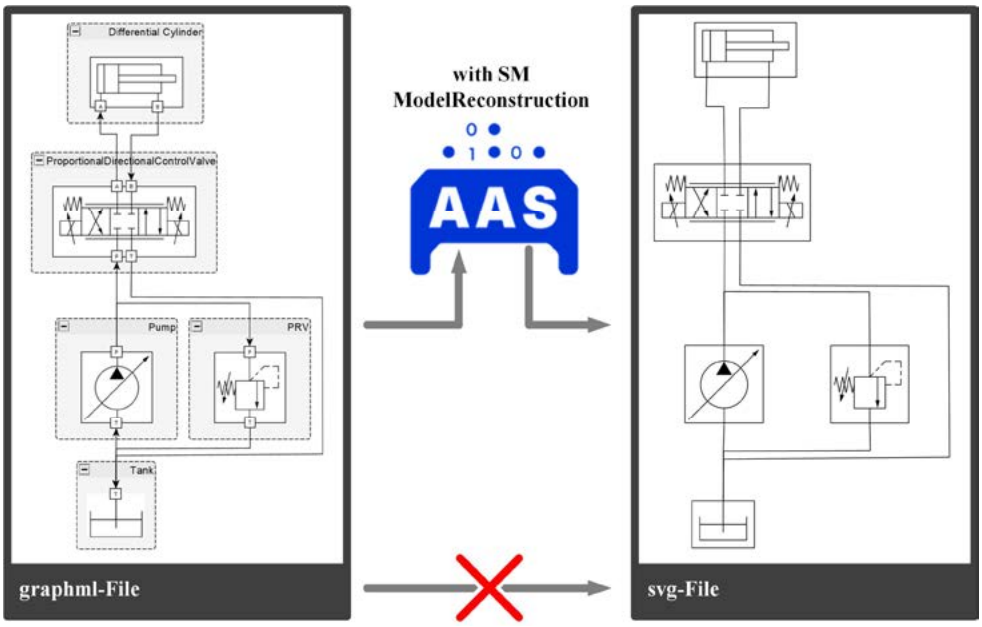


Figure 5.2. Vector graphic of a hydraulic system (right) created by the data of the Submodel Model Reconstruction inside the AAS of the system

A query request requires a payload specifying the constraints of the query. Various definitions can be integrated to specify the necessary properties that the AAS of the components need to provide in combination with the sought values. For example: *"Filter for all component AAS which represents a hydraulic pump with a displacement of 100 cm³ or higher on this server."* By intuitively using logical and comparison operators in the syntax

of the Query Language, this technology can be used to extract the relevant AAS from a large pool of assets. After the relevant components have been identified, more information can be retrieved for downstream purposes. [9] has been also using the query language to filter the available assets to select the fitting components, including their parameters, for the simulation model.

User stories III, IV and V have some strong similarities. In all of them, the machine AAS is used to collect the demand data for specific customers and the connections between all components. A specific advantage of user story V is described in more detail below.

5.2. Control Development for a miniature hydraulic Wheel Loader

The implementation of User Story V is explained using a demonstrator developed in Fluid 4.0. For this purpose, a reduced and simplified version of a hydraulic wheel loader has been designed to represent the advantages of the machine AAS, which links to their component AASs. The demonstrator is shown in Figure 5.3. The main purpose of this demonstrator is to visualize the control development and highlight the advantages of using the AASs in combination with the new SM *Characteristic Curves* (refer to chapter 4.3.1).

To demonstrate this effect interactively, a user interface has been designed as visualized in Figure 5.3. A task was set to insert missing parameters to map a curve of a proportional flow control valve, which has been depicted in red colour inside the hydraulic scheme. This curve describes the amount of flow that passes through the valve, depending on the electrical current applied to the magnet from the control unit. The valve is responsible for the translational speed of the hydraulic cylinder.

As described in chapter 3 before, the values of the lookup table representing the curve need to be retrieved manually from a graphic inside the data sheet today. By applying the new SM *Characteristic Curves*, which provides the data machine-readable in a clearly defined syntax, the required information can be inserted automatically by the application with the click of a button. This leads to the desired movement of the wheel loader kinematics.

To demonstrate the occurrence of errors, a graphic for a different-sized valve has been prepared. Even though there are slight differences between the curves, the movement of the wheel loader changes tremendously. In the end, the demonstrator envisions the advantages of reducing the effort of implementing parameters from graphics and reducing the risk of mistakes. Both ways of parametrization of the control software, the manual way today and the automatically one based on component AAS with SM characteristic curve of the valve, can be compared interactively. The benefits are directly noticeable. The first concepts of this demonstration were shown at the special show during Hannover Messe 2025. The feedback from the visitors were fed back into the development steps of the workflow presented within this contribution.

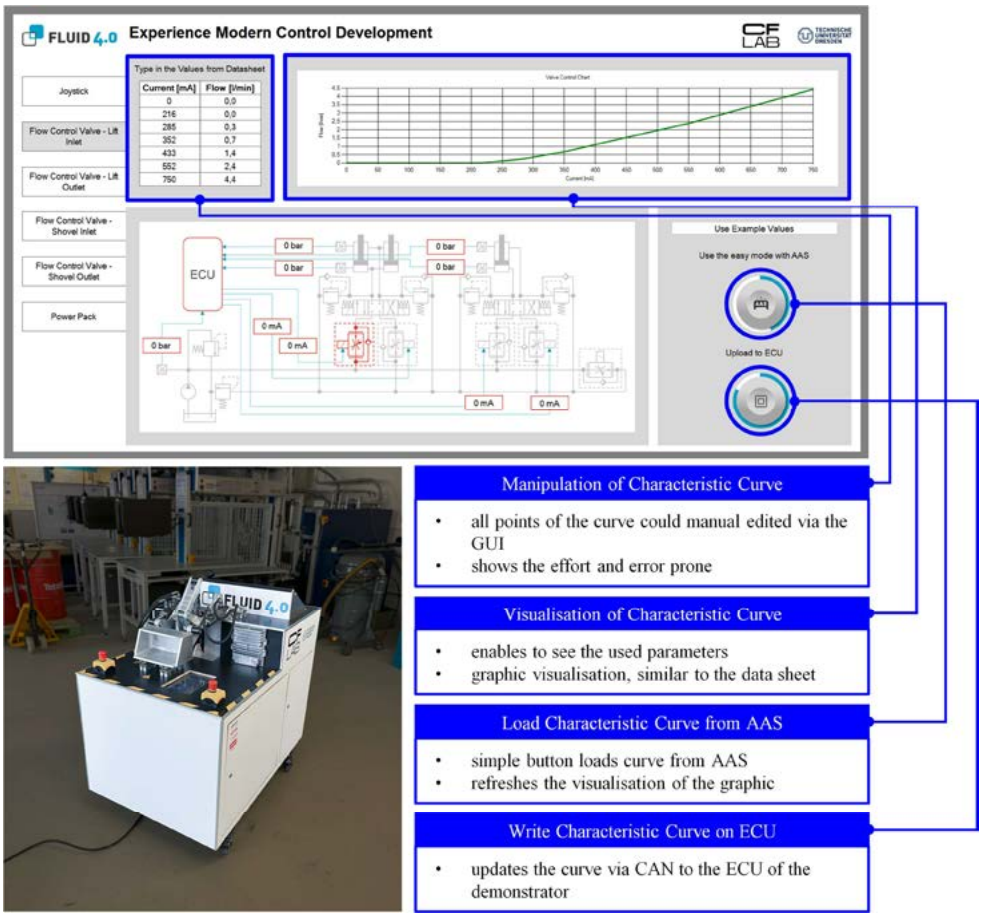


Figure 5.3. Miniature hydraulic wheel loader as demonstrator (bottom left) and interactive User Interface (top) to adjust control software manually or with the SM Characteristic Curve automatically

6. CONCLUSION AND OUTLOOK

This article highlighted the current problems in collaborative engineering of complex mechatronic systems such as hydraulic drive technology in mobile machines. Based on industry-relevant User Stories, an overall concept was developed using the AAS, and a total of five client applications. The essential building blocks for implementing collaborative engineering in fluid technology systems are information modelling, infrastructure, interfaces, and tools for automated signal flow. The administration shell technology, combined with the interaction possibilities offered by established web technologies and interfaces, forms a crucial basis for implementing these building blocks in the context of cross-company communication via dataspaces. The semantic description of an asset's properties using the existing SMs allows the implementation of basic User Stories. However, there has been a lack of specific implementations for User Stories in collaborative

engineering like “Machine AAS with physical based linkage of component AAS”, “Semantic search within the dataspace” or “Automatic control software parametrisation”. With the help of the SMs “*Model Reconstruction*,” “*Generic Frame for Technical Data for Industrial Equipment in Manufacturing*,” and “*Characteristic Curves*” the most important User Stories in fluid engineering system design can be implemented. The first prototypes have already been built to carry this out.

The results of these investigations have already been incorporated into the IDTA standardization processes for partial model development. In the following work, User Stories III and IV will be implemented and the overall process will then be evaluated. The software modules will be made available as open source repositories on GitHub until the end of the project: https://github.com/fluid40

7. ACKNOWLEDGEMENT

This work has been funded by the European Union (EU) and supported by the German Federal Ministry for Economic Affairs and Energy (BMWE) in the project Fluid4.0 – Revolutionizing Fluid Engineering: Implementation of digitalization for fluid power 4.0 in the cross-industry and cross-manufacturer data space using Asset Administration Shells, Submodels and demonstrators (project numbers 13IK039L and 13IK039C). The views and opinions expressed are solely those of the authors and do not necessarily reflect the views of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.



8. REFERENCES

[1] T. A. Abdel-Aty, E. Negri, and S. Galparoli, ‘Asset Administration Shell in Manufacturing: Applications and Relationship with Digital Twin’, presented at the 10th IFAC Converence on Manufacturingand Control MIM 2022, Nantes, France: Elsevier, Jun. 2022.

[2] VDI 2206, *Entwicklungsmethodik für mechatronische Systeme*, Jun. 2004.

[3] D. Ritz, M. Becker, J. Weber, and K. Schmitz, ‘Accelerating Product Development: The Effort of Asset Administration Shells for aggregated Systems’, presented at the The 19th Scandinavian International Conference on Fluid Power, SICFP’25, River Publishers, 2025, pp. 397–409. doi: 10.13052/rp-9788743808251A29.

[4] T. Famic, ‘Live Manifold - Automation Studio™’. Accessed: Oct. 07, 2025. [Online]. Available: https://www.famictech.com/en/Products/Automation-Studio-Live-Manifold/Product-Overview

[5] IEEE, ‘IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries’, *IEEE Std 610*, pp. 1–217, 1991, doi: 10.1109/IEEESTD.1991.106963.

[6] K. Chen and D. Schaefer, ‘MCAD-ECAD integration: Overview and future research perspectives’, vol. 3, Jan. 2007, doi: 10.1115/IMECE2007-41705.

[7] S. Rongen, N. Nikolova, and M. van der Pas, ‘Modelling with AAS and RDF in Industry 4.0’, *Comput. Ind.*, vol. 148, no. C, Jun. 2023, doi: 10.1016/j.compind.2023.103910.

[8] F. M. for E. Affairs and Energy (BMWi), ‘Usage View of Asset Administration Shell’. 2019. [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/2019-usage-view-asset-administration-shell.html

[9] R. Alt *et al.*, ‘SEAMLESS INTEGRATION OF DEVICE AND FIELD DATA INTO THE SYSTEM SIMULATION OF A HYDRAULIC SERVO-PRESS USING AAS’, presented at the 14th International Fluid Power Conference, Dresden, Germany, Mar. 2024.

[10] S. Heppner, T. Miny, S. Kleinert, M. Becker, K. Schmitz, and R. Alt, ‘Asset Administration Shells as Data Layer for Enabling Automated Simulation-based Engineering’, presented at the 28th International Conference on Emerging Technologies and Factory Automation (ETFA), Sinaia, Romania: IEEE, Sep. 2023.

[11] IDTA, ‘Specification of the Asset Administration Shell - Part 1: Metamodel’. Industrial Digital Twin Association, 2025. doi: https://doi.org/10.62628/IDTA.01001-3-1-1.

[12] IDTA, *Generic Frame for Technical Data for Industrial Equipment in Manufacturing*, 02003, Frankfurt Am Main, Germany., Mar. 2025. Accessed: Oct. 07, 2025. [Online]. Available: https://industrialdigitaltwin.org/wp-content/uploads/2025/03/IDTA-02003_Generic-Frame-for-Technical-Data.pdf

[13] IDTA, ‘Specification of the Asset Administration Shell - Part 5: Package File Format (AASX)’. Industrial Digital Twin Association, 2025. doi: https://doi.org/10.62628/IDTA.01005-3-1.

[14] IDTA, ‘Specification of the Asset Administration Shell - Part 2: Application Programming Interfaces’. Industrial Digital Twin Association, 2025. doi: https://doi.org/10.62628/IDTA.01002-3-1-1.

[15] IDTA, ‘Specification of the Asset Administration Shell - Part 4: Security’. Industrial Digital Twin Association, 2025. doi: https://doi.org/10.62628/IDTA.01004-3-0-1.

[16] H. Pettenpohl, M. Spiekermann, and J. R. Both, ‘International Data Spaces in a Nutshell’, in *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*, B. Otto, M. ten Hompel, and S. Wrobel, Eds, Cham: Springer International Publishing, 2022, pp. 29–40. doi: 10.1007/978-3-030-93975-5_3.

[17] S. ; et. al. Steinbuss, ‘Usage Control in the International Data Spaces’, International Data Spaces Association, 2021. doi: https://doi.org/10.5281/zenodo.5675884.

[18] M. Bacco, A. Kocian, S. Chessa, A. Crivello, and P. Barsocchi, ‘What are data spaces? Systematic survey and future outlook’, *Data in Brief*, vol. 57, p. 110969, 2024, doi: https://doi.org/10.1016/j.dib.2024.110969.

[19] IDTA, *Digital Nameplate for Industrial Equipment*, 02006-3–0, Frankfurt Am Main, Germany., Nov. 2024. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2024/11/IDTA-02006-3-0_Submodel_Digital-Nameplate.pdf

[20] IDTA, *Handover Documentation*, 02004-1–2, Frankfurt Am Main, Germany., Mar. 2023. Accessed: Oct. 07, 2025. [Online]. Available: https://industrialdigitaltwin.org/wp-content/uploads/2023/03/IDTA-02004-1-2_Submodel_Handover-Documentation.pdf

[21] ECLASS, 'ECLASS Standard', What is ECLASS? [Online]. Available: https://eclass.eu/en/eclass-standard

[22] ECLASS, 'IRDI'. Accessed: Oct. 10, 2025. [Online]. Available: https://eclass.eu/support/technical-specification/structure-and-elements/irdi

[23] IDTA, *Hierarchical Structures enabling Bills of Material*, 02011-1–1, 06/24. Accessed: Feb. 23, 2025. [Online]. Available: https://industrialdigitaltwin.org/wp-content/uploads/2024/06/IDTA-02011-1-1_Submodel_HierarchicalStructuresEnablingBoM.pdf

[24] M. Ristin, *aas-core*. Python. [Online]. Available: https://github.com/aas-core-works

[25] yworks, 'yEd - graph editor', yEd - graph editor High-quality diagrams made easy. [Online]. Available: https://www.yworks.com/products/yed

Biographies



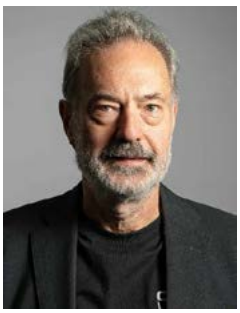
Paul Lukas Remde received the diploma degree in mechanical engineering from TU Dresden in 2024. He is working as scientific research assistant at Construction Future Lab gGmbH since 2024 and is currently a developer inside the Fluid 4.0 project. His main task is the application of I4.0 technologies by developing client applications and demonstrators.



Benjamin Beck received the diploma degree in mechanical engineering from TU Dresden in 2013. He worked as research assistant and group leader for mobile hydraulic systems at the chair of Fluid-Mechatronic Systems at TU Dresden from 2013 to 2023. His research areas include functional safety, energy efficiency and function development of hydraulic drive systems with independent metering for mobile machines. Now he works in addition as strategy and sales lead at Construction Future Lab gGmbH and coordinates the joint research project Fluid 4.0.



Denis Ritz completed his mechanical engineering studies at TU Dresden in 2018 and is currently a research assistant at the Chair of Fluid Mechatronic Systems at TU Dresden. His research centers on control strategies and the development of mobile hydraulic drive systems, with a strong emphasis on systems engineering. Prior to his academic studies, he completed a vocational training as a farmer, providing him with practical insights into mobile machinery.



Jürgen Weber has studied mechanical engineering at the TU Dresden and successfully finished his doctorate in 1991. Until 1997, he was the active senior engineer at the former chair of Hydraulics and Pneumatics. This was followed by an approximately 13-year

industrial phase. He was active in various positions in the R&D department of the agricultural and construction machinery manufacturer CNH. Effective March 1st, 2010, Dr.-Ing. Jürgen Weber has been appointed university professor and chair of Fluid-Mechatronic System (Fluidtronics) at the TU Dresden. Since July 2018 he is the leader of the Institute of Mechatronic Engineering. He is also one of the founders and managing director of the Construction Future Lab gGmbH.



Tom Gneuß received the degree *Master of Science* from the TU Dresden in 2024. Since then, he is working as a research associate at the Chair for Industrial Communications. His main focus lies on the integration of brownfield devices into interoperable Industry 4.0 ecosystems.



Nico Braunisch received his diploma from the TU Dresden in 2019. Since then, he is part of the research group at the Chair for Industrial Communications. During his research, he has gained extensive knowledge in the field of IIoT, Industry 4.0, and automation.



Martin Wollschlaeger studied Electrical Engineering at Otto-von-Guericke University Magdeburg, earning his doctorate in 1991 and Habilitation in 2001. After working at ifak

Magdeburg, he became professor and chair of Industrial Communications at TU Dresden in 2003, where he also directs the Institute of Applied Computer Science. His research focuses on industrial communication and automation networks, including Ethernet-based systems, information modeling, and semantic integration. He contributes to standardization (DKE, IEC) and leads national and European research projects in areas such as Industrie 4.0 and 5G.

Vertical Integration of Energy Data – an Architectural Evaluation of Asset Administration Shell (AAS)

M.Sc. Fabian Geibel[1], Dr-Ing. Karin Tischler[2]

[1]*Bosch Rexroth AG, Ulm, fabian.geibel@boschrexroth.de*

[2]*Bosch Rexroth AG, Lohr am Main, karin.tischler@boschrexroth.de*

Abstract.

Understanding the energy consumption of a process is key to optimize the economic and ecological performance. Today's data integration process still comes with manual effort especially during the initial set-up. The Asset Administration Shell (AAS) is a framework of open standards for normalized, referenced and machine-readable information that enables the exchange and collaboration among different players within a broad ecosystem. This research investigates the technological aspects of the AAS and provides a detailed concept on how to implement an AAS architecture in a practical energy monitoring use case. The test results show that AAS has the potential to reduce the manual effort during the vertical integration of energy data. Nevertheless, AAS needs to be adopted by a broad range of hardware and software suppliers to scale this approach. A reflection regarding the impact of Artificial Intelligence (AI) reveals further potential.

Keywords. Asset Administration Shell (AAS), Energy Data Integration

1. MOTIVATION

The vertical integration of energy data from machines into IT applications is crucial for energy monitoring and reporting. Furthermore, it is the basis for use cases with higher economic and ecological impact like optimizing energy consumption, managing systems, quantifying the efficiency of applied measures and various other applications within industrial settings to achieve an optimized operation.

The integration of energy data is a typical task of machine operators. While the general procedure is well-known it comes with high manual effort. Each machine has different interfaces and individually specified data. There is an continuously increasing number of machine types that can be rather complex and specific. However, the energy consumption should be provided as a meaningful dataset. Currently, each machine builder or sensor

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

supplier provides his own solution(s) that must be integrated. The initial set-up of the communication services includes the search for the right interface description and the extraction of the configuration information. Due to supplier- and device-specific variations of common standards the configuration and mapping has to be repeated for each interface and in many applications for each datapoint again. This work is evaluating how the concept of the Asset Administration Shell (AAS) can address these challenges and supports the vertical integration of energy data.

Once the integration is established, the understanding of the data is key to generate benefit. Therefore, the data itself has to be provided with meta-data e.g., regarding origin, type and quality. Even more value comes with context information, like measurement location and process details or product information. This work also shows how AAS can support the combination of such information.

2. CONTEXT

This research focuses on the evaluation of AAS technology applied to the vertical integration of energy data in the context of established standards. A hydraulics demonstrator of an injection molding machine from Bosch Rexroth [1] serves as an example and the basis for the implementation (Figure 1). The hydraulics demonstrator includes a physical energy meter as well as a database service. The concrete components and the implementation are described in chapter 5.



Figure 1: Hydraulics demonstrator of an injection molding machine – Bosch Rexroth AG [1].

3. AS-IS PROCESS AND CHALLENGES

Performing energy management strongly relies on historic time-series data. A first major step to establish the functionality is the configuration of the communication between the participants. In the context of this research, the data is related to a physical system, a machine or device, referred to as an *Asset*. Assets can contain further Assets e.g., an energy meter for monitoring. Services that allow an interaction with the data are classified as an *Asset service* if located on a physical part of the Asset (e.g., an OPC UA [2] server running on the PLC of the machine), or as *Asset related service* running on another machine or IT system (e.g., an influxDB [3] database running on a separate device).

This research investigates the processes and challenges based on two tasks (Figure 2):

- Task 1: Store real-time data (RT) in a database (DB).
- Task 2: Read time-series data (TS) from a database (DB).

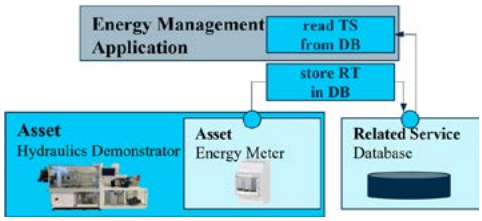


Figure 2: Overview task 1 and task 2. Task 1: Store real-time data (RT) in a database (DB). Task 2: Read time-series data (TS) from a database (DB).

After defining the term "Communication interfaces of participants" in chapter 3.1, the two main tasks and their processes are described in chapter 3.2 and chapter 3.3. Based on today's processes the main challenges are stated in chapter 3.4. A simplified example in chapter 3.5 shows how the configuration effort for the vertical integration of energy data can be assessed.

3.1. Communication interfaces of participants

Each participant (data sources and data sinks) provides a communication interface. The interface follows specific protocols and offers data in a structured way. Descriptions of the interfaces are needed for the configuration of the communication between two participants.

Elements of the interface description are e.g., for an OPC UA [2] communication: the endpoint URL, the node ID or browse path, credentials and data types. For databases e.g., influxDB providing a SQL interface [3], additional information about tables, tags and fields are required. Additionally, further meta-data like units, valid ranges, applicable actions are required for a correct processing of the data.

3.2. Task 1: Store real-time data in a database

The task "Store real-time data in a database" is realized by a service with three computing steps (Figure 3). An initial manual configuration of the service to communicate with the interfaces is necessary. The service then runs in a loop to read from the source, process the data and write the result to the sink.

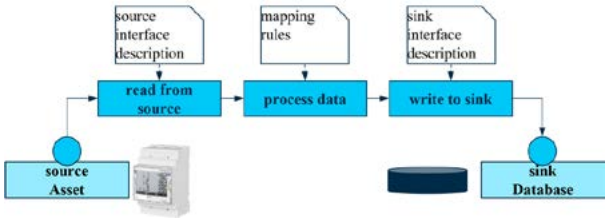


Figure 3: Task 1: Store real-time data (RT) in a database (DB) (dashed line: Initial configuration, continuous line: Running data process).

The overall process is shown in Figure 4. The user searches for the interface descriptions of the data source and the data sink from the providers or integrators of the source system (Asset) and the sink system (database). The user interprets the description and configures the runtime service for each source-sink element pair that needs to be mapped. This includes the set-up of a runtime service to communicate to the interfaces of the source and the sink (chapter 3.1).

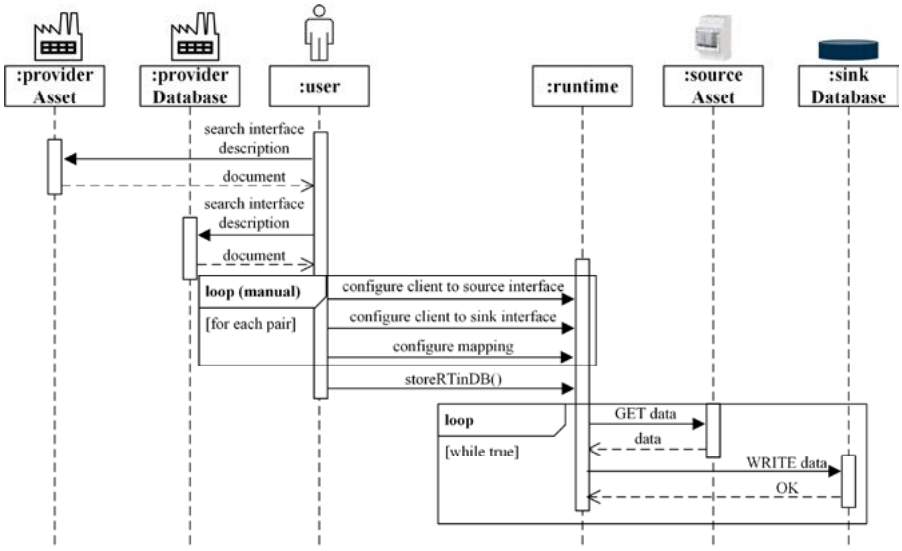


Figure 4: Process to initiate and perform task 1: Store real-time data (RT) in a database (DB).

3.3. Task 2: Read time-series data from a database

The task "Read time-series data from a database" is realized by one computing step, part of a runtime e.g., an energy management application (Figure 5). An initial configuration of the service is necessary to communicate to the interface of the source (chapter 3.1). The application defines how often the service performs the task.

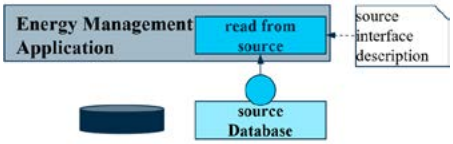


Figure 5: Task 2: Read time-series data (TS) from a database (DB) (dashed line: Initial configuration, continuous line: Running data process).

The overall process is shown in Figure 6. The user searches for the right interface description of the data source at the provider or integrator of the source system. The user interprets the

description and configures the runtime service. This includes the set-up of a client to communicate to the interface of the source (chapter 3.1).

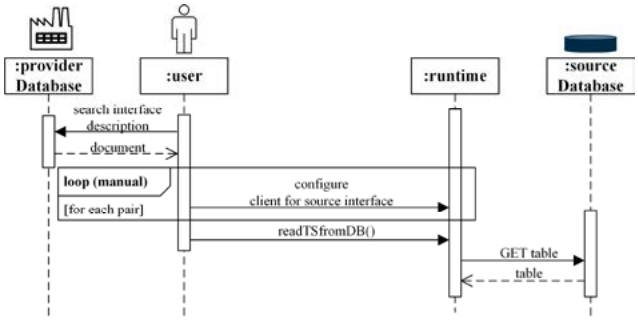


Figure 6: Process to initiate and perform task 2: Read time-series data (TS) from a database (DB).

3.4. Challenges

The manual steps during the integration process of energy data are identified facing the following four challenges:

Challenge 1: Find the right interface description for the Asset.

The interface description usually is not directly linked to the Asset. Typically, the interface description is provided as part of the product documentation. The documentation is provided via different and supplier-specific channels like printed manuals or on the manufacturer's website. In some cases, a detailed interface description can be retrieved from a webserver on the Asset itself. The identification of the information source with the correctly associated file and version might be difficult.

Challenge 2: Use the interface description to configure the service.

The interface description is often provided in a non-machine-readable way, such as tables in different formats or with a vendor specific variation of a standard. The configuration information is manually transferred into the services which is time-consuming and error-prone.

Challenge 3: Decouple and reuse code and configuration.

The configuration of the services that communicate to the interfaces and the mapping of the data between the interfaces is done directly inside the code or with a service specific configuration file. Code and vendor specific configuration information can't be reused easily. Additionally, different protocols need different connector services.

Challenge 4: Understand the data.

Meta-data and context information to understand the data coming from the interfaces, like semantic descriptions, measurement location, parameters and detailed data of the production process or the Asset itself are needed for the right interpretation of the energy data. For example, an increase of energy consumption might be related to a higher demand of the process, different production material, changes in environmental conditions, inappropriate machine configuration or machine wear. Depending on the correct understanding, it is possible to derive suitable measures or classify a change "as expected".

3.5. Evaluation example of the implementation effort

The following example shows how an assessment of the implementation effort for the vertical integration of energy data is possible (Figure 7). The effort must be quantified individually for each use case and set-up. The example consists of several Assets (A) (e.g., machines on a shopfloor), that offer the same amount of data points (D) (e.g., actual power consumption) via the same communication interface standard (e.g., OPC UA). Every data point from every Asset needs to be integrated into several user applications (U) (e.g., for energy management). The description of the communication interface is stated in each Asset's documentation.

The overall configuration effort (c) is based on the number of Assets (a), the number of user applications (u) and the number of data points per Asset (d). The effort is exemplarily defined as $c=a*d*u$. This example visualizes that with a rising number of Assets, data points and user applications, the configuration effort rises. In case of different communication interface standards, additional effort would be necessary.

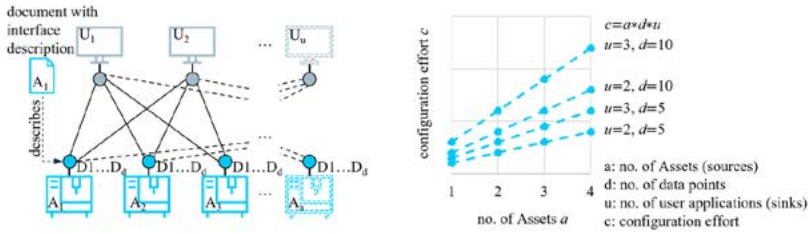


Figure 7: Integration example (left) and configuration effort (right) – providing configuration via documents.

4. AAS TECHNOLOGY TO ADDRESS THE CHALLENGES

AAS technology is characterised by open-source specifications, software development kits (SDK), simple interfaces and a community driving the interoperable data exchange standard. IEC 63278 [4] defines the *Asset Administration Shell* (AAS). The AAS consists of several open-source specifications and components relevant for this research (Figure 8):

- Asset: A virtual or physical product, machine or simple system of interest. The *Asset ID* is the globally unique identifier (ID) of an Asset.

- Asset Administration Shell [AAS]: Digital representation of an Asset. Each AAS is linked to one Asset by the Asset ID and reference to one or many *Submodels* [SM].
- Submodel [SM]: A data structure that is usually based on a *SubmodelTemplate* [SMT]. [SM] exist dynamically in a runtime as an object or statically as a .json or .xml format. Via the *AAS interface*, client applications are able to interact with a [SM] on a server. For this interaction, each [SM] has a unique *Submodel ID*. Each [SM] contains one or many *SubmodelElements* [SME].
- SubmodelElement [SME]: A structured data element. [SME] types relevant for this research are:
  - *SubmodelElementCollection* [SMC]: contains [SME]
  - *SubmodelElementList* [SML]: contains [SME] of the same type and definition
  - *Property* [PROP]: single value data element
  - *ReferenceElement* [REF]: defines a logical reference to another element
  - *RelationshipElement* [REL]: defines a relationship between two elements, contains two [REF] First and Second
- SubmodelTemplate [SMT]: A specified data structure. Each [SMT] has one unique *semantic ID*. For this research the two IDTA Submodel Templates *AssetInterfacesDescription* (AID) and *AssetInterfacesMappingConfiguration* (AIMC) are used.
  - [SM] AssetInterfacesDescription (AID) (IDTA 02017 [5]): It contains the description of the communication interfaces. Besides general information (URL, protocol type), a description of a specific property like the href or browse path, the unit, the data type or min and max values.
  - [SM] AssetInterfacesMappingConfiguration (AIMC) (IDTA 02027 [6]): It contains several [REL], that reference a property from the source [SM] AID to a property of the sink [SM] AID.
- AAS interface: A server hosting [AAS] or [SM] provides a specified REST API. For this research, the [SM] repository API with the following method is used: GET {URL}/submodels/{[SM]ID64}. Where the {URL} is the server address of the [SM] repository and the {[SM]ID64} is the base64 encoded [SM]ID.
- AAS user application: An application that communicates with the AAS or SM via the AAS interface.

[AAS] and [SM] data can be exchanged as a static file, a .json, .xml or within the .aasx package format (AAS Type 1) or, hosted on a server, via an API (AAS Type 2).

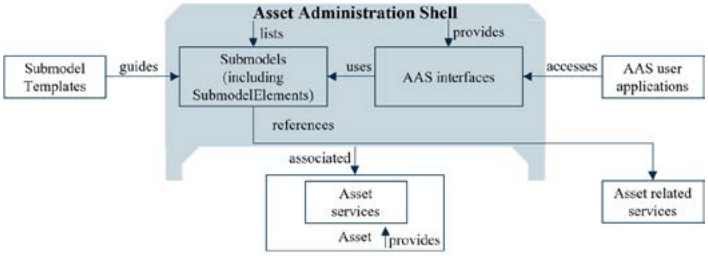


Figure 8: Relevant components of the Asset Administration Shell (based on IEC 63278 [4]).

4.1. Concept to apply AAS

The following concept is developed to apply the AAS technology in regards to the four challenges of the vertical data integration.

Challenge 1: Find the right interface description for the Asset.

The [AAS] is linked via the Asset ID to one Asset and refence the [SM] containing the information describing the Asset. An [SM] based on an [SMT] contains the semantic ID of the [SMT]. The AAS standard includes discovery, registry and repository specifications. Knowing the Asset ID and discovery URL (e.g., from the product type plate following the ID-Link standard (IEC 61406 [10])) as well as the semantic ID of the [SMT] AID, the [SM] AID can be retrieved automatically.

Challenge 2: Use the interface description to configure the service.

The [SM] AID contains all data relevant to configure a service to communicate with an Asset or Asset related interface. For each protocol, the [SM] structure follows a standard. The [SM] AID and its [SME] are accessed in a machine-readable way. Once the logic of the [SMT] is integrated in an application, the configuration process can be automated.

Challenge 3: Decouple and reuse code and configuration.

The communication interfaces can be described via [SM] AID and the mapping of the source-sink relationship via [SM] AIMC. With this concept, the configuration is provided in a neutral, standardised .json format and can be used by several applications.

Challenge 4: Understand the data.

The [AAS] represents with the [SM] different views on the Asset along its lifecycle. AAS technology is capable of structuring information that describes the Asset like provision of documentation or provision simulation models. Another aspect are relationships between Assets e.g., the connection of a valve with a pump. Here the RelationShipElement [REL] and ReferenceElement [REF] of the AAS are used. Furthermore, the context of different perspectives like production data of a part manufactured on a machine can be structured as well. In case of timeseries data, the time stamp is the main reference to synchronize different data sources.

This research is conducted as part of the funded project Fluid 4.0 [7][6]. Components to automate the communication channel between a runtime and an Asset interface are currently under development. An AAS API client [8], as well as a standard parser [9], including a handler for the [SM] AID and [AIMC], have already been published.

5. AAS TEST SET-UP, IMPLEMENTATION AND PROCESS

To evaluate the vertical integration of energy data by using an Asset Administration Shell (AAS) architecture, a test set-up was designed and implemented. The building blocks of the architecture are described in chapter 5.1. The processes to initiate and perform the two tasks "Store real-time data in a database" and "Read time-series data from a database" are described in chapter 5.2.

NOTE: for this research all software components are running on a local Windows 11 machine.

5.1. Building blocks

The AAS architecture consists of three building blocks, "Asset", "AAS Infrastructure", "Runtime" (Figure 9) that are described in more detail below.

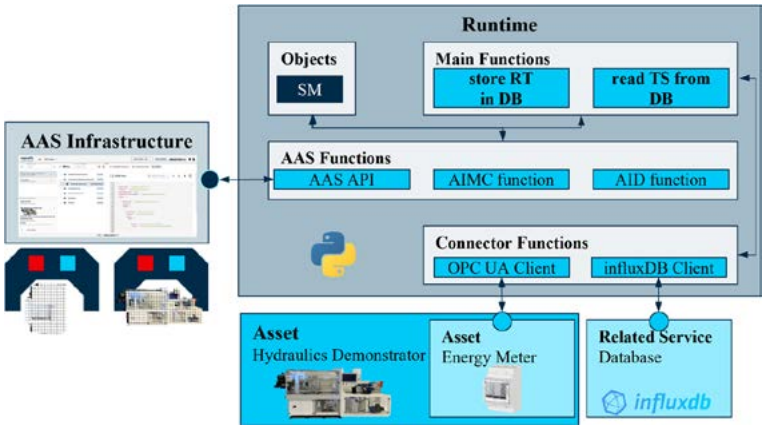


Figure 9: AAS architecture consists of three building blocks: "Asset" (bottom), "AAS Infrastructure" (left), "Runtime" (centre).

Asset

An hydraulics demonstrator of an injection molding machine from Bosch Rexroth [1] is modelled as the main Asset. The hydraulics demonstrator has an energy meter integrated, modelled as another Asset. The energy meter is providing real-time data of voltage, current and power via an OPC UA [2] interface (Figure 9, bottom). The OPC UA server is defined as an Asset service.

For the detailed testing of the AAS implementation, an additional virtual environment was established. In a local python environment, the OPC UA server of the energy meter is running with simulated data.

An InfluxDB 3 Core [3] database is storing the real-time data from the energy meter and providing the time-series data via a SQL interface. The influxDB is not part of the demonstrator and thus an Asset related service.

AAS Infrastructure

The AAS infrastructure is the Eclipse BaSyx Java implementation [11] running inside a local Docker [12] environment (Figure 9, left) The [AAS] are initialised by .aasx package files. There is one [AAS] for the hydraulics demonstrator and one for the energy meter. The AAS infrastructure offers several APIs. For this research, only the following AAS SM repository API method is used: GET {URL}/submodels/{[SM]ID64}. The return is a .json payload.

Design Decision: There is an API method to retrieve a specific [SME] inside the [SM] by a specific path: GET {URL}/submodels/{[SM]ID64}/submodelelements/{PATH}. To keep the number of requests between the runtime and the AAS Server to a minimum, the [SM] itself is retrieved as a single .json payload.

Figure 10 shows the concrete relationship between the Assets, the [AAS] and the [SM]. The [AAS] Hydraulics Demonstrator contains the [SM] AID to describe the influxDB interface. The [AAS] Energy Meter contains the [SM] AID to describe the OPC UA interface. Furthermore, the [AAS] Hydraulic Demonstrator contains the [SM] AIMC that refers to [SMC] inside the two [SM] AID.

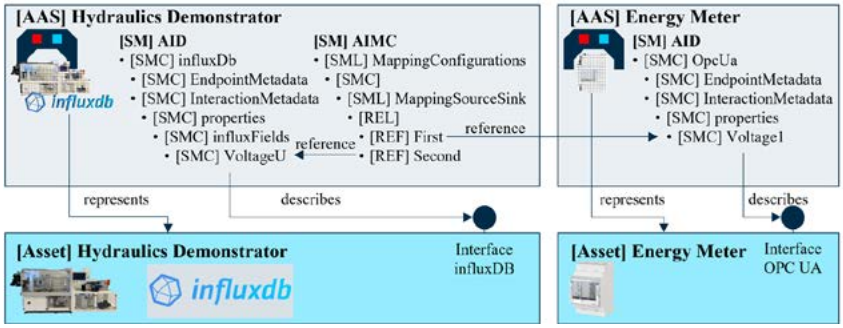


Figure 10: Relationship between Assets, AAS and Submodels.

Runtime

The runtime is implemented with Python. It contains three function modules “Main Functions”, “AAS Functions”, “Connector Functions” and the [SM] object storage (Figure 9, centre).

The functions and the object storage are described in more detail.

Main functions that fulfil the main tasks:

- Store real-time data in a database: this function starts the process to read from the source (OPC UA server) and writes to the sink (influxDB).
- Read time-series data from a database: this function starts the process to read from the source (influxDB).

AAS functions to interact with [AAS] and [SM] repository API and [SM] specific functions.

- AAS API: client to interact with the [AAS] and [SM] repository API. Relevant for this research: GET {URL}/submodels/{[SM]ID64}. The status code and the .json payload is returned.
- AIMC function: process the [SM] AIMC to set up two mapping vectors with information from the source and sink [SM] AID.
- AID function: process the [SM] AID to configure the clients.

Connector functions to set-up communication to the interfaces:

- OPC UA client: service to read data from an OPC UA server.
- influxDB client: service to write to and read from an influxDB.

SM object storage to store [SM] as a python object:

- The .json payload, returned from the [SM] repository server, is transformed via the AAS Core Works Python SDK [13] to a python object. The runtime interacts with this object instead of performing several API calls to the [SM] repository server.

5.2. Processes

NOTE: it is assumed that the suppliers or integrators, that set-up the interface and data structure offered by these interfaces, provide the interface descriptions as [SM] AID, as well as the URL to the corresponding repository and the ID of the [SM].

Task 1: Store real-time data in a database

The overall process to “store real-time data in a database” is described in Figure 11. The user effort for the configuration is limited to define the [SM] AIMC. That incudes to add for each mapping a [REL] element. The [REF] First points to the [SMC] within the [SMC] influxFields containing the description of the source element (e.g., [SMC] VoltageU in Figure 10). The [REF] Second points to the [SMC] within the [SMC] properties containing the description of the sink element (e.g., [SMC] Voltage 1 in Figure 10). To initiate the computing process, the user calls the runtime main function “Store Real-Time Data in Database” with the ID of the [SM] AIMC, the [SML] index of the [SMC] containing the concrete mapping and the URL of the corresponding repository.

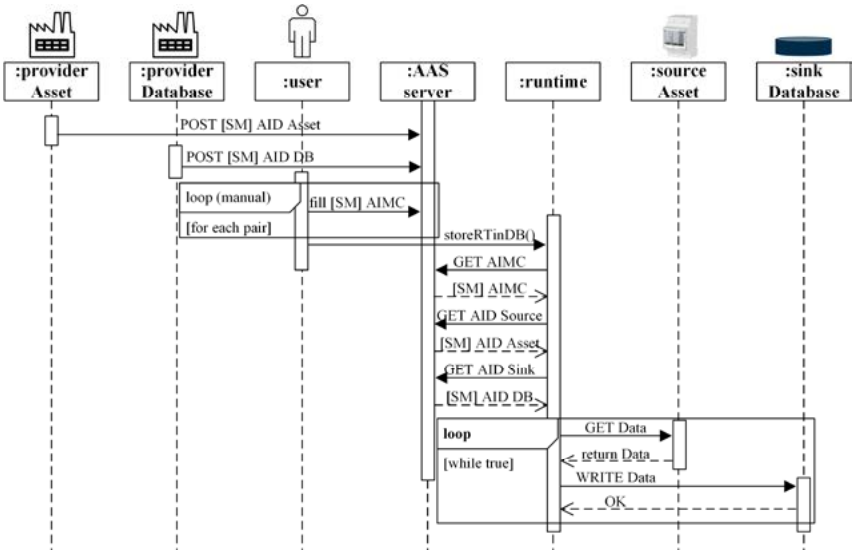


Figure 11: Process to initiate and perform task 1 with AAS: Store real-time data (RT) in a database (DB).

Task 2: Read time-series data from a database

The overall process to "Read time-series data from a database" is described in Figure 12. To initiate the computing process, the user calls the runtime main function "Read Time-Series Data from Database" with the ID of the [SM] AID, the name of the [SMC] for one concrete interface of the Asset (e.g., "OpcUa" in Figure 10) and the URL of the corresponding repository.

NOTE: it is assumed that all properties in the [SMC] for one concrete interface need to be integrated.

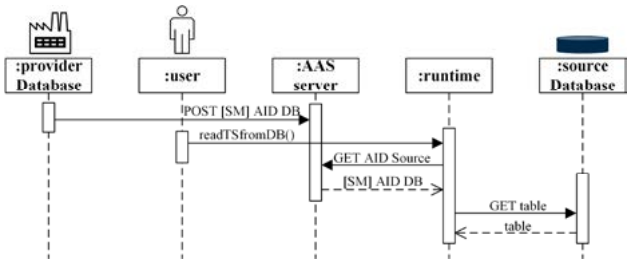


Figure 12: Process to initiate and perform task 2 with AAS: Read time-series data (TS) from a database (DB).

6. EVALUATION

The four challenges (chapter 3) of the vertical integration of energy data are evaluated based on the implemented test set-up of the Asset Administration Shell (AAS) architecture (chapter 5) in chapter 6.1.. It is important to note that the AAS is used in this evaluation to provide the static configuration information. The AAS is not used to handle the dynamic data directly. The data handling is done by runtime clients designed to communicate with the specific interface standards (e.g., OPC UA) directly. In chapter 6.2, the example from chapter 3.5 used again to show the potential of the AAS approach.

6.1. Evaluation of challenges

The evaluation focusses on the manual steps that a user needs to perform for the configuration and the interpretation of a communication interface. For each challenge, the limitations of the implementation are shown as well. The progress in the field of Artificial Intelligence (AI) will lead to new solutions in IoT applications. These opportunities in combination with the developed AAS approach are discussed as further potential.

Challenge 1: Find the right interface description for the Asset.

The process is reduced from searching the interface description in different supplier documents and websites to the search for the [SM] ID and URL of the Submodel repository. Furthermore, when [AAS] discovery and registry mechanisms are implemented, the Asset ID (following the ID-link standard) is sufficient to find and choose the right [SM]. The Asset ID can be encoded e.g., in the QR-code on the Asset.

Further potentials:

To scale this approach, it is necessary that a broad range of vendors create and provide AAS for their Assets. If vendors offer an AAS API, customer applications could retrieve the AAS automatically. As this is not yet common standard the scaling is an open question.

An AI approach could help transforming the data from a vendor specific standard, or a document to the AAS specification. This would reduce the initial effort for each participant. Otherwise, AI might search for the interface descriptions on websites and in documents – even without AAS, but with a risk of misinterpretation and recurring controlling efforts by a user in each case.

Challenge 2: Use the interface description to configure the service.

The process is reduced from taking over the configuration information manually into the application to a fully automatic configuration. This is due to the standardised machine-readable and vendor-independent .json format for the [SM] AID.

Further potentials:

Consuming applications could provide a ready-to-use service that configures the communication to the Asset services automatically based on the [SM] AID and [SM] AIMC.

The AAS specifications can be developed further to describe more different kinds of interfaces.

An AI approach would be suitable to take over the mapping process providing the [SM] AIMC, based on the source and sink [SM] AID.

Challenge 3: Decouple and reuse code and configuration.

Redundant integration effort is reduced significantly as a typical use case set-up and software services can be reused easily. Once the interface descriptions and the mapping are provided as [SM] AID and [SM] AIMC and the software can interpret the standard structure, the complete content can be transferred to different Assets. Changing or upgrading a single Asset within an IT infrastructure is possible with little implementation effort.

Further potentials:

The development of further open-source components for AAS is ongoing [8], [9], [11], [13], [14]. This enables different users and applications to adopt the AAS technology faster in their use cases.

Challenge 4: Understand the data.

The AAS provides easier access to structured information beyond the basic sensor measurement values. Information like units, valid ranges, accuracy, semantic IDs and additional descriptions can be used by different applications and use cases. Client applications are able to retrieve additional information when needed.

The AAS forms a framework in which the information of different domains is structured semantically and handled together. Therefore, the enrichment of a single use case with other perspectives can be conducted easily. The access via an API is also key for the interaction with AI-based applications. To gain deeper insights, information about the Assets, the manufacturing process or the production plan shall be combined.

Further potentials:

Today, [SMT] are able to describe location and relation information (topology) as well as process data and information about the Asset that is currently produced. By this, a big data ecosystem is evolving over time that supports information sharing. A concept how Assets, Submodels, AAS infrastructure and applications shall be structured and interact efficiently is needed, as well as the associated data ownership and privacy. Deriving [SMT] fitting to specific use cases would lead to data redundancy (also to other established standards like OPC UA companion specifications). To keep the framework efficient and comply with the single source of truth principle is still a challenge for cross-vendor interoperability.

6.2. Evaluation example of the implementation effort – AAS compared to document

To demonstrate the different integration effort when the communication interface description is provided as the AAS [SM] AID instead of a document for each Asset, the

integration example from chapter 3.5 is used again. When implementing the logic to process the [SM] AID, not every data point (D) needs to be configured manually anymore. For this reason, the number of data points (d) has no impact on the configuration effort (c). However, there is an initial effort to integrate the logic behind the [SM] AID. Therefore, to provide a meaningful comparison between [SM] AID and the document-based configuration effort, the initial configuration effort for the [SM] AID implementation has a positive offset of c_0.

The overall configuration effort (c) changes from $c=a*d*u$ (document-based interface description), to $c=a*u+c_0$ (AAS [SM] AID based interface description). This example visualizes that the integration effort rises with a higher number of Assets (a) and user applications (u), but more slowly by using the AAS [SM] AID in comparison to use a document-based interface description. Nevertheless, the break-even in the vertical integration of energy data is depending on the use case specific effort for the implementation. When software vendors start to implement the logic behind the AAS [SM] AID, the offset c_0 is assumed to decrease and the break-even is reached earlier.

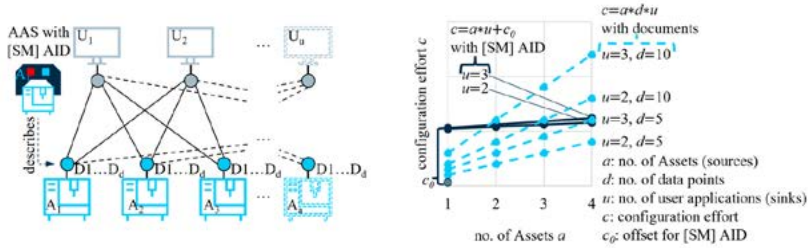


Figure 13: Integration example (left) and configuration effort (right) – providing configuration via [SM] AID (continuous lines) and via documents (dashed lines).

7. SUMMARY

This research addresses the significant manual effort involved in the vertical integration of energy data from diverse industrial Assets, a crucial bottleneck for optimizing economic and ecological performance. We investigate the Asset Administration Shell (AAS) as a framework to standardize and streamline this process. Our methodology involves developing and testing a practical AAS architecture for energy monitoring, focusing on two key tasks: “Store real-time data in a database” and “Read time-series data from a database”. Key findings demonstrate that AAS significantly reduces manual configuration and interpretation efforts by providing machine-readable, vendor-independent interface descriptions and enabling the reuse of code and configuration across different Assets. The AAS framework also facilitates a better understanding of data through structured metadata and contextual information. The main conclusion is that while AAS holds substantial potential to automate energy data integration, its widespread adoption by hardware and software suppliers is essential for scalability. The study also highlights promising future directions, particularly integrating Artificial Intelligence to further enhance configuration and data interpretation within the AAS ecosystem.

8. ACKNOWLEDGEMENT

This work has been funded by the European Union (EU) and supported by the German Federal Ministry for Economic Affairs and Energy (BMWE) in the project Fluid4.0 – Revolutionizing Fluid Engineering: Implementation of digitalization for fluid power 4.0 in the cross-industry and cross-manufacturer data space using asset administration shells, submodels and demonstrators (project number 13IK039L). The views and opinions expressed are solely those of the authors and do not necessarily reflect the views of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.



9. REFERENCES

[1] Bosch Rexroth AG, 'Virtual World – injection molding machine', accessed Dec. 2025 on home page
[2] OPC Foundation, 'OPC Unified Architecture - Part 1: Overview and Concepts, OPC 10000-1', Version 1.05.04, Nov. 2024
[3] InfluxData Inc, 'influxDB 3 Core – Documentation', accessed Sep. 2025 on vendor home page
[4] International Electrotechnical Commission (IEC), 'IEC 63278-1 – Asset Administration Shell for industrial applications – Part 1: Asset Administration Shell structure', edition 1.0, Dec. 2023
[5] Industrial Digital Twin Association (IDTA), 'IDTA 02017 – Asset Interfaces Description', Version 1-0, Jan. 2024
[6] Industrial Digital Twin Association (IDTA), 'IDTA 02027 – Asset Interfaces Mapping Configuration', Version 1-0, Jun. 2024
[7] Fluid 4.0, 'Project – overview, accessed Dec. 2025 on home page
[8] D. Klein – :em engineering methods AG, 'AAS HTTP Client – PyPi package', accessed Dec. 2025 on PyPi
[9] D. Klein , 'AAS Standard Parser – PyPi package', accessed Dec. 2025 on PyPi
[10] Deutsches Institut für Normung (DIN), 'IEC 61406-1:2022 Identification Link – Part 1: General requirements', Dec. 2023
[11] Eclipse BaSyx, 'BaSyx WiKi – Quick Start Guide', accessed Sep. 2025 on home page
[12] Docker Inc., 'Docker Desktop', accessed Oct. 2025 on home page
[13] aas-core3.0-python AUTHORS, 'aas-core-works - aas-core3.0-python', accessed Sep. 2025 on GitHub
[14] Fluid 4.0, 'GitHub Project – open-source components', accessed Oct. 2025 on GitHub

10. BIOGRAPHIES



Fabian Geibel received his master's degree in mechanical engineering from RWTH Aachen University in 2019. After working as a research assistant at the Institute of Power Plant Technology, Steam and Gas Turbines in Aachen, he started a graduate program at the Bosch Rexroth AG. His profession is in the field of standardised data exchange and the application of digital twin technologies in industrial use cases.



Dr. Karin Tischler received her Diplom in Mechanical Engineering from the Universität Karlsruhe (TH) in 2002 followed by a Doktorin der Ingenieurwissenschaften from KIT after serving as a Research Assistant in the Department for Measurement and Control Systems. Her foundational work was focussed on information fusion for autonomous driving. Since 2008, she has held various positions at Bosch Rexroth AG, including Project and Group Manager for Innovation and System Application in Mobile Controls. Responsible leader in the central development of systems & solutions she is a recognized expert in system optimization. As Innovation Cluster Owner she drives new business initiatives solving challenges regarding energy transition and digitalization.

System Simulation Meets Asset Administration Shell – Potential and Challenges of Standardized, Model-Based Development of Fluid Power Systems

Heiko Baum[1], Micha Koebele[1], Denis Ritz[2], Heiko Roecher[3], Jürgen Weber[2]

[1] *FLUIDON GmbH, heiko.baum@fluidon.com, micha.koebel@fluidon.com ,*
[2] *TU Dresden, Chair of Fluid-Mechatronic Systems, denis.ritz@tu-dresden.de, fluidtronik@mailbox.tu-dresden.de*
[3] *Röcher GmbH & Co. KG, heiko.roecher@roecher-pressen.de*

Abstract.

Dynamic simulation is an established tool in the development of fluid engineering systems. In combination with the Asset Administration Shell (AAS), new potential arises for more efficient, standardized, and automated modeling and parameterization. Using the example of a hydraulic forming press, this article shows how AAS-based data structures accelerate system simulation and improve its quality – and where existing AAS standards still need to be expanded in order to meet the requirements of fluid technology practice.

Keywords: Fluid 4.0, digitalization, asset administration shell (AAS), collaborative engineering, dataspace, sub-models, industry-relevant use cases

1. INTRODUCTION

The *dynamic system simulation method* [1], has established itself as a recognized tool in the development of technical systems. It enables complex physical relationships between individual components and their interactions to be visualized early in the development process – long before prototypes can be built or components procured. This means that digital prototypes are available for testing during the development phase. As a result, functional principles can be evaluated, design decisions can be validated, and potential system problems can be identified at an early stage.

1.1. Significance of dynamic system simulation in fluid technology

Despite its well-known advantages, system simulation has not yet been established as a standard tool in everyday development work in many companies. Instead, it is often used only selectively, for example to analyze specific issues that arise during the commissioning of a plant, but not continuously throughout the entire development process.

1.2. Challenges in practical application

One of the main reasons for the limited use of simulation-based methods is the high barrier to entry when it comes to model creation. In practice, the construction of simulation-capable models still involves considerable manual effort. In particular, the selection of suitable model components, their correct parameterization, and the documentation of relevant assumptions and simplifications require experience, specific domain knowledge, and a considerable amount of time.
These challenges not only affect the efficiency of the simulation process but also make it difficult to trace and reuse models. Decisions made during model creation are often implicit and inadequately documented, which makes it considerably more difficult to reuse or adapt existing models later on.

1.3. Administration shell as an enabler for automated simulation

Against this backdrop, the combination of dynamic simulation and Asset Administration Shell (*AAS*) [2] offers new perspectives for simulation-based development processes. AAS is a central concept of Industry 4.0 and enables the standardized digital description of technical assets throughout their entire life cycle. *Ritz* [3] explains the potential advantages and several norms and standards to be considered for the application of the administration shell throughout the product life cycle for hydraulic systems.
The structured and semantically described provision of information and metadata in machine-readable form allows numerous steps in model creation that were previously performed manually to be automated or at least systematically supported. This not only facilitates the initial model setup, but also improves the consistency, traceability, and reusability of simulation models across different use cases.

1.4. Objective and structure of the article

This article examines the interaction between system simulation and administration shells from a practical perspective. Using the example of simulating a hydraulic forming press, it shows how fluid power drive systems can be converted into a simulation-capable model in a largely automated manner with the help of AAS-based data structures. **Figure 1.1** shows the press in the simulation environment on the left side of the image and the list of administration shell submodels used in the application example on the right side. The legend distinguishes between submodels that have been standardized by the *IDTA* [4] and those that have been newly developed or expanded as part of the *Fluid 4.0* research project [5].
Using a simulation workflow from the *Fluidon Cube* development environment [6], the sequence of transformation steps from the machine's administration shell to the simulation model, its application in the design of the hydraulic system, and the return of results and models to the administration shell are explained.



Figure 1.1. Fluid 4.0 example simulation of a hydraulic forming press

In addition, the article highlights where existing AAS standards are currently reaching their limits and which additional semantic descriptions and submodels are required to comprehensively support simulation-based engineering processes. The concepts developed in the Fluid 4.0 project are intended to provide impetus for the further development of existing AAS standards and contribute to establishing digital system simulation as a universal tool in mechanical and plant engineering.

2. STATE OF THE ART IN THE SIMULATION OF FLUID TECHNOLOGY SYSTEMS

2.1. Starting point: Fluid engineering circuit diagram and system complexity

In practice, a simulation project usually begins with a fluid technology circuit diagram containing all the components required to perform the drive task – often far more components than would be necessary for the actual simulation. One reason for this is that circuit diagrams are also used to generate the parts list. Accordingly, they contain additional components that are required for secondary functions but have no relevant influence on the dynamic behavior of the system.

2.2. Model design and selection of relevant system components

Figure 2.1 shows the typical steps in a simulation project, based on published descriptions [7], [8], [9].
Starting from the problem (see Figure 2.1-1) to be solved by means of simulation, the project engineer first has to ask himself the following question in the design phase (see Figure 2.1-2): Which components are actually relevant for solving the respective problem with the help of simulation?

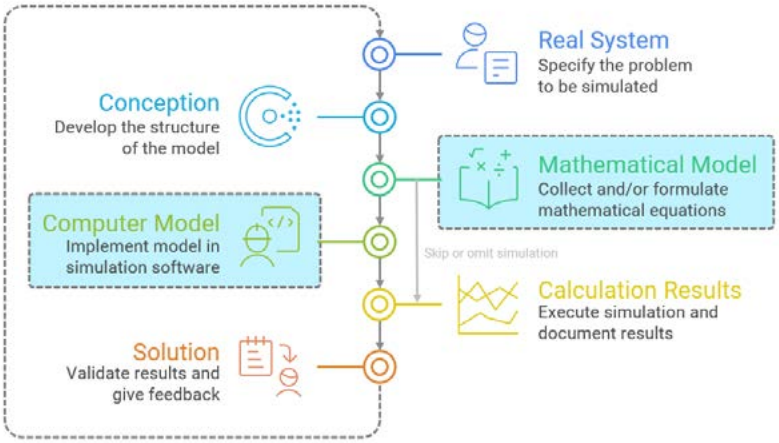


Figure 2.1. Sequence of work steps in a classic simulation process

This selection requires not only a deep understanding of the system but also experience in using the simulation environment and its model libraries and elements. In practice, this often leads to two typical error patterns. Either too many components are taken from the circuit diagram, resulting in unnecessarily large, computationally intensive, and numerically demanding models, or important components are omitted, meaning that crucial physical effects are not represented in the simulation.

2.3. Classic model creation process

Once the model concept has been defined, the classic simulation process begins with the creation of the mathematical model (see Figure 2.1-3) – regardless of whether a simulation will be performed later or not. The underlying equations are also required if the system is to be evaluated manually or with the aid of simple computer-aided design scripts, sometimes in the form of spreadsheets. In the past, the subsequent preparation of the equations for numerical calculation on a computer using a dedicated simulation program (see Figure 2.1-4) presented an almost insurmountable hurdle for many project engineers. As a result, simulations were often not used.

2.4. Modern simulation environments and model libraries

Nowadays, designers no longer have to manually create and code simulation models, as modern simulation environments take care of this task. Their model libraries contain many ready-made component descriptions that represent the dynamic properties of common components. These two process steps are therefore highlighted in color in Figure 2.1.

However, the designer must check whether the required components are available in the model library and whether their mathematical description meets the requirements. There is a high potential for error here, especially for occasional users, as it is usually difficult to choose between similar model types without in-depth background knowledge. Occasional users therefore often carry out additional, time-consuming test simulations to verify the behavior of individual component descriptions before using them in the overall system.

2.5. Parameterization of model components as a critical step

Another time-critical and error-prone step is the parameterization of the model (see Figure 2.1-5). As a rule, the required parameters must be manually researched and transferred from data sheets. For components that are mainly described by their geometry – such as cylinders – this can be done quickly. However, parameterization is much more complex for components with their own dynamics, such as control valves, regulated variable displacement pumps, or pilot-operated pressure valves. In practice, parameters such as step response times or frequency responses are often only available in graphical form. Their use in simulation therefore requires manual interpretation and conversion of the information obtained from diagrams, some of which are logarithmically scaled, into suitable model parameters.
In practice, this regularly leads to a time-consuming search for technical documentation on manufacturer websites and the need for internal consultations and validations using test simulations. However, the problem remains that some parameters and values are not provided by the respective manufacturers.

2.6. Motivation for an AAS-based simulation approach

The steps described above illustrate that, despite powerful simulation tools, the current state of the art in the simulation of fluid technology systems continues to be strongly influenced by manual activities, implicit expert knowledge, and individual decisions made by the project engineer. In particular, the selection of relevant system components, the interpretation of manufacturer information, and the parameterization of complex components are time-consuming and error-prone process steps.
At the same time, much of the information required for this – such as system structure, parts lists, or technical specifications – is available in digital form, but is often not structured consistently, lacks a unique semantic description, or is not linked in a machine-readable form. As a result, the consistent, automated use of this data throughout the entire simulation process has been limited to date. Concepts such as AAS open up new perspectives for data-driven, consistent, collaborative, and reproducible system simulation.
The following chapter shows how administrative shell data can be used to automate key steps in modeling and parameterization and to systematically support the simulation process.

3. SYSTEM SIMULATION REIMAGINED: AAS-BASED

In order to establish dynamic simulation as a standard tool across a wide range of fluid technology development processes, model design, model construction, and subsequent parameterization must be automated to a much greater extent and systematically supported. Only in this way can the full potential of simulation-supported development be exploited – both in early concept phases and in the detailed design and optimization of complex fluid engineering systems, as well as in the later further development of series products and the updating of machines designed in "batch size 1" for special applications.

The AAS-based approach opens up new possibilities here by structurally accessing existing engineering data and making it directly usable for the simulation process. **Figure 3.1** provides an overview of the essential steps in the AAS-based simulation process.

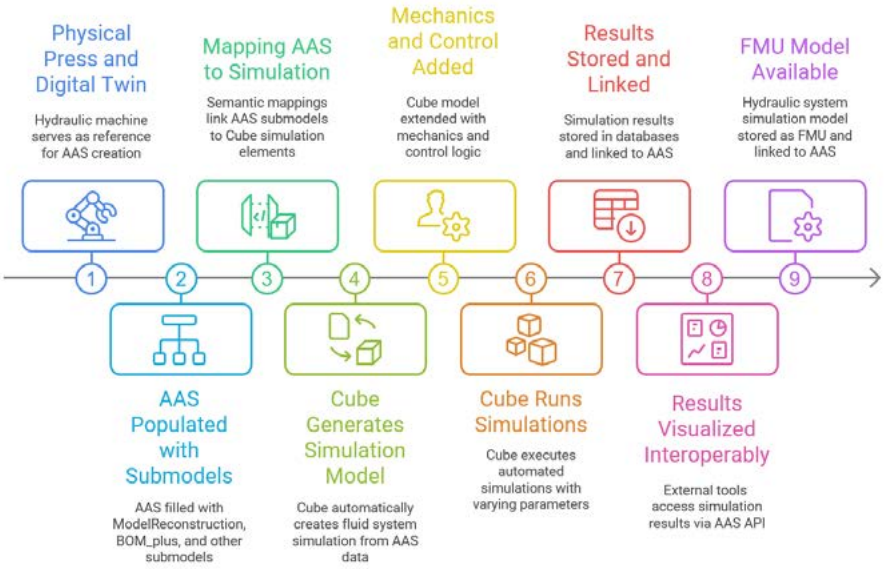


Figure 3.1. Digital Twin Workflow for a Hydraulic Forming Press

3.1. AAS-based system description

The starting point for the AAS-based simulation process here is, for example, the administration shell of a hydraulic forming press (see Figure 3.1-1), which the designer can use to access all relevant submodels. In addition to standardized submodels, the submodels newly developed or expanded in the *Fluid 4.0* project are also used (see Figure 3.1-2). This is necessary because existing IDTA submodels currently only provide incomplete simulation-relevant information.

The *ModelReconstruction* submodel plays a central role within the process. It contains the classification of components standardized in *ECLASS* [10] as well as their interconnection via hydraulic connections, thus defining the underlying system structure. The submodel is also linked to the parts list, which is mapped in the *BOM_plus* submodel. The *BOM_plus* submodel is the extension of the already standardized *BOM* submodel [11] proposed in the *Fluid 4.0* project.

The system structure and parts list for the application example were created using an external project planning tool (see [12]) and then stored in the administration shell of the hydraulic forming press. **Figure 3.2** shows the view of the *BOM_plus* submodel of the administration shell on the AAS repository server of the *Fluid 4.0* research project.

In the *BOM_plus* submodel, all components of the system are referenced to real existing components. The associated administration shells of these components in turn contain the parameters, model information, and settings required for simulation. This ensures that simulation-relevant information is always clearly assigned to a real, physical asset and can be used consistently across different engineering tools.

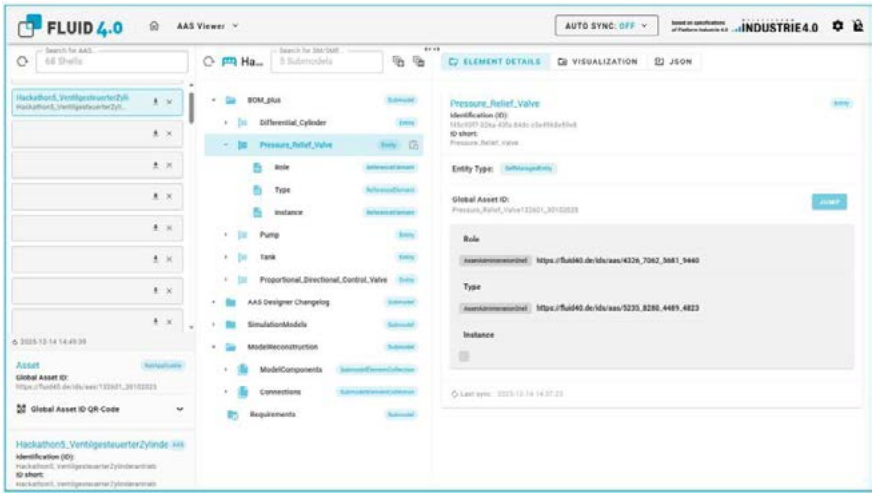


Figure 3.2. Administration shell of the hydraulic forming press

The classification into *role*, *type*, and *instance* describes the respective level of detail of a component in the development process. *Role* refers to an abstract component that merely represents the generic component class assigned in the *ModelReconstruction* submodel, for example, a pump. *Type* concretizes this abstract description to a real existing variant of the component class, for example, a gear, vane, or piston pump, which can be assigned to a specific type in a specific design from a manufacturer. If specific settings are then defined for such a component by the designer, this results in an *Instance* that is later integrated into the administration shell of the specific, real component. This shell is uniquely associated with a real, physical component with a unique serial number.

3.2. Mapping component behavior in the simulation

For each component contained in the system, there are basically two ways to map its fluid technical behavior in the simulation model:

1. **Direct model provision by the manufacturer:** A Functional Mockup Unit (*FMU*) [14] provided by the manufacturer can be loaded via the submodel *ProvisionOfSimulationModel* [13]. This contains an encapsulated mathematical description of the component and its parameterization and enables its use in the simulation – while preserving the manufacturer's know-how.
2. **Parameterization based on technical data:** If no FMU is available, the necessary parameters can be read from the *TechnicalData* submodel [15] and applied to a generic model from the library of the simulation environment used, which maps the real dynamic behavior as accurately as possible.

In the application example, the second option was deliberately chosen in order to demonstrate how flexibly technical component data can be assigned to existing generic mathematical model descriptions. **Figure 3.3** shows the *TechnicalData* submodel of the administration shell of a pump as an example. The parameters *NominalRotationalSpeed*, *Displacement,* and *MaxOperatingPressure,* among others, are relevant for the simulation.

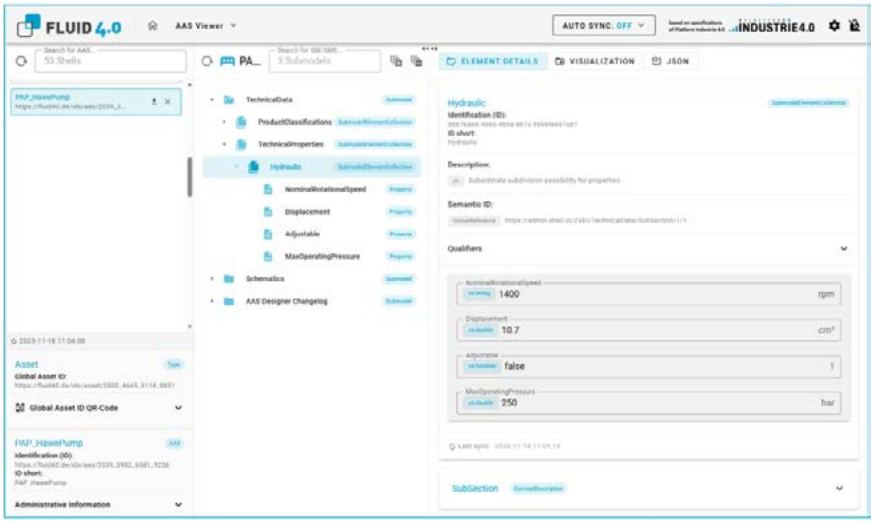


Figure 3.3. TechnicalData submodel of the administration shell of a pump

Access to this information and its further processing in the simulation workflow are explained in more detail in **Figure 3.4**.
First, the administration shell submodel is retrieved from the AAS server, read, and temporarily stored (see Figure 3.4-1). *ComponentName* references the unique component name from *ModelReconstruction*. In addition to name, value, and unit, each entry also contains a *GlobalReference* as an ECLASS-based semantic IRDI [16].

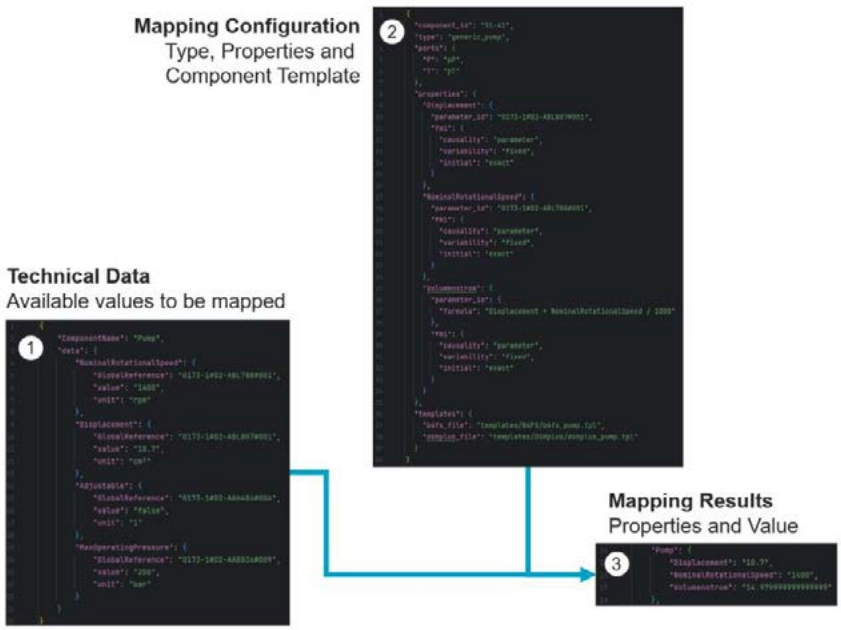


Figure 3.4. Mapping of technical data to the simulation model component

The declarative assignment of a real component type to a component from the simulation environment library is done via a mapping configuration (see Figure 3.4-2). In this configuration, the ECLASS IRDI is assigned to a simulation parameter as *parameter_id*.
Each parameter listed in the *properties* section also contains the semantic information *causality*, *variability,* and *initial*, whose meanings correspond to the definitions in the FMI standard [14].
The *templates* section references possible components of the simulation environment that can mathematically map the real component. The abbreviation *B4FS* refers to components that were developed in the *BaSys4FluidSim* project [17] to integrate component FMUs into the system simulation.

3.3. Formula-based parameter derivation in the mapping process

In the example of the pump, the real component is mapped to the simplest available pump model in the component library, which only provides a constant volume flow. However, the volume flow value required for this is not directly contained in the *TechnicalData* submodel. To enable consistent parameterization nonetheless, the Python-based mapping script allows formula-based derivation of missing parameter values (see Figure 3.4-2, line 27). Assuming that the pump is operated at its nominal speed, the required volume flow for the selected simulation model can be calculated in this way. The calculation rule for this is stored directly in the mapping configuration.
The results of this mapping step are again temporarily stored (see Figure 3.4-3) and transferred to the simulation model in a later process step.
Corresponding mapping configurations were created for all other components of the example system. For components not yet covered, users can independently define comparable mapping templates and integrate them into the workflow. Once created, these are then directly available for automatic model generation.

3.4. Automated model construction in the simulation environment

The actual model construction takes place within the *Cube* simulation environment (see Figure 3.1-4). A workflow script analyzes the structural information from the *ModelReconstruction* submodel, generates a network plan from it, and transfers it to a parameterized simulation model structure.
The entire process – from accessing the AAS repository to the fully parameterized simulation topology – is AAS-supported and requires no manual intervention. The hydraulic part of the simulation model is then immediately available for further analysis.

3.5. Extension to the multidisciplinary simulation model

However, for a meaningful evaluation – for example, in terms of energy efficiency, performance, or cycle times – the hydraulic simulation model must be supplemented with additional control and mechanical components. This extension is carried out manually by the user within the cloud environment (see Figure 3.1-5).
In the future, however, other domains, such as mechanics and control engineering, should also be mappable via corresponding administration shell submodels, so that fully automated multidisciplinary model generation will be possible at a later date.

3.6. Use of the generated model for simulation and optimization

With the completed simulation model (**Figure 3.5**), different simulation scenarios can then be carried out (see Figure 3.1-6). These range from the analysis of individual work cycles to automatic variant comparisons and optimization runs.

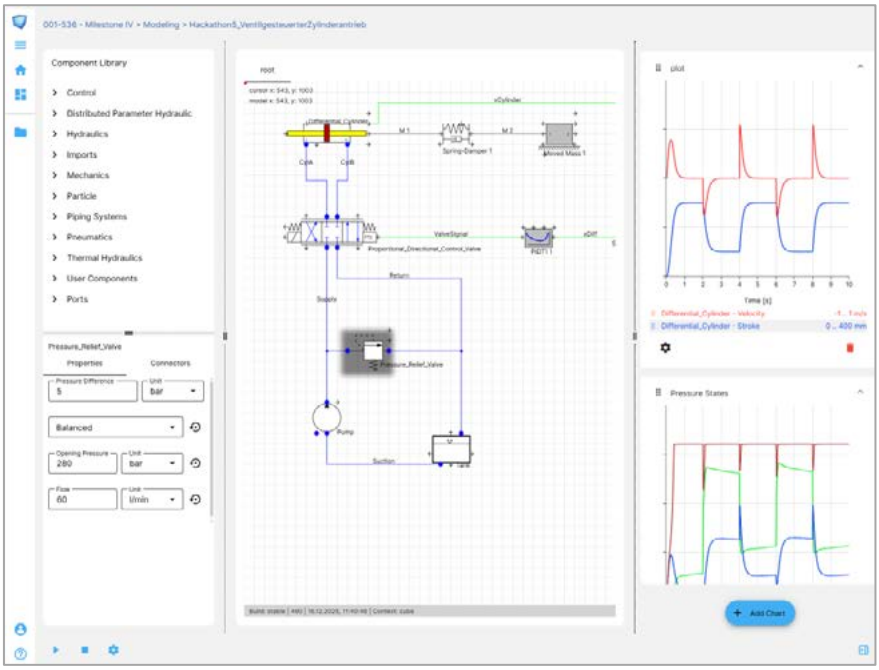


Figure 3.5. Simulation of the cylinder drive of the hydraulic forming press

The goal is to identify a system configuration that achieves the desired behavior while taking all technical constraints into account. At the same time, the simulation model should remain available as a consistently documented digital model in the AAS data ecosystem.

4. FEEDBACK OF RESULTS BACK TO THE ADMINISTRATION SHELL

Following the simulation, the simulation results are automatically checked and evaluated (see Figure 3.1-7).

4.1. Consistency check of the model parameters

A key step in this process is to check whether parameters that originally came from administration shells have been changed by the user. This check is made possible by the additional semantic information stored in the mapping configuration (see Figure 3.4-2). A distinction is made between *fixed product parameters* (e.g., piston diameter of a cylinder) and *freely adjustable settings* (e.g., opening pressure of a pressure valve or preload pressure of a hydraulic accumulator).
If the user changes parameters that are declared as *fixed*, the workflow generates a warning because the parameterization no longer matches the real component designed in *BOM_plus*. In the future – and already tested in a prototype as part of the *Fluid 4.0* research project –

the workflow can support an automated search for a component that matches the changed parameters. If a suitable alternative is found, *BOM_plus* can be updated accordingly so that the simulation and real project planning are consistent with each other again.

4.2. Rule-based verification of the system configuration

Beyond pure consistency checks, the Python scripts integrated into the workflow enable a rule-based evaluation of the system configuration. This checks whether parameter values comply with system-wide constraints and component-specific limits. **Figure 4.1** shows an example of this using the opening pressure of a pressure relief valve.

	ComponentName	MaxOperatingPressure	CrackingPressure
0	Differential_Cylinder	210	
1	Pressure_Relief_Valve	315	280
2	Proportional_Directional_Control_Valve	350	
3	Pump	250	

Figure 4.1. Checking the pressure resistance as a function of the selected system pressure

The opening (cracking) pressure is declared as *a variable* in the mapping configuration and can be set individually for each system by the project engineer in the real system. In the application example, a value of 280 bar was selected. However, a comparison with the maximum permissible pressure of the planned components shows that this value is not permissible for the cylinder and pump (see Figure 4.1).
Such rule-based checks could in principle also be implemented in today's project planning or circuit diagram creation tools, provided that they interact with the administration shell and use its data. In system simulation, however, this approach represents an innovation, as much of the component information required for it is not available in the classic simulation process. It is not necessary for numerical calculation and is therefore usually not represented in conventional simulation models.
In conjunction with dynamic simulation, the improved data available from the administration shell enables more advanced approaches to automated system evaluation. One example is the evaluation of the simulated cylinder speed, which checks whether permissible speed limits – for example, from the administration shells of the installed cylinder seals – are adhered to during the simulated work cycle.
Another example is the detection of critical pressure peaks in the simulated pressure curves, which can occur as a result of rapid reversal processes, for example. Such pressure peaks lead to material fatigue in hose lines or component housing. Based on the analyzed pressure curves, projections can be made regarding the number of critical load cycles in production operation. This information can in turn serve as a basis for condition-based maintenance strategies (predictive maintenance) or for analyses within the scope of the ongoing work on the Fluid 4.0 project on *the smart circular economy for the fluid industry* [18].

4.3. Feedback of results and models into the AAS data ecosystem

Once all tests have been successfully completed, the relevant results are automatically fed back into the administration shells. This includes, among other things:

- **Simulation documentation**, which is stored in the *HandOverDocumentation* submodel [19] and can be viewed by sales or the customer

- **Time series-based simulation results**, which are persisted in a time series database (e.g., *InfluxDB* [20]) and made available for detailed analysis via the *TimeSeriesData* submodel [21]
- **Settings determined in the simulation**, which are transferred to the *OperationalData* submodel developed specifically for this purpose as part of Fluid 4.0 and assigned to the instance AAS of the respective component so that they are available to the project engineer during commissioning
- **The generated simulation model**, which is stored as an FMU [14] on a Nextcloud instance [22] and referenced in the machine's AAS via the *ProvisionOfSimulationModel* submodel [13]. From there, it can be used, for example, for controller development or simulation-based optimizations.

The consistent feedback of all relevant data to the AAS ensures that digital project planning can be seamlessly integrated into downstream engineering processes. Project planning tools can then access the verified system configuration and its parameters to fully detail the plant – including all components required for implementation, such as screw connections, shut-off valves, or safety elements.
The administration shell thus functions as a continuous data hub between simulation, documentation, and actual machine construction.

5. CLASSIFICATION AND OUTLOOK

5.1. Simplification of the simulation process through AAS integration

The AAS-based system simulation presented in this article shows how the classic simulation process described in Chapter 2 can be fundamentally simplified and systematized. Many tool-specific and experience-based work steps that were previously performed manually are removed from the actual simulation process through the use of the administration shell and converted into standardized, interoperable data structures.
Figure 5.1 presents this new AAS-based simulation process, in which the remaining manual steps of the classic simulation process from Figure 2.1 are replaced by the automated steps presented in Chapter 3 and Chapter 4.

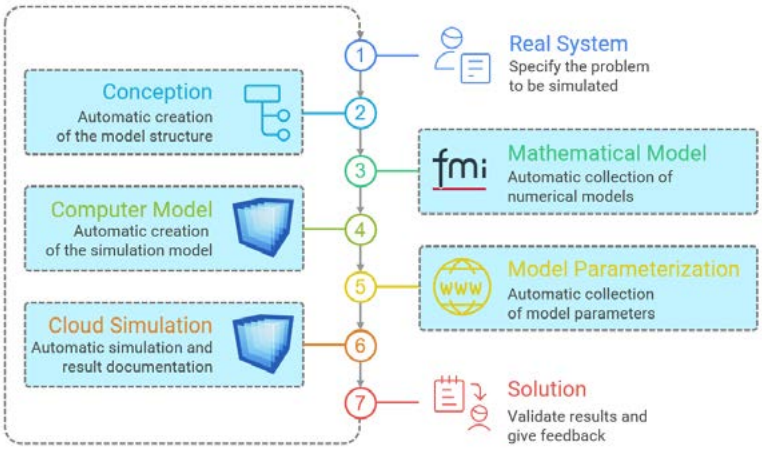


Figure 5.1. Automated simulation process with administration shell interaction

The central element of this simplification is the clear separation between *engineering data management* and *numerical simulation*. Information on the system structure, parts list, technical parameters, and semantic relationships is explicitly stored in the AAS and is available to the simulation workflow in a structured and machine-readable form. This makes model construction and parameterization largely automated and reproducible.
An example from the *Fluid 4.0* project (**Figure 5.2**) quantifies the difference to the classic approach (cf. Figure 2.1): A project engineer was tasked with researching, interpreting, manually reading, verifying component by component through test simulations, and transferring all component information required for the simulation from publicly available manufacturer data sheets based on the circuit diagram and parts list of the existing hydraulic forming press. The time required for this was recorded and forms the basis for comparison with the new AAS-based approach.

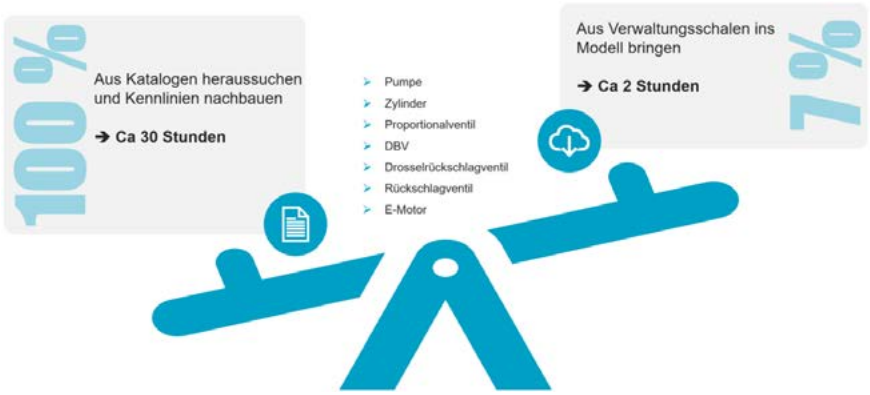


Figure 5.2. Time required for parameterization of components

Thanks to the AAS-based approach, the project engineer was able to begin his actual task – solving the problem or optimizing the system design – almost immediately after communicating the problem with the existing system design.

5.2. Insights gained

The practical implementation of the AAS-based simulation workflow in the *Fluid 4.0* project provides several important insights:

- Firstly, it shows that the combination of administration shell and dynamic system simulation goes far beyond mere automation of model creation. Feeding simulation results, settings, and models back into the AAS creates a consistent digital data space that links simulation, project planning, commissioning, and operation.
- Second, the improved data situation enables rule-based verification of the system configuration, which classic simulation approaches cannot achieve. Limit violations, inconsistent parameterizations, or potentially damaging operating conditions can already be identified and evaluated on the basis of simulated time sequences. This expands the scope of simulation from pure behavior analysis to an integral tool for system validation.
- Third, the AAS-based approach significantly improves the traceability and reusability of simulation-based development work. Decisions, assumptions, and results

are explicitly documented and clearly referable. This facilitates both internal collaboration and the transfer of engineering knowledge to downstream processes or external partners.

5.3. Gap analysis and contribution to standardization

The experience gained during the project reveals the following challenges and areas where action is needed:

- Available AAS submodels reach their limits in their current form and must be expanded in order to consistently map simulation-relevant information. For certain requirements – documentation of the system configuration or storage of plant-specific settings – there are currently no standardized submodels.
- The current AAS data situation for fluid engineering components is still insufficient. Many features relevant to simulation (e.g., efficiency data or characteristic curves) are either not included in current administration shells or are not described in the required granularity. In some cases, suitable semantic keys are missing in existing classification systems.

These gaps were closed in the *Fluid 4.0* project by means of project-specific extensions and newly developed submodels. All identified shortcomings and extensions deemed necessary will be systematically communicated to the IDTA and ECLASS after project completion so that they can be incorporated into future standardization activities. Among other things, the authors are already contributing their findings to the relevant IDTA submodel groups (Automation Engineering) or have already promoted the establishment and management of relevant working groups (Characteristic Curves) as a result of the project.

6. SUMMARY

This article shows how the dynamic system simulation of fluid power drive systems can be rethought and systematically simplified through the consistent use of the Asset Administration Shell. Based on the challenges of classic simulation processes, an AAS-based workflow was presented that integrates model design, parameterization, simulation, and result feedback in a continuous digital process.

Using the example of a hydraulic forming press, it was demonstrated how administration shell data can be used for automated model generation and how simulation results, validated settings, and simulation models themselves can be fed back into the AAS data ecosystem. This creates a consistent digital data space that establishes simulation not as an isolated tool, but as an integral part of modern engineering processes.

The results of the *Fluid 4.0* project highlight both the considerable potential of AAS-based system simulation and the need for further development of current standards. The insights gained and prototypical enhancements provide concrete impetus for standardization and contribute to establishing simulation-supported development as an integrated tool in mechanical and plant engineering.

7. ACKNOWLEDGEMENT

This work has been funded by the European Union (EU) and supported by the German Federal Ministry for Economic Affairs and Energy (BMWE) in the Fluid4.0 project – Revolutionizing Fluid Engineering: Implementation of digitalization for fluid power 4.0 in the cross-industry and cross-manufacturer data space using asset administration shells, submodels, and demonstrators (project numbers 13IK039E, 13IK039J, and 13IK039L). The views and opinions expressed are solely those of the authors and do not necessarily reflect the views of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.



8. REFERENCES

[1] 'Dynamical system simulation', *Wikipedia*. Feb. 23, 2025. Accessed: Jan. 03, 2026. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Dynamical_system_simulation&oldid=1277269006

[2] IDTA, 'Specification of the Asset Administration Shell - Part 1: Metamodel'. Industrial Digital Twin Association, 2025. [Online]. Available: https://industrialdigitaltwin.org/en/content-hub/aasspecifications/specification-of-the-asset-administration-shell-part-1-metamodel-idta-number-01001

[3] 'Accelerating Product Development: The Effort of Asset Administration Shells for aggregated Systems', in *The 19th Scandinavian International Conference on Fluid Power, SICFP'25*, Linköping University, Sweden: River Publishers, 2025.

[4] 'IDTA – The Standard for the Digital Twin', IDTA. Accessed: Jan. 03, 2026. [Online]. Available: https://industrialdigitaltwin.org/

[5] 'Joint Research Project Fluid 4.0 – Revolutionizing Fluid Engineering: Implementation of digitalization for fluid power 4.0 in the cross-industry and cross-manufacturer data space using asset administration shells, submodels and demonstrators.' Accessed: Jan. 03, 2026. [Online]. Available: https://fluid40.de/

[6] 'FLUIDON | Cube'. Accessed: Jan. 03, 2026. [Online]. Available: https://cube-lab.app/

[7] S. C. Brailsford, T. Eldabi, M. Kunc, N. Mustafee, and A. F. Osorio, 'Hybrid simulation modeling in operational research: A state-of-the-art review', *European Journal of Operational Research*, vol. 278, no. 3, pp. 721–737, Nov. 2019, doi: 10.1016/j.ejor.2018.10.025.

[8] S. Robinson, *Simulation: the practice of model development and use*. Chichester: Wiley, 2004.

[9] M. Landry, J.-L. Malouin, and M. Oral, 'Model validation in operations research', *European Journal of Operational Research*, vol. 14, no. 3, pp. 207–220, 1983, doi: https://doi.org/10.1016/0377-2217(83)90257-6.

[10] ECLASS, ‘ECLASS Standard’. Sep. 24, 2025. [Online]. Available: https://eclass.eu/en/eclass-standard

[11] IDTA, ‘AAS-SM - Hierarchical Structures enabling Bills of Material’. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2025/10/IDTA-02006-3-0-1_Submodel_Digital-Nameplate.pdf

[12] P. L. Remde *et al.*, ‘Digital fluid technology in the data space: From semantic component search to the development of mobile hydraulic systems using AAS’, presented at the 15th IFK 2026, Aachen, Mar. 2026.

[13] IDTA, ‘AAS-SM - Provision of Simulation Models’. Industrial Digital Twin Association, Frankfurt am Main, Germany, Oct. 2022. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2023/01/IDTA-02005-1-0_Submodel_ProvisionOfSimulationModels.pdf

[14] ‘Functional Mock-up Interface’. Accessed: Jan. 05, 2026. [Online]. Available: https://fmi-standard.org/

[15] IDTA, ‘AAS-SM - Generic Frame for Technical Data for Industrial Equipment in Manufacturing’. Frankfurt am Main, Germany, Mar. 2025. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2025/03/IDTA-02003_Generic-Frame-for-Technical-Data.pdf

[16] ECLASS, ‘IRDI’. Oct. 10, 2025. [Online]. Available: https://eclass.eu/support/technical-specification/structure-and-elements/irdi

[17] ‘Basic system Industry 4.0 | FLUIDON’. Accessed: Jan. 05, 2026. [Online]. Available: https://fluidon.com/knowledge-base/research-projects/basys4fluidsim

[18] ‘Fluid 4.0 – Smart Circular Economy for the Fluid Industry’, Supply Chain Services Division
of Fraunhofer IIS. Accessed: Jan. 05, 2026. [Online]. Available: https://www.scs.fraunhofer.de/de/referenzen/fluid-4-0.html

[19] IDTA, ‘AAS-SM - Handover Documentation’. Frankfurt am Main, Germany, Mar. 2023. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2023/03/IDTA-02004-1-2_Submodel_Handover-Documentation.pdf

[20] *influxdata/influxdb*. (Jan. 05, 2026). Rust. InfluxData. Accessed: Jan. 05, 2026. [Online]. Available: https://github.com/influxdata/influxdb

[21] IDTA, ‘AAS-SM - Time Series Data’. Frankfurt am Main, Germany, Mar. 2023. [Online]. Available: https://industrialdigitaltwin.org/en/wp-content/uploads/sites/2/2023/03/IDTA-02008-1-1_Submodel_TimeSeriesData.pdf

[22] ‘Nextcloud – The secure open source collaboration platform’, Nextcloud. Accessed: Jan. 05, 2026. [Online]. Available: https://nextcloud.com/de/

Biographies



Heiko Baum studied mechanical engineering at RWTH Aachen University and completed his doctorate in 2001. Since 2001, he has been the managing director of FLUIDON GmbH, where he focuses on model-based engineering, digital twins, and automated simulation workflows for hydraulic and pneumatic systems. His long-standing experience in the simulation and development of fluid power systems bridges industrial applications and research projects in the context of Industry 4.0 and the Asset Administration Shell.

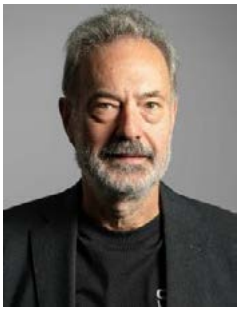


Micha Koebele received his M.Sc. degree from RWTH Aachen University in 2021. Since 2024, he has been working at FLUIDON GmbH as a development engineer, focusing on dynamic system simulation, model-based engineering, and automated simulation workflows for hydraulic and pneumatic systems. His work particularly addresses the integration of simulation models with digital twin concepts and Asset Administration Shell–based data structures in the context of Industry 4.0.



Denis Ritz completed his mechanical engineering studies at TU Dresden in 2018 and is currently a research assistant at the Chair of Fluid Mechatronic Systems at TU Dresden. His

research at centers on control strategies and the development of mobile hydraulic drive systems, with a strong emphasis on systems engineering. Prior to his academic studies, he completed vocational training as a farmer, providing him with practical insights into mobile machinery.



Jürgen Weber studied mechanical engineering at TU Dresden and successfully completed his doctorate in 1991. Until 1997, he was the active senior engineer at the former chair of Hydraulics and Pneumatics. This was followed by an approximately 13-year industrial phase. He was active in various positions in the R&D department of the agricultural and construction machinery manufacturer CNH. Effective March 1, 2010, Dr.-Ing. Jürgen Weber was appointed university professor and chair of Fluid-Mechatronic Systems (Fluidtronics) at TU Dresden. Since July 2018, he has been the head of the Institute of Mechatronic Engineering. He is also one of the founders and managing director of Construction Future Lab gGmbH.



Heiko Röcher is the managing director of Röcher GmbH & Co. KG, a family-owned company specializing in hydraulic presses and custom forming machinery. He combines engineering expertise with entrepreneurial leadership and is responsible for the strategic and operational development of the company, focusing on high-performance hydraulic press systems and customer-specific machine solutions.

Unlocking Semantic Interoperability in Industry with Large Language Models

Barkha Sharma, Malte Becker, Constantin Meier, Katharina Schmitz

RWTH Aachen University, Institute for Fluid Power Drives and Systems (ifas), Campus-Boulevard 30, 52074 Aachen, Germany

Abstract.

The digitalization of industrial processes within Industry 4.0 depends on the reuse of technical product data across systems, organizations, and lifecycle stages. In practice, the main challenge is not often the availability of data, but their heterogeneity, which prevents automated interpretation and semantic interoperability. This paper presents an automated pipeline for incrementally transforming existing heterogeneous technical data into interoperable representations. The approach builds on available industrial data and combines Large Language Models, retrieval-augmented generation, and the Asset Administration Shell to align manufacturer-specific descriptions with standardized vocabularies such as ECLASS. Using fluid power datasheets as an exemplary data source, the pipeline is introduced to demonstrate how heterogeneous engineering data can be made interoperable. The focus is on the pipeline concept itself, which aims to reduce manual mapping effort and support a gradual transition toward Industry 4.0-compliant digital engineering environments, rather than on the processing of datasheets. While demonstrated in the fluid power domain, the approach is transferable to other technical domains facing similar data heterogeneity challenges.

Keywords. Semantic Interoperability, Asset Administration Shell, ECLASS, Large Language Models, Retrieval-Augmented Generation, Digital Twin, Industry 4.0

1. INTRODUCTION

The ongoing digitalization of industrial processes has led to a steadily increasing availability of technical product data. In industrial practice, however, the fundamental challenge not the lack of data, but their significant heterogeneity. Technical data already exists in large volumes, yet it is described, structured, and maintained differently across departments, IT systems, and organizations. As a result, data cannot be automatically interpreted or reused, as a shared semantic basis is missing. Within the context of Industry 4.0, this lack of semantic interoperability limits the integration of technical product data into digital twins, automated engineering workflows, and cross-company data exchange [1].

Within companies, technical product data are typically distributed across ERP systems, PLM systems, engineering tools, spreadsheets, and document-based formats. Different departments often describe identical technical properties using varying terms, units, or levels of detail. Across organizational boundaries, this heterogeneity increases further due to

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

manufacturer-specific conventions and documentation practices. Although the underlying physical properties are equivalent, digital systems are unable to recognize these relationships [2]. Consequently, engineers are required to manually interpret, compare, and harmonize data, which leads to significant effort, introduces errors, and does not scale with increasing data volumes.

This challenge is particularly evident in the fluid power domain. Technical characteristics such as nominal pressure, maximum operating pressure, or flow rate are described inconsistently by manufacturers and even within individual organizations. For example, the maximum pressure of a hydraulic motor may be specified “maximum pressure”, “maximum operating pressure”, “p_max”, “peak pressure”, or “pressure limit”. While these descriptions refer to the same physical quantity, their semantic equivalence cannot be derived automatically. This lack of semantic interoperability significantly complicates the integration of fluid power components into digital engineering workflows, digital twins, and Industry 4.0 environments [3].

Standardized concepts such as the Asset Administration Shell (AAS) and semantic dictionary like ECLASS define suitable target structures for interoperable technical data. In practice, however, the transformation of existing heterogeneous data into these standardized representations remains largely manual. Industrial organizations cannot replace their existing data sources or enforce new data models across all systems. Instead, approaches are required that build on available data and enable a gradual transition toward semantic interoperability [4], [5], [6]. At the same time, recent advances in artificial intelligence, particularly Large Language Models (LLMs), have demonstrated the ability to capture semantic relationships within heterogeneous and partially unstructured data [7]. These capabilities provide new opportunities to support the semantic alignment of technical data without requiring fundamental changes to existing data sources [8], [9].

Against this background, this paper presents an automated pipeline that acts as an enabler for semantic interoperability. The proposed approach deliberately starts from existing technical data and allows organizations to retain their current data formats and structures. Using fluid power datasheets as a practical entry point, the pipeline combines LLMs, Retrieval-Augmented Generation (RAG) [10], and the AAS to incrementally align manufacturer-specific descriptions with standardized vocabularies such as ECLASS. In doing so, existing technical data can be progressively transformed into interoperable representations that support further digitalization efforts. While the approach is demonstrated in the fluid power domain, it is not limited to this field and can be transferred to other technical domains facing similar challenges related to heterogeneous product data.

The remainder of this paper is structured as follows: Chapter 2 discusses the theoretical foundations of the work. Chapter 3 reviews the state of the art and outlines the motivation for this work. Chapter 4 details the proposed methodology and implementation variants. Chapter 5 presents and discusses the evaluation results. Finally, Chapter 6 summarizes the findings and provides an outlook on future research directions.

2. THEORETICAL BACKGROUND

Achieving semantic interoperability is essential for digital manufacturing systems, allowing consistent data exchange and interpretation across heterogeneous environments. This

chapter introduces the theoretical foundations and core principles that form the basis of the proposed approach.

2.1. Interoperability

Interoperability is a key prerequisite for integrating heterogeneous systems in industrial and digital environments, enabling machines, software, and organizations to exchange data and effectively use the transferred information. In the context of Industry 4.0, interoperability supports collaboration across technical, organizational, and domain boundaries and forms the foundation of data-driven manufacturing [3].

As illustrated in Figure 1, interoperability can be understood as a layered concept in which higher-level coordination depends on capabilities provided by underlying layers.

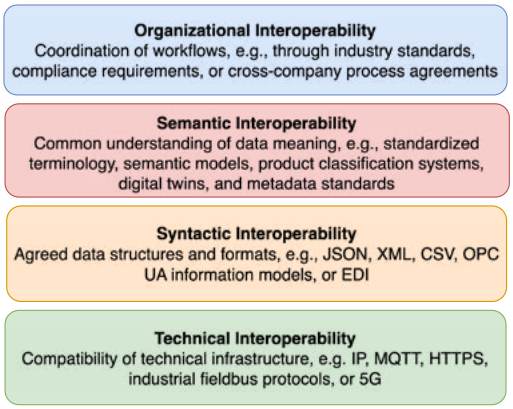


Figure 1. Levels of Interoperability [11]

At the organizational level, interoperability enables coordinated business processes and workflows across company boundaries. This coordination relies on semantic interoperability, which ensures that exchanged data conveys a consistent meaning across heterogeneous systems through shared vocabularies and semantic models [11].

The lower levels, namely technical and syntactic interoperability, provide the prerequisites for connectivity and structured data exchange but do not ensure a shared interpretation of information [11]. Consequently, semantic interoperability represents the critical transition from mere data exchange to meaningful information integration. This work therefore focuses on the semantic layer and explores how LLMs can support semantic alignment between unstructured technical data and standardized vocabularies.

2.2. Asset Administration Shell

The AAS is a core concept of Industry 4.0 that provides a standardized digital representation of physical or virtual assets. It acts as an interface between the physical and digital worlds, enabling interoperable management and exchange of asset-related data across the lifecycle. Within the Reference Architecture Model Industry 4.0 (RAMI 4.0) [12], the AAS constitutes the technical basis for digital twins. Its internal structure is defined by a formal metamodel specified by the International Digital Twin Association (IDTA) [4].

As illustrated in Figure 2, an AAS consists of one or more Submodels (SMs), each describing a specific aspect of an asset. Submodels are composed of SubmodelElements (SMEs), which represent individual data points or relationships. Core SME types include Properties, Files, ReferenceElements, and Ranges, which may be grouped using SubmodelElementCollections to represent complex structures. This hierarchical organization supports modular and scalable modeling of assets, ranging from individual components to complete systems [13], [14]y

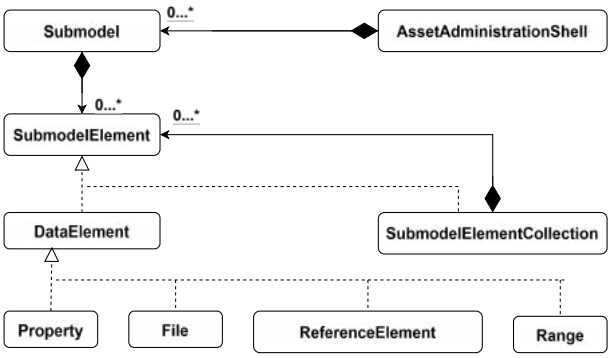


Figure 2. AAS hierarchy according to [5]

Depending on their functionality, AAS implementations can be categorized into three types:

- **Type 1 - Passive AAS:** Provides static information for documentation purposes.
- **Type 2 – Active AAS:** Enables automated data exchange via standardized interfaces.
- **Type 3 – Interactive AAS:** Supports bidirectional communication and autonomous behavior, effectively realizing a digital twin.

With increasing AAS maturity, both the volume and dynamics of exchanged data grow significantly. Consequently, consistent naming conventions and standardized semantic identifiers become essential to ensure reliable interoperability across tools and organizational boundaries, a requirement that is already becoming evident in fluid power research and industrial applications [15].

2.3. ECLASS Semantic Dictionary

Semantic interoperability requires a shared understanding of technical information across systems and organizations. A widely adopted approach to achieve this is ECLASS, a standardized framework for the classification and description of products and services. ECLASS provides language-independent, machine-readable representations of technical and commercial data, ensuring consistent interpretation across heterogeneous systems [15].

As illustrated in Figure 3, ECLASS organizes product information within a hierarchical classification structure, ranging from general product groups to specific classes, such as axial piston motors in the fluid power domain. Each class is associated with standardized properties and identified by International Registration Data Identifiers (IRDIs), enabling persistent and uniquely identifiable referencing of products, characteristics, and values across systems.

This structure supports interoperable and reusable data for applications such as digital twins, automated engineering, and cross-enterprise workflows.

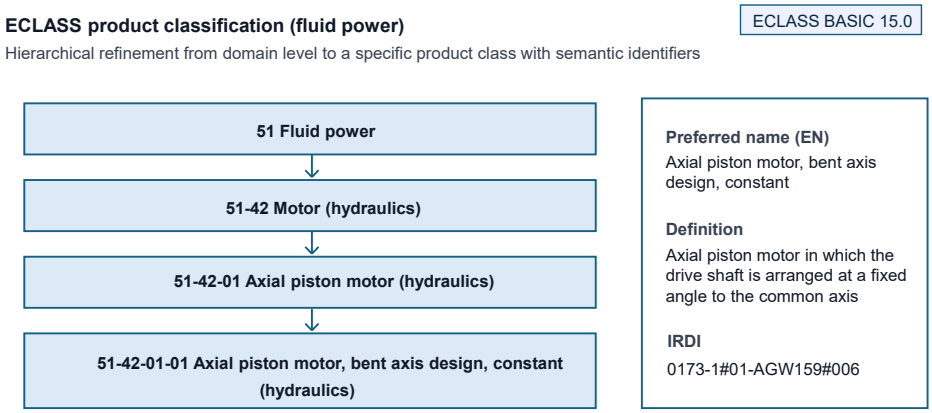


Figure 3. Example of hierarchical structure within the ECLASS classification [15]

Compared to conceptual standards such as IEC 61360 [16], which define general principles of data modelling, ECLASS provides a practical, domain-specific implementation that operationalizes these concepts for industrial use. Owing to its maturity and established adoption in domains such as fluid power, ECLASS serves as the semantic backbone for the semantic mapping approach presented in this work, particularly in the context of AAS. Further details on the internal ECLASS data model are outside the scope of this paper and are documented in the official ECLASS specifications [15].

2.4. Large Language Models

LLMs are based on the transformer architecture, which was introduced by Vaswani et al. [17]. The core element of this technology is the self-attention mechanism, in which each word is related to all other words in a text. As a result, it is possible to identify not only long dependencies but also complex semantic relationships [8], [10].

The development of LLMs is characterized by various fundamental learning approaches [18]:

- **Supervised learning:** Learning from labelled examples (e.g., a sentence is annotated with a category).
- **Unsupervised learning:** Learning from unstructured data without labels, only through pattern recognition.
- **Self-supervised learning:** The model generates the learning tasks itself, for example by predicting the next word or filling in missing text passages.

The development of LLMs typically follows three consecutive learning stages. First, during pretraining, the model learns general linguistic structures and statistical patterns from large volumes of text data using self-supervised objectives such as next-token prediction. Second, in the fine-tuning stage, the pretrained model is adapted to specific domains or tasks through supervised learning with curated datasets. Finally, in the alignment stage, instruction tuning or reinforcement learning from human feedback (RLHF) is applied to optimize the model's

responses for usefulness, safety, and factual accuracy. Together, these stages enable LLMs to combine broad language understanding with domain-specific reasoning capabilities [7], [9].

Although LLMs develop a broad understanding of language, they remain limited in two respects: Firstly, their knowledge is frozen at the level of the training data, and secondly, they may hallucinate information when context is insufficient. This is particularly problematic in industrial scenarios, such as when assigning technical data to standards, where precision and timeliness are crucial.

2.5. Retrieval-Augmented Generation

RAG is used to overcome these limitations. It extends a LLM with an external knowledge source that is consulted for every query, ensuring access to up-to-date and verifiable information. The RAG process follows three sequential steps [2], [19]:

1. **Retrieval:** In the first step, the user query is converted into a vector representation. This vector is compared against a knowledge base, such as technical documentation, standards, or AAS data sheets, to identify the most semantically similar text passages [19].
2. **Augmentation:** In the second step, the retrieved document excerpts are provided to the LLM as additional context. This augmented input serves as a factual reference for the model, anchoring its subsequent reasoning [19].
3. **Generation:** In the final step, the LLM produces an answer by combining its internal language knowledge with the externally retrieved information [19].

In this way, RAG combines two strengths: the language processing and generalization capabilities of LLMs with the timeliness and traceability of external knowledge sources. Transparency is particularly important in an industrial context as RAG allows answers to be provided with source references, which facilitates verification.

3. RELATED WORK

Achieving semantic interoperability remains a central challenge for realizing the Industry 4.0 vision. Reference architectures such as RAMI 4.0 and concepts like the AAS establish structured frameworks for linking physical assets with their digital representations. While these frameworks define harmonized data models and interfaces, the transformation of manufacturer-specific documentation into standardized vocabularies such as ECLASS or IEC remains largely manual in practice [12], [15]

In the fluid power domain, this challenge is particularly pronounced due to the high number of suppliers and the strong reliance on manufacturer-specific documentation practices. Hydraulic and pneumatic components are described using diverse terminologies, parameter definitions, and documentation formats. Although standards such as ISO 18582 and ECLASS define structured property sets for pumps, valves, and actuators, their consistent application across vendors is limited. As a result, semantic inconsistencies persist, hindering automated data integration, digital twin creation, and cross-company data exchange, which are key objectives of Industry 4.0 [20].

Recent research has explored the use of artificial intelligence to reduce manual effort in semantic mapping and interoperability tasks. LLMs, often combined with RAG, have

demonstrated the ability to extract, interpret, and structure data from unstructured technical documents [21], [22]. Xia et al. proposed an LLM-based approach for generating AAS from unstructured asset descriptions, showing that technical attributes can be identified and standardized with reduced manual intervention [14]. De Bellis investigated how semantic relations are encoded within LLM representations and how these can support ontology alignment tasks [23]. Wang et al. introduced Unify, a system for deriving structured semantics from heterogeneous industrial documents using LLMs [24]. In infrastructure engineering, Hoeltgen et al. demonstrated the enrichment of technical condition data through LLM-based semantic processing, illustrating the applicability of these methods beyond manufacturing domains [8].

In parallel, research on knowledge graphs and linked data emphasizes structured semantic representations as a foundation for interoperability across engineering systems [11]. While these approaches provide robust mechanisms for representing standardized relationships, they typically rely on manually curated data and predefined schemas, limiting their scalability and adaptability to heterogeneous legacy documentation.

Despite these advances, most existing approaches remain largely domain-agnostic and do not sufficiently address the specific terminology, parameter sensitivity, and contextual dependencies inherent to fluid power systems. Given the domain's reliance on precise physical parameters and cross-vendor component integration, automated semantic alignment tailored to fluid power documentation remains underexplored.

This work addresses this gap by proposing a domain-specific methodology that builds on existing heterogeneous technical data and supports their gradual transformation toward semantic interoperability. By combining retrieval-based context, LLM-supported reasoning, and established standards such as ECLASS and the AAS, the approach enables manufacturers and system integrators to retain their current data sources while systematically aligning fluid power documentation with standardized semantic models. In doing so, the methodology reduces manual mapping effort and supports digital twin creation and lifecycle-oriented data integration in accordance with Industry 4.0 principles.

4. METHODOLOGY AND RESULTS

This chapter introduces the proposed workflow for achieving semantic interoperability between unstructured industrial datasheets and standardized digital representations. The methodology is executed in parallel for a cloud-based and a fully local pipeline, both following the same processing stages and differing only in the execution environment.

The pipeline was implemented in Python due to its extensive ecosystem for document parsing, vector computation, and model integration. Its modular design supports flexible deployment in scalable cloud environments as well as local setups where data confidentiality is required. Figure 4 summarizes the shared workflow, consisting of five main stages described in the following subsections.

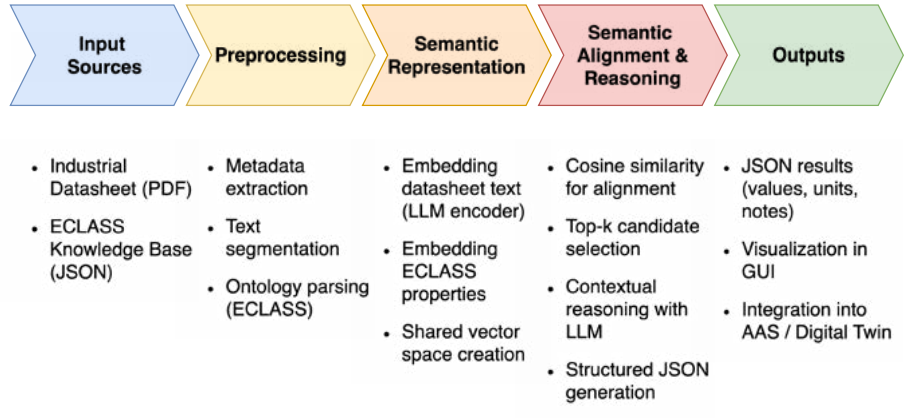


Figure 4. Workflow for achieving semantic interoperability

4.1. Input Sources

The workflow integrates two complementary data sources:

- **Technical datasheets (PDF):** These documents contain manufacturer-specific data on hydraulic and pneumatic products, such as nominal pressure or flow rate, but differ strongly in terminology, and structure. This heterogeneity makes automated processing difficult and motivates the need for a domain-independent mapping method.
- **ECLASS semantic knowledge base (JSON):** Using ECLASS ensures that the extracted data can be semantically aligned with established industrial standards and reused across systems, e.g., in AAS Submodels.

Integrating both sources forms the foundation for semantic interoperability. The datasheet delivers real-world technical content, while ECLASS defines the target semantic schema that standardizes its meaning.

4.2. Preprocessing

The preprocessing stage converts both the unstructured datasheets and the structured ECLASS ontology into machine-readable form. The workflow begins with data acquisition, where technical datasheets, usually PDF files containing textual descriptions, tables, and diagrams, are processed page by page. Text and metadata such as document titles or product identifiers are extracted using the Python library pdfplumber.

In the cloud-based setup, the same operation is performed through LangChain's PyPDFLoader, which supports API-driven processing. All extracted content is saved in JSON format, ensuring that each segment can be traced back to its original page and reused in later stages without reprocessing.

In parallel, the ECLASS ontology is loaded as a JSON file, providing the hierarchical structure of classes and standardized property definitions that serve as the reference vocabulary for later semantic alignment. Following extraction, the textual data undergoes normalization to correct common formatting issues in industrial PDFs, such as broken lines, hyphenations,

and layout artifacts. The cleaned text is then divided into overlapping sections using LangChain's RecursiveCharacterTextSplitter.

Unlike rule-based segmentation by headings or tables, this character-based approach is robust to inconsistent layouts and ensures that related data, such as a parameter and its corresponding unit, remains within a single segment. By the end of this stage, both sources are harmonized and structured, providing a consistent foundation for the embedding and semantic mapping steps.

4.3. Semantic Representation

In the third stage, the workflow converts textual and structured data into numerical vector representations, known as embeddings. This step allows semantic comparison between expressions even when they differ linguistically.

Each text segment t_i from the datasheet and each property s_j from ECLASS is mapped into an n-dimensional vector space through the embedding function:

$$E_T: T \to \mathbb{R}^n, E_S: S \to \mathbb{R}^n$$

where $E_T$and $E_S$represent the encoder functions for the datasheet and the semantic model, respectively.

Two embedding methods were implemented: The cloud-based pipeline employs the OpenAI model "text-embedding-3-large", which offers high semantic precision and context sensitivity. The local variant uses the Sentence Transformers library, specifically the model "all-mpnet-base-v2", which generates 768-dimensional embeddings and runs entirely offline. Both approaches use cosine similarity as their underlying measure of semantic proximity. Embeddings were selected over keyword-based methods because they capture conceptual meaning rather than literal matching. This enables the system to identify equivalent terms such as "pmax", "maximum pressure", and "max. operating pressure", even when phrased differently across manufacturers.

4.4. Semantic Alignment and Reasoning

In this stage, the embedded datasheet segments and ECLASS properties are compared within the shared vector space to identify semantic correspondences.
The similarity between a text segment $t_i$and an ECLASS property $s_j$is computed using the cosine similarity:

$$\text{sim}(a, b) = \frac{a \cdot b}{\| a \| \| b \|}$$

where $a = E_T(t_i)$ is the embedding vector of the text segment, and $b = E_S(s_j)$ is the embedding vector of the standardized property.
A higher value of $\text{sim}(a, b)$indicates a stronger semantic relation between both expressions. For each property, the top-k most similar text segments (typically $k = 3$-5) are selected as candidates for interpretation. Cosine similarity was chosen because it is scale-invariant and well suited for comparing high-dimensional sentence embeddings, providing more robust matches than keyword-based methods. The selected text segments are then processed by a LLM to extract structured data such as values, units, and contextual notes. This reasoning

step converts semantically aligned evidence into standardized outputs. In the cloud-based setup, reasoning is performed by OpenAI's gpt-4o-mini, while the local variant uses Meta-Llama-3-8B-Instruct with quantization for efficiency. Prompts follow a fixed JSON schema, ensuring syntactic consistency and easy post-processing. Together, the embedding and reasoning stages combine semantic precision with contextual understanding, enabling accurate and verifiable mappings between unstructured datasheets and standardized interoperability models.

4.5. Output

The workflow produces structured, machine-readable outputs that capture both the extracted data and its traceability. All results are stored as JSON files, each entry containing the standardized property, extracted value, physical unit, contextual notes, and a link to the corresponding page in the original datasheet. This design ensures that every mapping remains verifiable and can be reused for further processing.

4.6. Case Study

To evaluate the proposed framework, a case study was conducted using several publicly available datasheets from the fluid power domain. The documents originate from different manufacturers and describe hydraulic components using heterogeneous terminologies and specification formats. All datasheets were processed using the workflow described in Sections 4.1- 4.5. While multiple datasheets were used to assess general robustness, the quantitative evaluation focuses on one representative reference datasheet for which a fully manually curated ground truth mapping to ECLASS properties was available. This reference document comprises 114 technical features and enables a transparent and reproducible assessment of mapping quality.

Table 1 presents correctly mapped examples extracted from this reference datasheet, illustrating successful semantic alignment of heterogeneous fluid power specifications, where the data extraction and mapping were performed using an OpenAI-based variant (gpt-4o-mini).

Extracted Feature	**Assigned ECLASS Feature**	**ECLASS ID**	**Similarity Score**
Maximum operating pressure: 250 bar	Maximum operating pressure	0173-1#02-BAF016#005	0.90
Displacement volume: 45 cm^3/rev	Geometric displacement	0173-1#02-AAO677#004	0.85
Minimum ambient temperature: −20 °C	Minimum ambient temperature	0173-1#02-AAZ778#003	0.87

Table 1. Selected examples illustrating correct semantic alignment of heterogeneous fluid power datasheet descriptions with standardized ECLASS properties.

During preprocessing, textual content was extracted and normalized. Semantic embeddings were generated and compared with ECLASS feature embeddings using cosine similarity, and selected candidates were structured into feature–value representations by the reasoning

stage. A qualitative comparison with the expert reference confirmed that most mappings were technically plausible.

In addition to the qualitative assessment, the similarity scores obtained during the retrieval stage were analyzed in relation to the manually curated ground truth. Across all 114 mapped features, the average cosine similarity was 0.72. It was observed that correctly mapped features generally exhibited higher similarity values, while incorrect mappings occurred predominantly at lower similarity levels. Mappings with similarity scores above approximately 0.85 were found to correspond almost exclusively to correct semantic alignments with the expert reference. This value was not predefined as a hard decision threshold but emerged empirically from the evaluation as a high-confidence region, indicating stable semantic correspondence between heterogeneous datasheet descriptions and standardized ECLASS properties.

Despite the overall high mapping quality, mismatches and omissions were observed in specific cases. These can be attributed to the following factors:

- **Complexity of data extraction from industrial datasheets:** Fluid power datasheets often contain complex layouts, diagrams, characteristic curves, and hydraulic schematics that are not reliably captured by text-based extraction. Relevant features may therefore be partially extracted or missed before semantic mapping.

- **Variability in LLM reasoning performance:** Closely related technical concepts may be misinterpreted when contextual data is weak. A recurring example observed during evaluation is the confusion between nominal pressure and maximum operating pressure. Such cases were identified during qualitative comparison with the reference mapping and are not reflected in the correctly mapped examples shown in Table 1.

Overall, the case study demonstrates that automated semantic alignment of heterogeneous fluid power datasheets is feasible using the proposed workflow, while highlighting data extraction and reasoning robustness as key areas for further improvement.

5. DISCUSSION

This chapter discusses the main outcomes of the proposed approach, emphasizing its advantages, limitations, and validation perspective in the context of industrial data interoperability.

5.1. Advantages

The proposed workflow demonstrates that semantic data alignment in industrial domains can be automated largely without sacrificing interpretability. By embedding both datasheet content and ECLASS descriptors into a shared semantic space, the system can automatically recognize and assign technical data such as pressure, flow rate, or mounting type to standardized properties. This capability replaces a traditionally manual, time-intensive process with a consistent and reproducible mapping strategy. As a result, human involvement shifts from repetitive data entry to higher-level validation and quality assurance, which substantially increases efficiency and reduces the likelihood of human error.

Another notable strength lies in the method's scalability and robustness. Because the matching process operates on semantic similarity rather than keyword overlap, it generalizes well across heterogeneous datasheet formats, varying manufacturer terminology, and even multilingual documentation. The integration of ECLASS as a reference ontology ensures that the resulting data is compatible with AAS and digital twin implementations, thereby supporting interoperable data exchange across different systems and organizations.

Equally important is the flexible system architecture, which supports both cloud-based and local execution. The OpenAI-based configuration provides high inference speed and strong contextual understanding, making it suitable for rapid prototyping or large-scale processing. In contrast, the local pipeline based on Meta-Llama-3-8B-Instruct enables confidential, fully offline deployment, which is essential in regulated or security-critical industrial settings. Together, these two configurations demonstrate the approach's adaptability to different operational environments and data governance requirements.

5.2. Challenges

Despite the advantages of the proposed workflow, several challenges remain that affect its reliability. The most critical limitation is not the semantic mapping with LLMs itself, but the extraction of technical data from industrial datasheets. In practice, datasheets often contain complex layouts, scanned content, diagrams, characteristic curves, and hydraulic schematics that convey essential data but are not directly accessible through text-based extraction methods. As a result, relevant data may be partially extracted or missed entirely, which directly impacts all downstream processing steps.

Although preprocessing techniques such as OCR, normalization, and recursive text segmentation alleviate some of these issues, non-textual and weakly structured content remains a major source of errors. Improving document understanding at this stage therefore represents a key prerequisite before further improvements through larger or more advanced reasoning models can be expected.

Beyond data extraction, interpretive behaviour of LLMs constitutes an additional challenge. While larger, higher-parameter models generally improve contextual understanding and reasoning stability, they do not fully eliminate misinterpretations when technical descriptions are ambiguous. Moreover, increased model size introduces trade-offs in inference latency, computational cost, and deployment feasibility, particularly for fully local industrial deployments.

Finally, limitations in the coverage of existing interoperability schemas such as ECLASS can lead to unmapped attributes when manufacturer-specific or highly specialized parameters are encountered, highlighting the need for flexible hybrid approaches and future schema extensions.

5.3. Comparison and Validation Perspective

When compared to traditional manual annotation, the proposed workflow improves scalability and reduces processing effort. While manual mapping can achieve high precision for small datasets, it does not scale to large component portfolios or frequently updated documentation. The presented approach enables automated mappings that can be reviewed by domain experts in a human-in-the-loop setup, providing a practical balance between automation, reliability, and reduced manual effort.

Two deployment variants were evaluated: a cloud-based implementation using the OpenAI model gpt-4o-mini and a fully local open-source implementation based on Meta-Llama-3-8B-Instruct. The quantitative evaluation focuses on one representative reference datasheet with a complete manually curated ground truth comprising 114 technical features. Table 2 summarizes the quantitative comparison of both variants, including feature coverage, semantic similarity, inference time, and deployment-related aspects relevant for industrial application.

Criterion	OpenAI-based Variant (gpt-4o-mini)	Local Open-Source Variant (Meta-Llama-3-8B-Instruct)
Evaluated datasheet	1 reference datasheet	1 reference datasheet
Mapped features	114/114	114/114
Inference time per attribute	2.95s	5-10s
Total inference time per datasheet	5min 36s	10-19 min
Average Semantic Similarity	0.72	0.66-0.69
Data sovereignty	Requires external data exchange	Fully offline and confidential
Scalability	Easily scalable via cloud deployment	Limited by on-premise capacity

Table 2. Quantitative comparison of cloud-based and local deployment variants. Inference times reflect end-to-end execution and depend on deployment constraints.

Both deployment variants successfully mapped all 114 features contained in the reference datasheet. Mapping quality was assessed using three metrics. First, feature coverage was measured as the number of successfully detected and mapped features, resulting in full coverage for both variants. Second, semantic similarity was evaluated using cosine similarity during the retrieval stage. The cloud-based implementation achieved an average similarity score of 0.72, while the local implementation achieved values between 0.66 and 0.69. The observed differences in average semantic similarity between the cloud-based and local deployment variants are primarily caused by differences in the employed embedding models and execution configurations. The cloud-based variant utilizes higher-capacity embedding models with greater contextual sensitivity, while the local variant relies on smaller, quantized models optimized for offline execution. This results in slightly lower average similarity scores without fundamentally affecting mapping correctness for well-defined features.

Third, all generated mappings were compared against the expert reference. For clearly defined features, both implementations produced equivalent mappings, whereas differences primarily occurred for weakly specified or context-dependent attributes.

Correct mappings in both deployment variants were consistently associated with higher similarity values, while misassignments predominantly occurred in intermediate similarity ranges, particularly for closely related technical concepts such as nominal and maximum operating pressure. For the OpenAI-based configuration, similarity scores above approximately 0.85 emerged as an empirically observed high-confidence region for the evaluated

reference datasheet. This observation is model- and dataset-dependent and should not be interpreted as a strict or generally valid threshold. Future work will extend the evaluation to larger and more diverse datasets to further analyze similarity distributions and derive more robust confidence estimates.

Differences in inference time are primarily attributable to the execution environment rather than the semantic mapping methodology itself. The cloud-based variant benefits from highly optimized infrastructure, parallelized inference, and optimized model serving, whereas the local open-source variant prioritizes offline capability and data sovereignty and was executed on on-premise hardware with limited computational resources. Consequently, longer processing times represent an expected trade-off rather than a methodological limitation.

Overall, both deployment variants achieve comparable mapping quality for well-defined features. Differences are mainly observed in processing time and sensitivity to weakly specified input, indicating that stable semantic alignment depends not on model size alone but on the combination of structured preprocessing, retrieval-based context, constrained model interaction, and expert validation.

6. Conclusion and Outlook

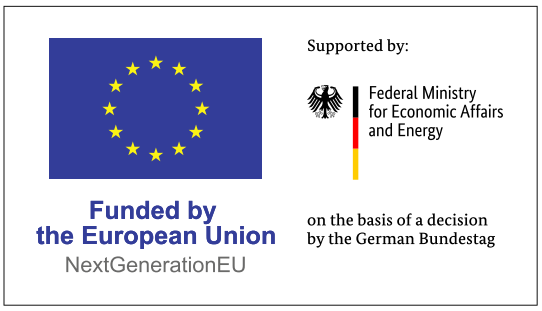
This paper demonstrated that LLMs can support semantic interoperability in the fluid power domain, where heterogeneous datasheets and inconsistent terminology continue to limit automated data integration. By combining embedding-based similarity search with structured reasoning and standardized vocabularies, the proposed workflow enables the transformation of existing manufacturer documentation into machine-readable representations suitable for digital engineering applications. The modular architecture allows the approach to be applied in different deployment scenarios. A cloud-based implementation supports scalable processing with low inference times, while a fully local configuration enables offline operation and compliance with data protection requirements, which are common constraints in industrial environments.

At the same time, the evaluation highlights that the main limitation is not the semantic mapping itself, but the extraction of technical data from industrial datasheets. Complex layouts, diagrams, characteristic curves, and schematic representations remain challenging and directly affect the quality of downstream semantic alignment. Future work should therefore prioritize improved data extraction and document understanding, including multimodal approaches, before extending the reasoning models or scaling the framework further. Additional integration with established interoperability frameworks such as the AAS may further support practical adoption.

Overall, the presented approach provides a feasible step toward improving semantic interoperability in fluid power domain by systematically bridging unstructured technical documentation and standardized digital representations, contributing to more consistent and reusable industrial data in Industry 4.0 contexts.

7. ACKNOWLEDGEMENT

This work has been funded by the European Union (EU) and supported by the German Federal Ministry for Economic Affairs and Energy (BMWE) in the project Fluid4.0 – Revolutionizing Fluid Engineering: Implementation of digitalization for fluid power 4.0 in the cross-industry and cross-manufacturer data space using Asset Administration Shells, submodels and demonstrators. (project number 13IK039L). The views and opinions expressed are solely those of the authors and do not necessarily reflect the views of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.



8. REFERENCES

[1] Bauernhansl, T. 'Handbuch Industrie 4.0', Springer Berlin Heidelberg, Berlin, Heidelberg, 2023. DOI: 10.1007/978-3-662-58532-0.

[2] Cartus, A. 'Interoperability of semantically heterogeneous digital twins through Natural Language Processing methods', https://proceedings.open.tudelft.nl/clima2022/article/view/143. DOI: 10.34641/clima.2022.143.

[3] Murrenhoff, H. '11th International Fluid Power Conference - 19th-21st of March 2018, Aachen, Germany', RWTH Aachen University, http://www.ifk2018.com/frontend/index.php_folder_id=470.html, Aachen, 2018.

[4] Rahal, J. R., Schwarz, A. et al. 'The asset administration shell as enabler for predictive maintenance: a review', Journal of Intelligent Manufacturing, Vol. 36, Nr. 1, pp. 19–33, 2025. DOI: 10.1007/s10845-023-02236-8.

[5] Zhang, J., Ellwein, C. et al. 'Digital twin and the asset administration shell', Software and Systems Modeling, Vol. 24, Nr. 3, pp. 771–793, 2025. DOI: 10.1007/s10270-024-01255-0.

[6] Salari, A. 'Details of the Asset Administration Shell - Part 2', https://digitalcollection.zhaw.ch/server/api/core/bitstreams/35301a19-fade-444f-bb40-8975bad47f48/content, 2021.

[7] Zhao, W. X., Zhou, K. et al. 'A Survey of Large Language Models', arXiv, 2023. DOI: 10.48550/arXiv.2303.18223.

[8] Höltgen, L., Zentgraf, S. et al. 'Utilizing large language models for semantic enrichment of infrastructure condition data: a comparative study of GPT and Llama

models', AI in Civil Engineering, Vol. 4, Nr. 1, 2025. DOI: 10.1007/s43503-025-00055-9.
[9] Minaee, S., Mikolov, T. et al. 'Large Language Models: A Survey', arXiv, 2024. DOI: 10.48550/arXiv.2402.06196.
[10] Zhao, P., Zhang, H. et al. 'Retrieval-Augmented Generation for AI-Generated Content: A Survey', arXiv, 2024. DOI: 10.48550/arXiv.2402.19473.
[11] Both, M. 'Semantische Interoperabilität digitaler Zwillinge als Basis für automatisierte Erkundungsmechanismen', 2024. DOI: 10.25673/116680.
[12] Technische Regel: 'DIN SPEC 91345:2016-04, Referenzarchitekturmodell Industrie 4.0 (RAMI4.0)', DIN Media GmbH, Berlin. DOI: 10.31030/2436156.
[13] Wei, K., Sun, J. Z., Liu, R. J. 'A Review of Asset Administration Shell', 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 1460–1465, 2019. DOI: 10.1109/IEEM44572.2019.8978536.
[14] Xia, Y., Xiao, Z. et al. 'Generation of Asset Administration Shell With Large Language Model Agents: Toward Semantic Interoperability in Digital Twins in the Context of Industry 4.0', IEEE Access, Vol. 12, pp. 84863–84877, 2024. DOI: 10.1109/ACCESS.2024.3415470.
[15] ECLASS: 'Fluid power', https://eclass.eu/eclass-standard/content-suche/, 2026.
[16] IEC 61360: 'Datenbank für elektrische/elektronische Bauteile', https://www.vde-verlag.de/iec-normen/iec-datenbanken/iec-61360-elekronische-bauteile.html, 2026.
[17] Vaswani, A., Shazeer, N. et al. 'Attention Is All You Need', arXiv, 2017. DOI: 10.48550/arXiv.1706.03762.
[18] Naveed, H., Khan, A. U. et al. 'A Comprehensive Overview of Large Language Models', arXiv, 2023. DOI: 10.48550/arXiv.2307.06435.
[19] Gao, Y., Xiong, Y. et al. 'Retrieval-Augmented Generation for Large Language Models: A Survey', arXiv, 2023. DOI: 10.48550/arXiv.2312.10997.
[20] Alt, R., Schmitz, K. 'Basic requirements for Plug-and-Produce of 14.0 fluid power systems', 2019 IEEE 8th International Conference on Fluid Power and Mechatronics (FPM), pp. 1440–1448, 2019. DOI: 10.1109/FPM45753.2019.9035813.
[21] Vaswani, A., Shazeer, N. et al. 'Attention Is All You Need', arXiv, 2017. DOI: 10.48550/arXiv.1706.03762.
[22] Zeid, A., Sundaram, S. et al. 'Interoperability in Smart Manufacturing: Research Challenges', Machines, Vol. 7, Nr. 2, pp. 21, 2019. DOI: 10.3390/machines7020021.
[23] Bellis, A. de: 'Structuring the unstructured: an LLM-guided transition', https://iswc2023.semanticweb.org/wp-content/uploads/2023/11/ISWC2023_paper_376.pdf.
[24] Wang, J., Li, Y. et al. 'Unify: A System For Unstructured Data Analytics', Proceedings of the VLDB Endowment, Vol. 18, Nr. 12, pp. 5287–5290, 2025. DOI: 10.14778/3750601.3750653.

Biographies



Barkha Sharma received a bachelor's degree in computer science from the University of Stuttgart in 2021 and a master's degree in computer science from the University of Stuttgart in 2025. She is a Research Associate in the Smart Systems group at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. Her research areas include machine learning, especially Large Language Models and AI planning, as well as data analysis.



Malte Becker obtained his Master of Science degree in Mechanical Engineering from RWTH Aachen University. Since 2021, he has been working as a research associate at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. His research focuses on the Industrial Internet of Things (IIoT) and Systems Engineering. Since 2025, he has been serving as Chief Engineer at ifas.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz's awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Best of Both Worlds – Load-Dependent and Load-Independent Hydrostatic Driving Switchable On Demand with Innovative Hybrid Electro-Proportional Control

Dipl.-Ing. (FH) Heiko Laffrenzen, M. Sc. Max Pfizenmaier,

Dr.-Ing. Matthias Liermann, Dr.-Ing. Robert Rahmfeld

Danfoss Power Solutions GmbH & Co OHG, Krokamp 35, 24539 Neumünster, hlaffrenzen@danfoss.com

Abstract.

When selecting a hydrostatic transmission, there is a choice between load-dependent and load-independent pump control. With certain limitations, the load stiffness characteristics can be tuned using software, although this requires sensors. In addition, a system controlled by an ECU can cause a small but noticeable delay for the machine operator. Danfoss now introduces the Hybrid Load Control (HLC) that combines load-dependent and load-independent control characteristics in one control valve without additional actuators and sensors. This paper explains the functionality and design of this innovative control valve, as well as its switching mechanism.

Keywords. Pump control, servo system, hydrostatic propel system, mobile hydraulics

1. INTRODUCTION

When designing hydrostatic propel systems, mobile machine manufacturers must select the machine's fundamental driving behaviour. State-of-the-art systems are categorized into two types.

The first is a load-independent behaviour, also known as speed mode. This is used for constant-speed applications or for scenarios where the system load should not affect the driving behaviour.

The second type is a load-dependent behaviour, also referred to as torque mode. This approach utilizes the power of the prime mover by creating a ratio between flow and pressure. In this mode, the flow is dependent on the system pressure p_{L}, equation 1.1.

$$q(p_{\mathrm{L}}) = \frac{V_{\mathrm{g,\,max}} \cdot n \cdot \tan\alpha(p_{\mathrm{L}}) \cdot \eta_{vol}}{\tan\alpha_{\mathrm{max}}} \qquad (1.1)$$

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

A new control valve developed by Danfoss combines both control characteristics, allowing the driver to switch between load-dependent and load-independent control behaviour [1]. The control valve has the same interface as an EDC valve. Additionally, the essential control loops are working hydromechanical. In detail, this means:

- no additional actuator
- no additional sensors
- fast control behaviour

The basic idea of this 'hybrid control' was introduced in [1,4,5]. The current paper presents the design realization of this new type of control, compares it to existing controls and elaborates how each mode of the control offers specific advantages in certain applications. Chapter 2 and 3 provide the context by describing the designs of the load-independent and load-dependent controls. Chapter 4 explains the design and function of the hybrid control. Chapter 5 presents simulation results of drive scenarios on a vehicle-level in a realistic 3D environment. These simulations include the propel hardware and software, which is also a crucial component for the overall performance.

2. EDC FUNCTION (LOAD-INDEPENDENT / SPEED MODE)

The Electric Proportional Displacement Control (EDC) is a type of Displacement Control (DC) for closed-loop pumps. In an EDC system, swashplate angle of the pump, and therefore its displacement, is proportional to the input current. The flow output is ideally proportional to the pump displacement; however, volumetric efficiency deteriorates the precise relationship between the input signal and the resulting flow. Overall, when a variable pump with an EDC is used in a propel application, the vehicle speed is primarily dependent on the input signal of the operator, and the speed of the prime mover.

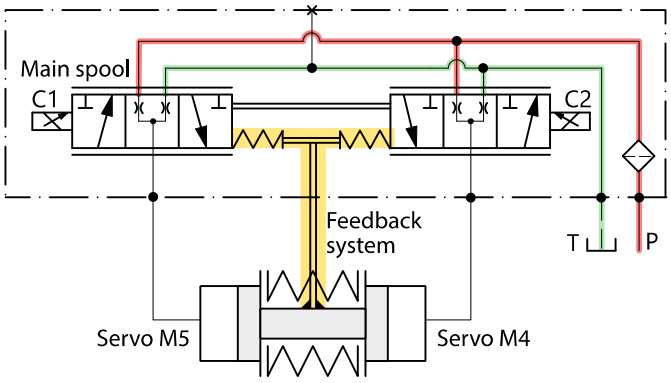


Figure 2.1: Schematic of EDC Control and servo system

The schematic of a Danfoss EDC control is depicted in Figure 2.1. The EDC incorporates a pair of proportional solenoids on either side of a common control spool, which is linked to the swashplate of the axial piston unit via a mechanical feedback system.

The solenoids convert the electrical input signal into force, moving the three-position, four-way control spool from its neutral position. This movement increases pressure on one side

of a double-acting servo piston, which rotates the swashplate until the forces from the mechanical feedback and the electrical input are balanced.

This hydro-mechanical feedback system is also referred to as a servo system. Figure 2.2 shows a schematic diagram of a servo system. Only one solenoid and one piston surface is depicted, sufficient to control displacement of the swashplate in one direction. The feedback spring is a soft spring intended to produce a force on the main spool equivalent of the swashplate displacement. In steady-state condition the main spool is closed and the solenoid force in perfect balance with the force produced by the feedback spring. The servo spring is a much stiffer spring, to make sure the swashplate angle returns to neutral at zero current condition. The axial piston unit produces reaction moments on the servo system which are dominantly pressure and speed-dependent but also vary with displacement and temperature. The force on the servo piston caused by swashplate reaction moments is shown as F_d.

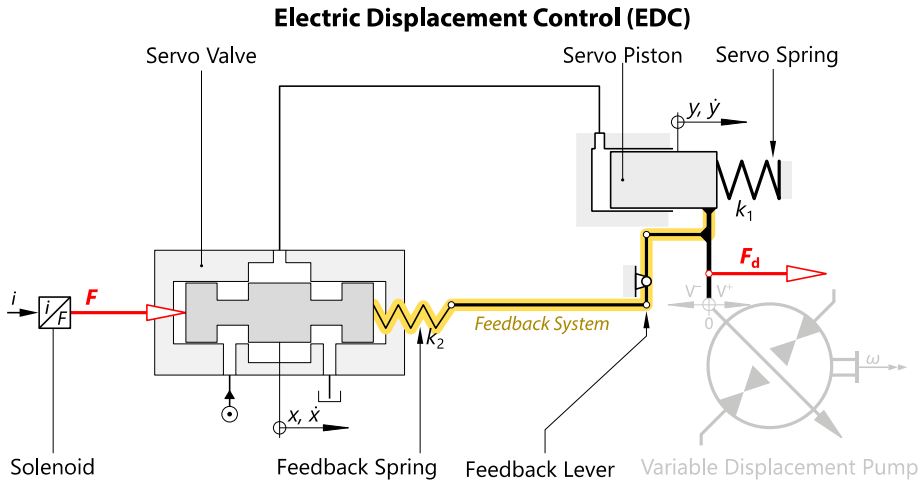


Figure 2.2: EDC with hydro-mechanical feedback system. The swashplate reaction force F_d is a disturbance.

3. NFPE FUNCTION (LOAD-DEPENDENT / TORQUE MODE)

The Non-Feedback Proportional Electric Control (NFPE) is a type of non-feedback control for closed-circuit pumps. It consists of two pressure control valves to control the pressure in each servo piston, see Figure 3.1. A simplified schematic is shown in Figure 3.2 with only one solenoid and a single servo cylinder chamber, which is sufficient to explain the action in one direction. It is clearly seen that there is no mechanical feedback between the servo cylinder and the control valve. The reaction force of the axial piston unit and the force of the servo springs result in a force F_d acting on the servo piston [2]. The preloaded servo springs produce a position-dependent force, and they are important for a fail-safe design to bring the pump back to neutral position. As a result, the servo system displacement is proportional to the valve current. However, because there is no mechanical feedback, the swashplate reaction moments can disturb the swashplate angle and render the pump load-dependent.

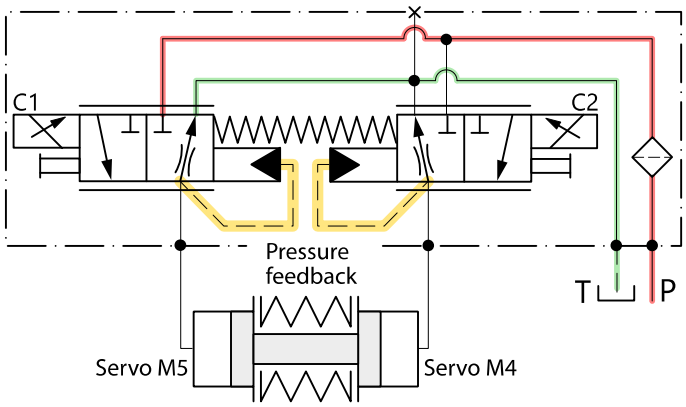


Figure 3.1: Schematic of NFPE Control and servo system

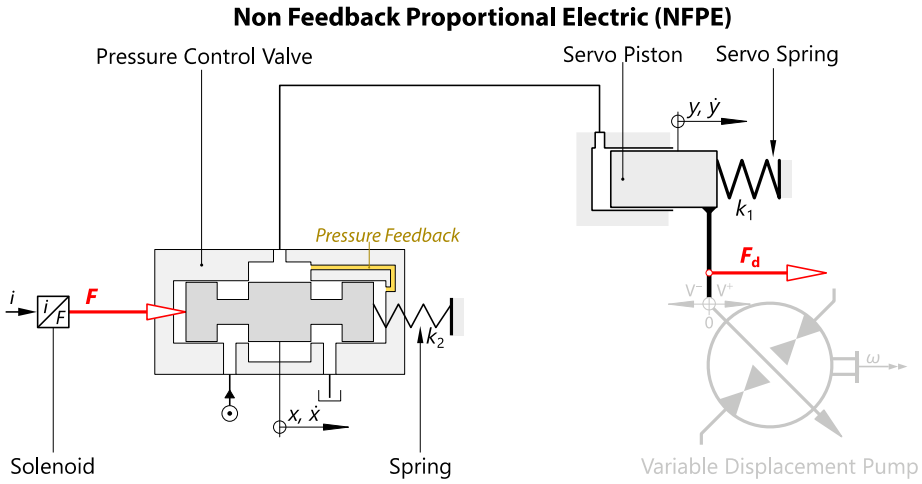


Figure 3.2: Simplified schematics of half of a non-feedback control system (NFPE). The valve realizes a pressure control. The servo piston finds an equilibrium position in relation to the external forces

If a variable pump with an NFPE control is used in a propel application, the oil flow and vehicle speed depend not only on the inputs of the operator. It is also influenced by the system pressure which affects the swashplate moments. The combination of system pressure and pump speed creates the load-dependent characteristic.

A pump with an NFPE control optimizes the use of available engine power by converting it into an ideal ratio of flow and pressure. This maximizes vehicle speed and tractive force without stalling the engine, making NFPE the preferred solution for applications with low engine power relative to the required corner power.

4. HYBRID LOAD CONTROL (HLC) FUNCTION

The patented Hybrid Load Control (HLC) is fundamentally an EDC with the added capability to create load-dependent behaviour, see Figure 4.1 [4,5]. It achieves this by allowing the servo pressure to superpose the mechanical feedback force. The primary difference between the HLC and a standard EDC is the control spool. It contains two nested inner spools. In their default positions, these inner spools open a passage for flow through an orifice cascade from the servo chambers M4/M5 to case. This orifice cascade acts as a pressure divider and lets the servo pressure produce a force on the control spool. On the side of the main spool which is pushed by the active solenoid the inner spool is pressed against the main spool and the orifice cascade is shut closed. It is the opposite side of the main spool, where pressure feedback can take place.

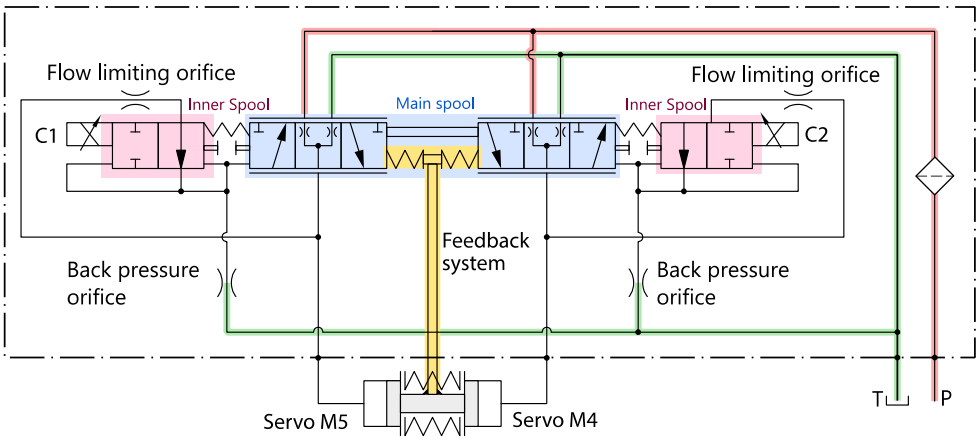


Figure 4.1: Schematic of HLC valve and system.

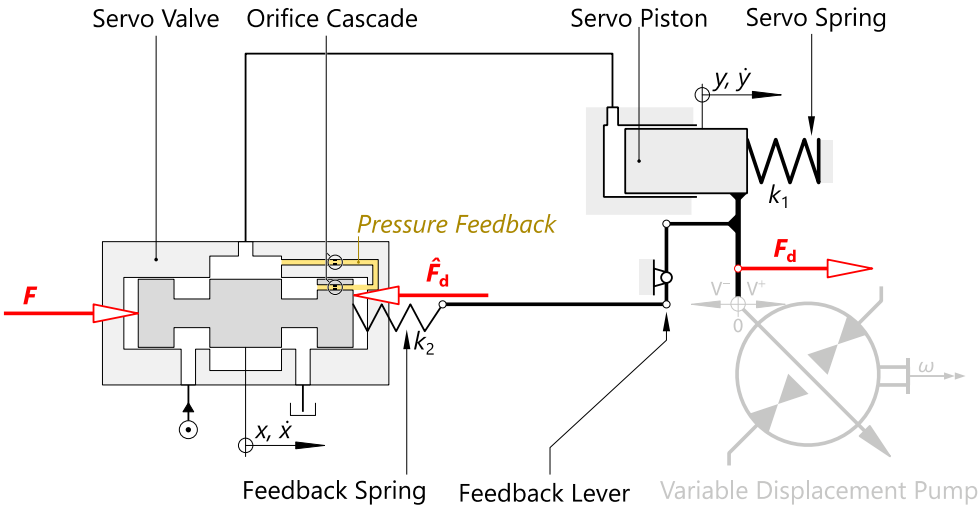


Figure 4.2: Simplified system schematic of half of HLC. In comparison with schematics of EDC and NFPE it includes both, pressure and position feedback. Pressure feedback can be switchted on or off.

The working principle can be explained by a simplified schematic in Figure 4.2. Here we consider the left side of the servo valve being pushed by the active solenoid. On the opposite side the orifice cascade ('pressure feedback') is seen, which transmits the servo pressure to the right side of the spool. It divides the servo pressure with a predefined ratio related to the ratio of orifice areas. The reduced servo pressure offsets the force produced by the solenoid and thereby reduces the set point of the servo control. Essential to the HLC is that the pressure feedback can be switched off. The orifice cascade is shut close, when the inner spool is slightly pushed against the main spool by the opposite (inactive) solenoid. It creates a disturbance for the active spool, but the force for closing the inner spool is small and can be easily compensated.

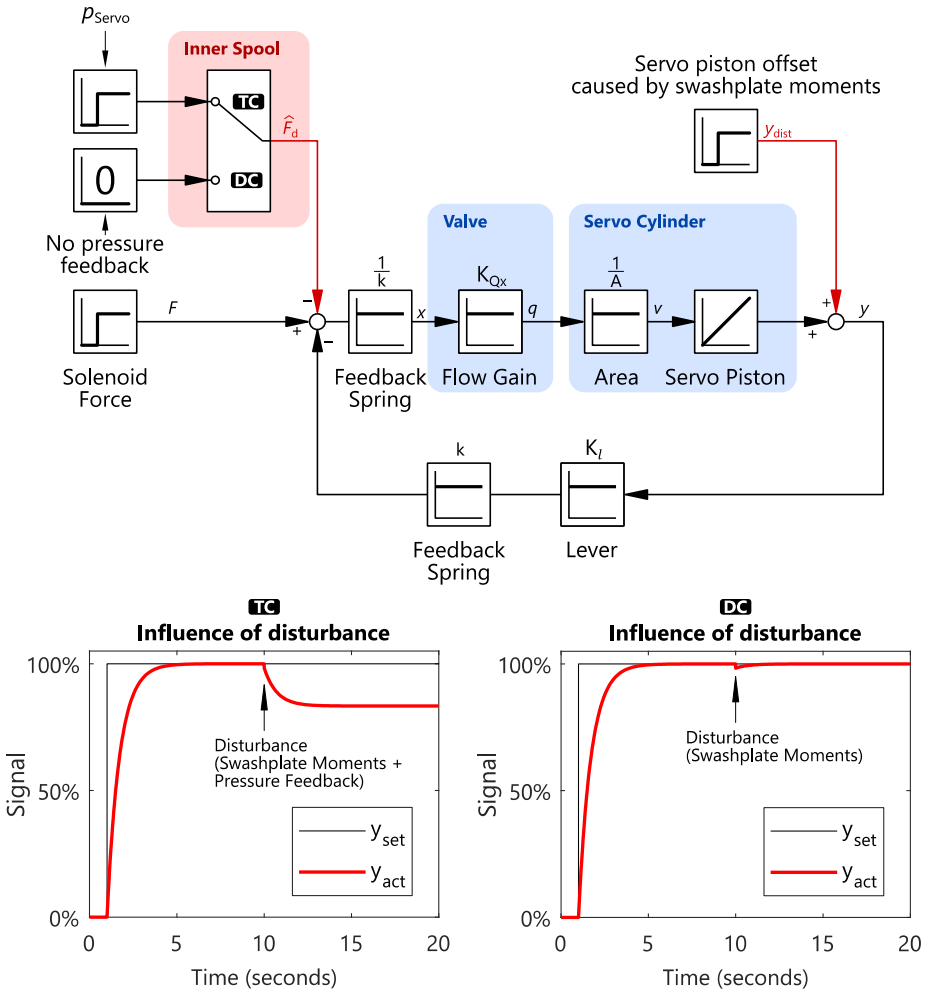


Figure 4.3: Block diagram of servo system with switchable pressure feedback and swash moment disturbance (top) and impact of disturbance (bottom).

Figure 4.3 shows a signal diagram of the control in the top. It can be used to derive the load stiffness with and without pressure feedback, compared in the bottom. For this analysis we

assume that the valve dynamics are much faster than the dynamics of the servo piston. Therefore, the valve is represented simply by a flow gain K_{Qx} that is proportional to the sum of the force components from solenoid, and mechanical and hydraulic feedback. The force component from the hydraulic feedback is indicated by the top left signal branch in red. Its switchable nature is indicated by the 'Inner Spool' switch. It is open in the default position.

In load-dependent mode, the reduced servo pressure is applied to the valve spool as a disturbance force $\hat{F}_{\mathrm{d}}$. It is a function of A_{Spool} spool area, and α ratio of the pressure divider:

$$\hat{F}_{\mathrm{d}} = A_{\mathrm{Spool}} \cdot \alpha \cdot p_{\mathrm{Servo}} \tag{4.1}$$

This disturbance force offsets the solenoid force and reduces the setpoint. As the servo pressure increases, the offset causes a de-stroking of the servo piston.

The closed-loop set-point transfer function can be written as

$$\frac{Y(s)}{F(s)} = \frac{\frac{K_{\mathrm{Qx}}}{kA}\frac{1}{s}}{1 + \frac{K_{\mathrm{I}}K_{\mathrm{Qx}}}{A}\frac{1}{s}} = \frac{\frac{1}{K_{\mathrm{I}}k}}{\frac{A}{K_{\mathrm{I}}K_{\mathrm{Qx}}}s + 1} \tag{4.2}$$

Where $F(s)$, $Y(s)$ are the Laplace transforms of spool force and servo position. The factor $\frac{A}{K_{\mathrm{I}}K_{\mathrm{Qx}}}$ represents the time constant of this first order transfer function. The steady-state relationship between force F and servo position y is given by

$$\lim_{s\to 0}\frac{Y(s)}{F(s)} = \frac{1}{K_{\mathrm{I}}k} \tag{4.3}$$

Therefore, the steady state relationship between servo pressure and servo displacement can be expressed as

$$\lim_{s\to 0}\frac{Y(s)}{p_{\mathrm{Servo}}(s)} = \frac{A_{\mathrm{Spool}}\alpha}{K_{\mathrm{I}}k} \tag{4.4}$$

And the load stiffness can be given as the inverse.

In load-independent mode, there is no disturbance by the servo pressure on the control spool. Or, in other terms, the orifice ratio α is 0. Reaction forces from the rotating unit that act directly on the servo piston do not affect the closed loop control.

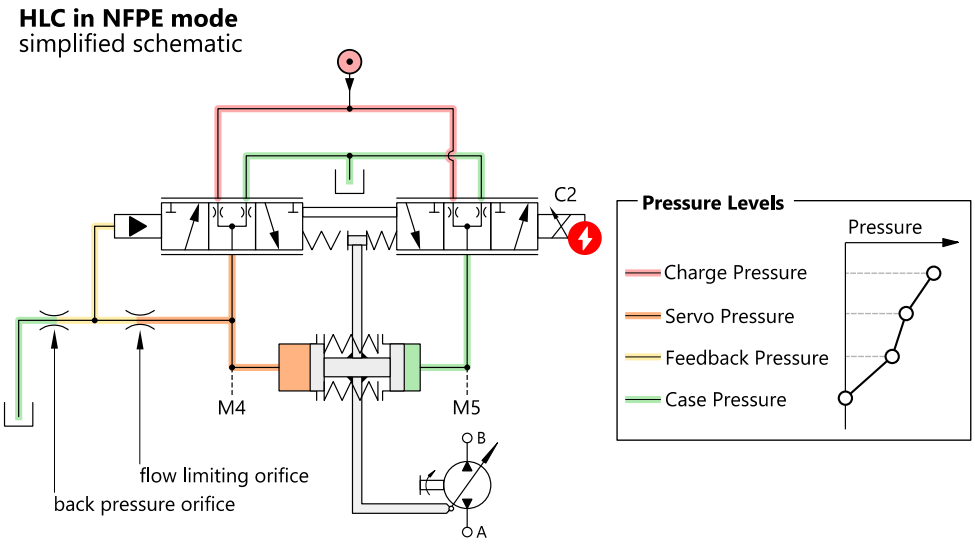


Figure 4.4: Pressure divider to superimpose reduced servo pressure on the valve spool.

Figure 4.4 shows a hydraulic schematic of the HLC, where solenoid C2 is activated. This leaves the orifice cascade on the opposite side open. Different pressure levels are indicated by colour. The servo pressure M4 is divided by the orifice cascade and produces a small force on the main spool. It opposes the force from solenoid C2. The set value is reduced, and as a result, the control behaves load-dependent.

To switch into load-independent (EDC) mode C1 is energized slightly with a current that does not affect the control action of C2 but is enough to close the orifice cascade via the small internal spool. The HLC does not require additional actuators or sensors to switch between modes. This is illustrated in Figure 4.5, which shows the control with both solenoids. The right solenoid is the active one and the left solenoid determines whether operation is in NFPE or EDC mode. The blue arrows indicate the current ranges when operating in load-dependent mode. The red arrows indicate the current ranges when operating in load-independent mode. The passive solenoid is given a small current that is below the dead-band threshold of the EDC control.

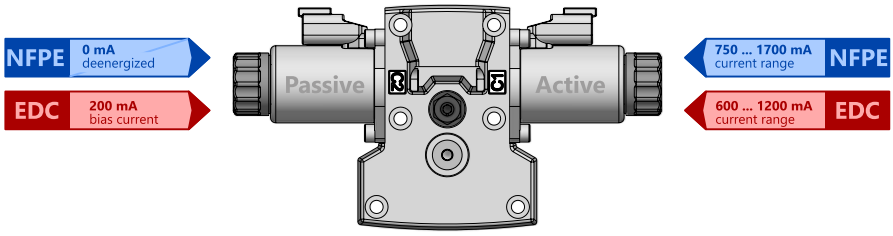


Figure 4.5: Actuation of Danfoss Hybrid Load Control

5. HLC OPERATION

Three driving scenarios are considered to illustrate the functionality of the HLC:

1. Driving a bucket into a pile of sand
2. Driving precisely over an obstacle
3. Downhill braking

For the first scenario, a load-dependent control characteristic is advantageous as this type of control prevents the diesel engine from stalling by de-stroking the pump. For the second scenario, a load-independent control characteristic is preferable because precise manoeuvring over an obstacle is required. A load-independent control characteristic regulates the displacement of the pump in a closed control loop, as mentioned above. The vehicle can thus maintain a nearly constant speed despite leaks. This enables precise manoeuvrability. In the third scenario, the braking behaviour between load-dependent control (NFPE) and HLC in load-dependent control mode is shown. It should be noted that load dependency is inactive during braking. During braking, the swashplate moment switches direction and now acts on-stroking. The control maintains the swashplate angle by pressurizing the opposite side. As can be seen from Figure 4.1 this opposite side cannot cause an offset to the servo spool because the orifice cascade is shut off by the active solenoid. To demonstrate the advantages of HLC in the scenarios mentioned above, a simulation was used that includes models of the hydraulic hardware, drive train components, a 3-D multi-body dynamic simulation of a vehicle and of the electrical control and software. Some parts of this real-time driver-in-the loop simulation setup are described also in [3]. The key data of the wheel loader and the drivetrain used in the simulation can be found in Table 1.

Table 1: Vehicle data.

Wheel rolling radius	0.6 m
Mass	8400 kg
Differential Gear ratio	16.51
Final differential gear ratio	15.24
Pump	H1P 89 cm³
Motor	H1B 160 cm³
Diesel engine max. power	103 kW

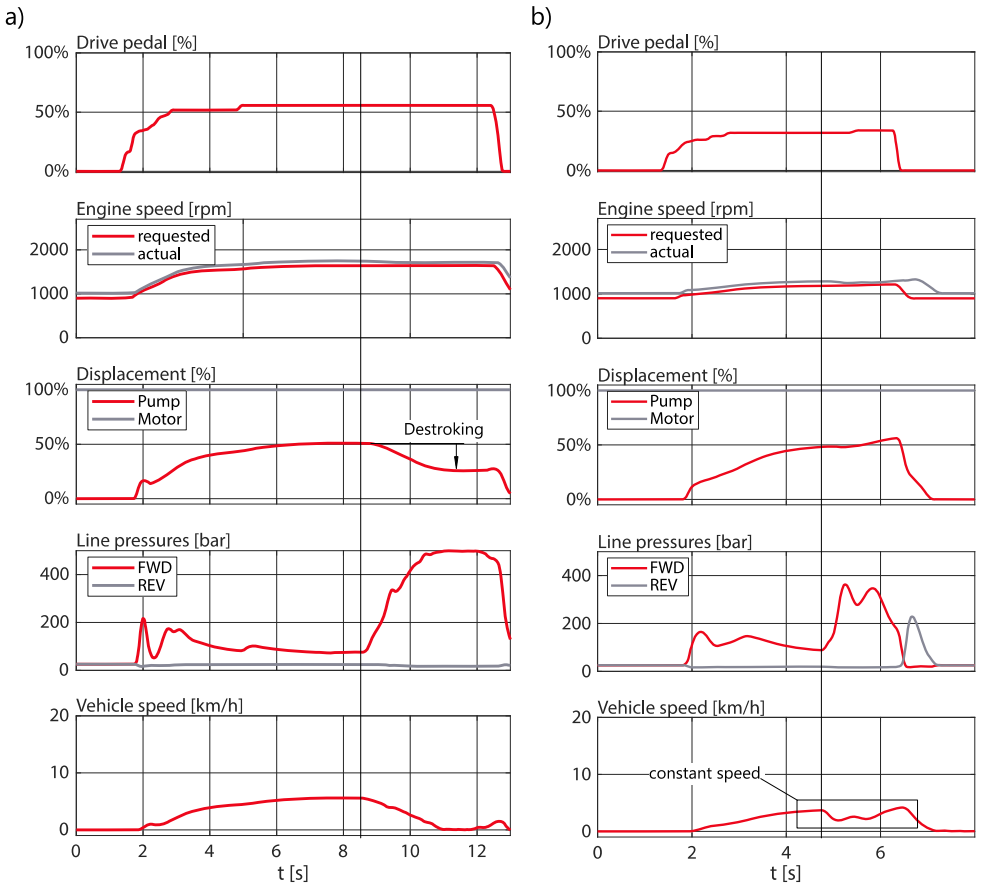


Figure 5.1: Simulation results showing two different driving scenarios with HLC operating in different modes. Left: Driving of bucket into pile (load dependent). Right: Precise maneouvre over obstacle (load independent)

Figure 5.1 shows simulation results for these driving scenarios. “Driving of bucket into a pile” (scenario 1) is shown on the left and “precise manoeuvre over obstacle” (Scenario 2) on the right. Considering scenario 1, the wheel loader approaches the pile of sand at almost constant speed. The drive pedal remains in approximately the same position until the bucket is full. As soon as the bucket reaches the pile of sand, the driving resistance increases, indicated by a steep rise in line pressure. Because of the servo pressure feedback, the pump de-strokes automatically and the diesel engine is protected from stalling.

In the second scenario, load-independent speed mode is selected because the driver wants to drive over an obstacle precisely and at almost constant speed. The corresponding signals are shown in Figure 5.1 on the right. As soon as the wheel loader reaches the obstacle, the load pressure increases. Due to the selected speed mode, the servo pressure cannot influence the control spool, and the vehicle speed remains approximately constant. The vehicle manoeuvres precisely over the obstacle.

The HLC provides a better braking performance, even in load-dependent mode. As stated before, the load-dependency is not active during braking. During braking the swashplate moments are on-stroking. Therefore, load-dependency, if it was active, would cause the pump to increase displacement during braking. This lowers the engine RPM and therefore reduces the available engine braking power. It can be demonstrated that with HLC, even in the load-dependent mode, braking performance is just as if the pump control was load-independent: fast and precise. Software can be used effectively to optimize the overrun operation of the diesel engine. Figure 5.2 shows a comparison of simulation results for downhill breaking between a vehicle that uses a standard NFPE pump control (left) and one that uses a HLC in load-dependent mode (right). The simulated vehicle is equipped with a 103 kW engine. Its braking power is the limitation for the achievable deceleration.

The driving scenario can be divided into three phases:

A. Acceleration on flat ground while approaching the incline,
B. Driving downhill a 10% slope,
C. Hydrostatic braking.

Braking starts at the same speed. As the upper plot in Figure 5.2 shows, with the HLC pump control, the software can better utilize the available engine braking power by controlling displacements to achieve maximum engine speed in the overrun operation. The vehicle with standard NFPE pump control stops after 9 s and 23.6 m, whereas the vehicle with HLC pump control stops after 7s and 20.8 m. It should be noted that the braking performance with NFPE control has also been optimized using software.

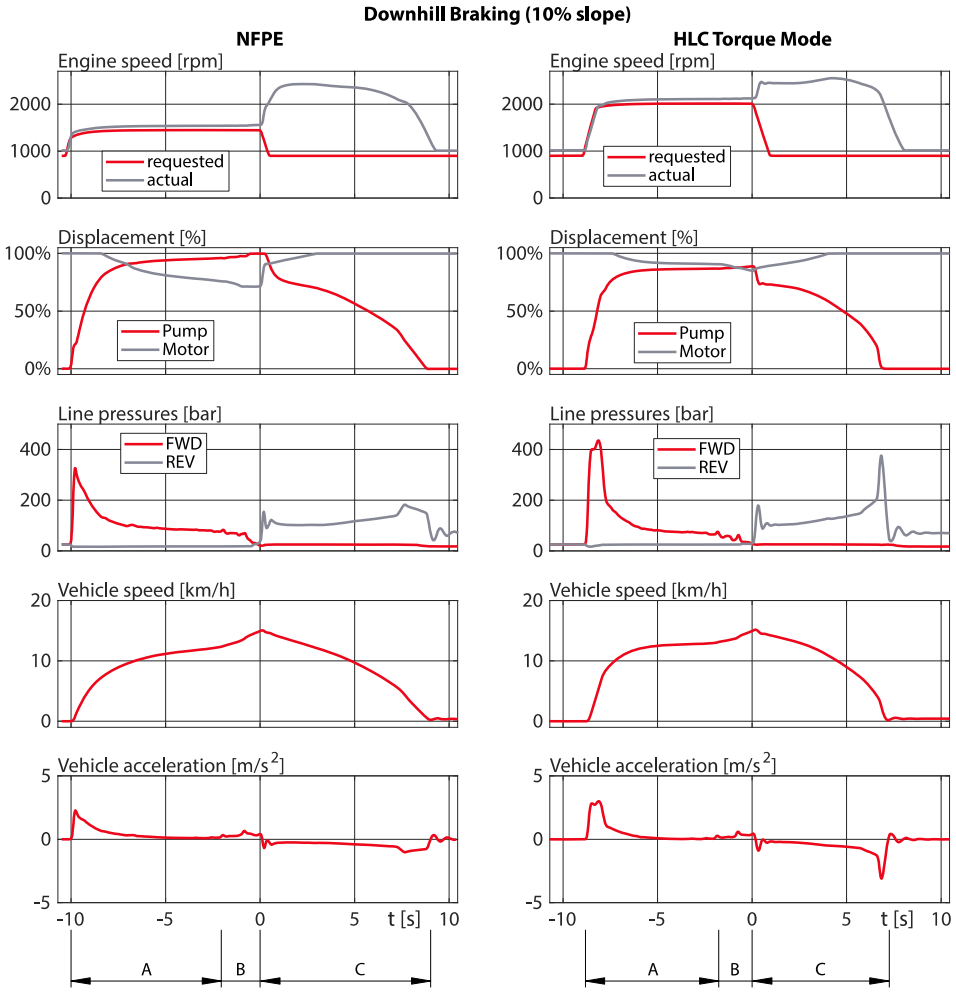


Figure 5.2: Simulation results of downhill braking. Left: Vehicle with standard NFPE pump control. Right: Vehicle with HLC pump control in load-dependent mode. A: acceleration on flat ground. B: driving downhill. C: hydrostatic braking.

SUMMARY AND CONCLUSION

The new Hybrid Load Control combines the best of both worlds, load-independent and load-dependent hydrostatic driving. By specifically using the dead band for a new function and making use of unused space, it was possible to combine two control valves into one. No additional actuators or sensors were required, nor was any additional space needed. The components that determine the characteristics of the control loops were adjusted so that the control characteristics of the HLC in the torque control mode are very similar to those of the current NFPE pump control. Since the control loops are still hydromechanical, the responsiveness of the established control valves is achieved and has been demonstrated on Danfoss Application Development Center. These validation results will be shown in a future publication.

NOMENCLATURE

Abbreviation	*Description*
DC	Displacement Control
EDC	Electric Displacement Control
FWD	Forward
HLC	Hybrid Load Control
NFPE	Non-Feedback Proportional Electric
REV	Reverse
TC	Torque Control

Variable	*Description*	*Unit*
A	Area	m^2
F_d	Disturbance force (from swashplate moments)	N
$\hat{F}_\mathrm{d}$	Force from pressure feedback	N
k	Spring constant	N/m
K_l	Feedback link ratio	-
K_Qx	Flow gain	m^2/s
p_Servo	Servo pressure	Pa
s	Laplace variable	s^{-1}
V	Volume	m^3
$V_\mathrm{g,max}$	Maximum displacement volume	$\mathrm{m}^3/\mathrm{rev}$
x	Set point	m
x_dist	Disturbance displacement	m
y_act	Actual servo piston position	m
y_dist	Servo piston offset due to disturbance	m
y_set	Desired servo piston position	m

REFERENCES

[1] A. Schumacher, R. Rahmfeld, H. Laffrenzen, 'High Performance Drivetrains for Powerful Mobile Machines', Proc. 10th International Fluid Power Conference, Dresden, 2016.

[2] J. Ivantysyn, M. Ivantysynova, Hydrostatic Pumps and Motors: Principles, Design, Performance, Modelling, Analysis, Control and Testing, Akademia Books International, 2001

[3] M. Liermann, C. Feller, F. Lindinger, D. Runge, 'Immersive 3d vehicle simulation for hardware-in-the-loop testing of mobile hydraulic controls', Proceedings of the 2019 Bath/ASME Symposium on Fluid Power and Motion Control, 2019

[4] M. Wüstefeld, R. Thoms, Danfoss Power Solutions GmbH & Co. OHG, Umschaltbare hydrostatische Verstelleinrichtung und zugehöriger Steuerkolben, German patent, DE 10 2014 206 460 B3, 2015 Jul 23

[5] C. Fiebing, R. Thoms, L. Belusa, H. Laffrenzen, M. Wüstefeld, Steuerkolbenzentrierung für hydrostatische Servoverstelleinrichtungen von Hydraulikmaschinen, German patent, DE 10 2011 079 691 B3, 2012 Aug 30

Biographies



Heiko Laffrenzen received the Dipl.-Ing. (FH) degree in mechanical engineering from University of Applied Sciences Flensburg in 2005. He is currently working as a Senior Manager Product Engineering for New Product Development for Hydrostatics at Danfoss Power Solutions based in Neumünster, Germany.



Max Pfizenmaier received the bachelor's degree in mechanical engineering from University of Applied Sciences Lübeck in 2017, the master's degree in mechanical engineering from University of Rostock in 2022. He is currently working as a Simulation Engineer for Dynamic Simulation for Hydrostatics at Danfoss Power Solutions based in Neumünster, Germany.



Dr. Matthias Liermann currently is Senior Manager Dynamic Simulation for Hydrostatics at Danfoss Power Solutions based in Neumünster, Germany. He studied mechanical engineering at RWTH Aachen University and received a Ph.D. in mechanical engineering under supervision of Professor Hubertus Murrenhoff at the Institute for Fluid Power Drives and Controls at RWTH Aachen in 2008. From 2009 to 2017 Dr. Liermann was an Assistant and Associate Professor at the Maaroun Seemaan Faculty of Engineering and Architecture at the American University of Beirut, Lebanon.



Dr. Robert Rahmfeld is the Senior Director Engineering for Hydrostatics at Danfoss Power Solutions, being technically responsible for hydrostatic axial piston units. He is with Danfoss over 20 years, holds several patents, and is author or co-author of more than 60 scientific journal and conference papers. Dr. Rahmfeld completed his PhD on new displacement controlled hydraulic linear actuators at the Technical University of Hamburg-Harburg in 2002 under supervision of Professor Monika Ivantysynova, and his MSc in mechanical engineering at Duisburg University in 1996. He was also holding a postdoc position in fluid power for 2 years at the Technical University of Hamburg-Harburg and Purdue University, Indiana, USA. Dr. Rahmfeld represents Danfoss within the VDMA fluid power board and is also the chairman of the corresponding fluid power research fund with more than 60 member companies. He was awarded in 2024 by Purdue University with the *Monika Ivantysynova Medal* for technology advancements of hydrostatic units and their controls, and for leadership in fostering industry - academic initiatives to promote fluid power.

High efficient Hydraulic Actuator Valve Sensor Unit without dedicated hydraulic supply (HAVSU)

Ralph Wassermann & Florian Hölzle

MAGNET-SCHULTZ, info@magnet-schultz.de

Abstract

MAGNET-SCHULTZ, specialists for electromagnetic solutions, has combined the advantages of fluid-power with the energy efficiency of sustainable electrified drivetrains, since purely electric actuators often lack power density and robustness in the demanding environment of applications in powerful modern electric drivetrains (EV, BEV and PHEV).

The highly innovative development resulted in an Actuator-Valve-Sensor-Unit (HAVSU) like not seen before in the vehicle industry. Adapted to the specific environment and requirements of electric drive axles, such a unit can actuate functions like disconnect and park-lock mechanics and supplement or substitute purely electric actuators with their well-known weaknesses in regards of robustness and power density.

The unit can provide force and stroke to the mechanism of the axle with high power-density and speed of hydraulics and is fully scalable to many other mobile and stationary applications.

1. GOALS

The main goal was to create an actuator providing all necessary functions for the typical applications in modern drivetrains and beyond, combining the proven advantages of hydraulics, like high power density, speed and robustness with an innovative layout with significantly reduced complexity, most efficient use of resources and small CO_2 footprint from cradle to grave.

Precisely the goals were:

(1) Stronger than purely electric driven actuators
(2) Less complex than common hydraulic systems

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

(3) Smaller and lighter than electric driven actuators
(4) More robust and durable in operation than electric actuators
(5) Less resource demanding, sustainable in production and more cost efficient
(6) Development according to the requirements of functional safety as defined by ISO26262
(7) Scalable to as many applications as possible (passenger car, truck & bus, off-highway, aerospace, defence, industrial)

2. INITIAL IDEA

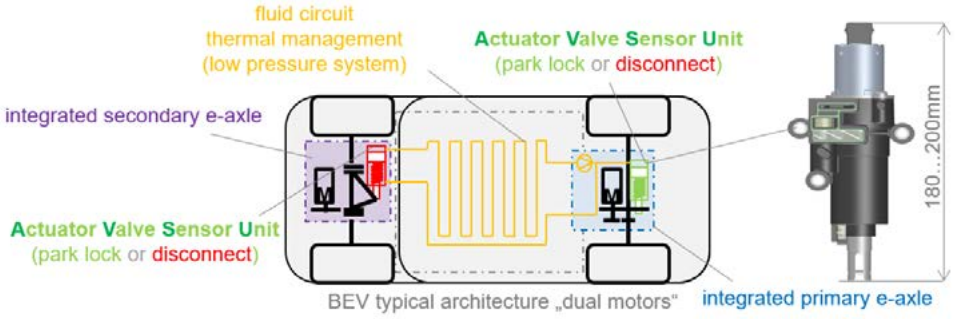


Figure 2.1: Typical architecture of a AWD EV

The invention has been inspired by the fact that e-axles generally need a thermal management or lubrication circuit to keep the system in an optimum state of operation. Thus, the creative idea was that exactly this fluid could be utilized in a second way to supply hydraulic actuators, even if initially not dedicated for this purpose.

But a major obstacle had to be overcome as a fluidic system primarily designed for thermal management is optimized to provide a high flowrate at low pressure levels and thus is not suitable to supply hydraulic actuators with sufficient pressure without boosting the fluid to higher pressure levels.

This innovative boost function is possible by blocking the hydraulic circuit for the very short actuation time of the actuator what leads to the pump working against a closed end. That on the one hand enables boosting the pressure up and on the other hand does not negatively affect the relatively slower responding thermal management system.

Once the actuator reaches its end position, the integrated solenoid-driven bistable locking mechanism is activated and secures the actuator stroke without the need of further supply pressure. After locking the actuator position, the integrated valve opens, and the thermal management is working freely with minimal pressure drop again.

To avoid an additional dedicated electrically operated pilot valve, the valve had to be part of the solenoid-driven locking mechanism, which is anyways necessary to lock the state of the hydraulically actuated cylinder. Thus, it became possible to actuate the bistable lockable hydraulic cylinder with the same solenoid as the pilot valve for pressure boosting. Summarized up and due to the dual use of the highly integrated subsystems, pressure boosting requires no additional actuation principle than the already existing from similar systems with a common hydraulic pressure supply.

3. FUNCTIONAL COMPONENTS

The unit consists of eight collaborating subcomponents.

(1) Hydraulic actuator, piston and cylinder rod – providing external stroke and force up to 800N (further scalable)
(2) Position sensor – monitoring the stroke-position of the cylinder rod
(3) Pilot valve – controlling the state of the main valve
(4) Main valve – switching the flow of the fluid in circuit ON/ OFF
(5) Locking mechanism – Locking the state of the hydraulic actuator
(6) Electromagnet – operating the pilot valve and locking mechanism
(7) Housing – providing mechanical interfaces and ensuring structural integrity
(8) Electrical connection – connecting all electrical subcomponents and building interface to the higher-level system (power supply, communication, sensing, diagnosis)

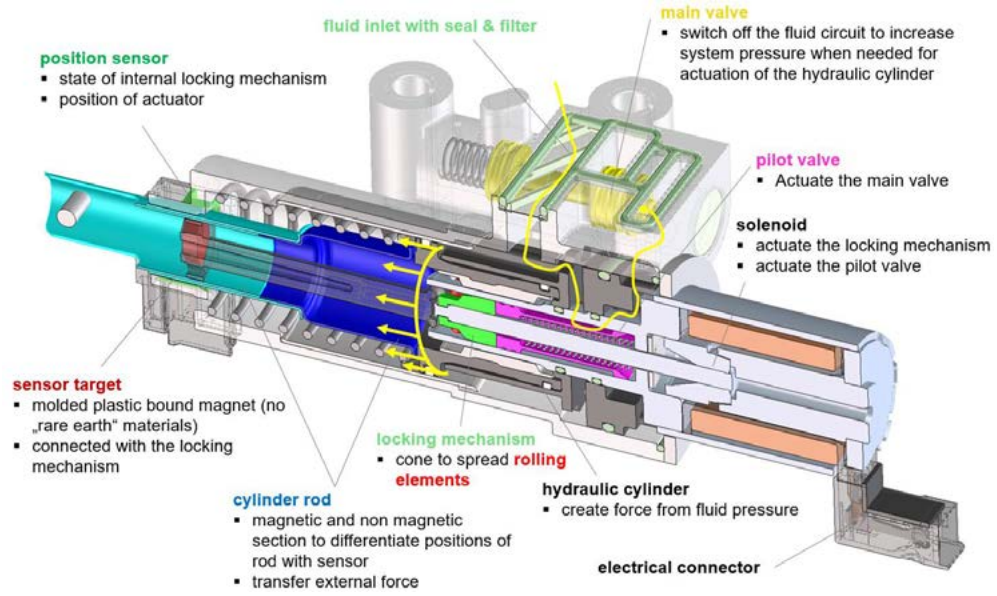


Figure 3.1: Integrated components and functions of the Hydraulic Actuator unit

The subcomponents are either controlled by the axle control unit (ECU) or with an onboard OBD-capable smart control unit as a ninth subcomponent.

All the required innovations of manufacturing technology to build those components were driven by MAGNET-SCHULTZ engineering. Especially the hybrid housing, spool valve and hydraulic manifold from thermoplastic materials and the two-component

ferromagnetic / non-ferromagnetic cylinder rod. All chosen materials have been selected to be free of per- and polyfluoroalkyl substances (PFAS).

The weight of the unit is 0.46kg with a length of 180mm in the retracted position of the cylinder rod.

4. FUNCTION

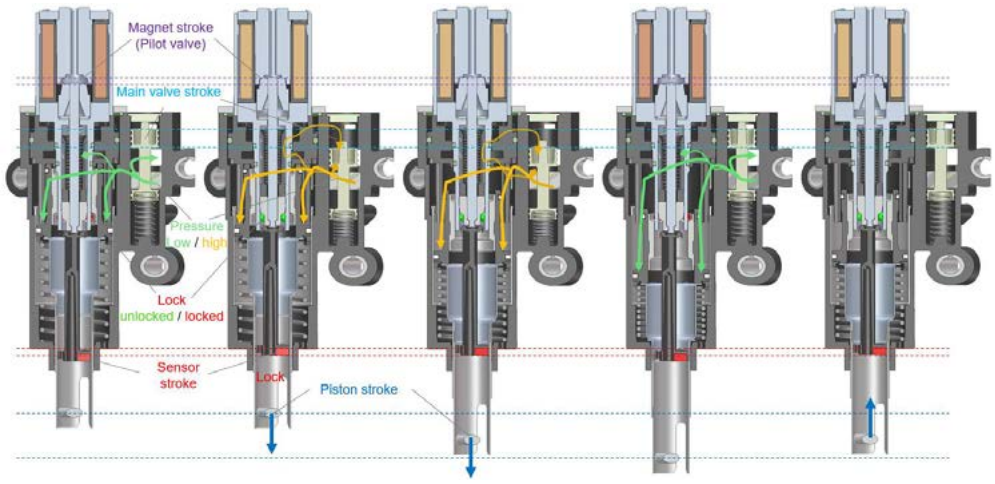


Figure 4.2: Main states of function of the Hydraulic Actuator unit

The HAVSU actuator can switch between five main states of operation.

First stage is the **Park State** (PS).

The solenoid is electrically OFF, no current flows through the coil. The locking mechanism is locked, the pilot valve is open. Pressure in the fluid circuit is low, and pressure drop in the main valve is very low, even if the pump is working at high flowrates. The hydraulic piston-rod remains retracted.

Second stage is the **Pressure Boost State** (PBS).

The solenoid is electrically ON, current is drawn. The locking mechanism, driven through the same armature rod as the pilot valve, is thereby released (unlocked) simultaneously with the opening of this pilot valve. Fluid is drawn from the main line and pilot pressure is applied to operate the main valve. The main valve closes to block the main circuit. Now the pump is triggered for increased performance to work against the dead end (closed line) what causes the pressure to rise immediately.

Third stage is the **Extraction State** (ES).

As soon as the pressure reaches the level needed to overcome the force of the return spring mounted on the cylinder rod, this cylinder rod starts to extract till its end stop. During this movement the boost pressure reaches its maximum.

Fourth stage is the **Drive State** (DS).

After the cylinder rod hits its end stop, the solenoid is switched OFF. The locking mechanism toggles to locked and the pilot valve releases the pilot pressure from the main valve; both through spring pretension. The cylinder rod is locked in the extracted position to remain in drive without pressurizing what means the boost pressure is no longer needed and falls back to normal mode.

Fifth stage is the **Retraction State** (RS).

If the retraction of the cylinder rod is desired, the first cycle “Boost stage” is repeated but with the pump switched OFF, so that the pressure reliefs. The locking mechanism is opened by the energized solenoid. With decreasing pressure and thus hydraulic force the return spring can retract the piston together with the cylinder rod to the retracted position. Afterwards the actuator returns to park or idle stage. With the solenoid OFF, the piston locked in retracted position, fluid passes the main valve with very low pressure drop if the pump is activated again to fulfill its supposed dedication.

Important note: Because thermal processes are slow compared to the few hundred milliseconds needed to extract the cylinder, this is usually easily acceptable for the main fluid circuit of the cooling-system to be interrupted for this short boost and extraction phase. This is why usually easily accessible cooling or lubrication systems are often appropriate for utilizing a HAVSU.

Dual Sensor Solution

Both independent strokes, the solenoid stroke (corresponds to locked / unlocked) including pilot valve state and that of the piston rod (park / not park) are monitored by only one 2D hall-effect position sensor with only one permanent magnetic sensor target. In known applications two separate sensors with two separate targets for each sensing function were necessary what offers plenty of advantages beginning from the often difficult supply chains of rare earth material for sensor targets or the chips itself and ending with quality by design.

This innovative design becomes possible, because the sensor target is connected to the solenoid on the one hand as well as to a tailored multi-material cylinder rod on the other hand. This cylinder rod is made of materials with different magnetic properties along its longitudinal axis, what enables it to act as a varying magnetic shield between the permanent magnetic sensor target, and the opposite positioned hall-effect sensor. Thus, the hall sensor is exposed to a stroke-dependent magnetic field as illustrated below:

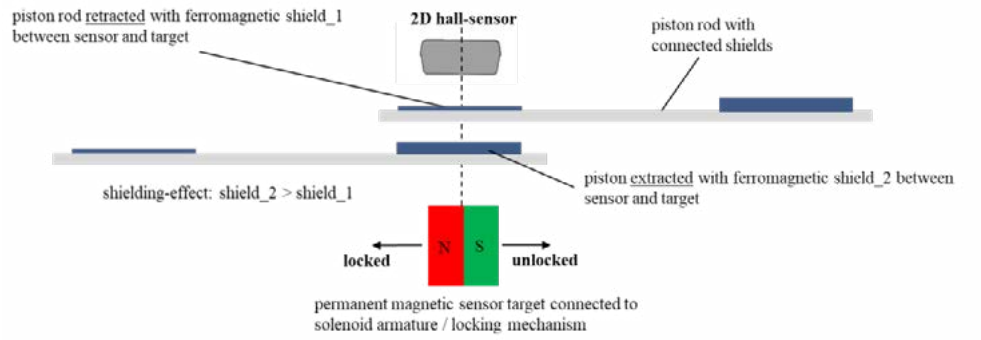


Figure 4.2: Dual sensor principle

To gain two independent positions signals of the 2D hall sensor, it is necessary to evaluate the sensor signal in two different ways. 2D hall sensors can measure the magnetic field strength in two planes perpendicular to each other (Bx, Bz; illustrated below), what is important for the dual sensor principle.

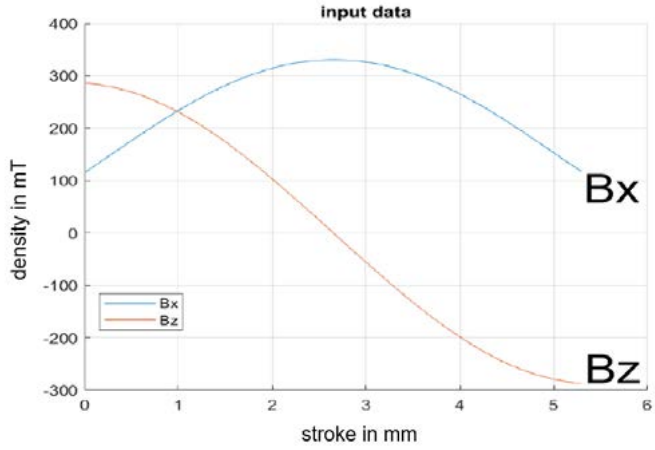


Figure 4.3: Raw sensor signal

For the evaluation of the stroke of the piston rod (retracted / extracted) it is necessary to calculate the stroke-dependent (through varying shield effect) absolute field strength of the target, which corresponds to the vector sum of the two evaluation planes of the 2D hall sensor:

$$B_{abs} = \sqrt{{B_x}^2 + {B_z}^2}$$

For the evaluation of the stroke of the solenoid (locked/ unlocked) it is necessary to calculate the stroke-dependent (through varying shield effect) angle between field strengths of the two perpendicular to each other evaluation. This angle is fully independent from the absolute field strength and thus from the shielding effect which is necessary for the before mentioned evaluation for the position of the piston rod. The angle corresponds to the trigonometrical relationship:

$$\alpha = atan2(B_x; B_z)$$

The Dual-Sensor Solution as well as the whole system Actuator-Valve-Sensor-Unit are patent pending.

5. PERFORMANCE DATA AND TEST RESULTS

The test results with the mock-up samples of the actuator-unit showed great success. The tests included basic functional test, durability test, environmental testing from -40 to +150°C and in-system tests.

The following figure 5.1 shows a typical operating cycle and table 5.1 gives an overview of the technical data measured.

Durability tests confirmed the expected life-length and environmental testing showed no malfunctions, e.g. in thermal cycling or under vibrations and mechanical shocks acc. to well-known automotive standards like LV124.

The successful results have been confirmed by the system-level test and were honored by customers.

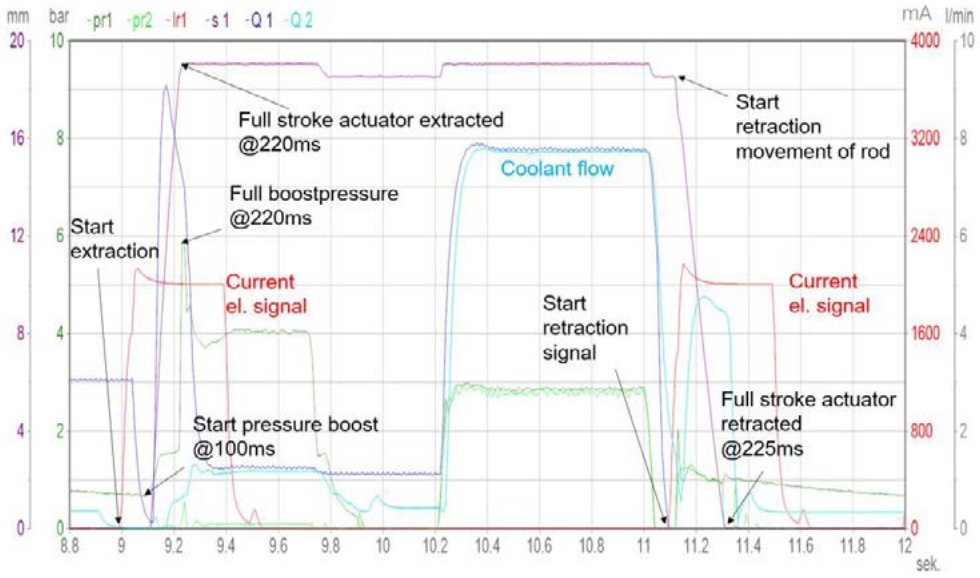


Figure 5.3: Measurement el. current, flow-rate, pressure and stroke over time @ 45°C

Characteristic	Unit	Typical spec.	MSM-HAVSU
Supply Voltage	V	24 / 48	24 / 48
Operating current	A	<3	1.5
Pull in current	A	<3	1.5
Holding current	A	<0,6	0.2
Switching time solenoid (locking mech.)	ms	<50	<20
Pressure build up time (boost)	ms	-	100
Stroke hydraulic	mm	15…50	19.5
Piston area (specific type; scalable at will)	mm²	-	780
Force potential (scalable at will)	N/bar	-	78
Burst pressure	bar	>60	>60
Max. operating pressure	bar	-	20
Max Force (scalable at will)	N	1000	1560
Switching hydraulic (extracting)	ms	<300	<100…220(T)
Switching hydraulic (retracting)	ms	<300	<100…220(T)

Flow-rate main valve (e.g. cooling circuit)	l/min	-	8…15
Pressure drop	bar	<0,1	<0,1
Leakage idle mode @ 1 bar	ml/min	-	<5
Leakage during extracting phase @ 10 bar	ml/min	-	<50
Weight (specific type)	kg	-	0.46

Table 5.1: Summary technical data

The actuator meets all typical specifications regarding dynamic stroke behavior. Especially the very low leakage in idle mode results in very high efficiency.

6. EXAMPLES FOR APPLICATIONS

Lockable linear actuator units including position feedback are highly suitable for applications requiring:

(1) High force density without a full hydraulic infrastructure.
(2) Precise positioning and closed-loop control.
(3) Fail-safe locking for safety-critical systems.
(4) Energy efficiency through bistable lockable hydraulic cylinder which only needs hydraulic actuation for change of state.

The unique advantage of this system is the operation without a dedicated high-pressure hydraulic supply through utilizing an existing low-pressure fluid circuit, e.g., for cooling or lubrication. Dedicated on-demand pressure boosting makes these actuators perfect for modern architectures in many next-generation platforms (e.g. trucks, buses, aerospace, defense, automotive) by moving towards more electric systems and reducing or removing centralized hydraulic supplies.

However, certain functions still require high forces that pure electromechanical actuators struggle to deliver efficiently. By tapping into an existing low-pressure circuit (e.g. 1–3 bar) and locally boosting pressure (e.g. up to 10–50bar), the actuator eliminates the need for a full hydraulic system with reduced weight, complexity and costs as well as minimizes leakage risk and maintenance combined with the typical high force capability and power density of hydraulic actuators.

Possible application fields and examples:

Automotive

- **Transmission Actuation**, Gearshift or clutch actuators in hybrid/EV drivetrains where no hydraulic system exists.
- **Parking Lock Mechanisms**, securely locking the driveline without relying on mechanical-only solutions. Figure 6.1 illustrates the HAVSU in one of our already a realized application.
- **Active Aerodynamics**, spoilers or air flaps that need high force and locking at extreme speeds without adding hydraulic complexity.
- **Cooling on Demand** solutions where the airflow through radiators is a controlled for expample by air flaps to optimize overall aerodynamic drag losses.

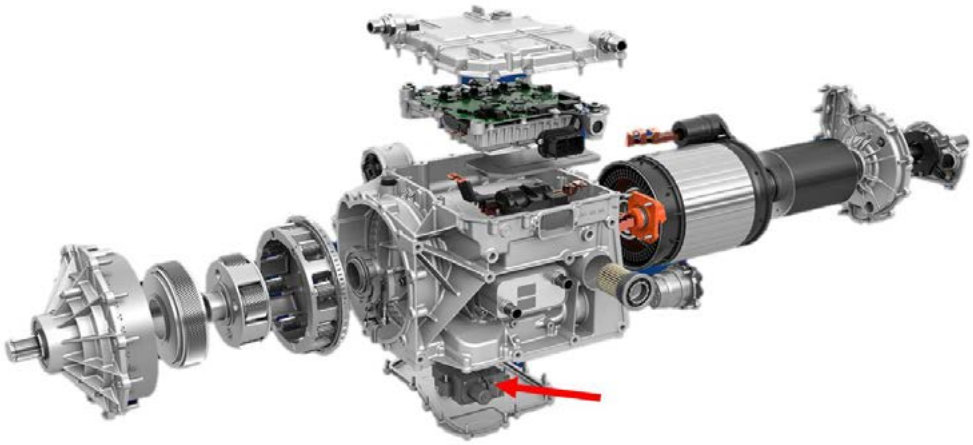


Figure 6.1: Actuator as park-lock-application in e-axle, image courtesy of ZF Friedrichshafen AG [1]

Truck & Bus, Rail and other On- and Off-Highway applications

- **Steerable Axles and Axle Locks,** for controlling auxiliary steering axles and lock them at highway speeds without a central hydraulic system.
- **Suspension Height Adjustment**, for suspension systems, actuators can provide precise height control and secure locking during loading and unloading.
- **Heavy-Duty Doors and Ramps**, passenger doors or wheelchair ramps require strong actuation and safe locking, ideal for local hydraulic boosting.
- **PTO clutches** for distributing mechanical power to auxiliary units.
- **Thermal flow control** especially for large heating circuits, e.g. bus passenger compartments

Aerospace

- **Door-actuators** could extend/ retract and lock doors in position, leveraging low-pressure cooling circuits for energy supply.
- **Flight Control Surfaces**, high-force actuation with fail-safe locking for flaps or stabilizers in "More Electric Aircraft" designs.
- **Cargo Doors**, large and heavy doors require strong actuation and locking, perfect for local pressure boosting.

Defense

- **Weapon Stations**, positioning and locking of turrets or launchers without a full hydraulic network.
- **Armored Vehicle Hatches and Ramps**, heavy hatches need secure locking and high force in compact systems.
- **Stabilization Systems,** radar, missile or laser weapon platforms requiring precise positioning and locking under harsh conditions.

7. BENEFITS OF THIS INNOVATION

The integration of all necessary functions is successfully combining the proven advantages of hydraulics (high power density, robustness) with an innovative layout providing significantly reduced complexity and CO_2 footprint of supply chain, system manufacturing and system operation.

The key advantages versus purely electric actuators can be summarized as follows:

(1) Higher force density: Hydraulic boosting enables compact actuators to handle heavy loads.
(2) Fail-safe locking: Integrated hydraulic lock ensures safety even in power loss scenarios.
(3) Energy efficiency: Uses existing low-pressure fluid, reducing electrical power demand.
(4) Simplified architecture: No need for centralized hydraulic systems, ideal for electrified platforms.

The benefit also includes the reduction of weight, overall size, number of parts and assembly steps as well as improved efficiency of electric energy and use of material resources in the system and whole supply chain.

The lean design with highly integrated functions result in a significant cost reduction compared to alternative systems that combine actuators with separate, external valves and sensors.

At least one electromagnetic valve and one sensor could be saved compared to alternative and less integrated solutions, even without taking the pump and further components of a dedicated system into account.

The innovation is showing great benefits and uniqueness for the markets by bridging the gap between great power density provided by hydraulics, functional robustness, energy-savings and cost effectiveness compared to standard actuator technologies.

Further the innovation has the potential to provide substantial competitive advantages in the mass market of next generation electric vehicles (NGEVs) launched in the second half of the decade and beyond, to make such cars affordable to a broader customer base.

The actuator is fully scalable to different force and stroke requirements as well as environmental conditions in automotive, all mobile or industrial applications.

Thus, a high degree of innovation has been achieved and provides high flexibility for additional applications and future innovations and platforms in a variety of industrial sectors.



Figure 7.4: 3D view of the Actuator-Valve-Sensor-Unit (HAVSU)

8. REFERENCES

[1] https://www.zf.com/products/de/cars/edrive/select.html

Biographies



Ralph Wassermann graduated in mechanical engineering in 2006, specializing in precision engineering and fluid systems. His early work focused on electro-pneumatic micro-valves, laying the foundation for his expertise in miniaturized actuator technologies.

Today as Director of Design and Development at MAGNET-SCHULTZ, he leads the creation of innovative electromagnetic components and systems for demanding industrial applications. His work bridges the gap between theoretical engineering principles and practical, production-ready solutions. He uses his ability to integrate mechanical design with fluid dynamics and electromagnetic functionality, driving forward the development of compact, energy-efficient actuators with high reliability and precision. His leadership fosters interdisciplinary collaboration across mechanical, electrical, and manufacturing teams. With a strong focus on quality, functionality, and system integration, he ensures that every product meets the highest technical standards while being cost-conscious and committed to sustainable resource use.

Ralph’s engineering philosophy combines deep technical expertise with strategic thinking and a passion for innovation. His contributions have helped position MAGNET-SCHULTZ as a technology leader in the field of actuator and sensor systems.



Florian Hölzle holds a master's degree in mechanical engineering (2014) and has over a decade of experience in the development and design of electromagnetically actuated hydraulic systems, with focus on automotive powertrain solutions.

Since joining MAGNET-SCHULTZ in 2014, he has combined deep technical expertise with leadership responsibilities, managing complex projects from concept to implementation. His role includes full technical accountability and customer-oriented project management, ensuring innovative, reliable solutions tailored to client requirements.

Florian is recognized for his ability to bring engineering excellence together with strategic customer engagement, driving successful outcomes in a highly competitive automotive sector.

EHAs in mobile machinery: Experimental validation of a high-speed EHA with recuperation and load-holding capability

Felix Figge, Katharina Schmitz

RWTH Aachen University, Institute for Fluid Power Drives and Systems (ifas)

felix.figge@ifas.rwth-aachen.de

Abstract.

The electrification of mobile machinery is advancing, with more machines now featuring battery-electric power sources. However, this shift often results in reduced energy availability and shorter operating hours before recharging is needed. This has led to increased interest in efficient systems driven by the need for energy conservation and emission reduction, alongside relevant legislation.

Traditional hydraulic systems in mobile machinery have demonstrated poor energy efficiency [1,2], highlighting significant improvement potential. Decentralized electro-hydraulic actuators (EHAs) consume less energy than conventional load-sensing systems [2,3], yet their adoption remains limited, possible due to higher costs, unwanted oscillations from mode switching, and increased spatial requirements.

To address these issues, an EHA with a high-speed internal gear unit and electronically controlled valves was developed. The design enables energy recuperation and load-holding while keeping costs low by using a single prime mover and hydraulic unit. High operational speeds reduce installation volume, and the used electronic valves facilitate a control strategy that synchronizes motor speed and valve positions as best as possible. Active hysteresis control and throttling minimize cylinder velocity deviations.

This paper presents the functional model of the EHA and its experimental validation through a hardware-in-the-loop (HiL) test rig. Predefined synthetic test cases are utilized to analyze quadrant switching, demonstrating effective velocity tracking, with minor deviations during transitions. Additionally, measured load cycles of an excavator indicate sufficient position tracking and high efficiency, especially when implemented on the boom.

Keywords. EHA, 4Q-operation, electrification, active throttling, hysteresis control

1. Introduction & State of the Art

Substituting the combustion engine is arguably the most significant change in mobile hydraulics in recent decades. Historically, most machines have utilized a centralized hydraulic pressure supply, with power distributed through fluid lines and valves to various

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

actuators. While this approach is cost-effective in terms of initial component cost, it is relatively inefficient in terms of energy consumption. The different pressure levels required by the actuators during operation, combined with a single pressure supply, inherently result in significant throttle losses. This contradicts the common understanding of the need to conserve energy and reduce emissions, particularly in the context of the lower energy density and higher costs associated with alternative energy sources. Therefore, the hydraulic system of mobile machinery must become more efficient and sustainable.

Decentralized electro-hydraulic actuators (EHAs) can significantly reduce energy losses by using a dedicated pump for each actuator, eliminating throttling losses between varying pressure levels. However, despite their increasing use in stationary industrial applications, EHAs have not yet gained traction in mobile machinery. This is mostlikely due to higher initial component costs, increased installation space requirements, reduced damping linked to lower system losses, and more complex control software. The additional prime mover and hydraulic unit contribute significantly to the higher costs, while the larger footprint complicates integration into existing machines. Latter can be partly counteracted by an increase in rotational speed, reducing the volume of the electric motor and the displacement volume of the pump. Some research even attempted to increase the motor speed further by use of an intermediate gear [4]–[8].
Consequently, this research aims to develop a high-speed closed-circuit EHA for mobile machinery, featuring recuperation and load-holding capabilities, while also establishing the foundation for the necessary software implementation [9]–[12].

Closed-circuit electro-hydraulic actuators (EHAs) with a single pump may display undesirable behavior during quadrant transitions, such as oscillations and frequent mode switches. Pressure feedback is often used to increase the overall damping of the system [13,14], which is typically too low for optimal performance. Williamson and Ivantysynova [15] developed a predictive observer for feed-forward control of pump displacement in skid-steer applications, addressing excessive phase lag that hampers feedback control during mode-switching oscillations. They identified the interaction between moving mass and compressible fluid as a key issue, comparing it to a mass-spring-damper system [13,14].
To mitigate oscillations and prevent mode switches, increasing system damping through pressure feedback is recommended. Michel calculated maximum allowable acceleration to further prevent mode switches [14]. Wang et al. [16] employed electrically actuated proportional valves with a control algorithm to introduce leakage under critical conditions, selectively enhancing damping at low loads. Çalışkan et al. proposed a mechanical solution using an underlapped shuttle valve instead of electrical component [17]. Imam et al. introduced the ‘limited throttling valve’ designed to restrict cylinder outflow at low loads near critical operating conditions, demonstrating no oscillations in simulations [18,19]. This valve is designed to restrict cylinder flow only at low loads near critical operating conditions. Their simulation results show no oscillations under any operating conditions. They also demonstrate that applying different charge pressures to both sides of the cylinder can reduce the critical zone and oscillation amplitudes [19]. Costa and Sepehri [20] introduced a new four-quadrant definition based on cylinder load pressure rather than pressure difference over the pump. They concluded that this was the primary reason for the poor performance of previous circuits and introduced a new circuit that implements their quadrant definition.

However, Gøytil et al. [21] showed that the proposed circuit can still experience unwanted mode switches under certain conditions due to dynamic factors. They recommend using a software-based observer to calculate a modified force estimate, which determines valve states and mitigates these switches. This approach excludes forces from valve switching and acceleration but requires a minimum accumulator pressure of 30 bar to prevent cavitation, which may not be feasible for all circuits. Kärnell and Ericson [22] investigated hysteresis control to prevent unwanted mode switches. They concluded that pressures should be sensed so that the difference remains unchanged after a mode switch. Nevertheless, switches can still occur, but active hysteresis control reduces the risk if the reservoir pressure is sufficient.

The control approach discussed in this paper, first introduced in [9], utilizes the cylinder load pressure definition from Equation 1.1 [17,20] and implements active throttling [19,22] to enable complete coverage of the four quadrants for excavators. To minimize mode switches caused by minor pressure fluctuations, active hysteresis is incorporated [22]. Additionally, the approach minimizes cylinder velocity spikes during quadrant transitions by decoupling the cylinder from the pump using the load-holding valves.

$$p_{load} = p_{pist} - \alpha \cdot p_{rod} \tag{1.1}$$

The paper is structured as follows: Section 2 presents the circuit layout and functional model of the EHA, along with a summary of the control scheme. Section 3 introduces the experimental validation setup, and Section 4 presents the results and discussion.

2. DESIGN & CONTROL OF EHA

In this section, first, the design of the developed EHA and the assembled functional model are introduced. Secondly, the hardware-in-the-loop test rig used for validation of the EHA in all quadrants (Q1 to Q4) under a wide range of operational conditions is presented.

2.1. EHA Design

The EHA design under investigation, introduced in [9], is specifically tailored for mobile machinery integration, particularly excavators. Key requirements included minimizing initial costs, achieving a wide range of operating conditions with high efficiency, and enabling complete stoppage of the prime movers. As shown in the circuit layout at the top of Figure 2.1, the design features only a single prime mover, one pump, and a standard differential cylinder. The accumulator stores the differential cylinder volume by switching the 'differential' valves ($Diff_{pist}$ and $Diff_{rod}$). Load-holding capability is realized by two electric actuated proportional valves denoted LH_{pist} and LH_{rod}. The use of proportional valves is required to allow throttling of the cylinder outflow and thereby expanding the range of Q1 and Q3 operation, as detailed in [9]. Otherwise, throttling losses from the load-holding valves prevent high velocities in combinations with low assistive loads, compromising full four-quadrant operation essential for mobile machinery's variable working conditions.

The functional model of the EHA is shown at the bottom of Figure 2.1, with key components labeled. A PMSM electric motor was chosen for its low operating voltage (48 V) and maximum rotational speed of 7,200 rpm, constrained by inverter dual-use regulations. Due to the limited availability of motors that meet these criteria, the motor is oversized with a continuous power rating of approximately 10 kW. In contrast, the peak power output during

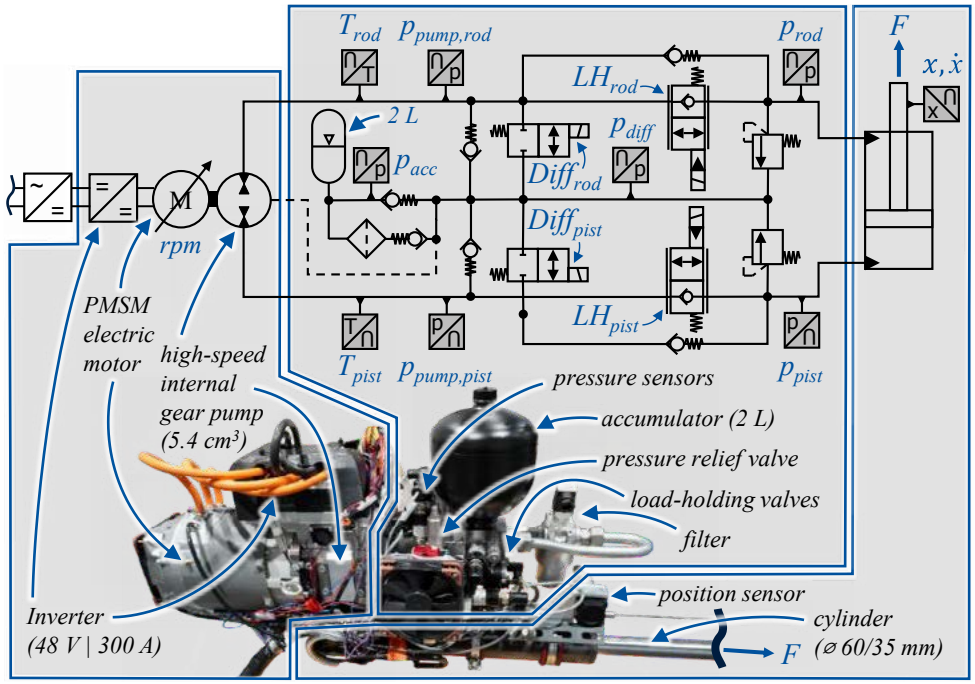


Figure 2.1. Circuit layout and functional model of the high-speed EHA

a dig and dump cycle on a 1.8-ton excavator's arm is around 4 kW. Consequently, software limits the motor's output to 6 kW to account for power losses. The hydraulic unit features an internal gear design with a displacement volume of 5.4 cm³ and is optimized for operation up to 10,000 rpm. The pump is flange-mounted to the valve block, compactly interconnecting it with the valves, sensors, and accumulator hydraulically. Additionally, check valves parallel to the load-holding valves prevent cavitation in the cylinder while another pair near the filter and accumulator directs flow through the filter.

To achieve high rotational speeds without cavitation, a minimum absolute inlet pressure of 1.2 bar is required at the pump. The accumulator serves two functions: it stores the cylinder's differential volume (0.358 L) and maintains this minimum pressure. However, the pump shaft seal limits the maximum pressure to 10 bar. To operate within a temperature range of -10°C to +60°C, the accumulator needs a nominal volume of 2 L, achievable only if check valves between the accumulator and pump inlet have minimal pressure drop. Higher pressure drops reduce the effective pressure difference, necessitating a larger accumulator volume for storage. Therefore, check valves are installed parallel to the load-holding valves to minimize pressure drop, allowing for a lower minimum accumulator pressure and increasing the maximum pressure difference, which reduces the required nominal volume.

2.2. Control scheme

The control scheme is presented and evaluated using simulations in [9]. The most important features are summarised briefly below.

The operating boundaries, which are shown in Figure 2.2, are based on cylinder friction ($p_{bound,fric}$), accumulator pressure ($p_{bound,acc}$), and the pressure drop over the load holding

valve ($p_{bound,valve}$) due to the circuit layout. The latter is problematic during motor operation due to the minimum pressure required for a given flow rate. High velocities cannot be achieved under low loads. However, since mobile machinery is used under a wide range of operating conditions that cannot be predicted in advance, an inoperative load-velocity combination is unacceptable. Therefore, the load pressure operating range of pump operation (Q1 and Q3) is expanded by actively throttling the cylinder outflow using proportional valves, in a manner comparable to metering-out systems. This enables displacement control in regions where valve losses would otherwise limit it. The aim is to maintain a fixed pressure difference over the load-holding valves using a characteristic field in combination with a PI controller. By engaging this only in regions close to the boundary, the influence on efficiency is minimized.

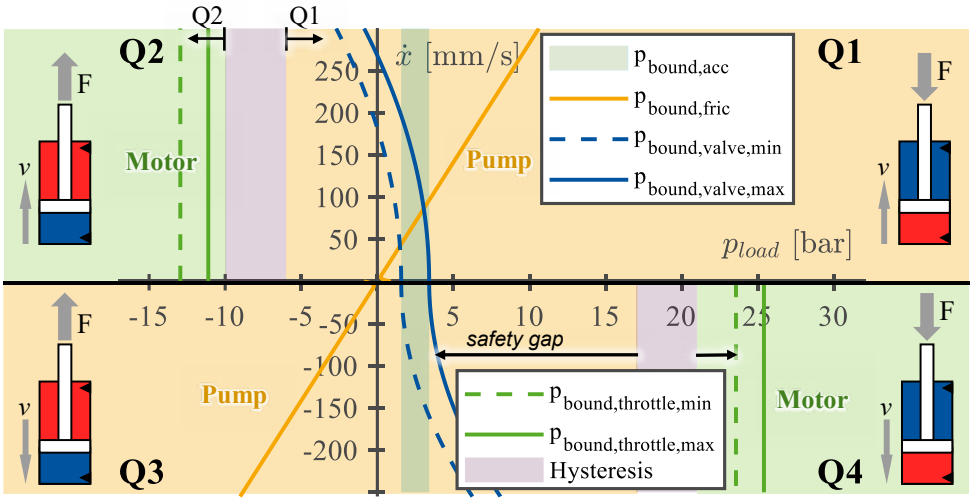


Figure 2.2. Quadrant definition used by the new control approach

However, switching from displacement to valve control still occurs at higher loads. This is denoted as the throttle boundary ($p_{bound,throttle}$) in Figure 2.2, where the 'min' and 'max' indices indicate dependency on the corresponding accumulator pressure limits. If the load pressure exceeds the current limit, the constant pressure difference request of the load-holding valves will cause the cylinder to accelerate. Therefore, switching to motor operation beforehand is necessary. Active hysteresis, marked in violet in Figure 2.2, is used to determine when the quadrant switch must be executed. During cylinder extension, mode Q1 is selected if the load pressure is above -10 bar. If the load pressure falls below this threshold, the system switches to mode Q2. Switching back to mode Q1 requires an increase in load pressure exceeding -6 bar. The positioning and width of the hysteresis band depend on the specific application and the optimization of performance. The configuration shown is for an excavator arm. Therefore, the hysteresis band is placed closer to the throttle boundary because the centre of mass generally does not cross top dead centre. This means that switches Q1→Q2 or Q3→Q4 generally do not occur. This placement enables operation under all working conditions and maximizes the safety gap, corresponding to the time between reaching a hysteresis edge and its corresponding boundary, during switches from Q2→Q1 and Q4→Q3. Thereby allowing more time for adjustment of valve positions.

The control scheme features two state machines. One determines the current operation quadrant based on the quadrant definition with active hysteresis, as shown in Figure 2.2. The other synchronizes motor speed and valve positions during quadrant switches. The hydraulic unit's volumetric efficiency, pressure feedback, and valve controllers are implemented to enable the different required valve modes, as well as to provide overload protection for the electric motor. A more detailed description is given in [9]. As described there, the control scheme used in this paper was improved to also allow for smooth transitions Q1↔Q3 with recuperation and improved pre-pressurization before motor operation.

3. EXPERIMENTAL VALIDATION

Experimental validation is performed using a hardware-in-the-loop test rig. The tests include synthetic test cases that focus on the quadrant switches, as well as different load cycles that validate the application's performance in excavators. Additionally, a quadrant-based efficiency map and the efficiencies under different load cycles are calculated.

3.1. Hardware-in-the-loop test rig

The hardware-in-the-loop (HiL) test rig, introduced in [23], primarily consists of two load cylinders and associated servovalves, as shown in Figure 3.1. The load cylinder rods are mechanically linked to the EHA cylinder rod via a movable crossbeam, enabling the application of pulling or pushing forces. The actual force on the EHA differs from the hydraulic force due to friction, which is measured by a load cell. While the friction forces from the load cylinders increase system damping compared to excavator applications and the moving mass of the Hil of roughly 100 kg is lower than that of an excavator arm, the HiL setup allows for repeatable and adjustable testing conditions for the EHA.

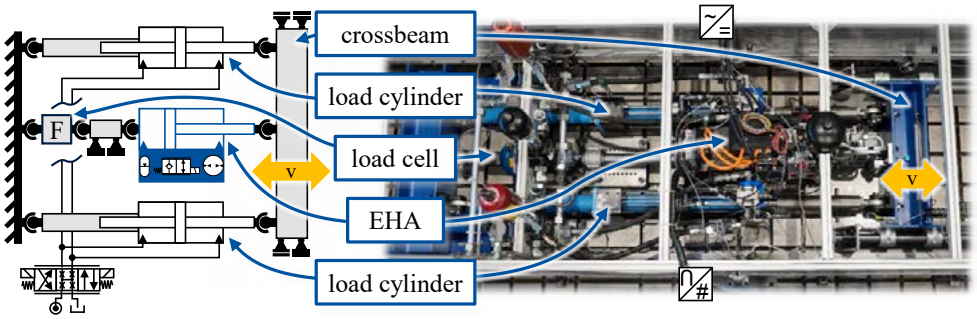


Figure 3.1. Hardware-in-the-loop (HiL) test rig (simplified hydraulic circuit layout)

Although controlling the HiL test rig is not the focus of this paper, a few important notes are provided below. The system is rather stiff due to the hydraulic tension between the load cylinders and the EHA, which complicates force control. Small errors in the flow rate of the valve-controlled cylinders result in significant force changes due to the system's high stiffness, which is caused by the bulk modulus of the oil. To achieve acceptable force control, it is necessary to predict the cylinder flow rate, and thus, cylinder velocity, to enable feed-forward control of the rough valve position. A PID controller then fine-tunes the position to achieve good force accuracy. However, for feed-forward control, simply calculating velocity by differentiating the position signal introduces lag due to the required

filtering. This lag results in temporary, large differences between the required and actual valve positions, leading to significant force deviations. Therefore, an approach is necessary that predicts cylinder movement based on pump speed and valve actuation. Valve actuation must be considered separately due to temporary switching to valve control of cylinder speed, for example, when stopping. A perfect prediction still was not achieved.

3.2. Synthetic test cases

The synthetic test cases examined are depicted in Figure 3.2 and are identical to those previously analyzed through simulations. The first test case is called 'velocity reversal', representing quadrant switches by reversing the direction of movement under constant load force F. The second test case entails a force reversal, which linearly alters the load force during movement at a constant velocity. Here, F_{mean} represents the midpoint of the hysteresis band, and ΔF is half the change in force over a duration of 0.5 seconds.

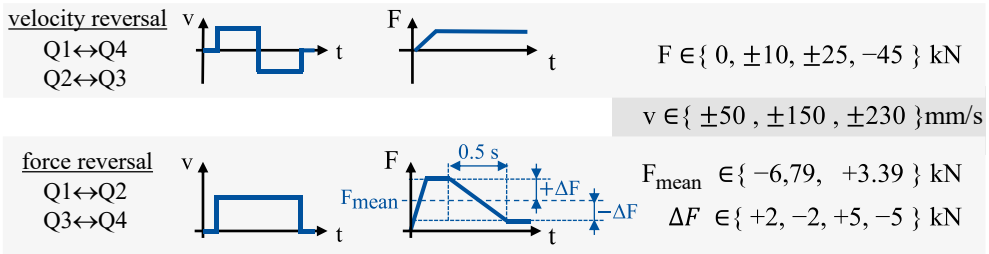


Figure 3.2. Synthetic test cases

3.3. Load cycles

The authors had access only to a measured dig and dump load cycle of a 1.8-ton excavator. To evaluate the EHA under a broader range of working conditions, load cycles from the STEAM research project were scaled to fit within operational limits:

1. The load pressure profile is scaled to not exceed the maximum operational pressure.
2. The cylinder position is filtered and differentiated to obtain the velocity profile, which is then scaled to remain within the maximum operational velocity.
3. By integrating the scaled velocity profile, a new cylinder stroke profile is generated.
4. If the cylinder end stops are reached, the time is adjusted to reduce the duration of EHA movement, thereby decreasing the maximum stroke.

It is important to note that this approach modifies not only the absolute values of the profiles but also their gradients. However, these profiles remain related to the measured data, and such alterations appear acceptable given that profiles can vary significantly between different machines as well as depending on current work conditions and operator actions.

4. RESULTS & DISCUSSION

The cylinder velocity in the following results is calculated by filtering the measured position signal. A significant challenge in evaluating performance is achieving operator acceptance during EHA operation, which is inherently subjective. The authors were unable to identify a meaningful method to assess this type of acceptance for mobile machinery and strongly recommend that it be evaluated separately. Consequently, the performance rating in this paper is based on the calculated velocity and the authors' considerations.

4.1. Test cases

Figure 4.1 displays the results of the velocity reversal test case. The results are divided into six subgroups based on the initial velocity. The top graph in each group displays the cylinder velocity, and the bottom graph shows the calculated hydraulic force based on the load pressure and the cylinder piston area. Positive force values represent a pulling force at the cylinder.

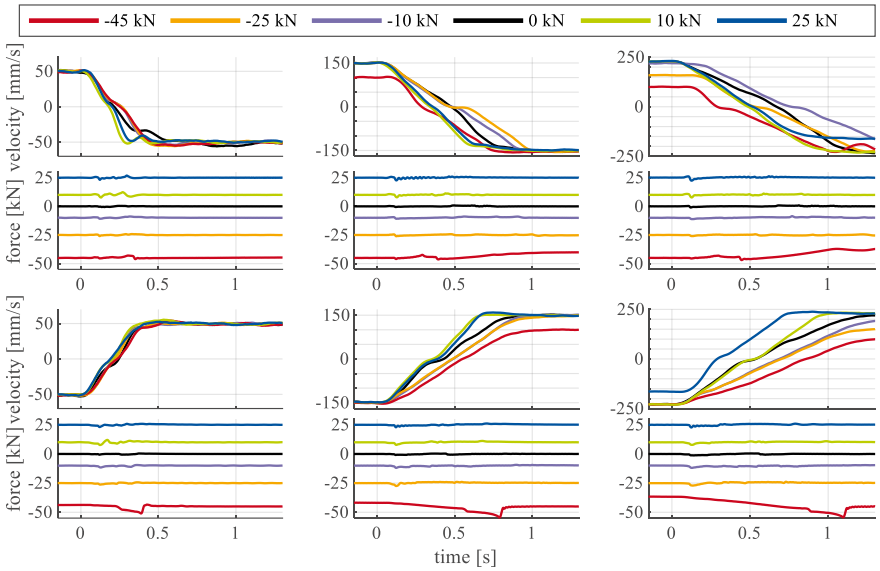


Figure 4.1. Experimental results for test case 'velocity reversal' (force based on p_{load})

As intended, the forces during the reversal are constant, except for the -45 kN case, which shows smaller deviations. As can be seen from the velocities, high velocity and load combinations result in a power limitation of the motor and, therefore, a smaller-than-requested cylinder velocity. Negative forces correlate with a loaded piston side, requiring higher flow rates and rotational speeds. Therefore, the torque capacity of the electric motor is reduced further, resulting in slower acceleration and deceleration. The top group of graphs represents positive initial velocities. As expected, a linear velocity change is achieved for positive pulling forces (Q2→Q3) and zero load (Q1→Q3). For negative forces (Q1→Q4), the designed rest phase, which allows time for LH_{rod}, $Diff_{rod}$, and LH_{pist} to open and close, respectively, can be seen at about zero velocity.
The bottom group represents negative initial velocities. For these, a negative pushing force (Q4→Q1) results in a linear velocity change. Positive forces (Q3→Q2) and no load (Q3→Q1) demonstrate the intended rest state, allowing time for LH_{pist} to open and LH_{rod} and $Diff_{rod}$ to close.

In conclusion, the velocity reversal test cases show the desired and expected behaviour. The velocity reversal is smooth and does not show strong oscillations.

The results for the 'force reversal' test case are displayed in Figure 4.2, grouped by the corresponding quadrant switch. The graphs on the left side show results for zero load. In contrast to the velocity reversal, larger deviations from the set force (dashed black line) are

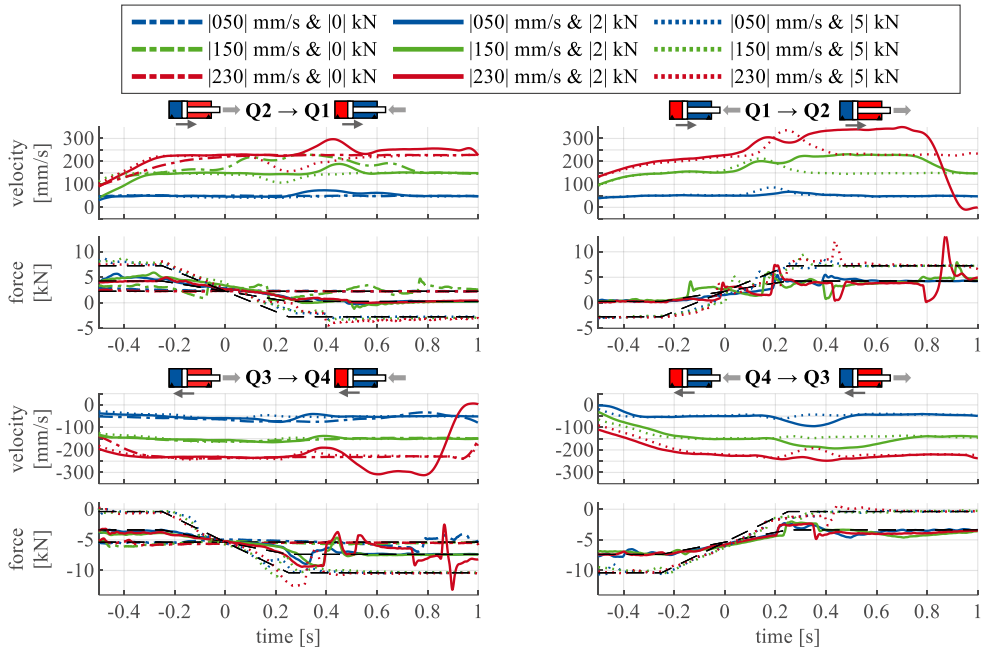


Figure 4.2. Experimental results for test case 'force reversal' (force based on p_{load}) (set force marked with a black dashed line)

evident in some test cases, particularly during Q1→Q2 and Q3→Q4 at high velocities. The interaction between the EHA and force control of the HiL during initial switches leads to deviations so significant that they can trigger additional switches.
As previously mentioned, force control in the HiL is challenging due to system stiffness combined with changing valve states, making it difficult to accurately identify the root cause of these deviations. However, despite these challenges, good velocity tracking is achieved when examining the velocity graphs of cases with effective force tracking.

For switches Q2→Q1, the no-load case at 150 mm/s stands out as an outlier in terms of velocity tracking, which can be attributed to mode switches resulting from poor force tracking. In contrast, the other no-load test cases demonstrate good tracking. When force changes of 2 kN occur, short and minor overshoots are observed due to the motor speed increasing while still supporting some cylinder load, coupled with the valve not being in the correct throttling position because of its dynamics. At higher change rates, the force begins to slow down the cylinder sooner, and the pump speed fails to increase quickly enough to supply the required power. Consequently, deviations increase with rising cylinder velocity due to heightened friction.
During the Q1→Q2 switches, initial overshoots are observed due to the assistive load accelerating the cylinder before displacement control is reestablished through a slowdown of the hydraulic unit. Faster force changes lead to greater deviations because of increased acceleration, while not under displacement control.
Q3→Q4 switches show small negative deviations at lower velocities and smaller load changes. This is attributed to the assistive load still being applied while the valve lags before throttling and slowing down the cylinder.

For Q4→Q3 switches, overshoots appear for small force change rates, again due to assistive load being applied and valve lagging before throttling can slow down the cylinder. The prolonged deviation at -150 mm/s with 2 kN can be traced back to an extended adjustment period of the valve position by the PI controller, likely resulting from inaccuracies in the characteristic field that necessitate longer adjustments.

The results for most cases qualitatively match the simulation results presented in [10]. Exceptions include the Q2→Q1 switches with slow change rates, which show an overshoot instead of the simulated undershoot, and the Q1→Q2 switches with slow change rates, where experimental results reveal an overshoot while simulations indicate almost no deviation. These discrepancies may stem from differences in moving mass and friction between the simulation and the hardware-in-the-loop (HiL) test rig.

In conclusion, the force reversal test cases demonstrate that quadrant switches, excluding those with problematic force application due to coupling between HiL and EHA, can be executed with only short and limited deviations in the velocity profile. Although perfect tracking is unattainable due to valve lag and rotational speed adjustments, it has been demonstrated that the developed control scheme can effectively mitigate the severity of these issues.

4.2. Efficiency map

Figure 4.3 shows the efficiency map of the EHA, which was calculated based on measurements taken under quasi-stationary conditions. The cylinder is loaded, accelerated, and stopped again. Due to the limited cylinder stroke, the 400 ms period before deceleration is evaluated. The efficiency is interpolated between the marked measurement points, resulting in anomalies at the outer edges of the efficiency field due to the algorithm used. Additionally, reduced motor performance is evident at higher speeds in Q1, Q3, and Q4. Efficiency is calculated using Equation 4.2, which considers mechanical (Equation 4.3), electrical (Equation 4.4), and hydraulic energy from the accumulator (Equation 4.5). Therefore, all system losses, including those of the inverter, electric motor and cylinder, but excluding those of the valve actuation, are considered. Depending on the sign of the total transferred energy, the different parts are considered output or input in Equation 4.2.

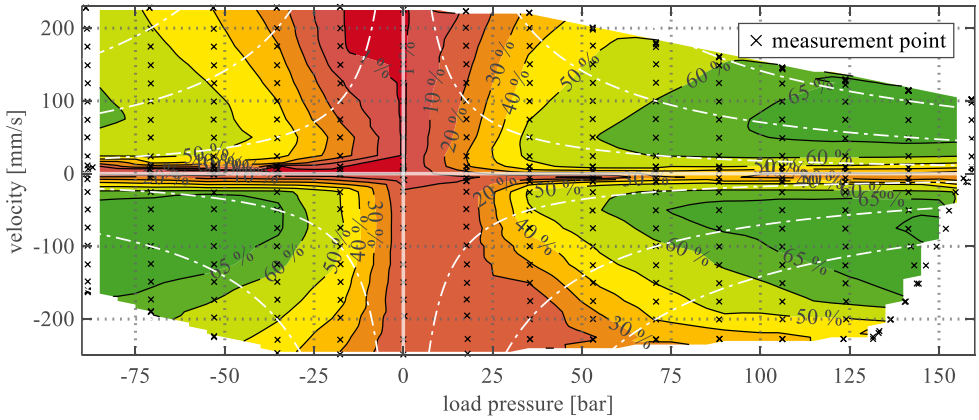


Figure 4.3. Efficiency map of the EHA (isolines for cylinder power neglecting friction of 0.5, 2 & 4 kW)

$$\eta = \frac{E_{out}}{E_{in}} \tag{4.2}$$

$$E_{mech} = \int F_{cyl} \cdot v_{cyl}\, dt \tag{4.3}$$

$$E_{elec,powersupply} = \int U_{sup} \cdot I_{sup}\, dt \tag{4.4}$$

$$E_{acc} = \int Q_{acc} \cdot p_{acc}\, dt \quad \text{with} \quad Q_{acc} = (A_{pist} - A_{rod}) \cdot v_{cyl} \tag{4.5}$$

Except for Q2, the highest efficiencies per quadrant exceed 65% for moderate velocities and high loads. The added isolines for cylinder power outputs of 0.5 kW, 2 kW, and 4 kW indicate that areas with poor efficiency correlate with low power transfer, resulting in minimal absolute power loss. The regions of low efficiency at low loads in Q2 and Q4 correspond to the expanded areas in Q1 and Q3, where a significant portion of energy input is lost due to active throttling. In Q2 at high velocities, there is no power output from the cylinder due to friction; instead, all power input from the pump and accumulator is converted to heat. In contrast, Q4 exhibits higher efficiency because the accumulator is charged.

The impact of load-holding valve losses, along with cylinder and pump friction at higher speeds, is evident. The pressure difference across the load-holding valves is particularly relevant in Q1 and Q3, as it remains constant, leading to linear increases in losses with increasing cylinder velocity. Interestingly, efficiencies in Q1 and Q4 are comparable; however, this is not the case for Q2 and Q3. This discrepancy most likely arises from the interplay between the hydraulic unit's efficiency and the efficiency of energy storage in the accumulator. In Q2, these effects negatively combine, while in Q4, the higher efficiency of energy storage counteracts the lower motor operation efficiency.

4.3. Load cycles

Figure 4.4 shows the results of the 1.8-t dig and dump cycle. The results are limited to the first 13 seconds to allow for a higher level of detail. A simple outer PI controller was added to adjust the requested speed, which was calculated offline beforehand, based on the deviation between the requested and current position during the tests. These values, as well as accumulator pressure, are shown in graph 1 (see grey box in Figure 4.4). A maximum RMSE of 75.45 mm and an ITAE of 34 ms^2 are achieved over the 55 second cycle. The deviations can be traced back to the cylinder velocities shown in graph 5. There is a delay between the request and the actual velocity. Examining the deviation between the cylinder velocity based on the actual pump speed, neglecting volumetric efficiencies (f(rpm)), and the current cylinder velocity, it can be concluded that the displacement control is functioning as intended. However, especially during deceleration phases, which are not subject to delays by the control scheme, a delay of up to 200 ms can be observed between the cylinder speed request and the rotational speed adjustment. Approximately 50 ms of this delay is related to the time between the speed request and the motor's rotational speed adjustment.

Graph 2 illustrates the pressures in the cylinder chamber and the hydraulic unit ports. Following each entry into a quadrant (as shown in Graph 3), higher pressure levels are initially observed due to the delayed adjustment of the load-holding valve position. This phenomenon is further evidenced in graph 4, which displays the adjustment of current over

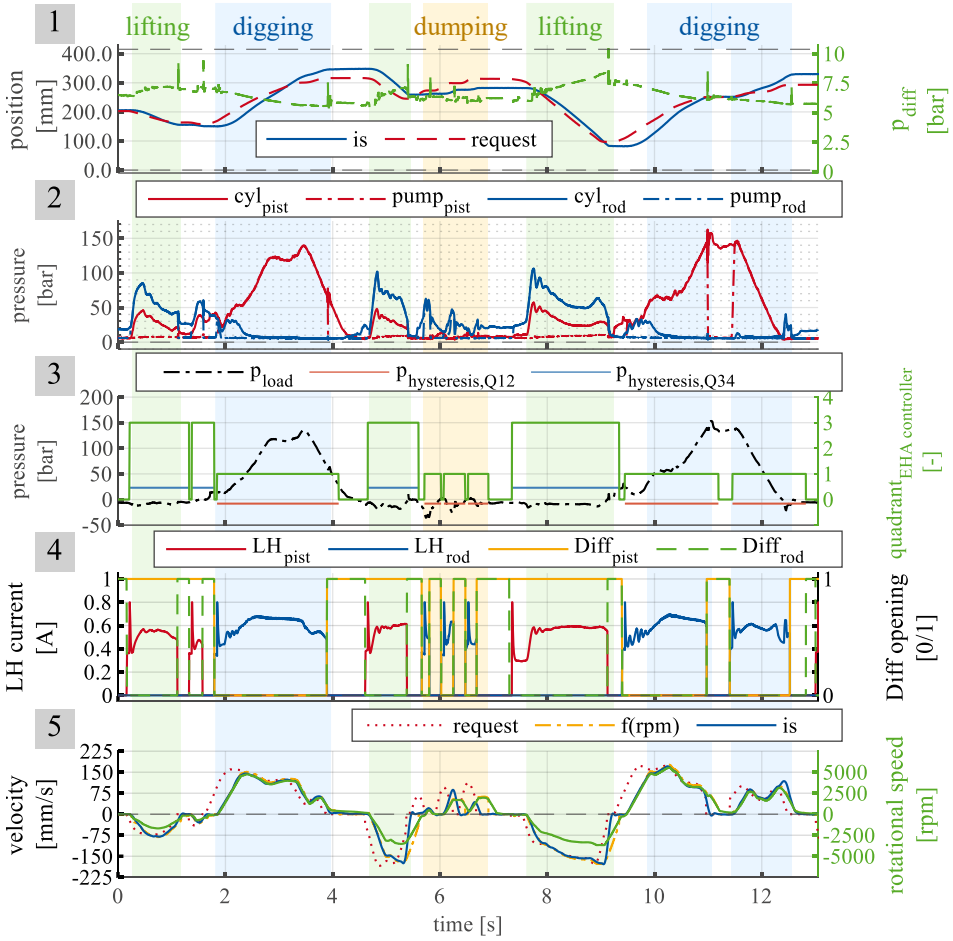


Figure 4.4. Experimental results of the 1.8-t excavator dig and dump cycle with outer position controller

time alongside the actuation signal of the differential valves. The initial peak in the actuation signal for the load-holding valves at maximum current is designed to open the valve promptly, preventing hydraulic tensioning of the cylinder caused by obstructed outflow.

Shortly after 6 seconds and again at 12 seconds, peaks in the cylinder velocity request are observed. At these moments, the cylinder speed exceeds that based on pump speed, indicating a loss of displacement control. During these instances, the load pressure drops below the hysteresis band between Q1 and Q2, as shown in Graph 3. This triggers a mode switch towards the motoring quadrant (Q1→Q2), which is associated with an overshoot, as previously illustrated in Figure 4.2.

It is important to note that the load cycle was measured in a valve-controlled system. The measured load is applied based on time without feedback from the cylinder position. Consequently, some working conditions may differ when implemented on a real excavator due to changes in system response. Nevertheless, the overall tracking of the reference load

cycle is achieved, albeit with some deviations primarily attributed to initial delays inherent in the control scheme and partially due to delays in the motor response.

The authors acknowledge their lack of a meaningful metric for assessing performance in the context of mobile machinery. A review of existing research on teleoperation, as referenced in [24], indicates that latencies below 250 ms or 300 ms are acceptable, with less than 170 ms being preferable for autonomous vehicles. Ideally, the time between input and output should be under 80 ms. While direct transfer may not be entirely suitable, the delays experienced by the EHA generally fall within an acceptable latency range; however, input-output delays occasionally exceed this threshold.
Given that operators can adjust for these delays, the successful application of this technology in an excavator appears feasible. Nonetheless, further testing with real human operators in an excavator is necessary and recommended to gather subjective feedback.

The efficiency of the developed EHA is evaluated based on the available load cycles. The energy transfers considered include electrical energy from the DC side of the power supply, hydraulic energy from the accumulator, and mechanical energy from the cylinder. Energy transfer is aggregated for input and output per system boundary by calculating power transfer at each sample point, multiplying it by the time difference between samples, and then summing all negative and positive values separately (eq. (4.3)-(4.5)). Since operation during the cycles is not quasi-stationary, the flow rate is estimated based on current quadrant operation and valve actuation, which introduces some inaccuracies due to partially open valves and volume changes resulting from compression.

As discussed in [11], there is currently no consensus on how to calculate the efficiency of systems with recuperation capability. Therefore, Table 4.1 presents three different definitions based on the experimental results from the adjusted STEAM load cycles, along with the calculated energy transfers and total energy loss. The very limited bucket movement during air grading leads to questionable efficiency results.

Table 4.1. Transferred energies (negative values exiting the system), efficiencies and dissipation ratios for the adjusted 18-t STEAM load cycles per actuator

	Dig and dump			**Air grading**			**Excavating**		
	boom	arm	bucket	boom	arm	bucket	boom	arm	bucket
$E_{cyl,out}$ [kJ]	-144.3	-24.9	-50.5	-55.0	-0.7	-0.04	-69.2	-41.1	-34.3
$E_{cyl,in}$ [kJ]	113.8	21.5	9.5	48.6	7.6	0.26	65.1	4.5	10.0
$E_{elec,in}$ [kJ]	237.2	122.5	168.0	88.9	51.2	0.66	127.0	129.5	115.0
$E_{elec,out}$ [kJ]	-53.9	-21.3	-7.7	-30.2	-7.9	-0.08	-39.3	-10.1	-6.3
$\sum E_{acc}$ [kJ]	-0.88	-2.04	-0.75	-0.06	-0.88	-0.05	-0.50	-0.38	-0.66
E_{loss} [kJ]	151.8	95.7	118.5	52.3	49.3	0.75	83.2	82.3	83.8
$\eta_1 = \frac{\lvert E_{cyl,out} \rvert}{\sum E_{elec} + \sum E_{acc}}$	79.1 %	25.1 %	31.6 %	93.7 %	1.6 %	*8.1 %*	79.3 %	34.5 %	31.7 %
$\eta_2 = 1 - \frac{E_{loss}}{E_{elec,in} + E_{acc,in} + E_{cyl,in}}$	57.1 %	33.9 %	34.2 %	62.3 %	16.6 %	*19.3 %*	57.1 %	39.3 %	33.9 %
$DR = \frac{E_{loss}}{\lvert E_{cyl,out} \rvert + \lvert E_{cyl,in} \rvert}$	0.59	2.06	1.98	0.50	5.98	*2.48*	0.62	1.81	1.89

The first definition, η_1, considers only the positive cylinder output in relation to the total energy input. The second definition, η_2, incorporates the energy input of the cylinder by

relating energy loss to the sum of positive energy inputs. Lastly, a new key performance indicator, termed ‘dissipation ratio’ (DR), is introduced as the ratio of power loss to the absolute value of transferred cylinder energy [10]. The dissipation ratio effectively addresses issues associated with the cancellation of energy transfers in other definitions, which can result in negative or misleading efficiency values. It is important to note that smaller dissipation ratio values indicate higher efficiency due to reduced energy loss, while ratios exceeding one may also occur.

The results clearly indicate that the highest efficiencies, regardless of the definition used or the load cycle analyzed, are achieved for the excavator boom. The boom also exhibits the highest energy inputs from the cylinder for all load cycles, likely due to the supported mass consisting of the boom, arm, and bucket. The substantial energy inputs correspond with high electric energy outputs. However, this very electric energy output is responsible for the previously mentioned cancellation in the denominator of efficiency definition one, which can theoretically lead to an efficiency value exceeding one.

Table 4.2 presents the efficiency η_2 and dissipation ratio (DR) results for various linear scalings of the velocity and load of the 1.8-ton dig and dump cycle. For both definitions, efficiency increases with higher loads, as smaller DR values correspond to higher efficiency. This trend aligns with the efficiency map shown in Figure 4.3, which indicates that efficiency improves with increasing loads. Conversely, a decrease in efficiency is typically expected at higher velocities; however, η_2 shows an increase instead. This phenomenon occurs because energy is stored in the drive train's inertia during acceleration, resulting in a higher electric energy input (denominator). During deceleration, this energy is converted back into electric output (numerator), resulting in a shift toward higher ratios. This highlights a misleading result that can arise from current efficiency definitions. In contrast, the dissipation ratio behaves as expected, decreasing with greater loads and increasing with higher velocities, except for a minor decrease at higher loads and lowest velocities, which might be the result of higher friction at lower speeds as well. Thereby validating its role as an appropriate key performance indicator for systems with recuperation capability.

Table 4.2. Efficiency and dissipation ratios for the scaled 1.8-t dig and dump cycle

	$\eta_2 = 1 - \frac{E_{loss}}{E_{elec,in} + E_{acc,in} + E_{cyl,in}}$				$DR = \frac{E_{loss}}{\lvert E_{cyl,out}\rvert + \lvert E_{cyl,in}\rvert}$			
	25 % F	50 % F	75 % F	100 % F	25 % F	50 % F	75%	100 % F
25 % v	14.7%	24.1%	32.1%	37.6%	6.893	3.252	2.090	1.609
50 % v	16.4%	26.5%	34.3%	40.3%	7.385	3.197	2.009	1.526
75 % v	17.6%	27.0%	34.3%	41.0%	8.499	3.440	2.152	1.556
100 % v	18.4%	27.9%	34.9%	41.0%	8.880	3.550	2.215	1.612

5. CONCLUSION & OUTLOOK

This study presents a newly developed EHA that incorporates a high-speed internal gear unit and a differential cylinder, enabling recuperation and load-holding capabilities. The EHA was experimentally tested in a hardware-in-the-loop test rig and specifically sized for use in an 1.8-ton excavator. The design focuses on a novel control approach that aims to minimize velocity deviations during quadrant switches through active throttling and active hysteresis.

Performance validation during quadrant switches was achieved using specially designed test cases. Under varying loads, the actuator demonstrated smooth cylinder velocity without oscillations when changing direction of movement. When the load direction changed during a constant velocity request, corresponding to an operational quadrant shift, only short and limited velocity deviations occurred, despite adjustments to valve positions and motor speed.

The tracking of a measured dig and dump cycle for the 1.8-ton excavator is sufficnet for use in excavators, although some deviations between the requested and actual positions were noted due to delays inherent in the control scheme. The efficiency of the EHA under tested working conditions reached 65% in three of the four quadrants, with an overall maximum efficiency of 70%. Analysis based on three downscaled load cycles (dig and dump, air grading and excavating) for an 18-ton excavator load cycles concluded that the highest efficiency, regardless of load cycle or energy efficiency definition, was achieved by the boom actuator.

Additionally, it was highlighted that current efficiency definitions often neglect energy stored in the drive train, resulting in misleading results. In contrast, a new key performance indicator introduced by the authors, termed ‘dissipation ratio’, more accurately reflects true system performance.

This contribution shows the high potential of the new EHA design and control approach. Future works include the implementation of the EHA on a real excavator to assess operator acceptance, address challenges related to force control in hardware-in-the-loop testing, and evaluate the interaction between excavator kinematics and the EHA.

6. ACKNOWLEDMENT

The IGF research project ‘22246 N / 1’ of the research association Forschungskuratorium Maschinenbau e. V. – FKM, Lyoner Straße 18, 60528 Frankfurt am Main was supported from the budget of the Federal Ministry of Economic Affairs and Climate Action through the DLR within the scope of a program to support industrial community research and development (IGF) based on a decision of the German Bundestag.

7. REFERENCES

[1] Love, L. J., Lanke, E., Alles, P. 'Estimating the Impact (Energy, Emissions and Economics) of the U.S. Fluid Power Industry - Tech. Rep. ORNL/TM-2011/14.', 2012.

[2] Zimmerman, J. D., Pelosi, M. et al. 'Energy Consumption of an LS Excavator Hydraulic System', Volume 4: Design, Analysis, Control and Diagnosis of Fluid Power Systems, pp. 117–126, 2007. DOI: 10.1115/IMECE2007-42267.

[3] Gøytil, P. H., Padovani, D., Hansen, M. R. 'On the Energy Efficiency of Dual Prime Mover Pump-Controlled Hydraulic Cylinders', ASME/BATH 2019 Symposium on Fluid Power and Motion Control, 2019. DOI: 10.1115/FPMC2019-1642.

[4] Pietrzyk, T., Schmitz, K. 'Messtechnische Erprobung einer elektro-hydraulischen Linearachse mit High-Speed Antrieb', Hybride und energieeffiziente Antriebe für mobile Arbeitsmaschinen: 8. Fachtagung, 23. Februar 2021, Karlsruhe.

[5] Pietrzyk, T., Schmitz, K. et al. 'Entwicklung und Auslegung einer elektro-hydraulischen Achse mit einem 48 V High-Speed-Antrieb zur Dezentralisierung der Arbeitshydraulik eines Kompaktbaggers', Hybride und energieeffiziente Antriebe für mobile Arbeitsmaschinen: 7. Fachtagung, 20. Februar 2019, Karlsruhe.

[6] Pietrzyk, T., Roth, D. et al. 'Design of a High Speed Internal Gear Pump to Increase the Power Density of Electro Hydraulic Actuators (EHA) in Mobile Applications', Volume 7: Fluids Engineering, 2019. DOI: 10.1115/IMECE2019-10351.

[7] Pietrzyk, T., Roth, D. et al. 'Design study of a high speed power unit for electro hydraulic actuators (EHA) in mobile applications', 11th International Fluid Power Conference, 2018. DOI: 10.18154/RWTH-2018-224632.

[8] Pietrzyk, T. 'Entwicklung einer Hochdrehzahl-Innenzahnradpumpe für die Elektrifizierung mobiler Anwendungen am Beispiel einer autarken dezentralen elektro-hydraulischen Achse', 2022. DOI: 10.18154/RWTH-2022-00767.

[9] Figge, F., Schmitz, K. 'Simulative controller evaluation for an electro-hydrostatic excavator actuator with load-holding capability using active hysteresis and throttling control', The 19th Scandinavian International Conference on Fluid Power, SICFP'25, 2025.

[10] Figge, F., Schmitz, K. 'EHAs in mobile machinery: Simulative evaluation of a control scheme including load-holding capability using active hysteresis and throttle control', International Journal of Fluid Power, accepted for publication on 18 December 2025.

[11] Figge, F., Reinertz, O., Schmitz, K. 'Review of efficiency definitions for systems with energy recuperation and introduction of the dissipation ratio', Proceedings of the ASME 2025 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2025.

[12] Figge, F., Schmitz, K. 'Simulative Comparison Approach of Electro-Hydrostatic Actuator Circuits in Excavators', ASME/BATH 2023 Symposium on Fluid Power and Motion Control, 2023. DOI: 10.1115/FPMC2023-111798.

[13] Williamson, C., Ivantysynova, M. 'Stability and motion control of inertial loads with displacement controlled hydraulic actuators', 6th FPNI-PhD Symp, pp. 499–514, 2010.

[14] Michel, S. 'Elektrisch-hydrostatische Kompaktantriebe mit Differentialzylinder für die industrielle Anwendung', Dissertation, Dresden, 2020. DOI: 10.25368/2021.86.

[15] Williamson, C., Ivantysynova, M. 'Pump mode prediction for four-quadrant velocity control of valveless hydraulic actuators', 7th JFPS International Symposium on Fluid Power, 2008.

[16] Wang, L., Book, W. J., Huggins, J. D. 'A Hydraulic Circuit for Single Rod Cylinders', Journal of Dynamic Systems, Measurement, and Control, Vol. 134, Nr. 1, 2012. DOI: 10.1115/1.4004777.

[17] Çalışkan, H., Balkan, T., Platin, B. E. 'A Complete Analysis and a Novel Solution for Instability in Pump Controlled Asymmetric Actuators', Journal of Dynamic Systems, Measurement, and Control, Vol. 137, Nr. 9, 2015. DOI: 10.1115/1.4030544.

[18] Imam, A., Rafiq, M. et al. 'Design, Implementation and Evaluation of a Pump-Controlled Circuit for Single Rod Actuators', Actuators, Vol. 6, Nr. 1, pp. 10, 2017. DOI: 10.3390/act6010010.

[19] Imam, A., Tolba, M. S., Sepehri, N. 'Improving performance of pump-controlled hydraulic circuits for single-rod actuators: conceptual study', Journal of Physics: Conference Series, Vol. 2616, Nr. 1, pp. 12015, 2023. DOI: 10.1088/1742-6596/2616/1/012015.

[20] Koury Costa, G., Sepehri, N. 'A Critical Analysis of Valve-Compensated Hydrostatic Actuators: Qualitative Investigation', Actuators, Vol. 8, Nr. 3, pp. 59, 2019. DOI: 10.3390/act8030059.

[21] Gøytil, P. H., Padovani, D., Hansen, M. R. 'A Novel Solution for the Elimination of Mode Switching in Pump-Controlled Single-Rod Cylinders', Actuators, Vol. 9, Nr. 1, pp. 20, 2020. DOI: 10.3390/act9010020.

[22] Kärnell, S., Ericson, L. 'Hysteresis Control in Pump-Controlled Systems—A Way to Reduce Mode-Switch Oscillations in Closed and Open Circuits', Energies, Vol. 15, Nr. 2, pp. 424, 2022. DOI: 10.3390/en15020424.

[23] Haas, C., Schmitz, K. 'Control Strategy for a Hardware in the Loop Test Bench', Advances in Hydraulic and Pneumatic Drives and Control 2023, 2024. DOI: 10.1007/978-3-031-43002-2_9.

[24] Kamtam, S. B., Lu, Q. et al. 'Network Latency in Teleoperation of Connected and Autonomous Vehicles: A Review of Trends, Challenges, and Mitigation Strategies', Sensors (Basel, Switzerland), Vol. 24, Nr. 12, 2024. DOI: 10.3390/s24123957.

Biographies



Felix Figge received the bachelor's degree in mechanical engineering with a specialization in design methodology in 2019 and the master's degree in mechanical engineering with a specialization in product development in 2021, both from RWTH Aachen University. He is currently working as a Research Associate at the Institute for Fluid Power Drives and Systems (ifas), RWTH Aachen University. His research encompasses fluid power systems in general but focuses on the development of an electro-hydrostatic actuator for use in mobile machinery.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated in 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz’s awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

Beyond PFAS: Challenges and Opportunities in Material Innovation for replacing PTFE in Hydraulic Seals

Niklas Bauer, Gonzalo Barillas, Jürgen Hieber, Martin Goerres

Freudenberg Sealing Technologies GmbH, Corresponding Author: niklas.bauer@fst.com

Abstract.

The European Chemicals Agency (ECHA) is currently evaluating a proposal to eliminate all nonessential uses of per- and polyfluoroalkyl substances (PFAS) as many of these substances are classified as potentially carcinogenic. A partial or complete ban on PFAS could severely impact many branches of industry as PFAS are widely used in a variety of applications. Nowadays, PFAS are also used in critical components of fluid power systems. One of the most important PFAS in fluid power applications is PTFE which is used e.g. in many piston and rod seals of hydraulic cylinders. The PFAS ban which is currently being discussed could also lead to severe restrictions on using PTFE seals in fluid power applications.

As a sealing material, PTFE combines many favorable properties, especially a very high temperature resistance, low stick-slip friction, low wear and excellent media compatibility. Therefore, a ban of PTFE seals would deprive hydraulics of one sealing material with a unique combination of properties.

In this paper, we present an overview of the challenges associated with replacing PTFE with an alternative material and discuss polyurethane as an alternative material for piston and rod seals in hydraulic cylinders. An evaluation of a PU and a PTFE piston seal shows that the investigated PU seals reached superior friction, leakage and endurance behavior for the investigated operating conditions, indicating that the selected polyurethane is a viable alternative even for hydraulics with high pressures and high temperatures.

Keywords. Hydraulics, PFAS, Piston Seal, Polyurethan, PTFE, Tribology.

1. INTRODUCTION

Since 2023, the proposal by the European Union and other international authorities to regulate PFAS chemicals has been causing discussions about the impact of upcoming regulations in almost all industrial sectors. PFAS are part of many industrial and everyday products like frying pans and raincoats. For fluid power PFAS can occur among others in lubricants, hydraulic oils and seals. Seals are functionally critical and often life time determining components of all fluid power systems. Therefore, this paper focuses on

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

alternatives to PFAS materials on hydraulics seals. The scope of this paper is the evaluation of an alternative material to PTFE as one highly performant sealing material.

This paper starts with a brief overview of the current status of the regulations. After that, the unique properties of PTFE are outlined and discussed with respect to the demand of hydraulic seals. After that, polyurethane (PU) is proposed and investigated as an alternative material for hydraulic applications. The performance of different PU materials regarding temperature and media resistance for hydraulic applications is compared to select one especially suitable material for experimental comparison to PTFE in terms of friction, leakage and endurance.

2. PFAS IN THE CONTEXT OF FLUID POWER SEALS

In this chapter, a brief overview of the timeline of the PFAS regulation in the EU will be given. Afterwards, the possibly affected substances in hydraulic sealing systems and their unique properties will be briefly highlighted. Focusing on PTFE, as an exemplary component and material, typical hydraulic seals within a cylinder are discussed with respect to the required material properties.

2.1. Regulatory Background

The European Chemicals (ECHA) has proposed a regulation for the substance class of per- and polyfluoroalkyl substances (PFAS). PFAS denote “any substance that contains at least one fully fluorinated methyl (CF3-) or methylene (-CF2-) carbon atom (without any H/Cl/Br/I attached to it)” [1]. Since then, an updated restriction proposal has been published [2], in which sealing applications are named as an evaluated sector. Outside the EU, different PFAS regulations are also in discussion or under effect, e.g. in Japan, China, Canada and different states of the USA – the timeline and affected substances differ from region to region. The timeline of the PFAS regulations in the EU is shown in Figure 2.1.

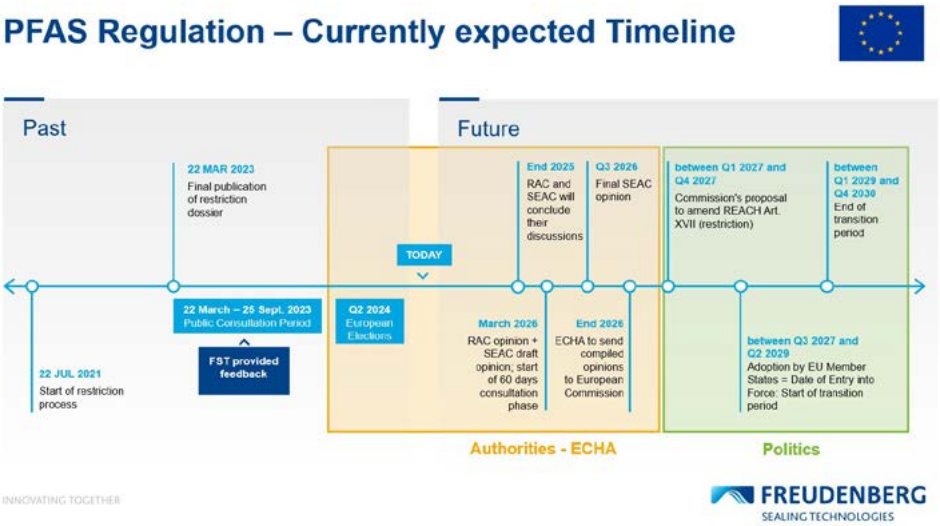


Figure 2.1. Expected Timeline for PFAS regulations [3] as of end of 2025.

As shown in the timeline above, the restriction process takes several years to go from initial proposal until it takes effect. This is partially due to the high number and huge diversity of different existing substances. The OECD discussed and propose different categories to distinguish between different groups of PFAS [4]. A discussion of different grouping strategy is discussed by Cousins et al [5]. They compare several approaches for grouping PFAS and evaluating the effects and adequate measures for restriction of these groups. The discussed grouping strategies range from a general approach covering all PFAS up to approaches differentiating between intrinsic properties such as mobility in water or approaches based on risk assessment such as toxicity.

One common differentiation that is often used in the public discussion between manufacturers and customers is the classification of PFAS based on polymeric and non-polymeric PFAS. For most sealing applications, polymeric PFAS are covering the most important materials such as plastics like PTFE, PVDF, PFA, rubbers like FKM, FFKM, FVMQ and base oils like PFPE. The category of non-polymeric PFAS covers substances with small molecules such as gases for refrigerants, foaming agents and liquids like fire fighting foams or coatings in textiles. The most important non-polymeric PFAS for sealing applications are fluorinated surfactants used as process aids in the manufacturing of polymeric PFAS like FKM. In case of a differentiation between polymeric and non-polymeric PFAS, where only non-polymeric PFAS would be affected, it is possible to adapt the manufacturing processes in order to evade the usage of these non-polymeric PFAS. This transitioning is already being made by suppliers of FKM materials.

2.2. Use of PFAS in Sealing Systems

In fluid power sealing systems, the most important substances are the fluorinated elastomers FKM and FFKM, the fluorinated silicones FVMQ as well as Polytetrafluoroethylene (PTFE). Their unique combination of high temperature stability media resistance and low friction and wear makes them an attractive material for a variety of applications ranging from transportation, food & beverage to fuel cells or batteries [6]. The material compatibility of materials and fluids is typically evaluated using insertion tests as shown for hydraulic fluids and bio-hybrid fuels e.g. in [7] and [8]. A comparison of the temperature stability and swelling behavior is given in Figure 2.2.

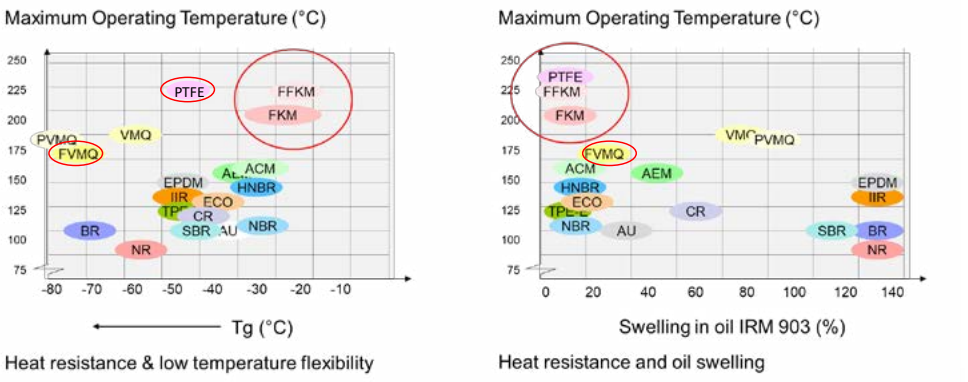


Figure 2.2. Classification of elastomer materials according to ASTM D2000 / SAE J200 PFAS substances are highlighted [6].

As seen in Figure 2.2, simply replacing all PFAS containing materials with other currently available materials would cause a loss in functionally relevant properties potentially leading to shorter maintenance intervals, premature failure of sealing systems as well as leakage of potentially hazardous fluids into the environment. Especially in the aerospace and chemical industry e.g. for processing aggressive chemicals or in the pharmaceutical and medical industry where in dosing applications the seals have to withstand both the medications and the fluids and conditions used in the cleaning and sterilization cycles. In addition, seals in medical applications have to undergo an extensive certification according to international standards such as EN ISO 10993-5:2009. [6]

The beneficial properties of PFAS containing materials for sealing applications can be summarized as follows: [6]

- High wear resistance
- Low friction coefficient
- High temperature stability
- High media resistance
- Low swelling in contact with fluids
- Permeation tightness for gases
- Compliance with industrial standards for medical safety

2.3. Typical Fluid Power Seals

In typical fluid power applications, several seals can currently contain PFAS materials. These include both rod and piston sealing package in hydraulic cylinders (wipers, rod seals, guide rings, buffer seals, piston seals), see Figure 2.3, as well as static seals and rotating seals (e.g. swivel seals and rotary shaft seals). For pneumatic applications, similarly both static and dynamic seals can be affected by possible regulations.



Figure 2.3. Typical sealing system of a hydraulic cylinder.

The main materials include FKM and FFKM, e.g. for static seals and preloading element in two-piece hydraulic seals, and PTFE mainly used in profile rings for profile rings in two-piece hydraulic seals. For illustration purposes, some typical designs[1] of hydraulic rod seals are shown in Figure 2.4. The first seal consists of just one piece which gets deformed when exposed to a pressure difference so that the contact pressure increases at higher operating pressures. The preloading of the seal results from the interference of the flexible seal with housing and rod. The second seal consists of two parts: a profile ring which has a similar function as the aforementioned one-piece seal and the back-up ring which reinforces the seal against gap extrusion. The third design shown also consists of two parts. However, here the profile ring gets reinforced by a preloading element (such as an O-Ring or a metal spring).

The first and second of the shown designs typically use polyurethane for the profile ring because of the required flexibility to preload the sealing contact. The back-up ring typically consists of a thermoplastic polymer. Thus, the first and second kind of seals do usually not contain PFAS materials. For the third option shown, PU, PTFE or thermoplastic polymers such as PA can be used for the profile ring. The preloading element typically consists of an elastomer part made from NBR or FKM or a metal spring. In conclusion, mainly the third kind of rod seals could be affected by PFAS regulations, as here both profile ring and preloading element can contain PFAS material if they are made from PTFE or FKM, respectively. Due to the lack of flexibility of PTFE, it is not used in seal designs where flexibility of the profile ring is required.

[1] These three designs are the most typical designs to occur in hydraulic applications with average diameters and operating conditions. For special applications e.g. special media, high temperatures, large diameters or special requirements, other seals like PTFE V-packings are also used but will not be covered in this paper.

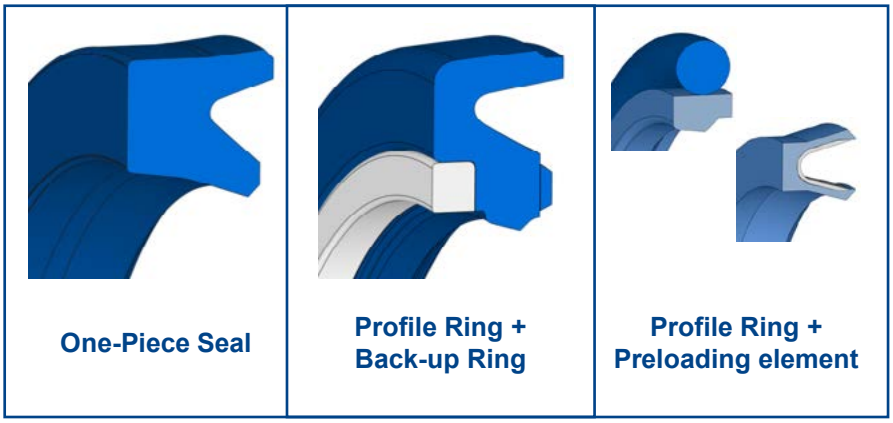


Figure 2.4. Different Seal Designs for Rod Seals.

Although slightly different in shape, similar designs to the ones shown in Figure 2.4 are also used for piston seals. For wipers, PU and elastomers like NBR and FKM are the most common material, however two-piece versions with PTFE and an elastomeric preloading element are also used. For guide rings, either PTFE or fabric-based laminate e.g. consisting of cotton-based fabric and a phenolic resin is used.

2.4. Material Requirements of Fluid Power Seals

When comparing the abovementioned strengths of PFAS containing materials with typical demands of fluid power applications (hydraulics and pneumatics) it becomes evident that not all of the properties are equally relevant for hydraulic applications. Whilst high wear resistance and low friction are still desirable properties of a seal in a hydraulic system, the importance of the other benefits is typically lower than in the previously mentioned highly demanding applications:

The temperature range in a hydraulic system is considerably lower, also because other components limit the ideal operating conditions (e.g. the degradation of mineral oils due to high temperatures as well as the temperature stability of the hoses and accumulators). Thus, the recommended maximum system temperature of a typical hydraulic system is less than 80°C [9], whereas typical stationary temperatures for mobile and industrial hydraulic applications range at around 40 to 60°C [10]. For aerospace applications, the occurring temperatures may vary.

The media resistance and low swelling has to be considered when choosing seals in fluid power applications. Similar to hydraulic systems, in pneumatic systems the seal only gets in touch with the lubricating grease[2]. For pneumatic systems, a good compatibility with the grease is important. In hydraulic systems the only media in contact with the seals is the hydraulic fluid itself (potentially containing entrained air and water). Due to the high effort

[2] Some greases in pneumatic applications contain PFAS themselves. For designing a sealing system without PFAS containing materials, the interaction of grease and sealing material has to be carefully considered. Due to the scope of this paper, only PFAS in the sealing material itself are taken into account here.

of changing the oil type of a complete hydraulic system, a hydraulic system mostly sticks to one type of oil. This means that far less media get in contact with hydraulic seals during their lifetime than seals in other applications e.g. in chemistry. However, the compatibility of sealing material and fluid is still crucial and cannot necessarily be taken for granted for all combinations of hydraulic fluids and sealing materials. Experimental results for the media compatibility of selected materials are shown in chapter 3.

The permeation tightness for gases is not important in typical fluid power applications. The compliance with medical safety is not important for most hydraulic systems. For pneumatic systems, the importance of compliance heavily depends on the application.

3. POLYURETHANE AS AN ALTERNATIVE FOR HYDRAULIC SEALS

In this chapter, the basic differences between PTFE and PU are compared from a material point of view. As a first step, the general material properties are compared and evaluated regarding their implications with respect to hydraulic applications. In the next step, the resistance to hydraulic fluids, one functionally critical property of hydraulic seals, is evaluated for six different polyurethanes.

3.1. Comparison of Material Properties for Hydraulic Seals

Besides the aforementioned favorable temperature and media resistance, PTFE also brings some unique properties that have to be paid attention to when designing hydraulic seals. First, in comparison with elastomers or PU, PTFE is not an elastic material and tends to deform plastically rather easily. The lack of elasticity means that PTFE seals typically use simpler geometries than PU seals because complex geometries are more difficult to manufacture with PTFE and also tend to get “simplified” due to plastic deformation in the application. The potential benefit of complex geometries is shown in the experimental data presented in chapter 4.

One additional consequence of the plasticity and softness of PTFE is that it is rather prone to getting damaged during assembly such as due to chamfers, threads and improper handling. Additionally, for the assembly in tight spaces, PTFE often needs to be pre-heated to around 80°C and adjusted in diameter (i.e. widened for piston seals and made smaller for rod seals). After assembly, the diameter of the seal has to be re-calibrated by plastically deforming the seal with a cone. In comparison, the assembly of elastic materials is generally easier and less prone to damaging the seals.

Disadvantages of PU in comparison to PTFE are a higher friction coefficient (especially relevant for breakaway friction) and a smaller tolerance towards the immersion in different fluids at high temperatures. In the next section, the material compatibility of different PUs with typical hydraulic fluids will be investigated.

3.2. Changes of Tensile Strength of Different Polyurethanes in Hydraulic Fluids

The material compatiblity is shown in the Figure 3.1, Figure 3.2 and Figure 3.3 for six different polyurethanes all of which are used in hydraulic sealing applications. The loss of tensile strength is chosen as one representative material property that can changes during exposure to fluids at high temperatures. As reference fluids three commercially available hydraulic oils were chosen. The first of which is a mineral oil (HVLP), the second a fire-resistant water-glycol solution (HFC) and the third a synthetic esther based (HEES). The

temperatures were chosen according to typically allowed maximum operating temperatures for polyurethanes. For assessing the material compatibility, the polyurethanes were exposed to the respective fluid and temperature for the indicated number of weeks. After the storage, the tensile strength was determined and compared to the value of a non-exposed specimen.

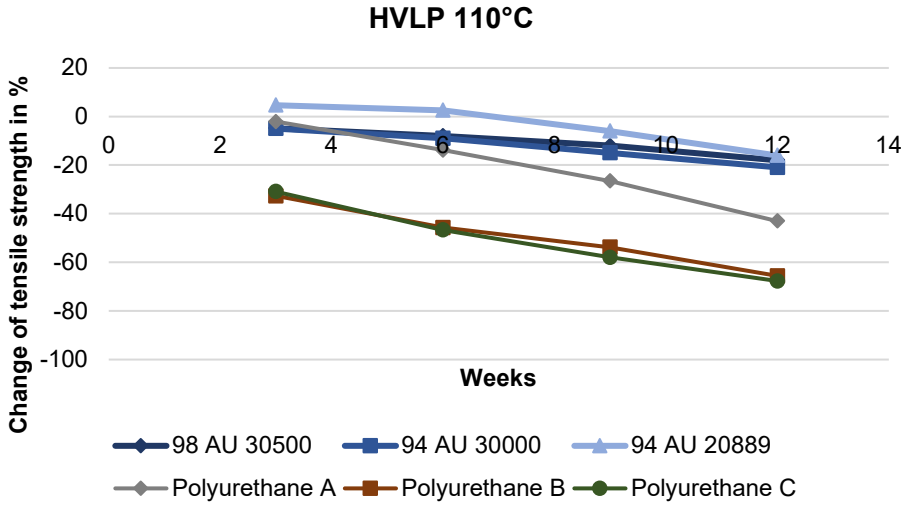


Figure 3.1. Change of tensile strengths for different polyurethanes in HLVP at 110°C.

It can be seen that the loss of tensile strength depends heavily on the chosen fluid and temperature. In other experiments (not displayed here) it could be shown that the exposure at lower temperatures generally had lower effects on the change of tensile strength. Most of the materials lose tensile strength during exposure to the fluids and temperatures. A too high loss of tensile strength can indicate that the material might perform worse in the practical application. In the worst case, the material can become so brittle that the seal specimens break by slight loads applied with bare hands.

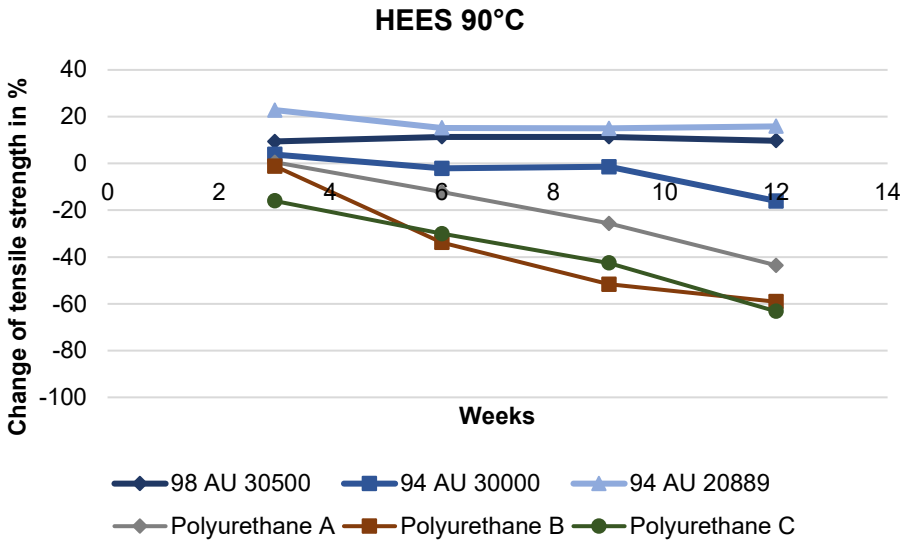


Figure 3.2. Change of tensile strengths for different polyurethanes in HEES at 90°C.

Even though the behavior when exposed to different fluids varies for each combination of fluid and material, it becomes evident that three of the investigated materials (98AU30500, 94AU30000 and 94AU20889) show particularly low changes of tensile strength for all three liquids.

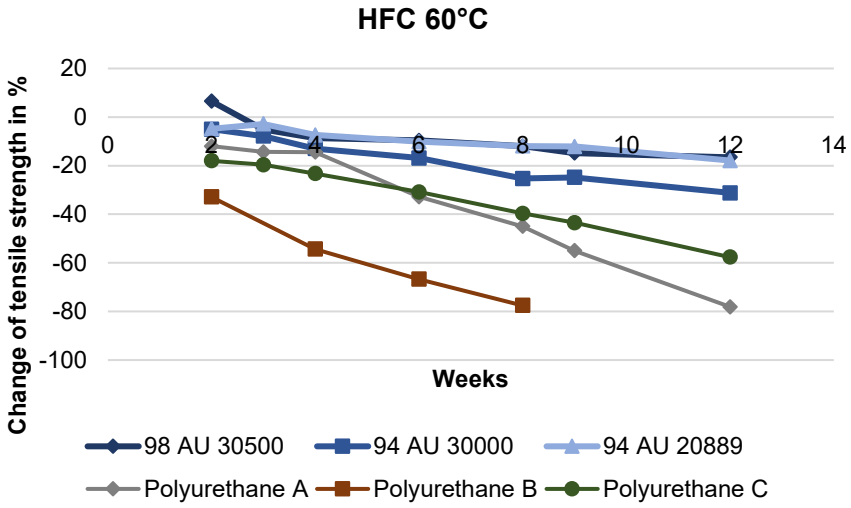


Figure 3.3. Change of tensile strengths for different polyurethanes in HFC at 60°C.

Concluding the immersion tests, it was shown that some combinations of fluids and high temperatures can cause huge changes in the material properties of polymers. However, despite the large differences between the chemical structure of the three investigated fluids,

some polyurethanes performed especially well by showing lower changes in tensile strength than the other investigated materials. This indicates that there are chemically resistant polyurethanes that can indeed be used in a variety of hydraulic applications with long-term exposure to different media at high temperatures.

4. EXPERIMENTAL RESULTS

In this chapter, the performance of PU seals is assessed in comparison to PTFE seals using experimental investigations on a hydraulic test rig. For the scope of this paper, piston seals are chosen as the reference. In the first section, the test rig is introduced briefly. In the following sections, friction, leakage and lifetime are assessed. Goal of the investigations is to show how PU seals compare with commercially available PTFE solutions.

4.1. Hydraulic Piston Seal Test Rig

For the investigations presented in this chapter, a hydraulic piston test rig was used. The test rig uses five separate modules. Each module consists of a piston, a hollow rod and a tube. Each module tests up to three seals at the same time, all of which are mounted on the piston as shown in Figure 4.1. During operation, a linear drive moves the piston back and forth within the tube. The two chambers between seals 1/2 and 2/3 are pressurized individually through channels in the hollow rod. Pressurization is chosen in accordance with the movement to imitate the operating conditions occurring in a differential cylinder for the middle seal 2.

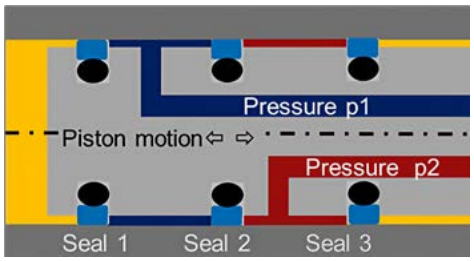


Figure 4.1. Schematic of one Module of the Piston Seal Test Rig.

A force sensor measures the friction force of the piston in parallel. The leakage of the outer seals 1 and 3 is assessed by gathering the droplets at both ends of the tube. The reciprocating motion of the piston is repeated until the chosen number of cycles is reached or when any of the seals is damaged in such a way that the chambers between the seals cannot be pressurized anymore.

4.2. Comparison of Measured Friction Forces

To compare the friction force of a PU and a PTFE piston seal, these seals were installed on the test rig described in section 4.1. Both seals are commercially available as double acting piston seals. The PU seal is a Simko 300 and is made of 98 AU 928. The geometry was optimized to gain a good combination of low leakage and low friction force. As investigated parameters, a velocity of 0.3 m/s and an oil temperature of 60 °C were chosen. The friction forces were evaluated at 50, 100, 200, 300 and 400 bar. The measured average friction forces of the two seals are shown in Figure 4.2, all normalized to the same reference force F_{Ref}. It

should be taken into consideration that the test rig measures the friction force of one module with three seals at once as shown in Figure 4.1.

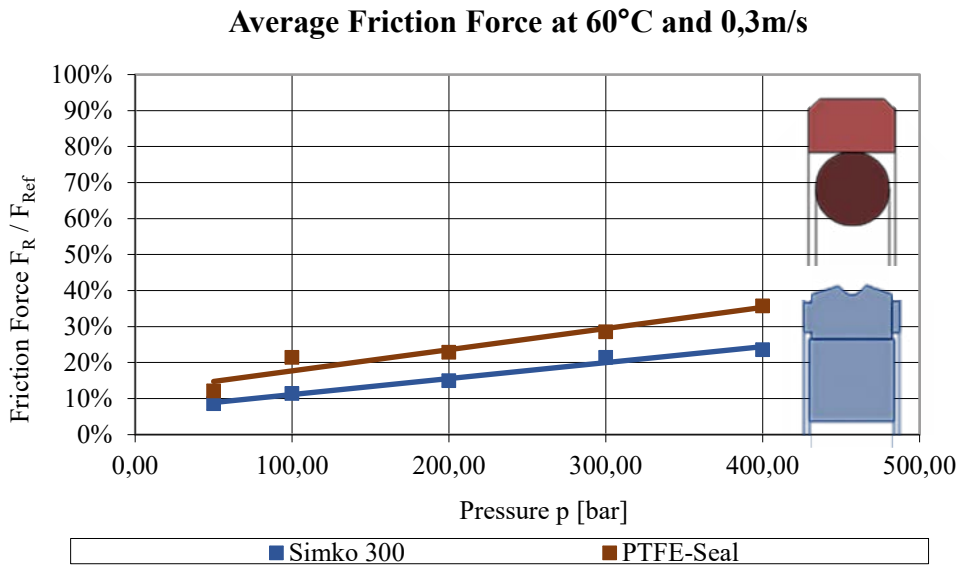


Figure 4.2. Comparison of the Friction Forces between a PTFE and a PU seal for Different Operating Pressures.

When evaluating the results, it should be noted that the friction forces during operation vary with different oils, temperatures, pressures and kinematics. However, the comparison of the two seals shows that the friction force of the PU seal at these operating conditions is lower for all investigated pressures between 50 and 400 bar. This indicates that although the friction coefficient for PTFE/steel is typically lower than for PU/steel, this does not necessarily cause higher friction forces for PU seals than for PTFE seals. The better performance of the PU seal can be attributed to the more favorable geometry of the PU seal, which has been designed to reach optimal hydrodynamic lubrication between seal and countersurface to deliver a good balance between low friction and low leakage. The geometry used for the PU seal is typically not feasible for PTFE seals due to difficulties in manufacturing and plasticity of the sealing material, which is why PTFE seals often use more simple geometries than PU seals.

4.3. Comparison of Leakage Measurements

The leakage of the two seals investigated in the previous section, the same test as shown before have been used. The seals designs are the same as in the previous section. The leakage has been determined after the test.by measuring the amount of leakage from seals 1 and 3 as shown in Figure 4.1. It should be noted that the leakage of seal 2 cannot be detected with the used test rig. The results of the leakage measurement are shown in Figure 4.3. The values for both seals were determined in the same operating conditions on the same surfaces and are given in the unit ml per 1000 doubles strokes (DS).

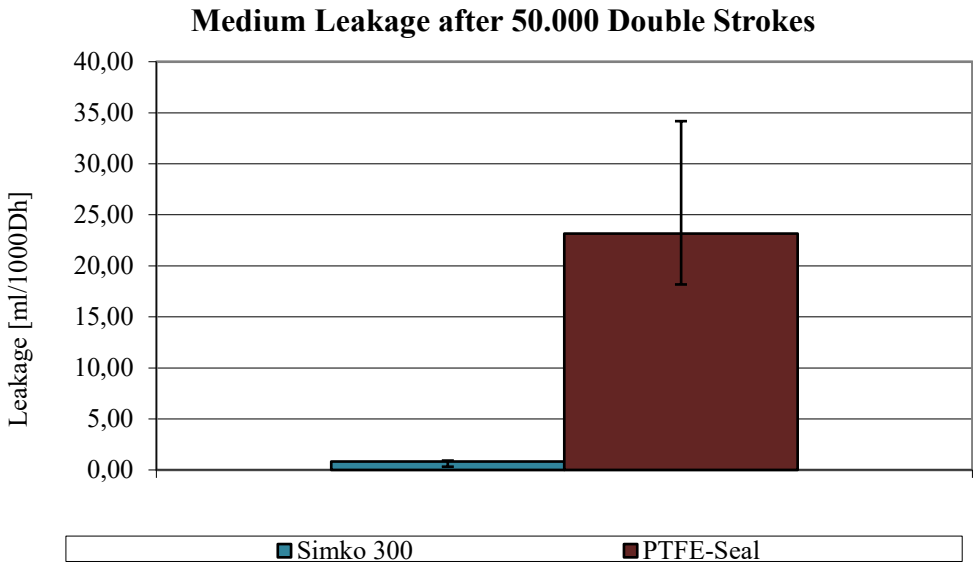


Figure 4.3. Comparison of the leakage of a PTFE and a PU seal.

The comparison of the measured leakage shows that the PTFE seal shows about 20 times higher leakage than the PU seal. This is especially remarkable considering that the friction force of the PTFE seal was higher than the friction force of the PU seal, indicating that the chosen geometry of the PU seal caused more favorable hydrodynamics than the geometry of the PTFE seal.

4.4. Endurance Test

Finally, after evaluating friction and leakage, the endurance of PU seals will be evaluated. Even more than friction and leakage, the lifetime of a seal is highly depending on the interaction of various operating conditions. The same seal design and material can reach the end of its life time in many different ways depending on the operating conditions. Some common failure modes are excessive abrasive wear, chemical deterioration of the material and gap extrusion. The latter mechanism will be the dominating failure mechanism for the operating conditions chosen for the investigation in this paper. Note that the seals and materials tested here are from the ones shown in the previous two sections.

For quickly evaluating the durability of piston seals, the formerly introduced test rig is used. As operating conditions, the combination of high pressure of 400 bar and a high temperature of 100°C are chosen, because these conditions typically cause many commonly available seal designs and materials to fail within less than 50.000 double strokes. The results of one PU seal (OMK-PU) made from the hydrolysis resistant 98 AU 30500, 4 other PU and one PTFE seal are shown in Figure 4.4.

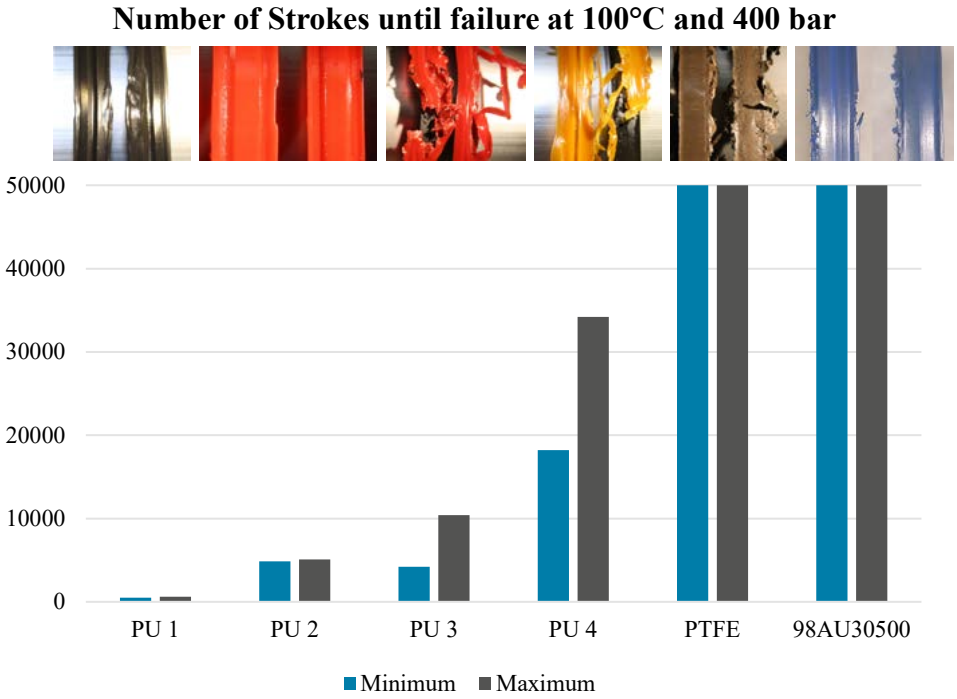


Figure 4.4. Comparison of the Life Time of one PTFE and Multiple PU Seals.

All investigated PTFE seals survived the endurance test. Out of the five PU seals, only the seal made from 98 AU 30500 managed to survive the endurance test with all five specimens. Only one of the other four PUs managed to survive with one module up to 50.000 double strokes. The reason for the premature failures can be observed in the pictures of the seals after tests. Severe extrusion occurred in all prematurely failed seals, some of which even broke apart. The results are in line with the material compatibility tests shown in section 3.2, where huge differences between the different investigated PUs occurred.

It is worth comparing the PTFE and the 98 AU 30500 seals after the endurance test. While PTFE seal already shows severe extrusion patterns, the PU seal only shows minor damage. The presence of large extrusion damage usually indicates that only limited life time remains. In addition, if parts of the extrusion damage breaks lose from the seal, they can get stuck in other hydraulic components like valves causing down time and damage.

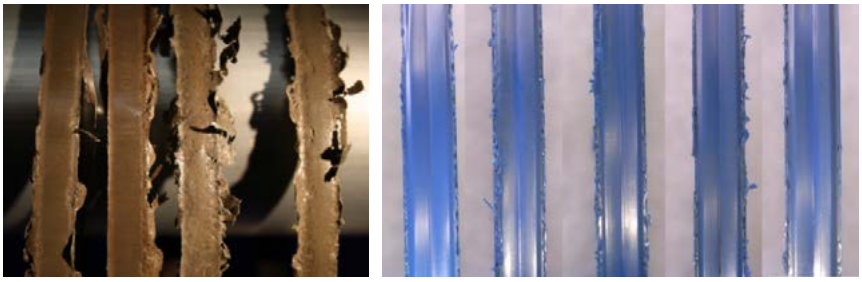


Figure 4.5. PTFE (left) and PU Seal (98 AU 30500, right) after the Endurance Test.

5. SUMMARY AND OUTLOOK

Motivated by international discussions about increasing regulations for PFAS substances, this paper investigated where PFAS containing materials are used in hydraulic sealing systems. For the scope of this paper, the focus of further investigation was on PTFE and its unique properties as a sealing material. After evaluating the beneficial material properties of PTFE, the paper discussed the potential of using polyurethane (PU) as an alternative material for hydraulic seals.

Some main advantages of PTFE over PU include lower friction and wear coefficients, better temperature and media resistance. To investigate whether PU can serve as a viable alternative, the practical consequences of these advantages in hydraulic applications were investigated using both material and seal tests.

In terms of material compatibility, it could be found that some PUs show severe degradation of material properties even after few weeks of immersion in typical hydraulic fluids at temperatures between 60 and 110°C. However, there are also hydrolysis resistant PUs like 98 AU 30500 which showed stable behavior during the immersion even after 12 weeks of severe exposure to temperature and different media. This indicates that there are PUs with sufficiently high media and temperature resistance for the majority of hydraulic applications.

For comparing the performance of these materials used in a hydraulic piston seal, a test rig was used. While some PU seals showed severe degradation even after less than 10.000 double strokes during an endurance test at 400 bar and 100°C, the material 98 AU 30500 reached the full 50.000 double strokes with less apparent damage than the PTFE seal. In terms of leakage and friction the investigated PU seal Simko 300 also showed better behavior than the PTFE seal.

Before concluding this paper, it should be noted that sealing systems are highly complex tribological systems. Therefore, especially in demanding applications, the selection of sealing design and material should be carefully adjusted to the conditions and needs. Thus, a general statement about whether “PTFE or PU is a better sealing material”, cannot be made. Especially where low break away friction forces and low stick slip probability are desired or high temperature and aggressive media is present, PTFE seals can rely on the core strengths of its material. However, the results presented here demonstrate that even at high pressures and temperatures, choosing a highly resistant PU like 98 AU 30500 as a sealing material can deliver not only comparable but superior results in terms of endurance, friction and leakage.

6. DECLARATION

For the creation of this paper, the authors used the assistance of AI tools for editing and revising parts of the text and for accelerating literature research. All AI outputs and references were reviewed, edited, and verified by the authors prior to inclusion.

7. REFERENCES

[1] ECHA, „Restrictions under Consideration - Per- and polyfluoroalkyl substances (PFAS),“ 22 03 2023. [Online]. Available: https://echa.europa.eu/de/restrictions-under-consideration/-/substance-rev/72301/term. [Accessed 18 01 2026].

[2] ECHA, „ECHA publishes updated restriction proposal,“ 20 08 2025. [Online]. Available: https://echa.europa.eu/de/-/echa-publishes-updated-pfas-restriction-proposal. [Accessed 18 01 2026].

[3] FST, "Statement on the European Chemicals Agency PFAS Restriction Proposal," [Online]. Available: https://www.fst.com/news-stories/pfas-update/. [Accessed 18 01 2026].

[4] Organisation for Economic Co-operation and Development (OECD), "Reconciling Terminology of the Universe of Per- and Polyfluoroalkyl Substances:," Recommendations and Practical Guidance, 09 07 2021. [Online]. Available: https://one.oecd.org/document/ENV/CBC/MONO(2021)25/En/pdf. [Accessed 24 01 2026].

[5] I. T. Cousins, J. C. DeWitt, J. Gluge, G. Goldenman, D. Herzke, R. Lohmann, M. Miller, C. A. Ng, M. Scheringer, L. Vierke und Z. Wang, „Strategies for grouping per- and polyfluoroalkyl substances (PFAS) to protect human and environmental health,“ *Environmental Science: Processes and Impacts,* Nr. 22, p. 1444–1460, 2020.

[6] FST, „Statement on PFAS - EU restriction proposal,“ 08 08 2023. [Online]. Available: https://www.fst.com/-/media/files/news/pfas-letter_fst_non-confidential-final.pdf. [Accessed 18 01 2026].

[7] F. Schlegel, M. Hofmeister, D. Osmanovic und K. Schmitz, „Swelling, wear and property changes of high-performance polymers in oil-hydraulic tribological contacts,“ *Tribologie und Schmierungstechnik,* Bd. 72, Nr. 1, pp. 5-13, 2025.

[8] M. Hofmeister, F. Fischer, L. Boden und K. Schmitz, „Challenges in the Use of Bio-Hybrid Fuels As Drop-in Fuels.,“ in *roceedings of the ASME 2024 18th International Conference on Energy Sustainability collocated with the ASME 2024 Heat Transfer Summer Conference and the ASME 2024 Fluids Engineering Division Summer Meeting.* , Anaheim, California, USA, 2024.

[9] E. Thompson, „What is the temperature range for a standard hydraulic cylinder to operate normally?,“ IronClad Hydraluics, 04 11 2025. [Online]. Available: https://www.ironcladhydraulics.com/blog/what-is-the-temperature-range-for-a-standard-hydraulic-cylinder-to-operate-norma-1656872.html. [Accessed 18 01 2026].

[10] Harvard Filtration, „Navigating the Effects of Temperature on Hydraulic Systems,“ 01 02 2024. [Online]. Available: https://www.harvardfiltration.com/effects-of-temperature-on-hydraulic-systems/. [Accessed 18 01 2026].

BIOGRAPHIES



Niklas Bauer received the bachelor's and master’s degree in mechanical engineering from RWTH Aachen University in 2017 and 2019, respectively. From 2019 to 2025 he worked in various positions at the Institute for Fluid Power Drives and Systems (ifas) in Aachen, last as chief engineer. He graduated in 2024 with a Doctoral degree about elastohydrodynamic simulation of pneumatic sealing systems. In 2025, he joined Freudenberg Sealing Technologies as Head of Advanced Development, where he is responsible for the development of new products for new markets. He supports the University of Dresden as a lecturer on the fundamentals of sealing technology.



Gonzalo Barillas received his Diploma (Dipl.-Ing. (FH)) at the former Fachhochschule für Technik Esslingen, today's University of Applied Sciences Esslingen in 1984, his thesis thema was Tribology of Ceramic Materials. He started then in 1985 as product developer at sealing company Merkel in Hamburg, in 1990 he joined Freudenberg Sealing Technologies in Schwalmstadt, where he worked in several positions in the development and business development of seals. His last position before retiring in Dec. 2025 was Manager of advanced product development at the Lead Center Fluid Power Industry. He supports the University of Dresden as a lecturer on the fundamentals of sealing technology.



Jürgen Hieber received his Diploma in Chemistry in 1993 and his Doctorate (Dr. rer. nat.) in 1996 from the University of Tübingen, with a doctoral thesis entitled *"Electrically Conductive Polymers"*. Since 1996, he has been employed by Freudenberg Sealing Technologies, where he currently serves as Head of Material Development for the Fluid Power Industry. During his tenure, he has, among other responsibilities, led the construction of a new polyurethane granulate production facility. He has delivered several technical presentations on polyurethane materials for sealing applications.



Martin Goerres graduated in Mechanical Engineering at RWTH Aachen University in 1999. He worked as Research Associate at the Institute for Fluid Power Drives and Controls (IFAS) until 2005. He joined Freudenberg Sealing Technologies (formerly known as Merkel Freudenberg Fluidtechnic GmbH) in 2005 where he worked in various positions. As Director Engineering and Innovation he is globally responsible for the Engineering at the Lead Center Fluid Power Industry. He has been serving as a reviewer and committee member for the international fluid power and sealing community.

Percolation Analysis of Reciprocating Sealing Surfaces Based on the Statistical Evolution of HPD and PSD Induced by Wear

Jiehao Wang[1], Jing Li[2], Wei Shen[1], Katharina Schmitz[3]

1. Department of Mechatronics Engineering, University of Shanghai for Science and Technology; 2. School of Mechanical Engineering, Tongji University; 3. Institute for Fluid Power Drives and Systems (ifas), RWTH Aachen University

Abstract.

The reciprocating seal exhibits significantly different wear behaviours and surface damage patterns under various friction and operating conditions. Precisely describing the topography characteristics of the worn surface is of great significance for the analysis of oil flow and the study of the leakage mechanism. The high probability density (HPD) and the power spectral density (PSD) are commonly used for reconstructing the surface topography. It is particularly necessary to study the evolution laws of HPD and PSD during the wear process to reconstruct worn surface. This study is based on the polymer-steel reciprocating friction experiments and analyses the changes in HPD and PSD characteristics of the PTFE worn surface under different temperatures. A three-dimensional oil percolation simulation model is established to reflect the flow characteristics of the oil fluid on the sealing contact surface. Based on the actual worn surface, the evolution laws of the surface topography under different wear mechanisms are revealed. On this basis, a numerical algorithm based on HPD and PSD is proposed to realize the predictive reconstruction of the worn surface topography. The reconstructed surface can be applied to the oil percolation simulation to explore the influence of surface topography evolution on sealing performance and leakage characteristics. The research results provide a theoretical basis for understanding the wear mechanism of polymer, predicting the sealing life, and revealing the failure mechanism.

Keywords. Reciprocating seal, percolation analysis, surface reconstruction, polymer wear, oil leakage

1. INTRODUCTION

Hydraulic systems are widely used in key fields such as aerospace, shipbuilding, and high-end equipment manufacturing due to their high-power density, rapid response, and excellent reliability [1]. In hydraulic systems, the actuator is the core actuator component, and its performance and stability play a decisive role in the overall operation of the system. Among

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

them, the reciprocating seal mechanism is a key functional link in the actuator. The stability, reliability, and service life of its sealing performance are directly related to the working efficiency and operational safety of the actuator [2]. The failure of the seal will not only significantly reduce the transmission efficiency of the hydraulic system but also increase maintenance costs [3].

Reciprocating seal structures exhibit distinct multiscale coupling characteristics, integrating the functions of both a sealing pair and a friction pair. Their sealing interfaces often experience a complex mixed lubrication state. Although a fluid oil film effectively separates some contact areas, the film thickness is insufficient to completely cover the microscopic topography of the sealing surface, leading to unavoidable direct contact between surface roughness peaks [4]. Mechanical contact between these submicron-scale roughness peaks not only disrupts the continuity of the oil film but can also trigger complex wear mechanisms within the contact area, including localized stress concentration, severe plastic deformation, adhesive wear, and material delamination [5].

Due to the complex and variable working conditions of sealing pairs, traditional testing methods have difficulty accurately capturing their contact state and stress evolution. Current research on the failure modes and life prediction of aerospace actuator seals is primarily based on material fatigue damage criteria. However, the microscopic topography characteristics of the contact surface are crucial for revealing the mixed lubrication behaviour, as well as the friction and wear mechanisms during service [6, 7]. As the actuator continues to operate, the contact surface topography evolves, and the lubrication state also changes accordingly. The interaction between the two ultimately drives the degradation of seal performance and wear failure. Therefore, obtaining the characteristics of surface topography evolution during service and conducting an in-depth analysis of the bidirectional coupling mechanism between lubrication and wear are key to revealing the essential mechanism of seal failure [8].

To understand the contact behaviour of rough surfaces, Persson et al. proposed a multiscale contact theory model for the elastic-plastic contact behaviour of random rough surfaces [9, 10]. This model assumes that the material exhibits an elastic response below the yield stress and undergoes plastic flow above it. On this basis, a linear relationship between the actual contact area and the normal load is derived. In the absence of plastic deformation, the actual contact area of a rough surface with self-affine fractal characteristics shows a power-law relationship with the lateral scale of the nominal contact area. This model emphasizes the surface power spectrum density (PSD) as the core parameter to characterize the multiscale characteristics of the rough surface and establish a quantitative relationship between it and macroscopic properties such as contact area, pressure distribution, and friction [11, 12].

Based on this contact theory, Persson et.al further expanded their research, introducing a coupling model of percolation theory and contact mechanics to explore the leakage behaviour and microscopic mechanism of seals under static and dynamic conditions [13]. Through multi-scale analysis of the power spectrum (PS) of rough surfaces, theoretical and numerical simulations showed that the formation and evolution of percolation channels are the key driving factors for leakage at the sealing interface. On this basis, Lorenz et al. developed a leakage prediction method based on the effective medium theory (EMT), treating the oil flow at the sealing interface as propagating in multiple independent microchannels, and combined the surface power spectrum to analyze the statistical evolution

of the seal contact area and gap height [14]. Bottiglione et al. further integrated the seal geometry, applied load, and surface roughness parameters to construct a comprehensive numerical model that can simulate contact randomness and predict leakage rate [15]. Scaraggi et al. proposed a numerical method to study the oil extrusion behaviour between soft elastic bodies and random rough rigid surfaces based on the transient mixed elastohydrodynamic lubrication theory [16].

Peng et al. analysed the differences between the multi-asperity model and the magnification-based percolation model in predicting sealing performance [17]. The study focused on the numerical prediction of gas leakage in metal sealing structures. By comparing the simulation results with the experimental data, it was found that the magnification-based percolation model performed better in leakage prediction accuracy. The model can directly characterize the microscopic topography characteristics of the sealing contact surface and analyze the spatial distribution of the contact and gap areas based on the PSD function, thereby revealing the generation mechanism of the oil percolation channel. Therefore, the model shows advantages in evaluating the influence of surface topography on local leakage behaviour.

However, the percolation model still has certain limitations:

- Current percolation calculations are mainly based on the PSD function, making it difficult to effectively couple the effects of seal load, temperature, and dynamic wear under real working conditions;
- Existing percolation theories use one-dimensional (1D) PSD functions to describe surface roughness distribution, which cannot capture the anisotropic characteristics of the worn surface.

To address the limitations of existing percolation models in coupling to actual operating conditions and characterizing surface anisotropy, this paper proposes a method for simulating oil percolation based on numerically reconstructed surface topography. This method conducts seal wear tests under typical operating conditions to obtain actual act contact surface topography data at different service stages. Based on this data, the worn surface is numerically reconstructed, and parameter extraction is performed. Furthermore, the statistical distribution and evolution of characteristic wear topography parameters are analysed, and a correlation model between wear topography and percolation characteristics is constructed, which allows for quantitative prediction and mechanistic understanding of oil percolation behaviour on worn sealing surfaces.

2. WEAR TEST

2.1. Test method

The RTEC MFT-5000 multi-functional wear testing machine was used to conduct wear tests on the commonly used material PTFE (polytetrafluoroethylene) for the wear ring. To investigate the wear characteristics of the PTFE specimens under high-temperature conditions, the specimens were heated for 15 minutes before the wear test to ensure that the material of the specimens reached thermal stability. An infrared thermal imager was used to perform real-time imaging and temperature measurement of the PTFE specimens. The specific test parameters are shown in Table 2.1.

Table 2.1 PTFE material wear test parameters

Parameters	Value
Test material	PTFE
Wear type	Ball-on-disk
Roughness of specimen	3-4μm
Grinding piece	Cr15 ball with a diameter of 4 mm
Single stroke	2mm
Frequency	3Hz
Normal load	10N
Test time	10 mins
Test temperature	20, 60℃

The worn surface can be observed and measured using two methods. A digital microscope is employed for micro-imaging of the worn surface, capturing the details of wear marks, analysing surface damage modes, and identifying different wear mechanisms. A laser confocal profilometer is used to acquire data on wear tracks and local surface topography. This data is then imported into Mountains Map software for analysis to obtain high-precision measurements of wear depth and surface topographic features. A 3D surface topography model of the worn area is reconstructed, providing data support for further analysis of the worn surface characteristics. The equipment used for wear tests is shown in Figure2.1.

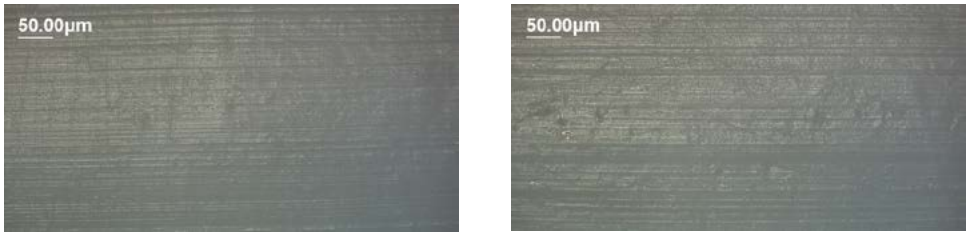


(a) MFT-5000 wear testing machine (b) Laser confocal profilometer

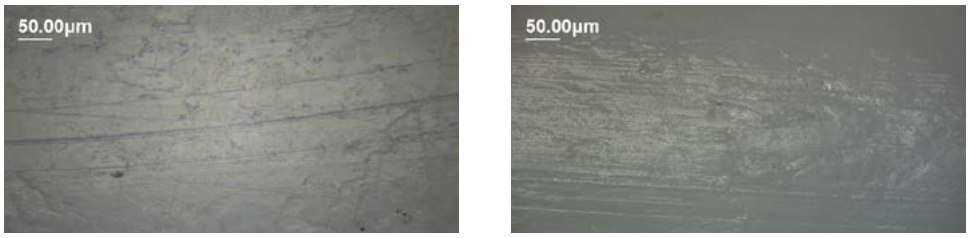
Figure 2.1. Wear test equipment

2.2. Test results

Figure2.2 shows the worn surface of PTFE at different temperatures and times. At 20°C, the PTFE surface exhibits typical scratch-like abrasive wear characteristics. The wear mechanism at this point is primarily driven by the cutting action of hard particles, resulting in a plowing effect. When the temperature rises to 60°C, the surface material starts to peel off, and localized depressions appear, predominantly exhibiting a ridge-like adhesive wear topography. The interfacial adhesion of the material is enhanced under high temperature conditions, making adhesive wear the dominant wear mechanism.



(a) 20℃, 10 mins (b) 20℃, 30 mins



(c) 60℃, 10 mins (d) 60℃, 30 mins

Figure 2.2. Worn surfaces of PTFE under different temperatures and durations

The variation of PTFE friction coefficient with wear time at different temperatures is shown in Figure2.3. From a wear mechanism perspective, abrasive wear occurs when hard abrasive particles embed themselves in the PTFE surface and scratch across it, creating a cutting or plowing effect. Adhesive wear primarily occurs when localized adhesion nodes form between the rough peaks of the two contacting surfaces, leading to material removal due to tearing caused by relative motion. At room temperature, PTFE is rigid and elastic, resulting in a small actual contact area and high local pressure, which in turn generates high interfacial shear forces. However, at high temperatures, the softening of PTFE reduces adhesion strength, reducing the adhesion between adhesion nodes.

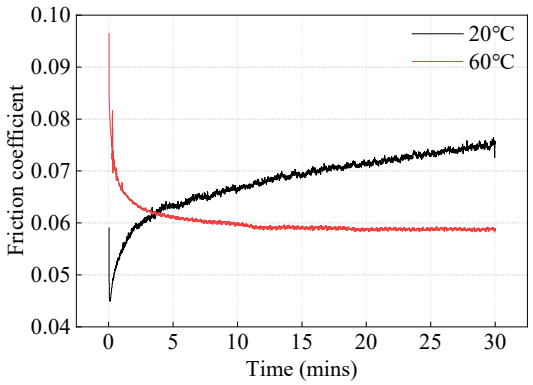
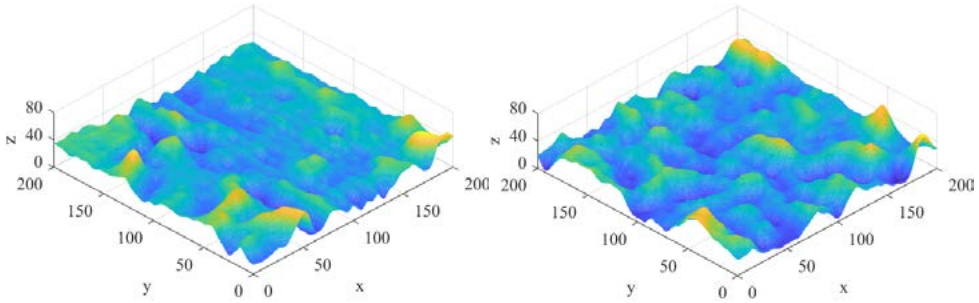


Figure 2.3. Friction coefficient variation of PTFE

The wear rate is calculated by scanning the average wear depth on the worn surface. The wear rates at 20°C and 60°C are $5.86 \times 10^{-5} mm^3/Nm$ and $9.08 \times 10^{-5} mm^3/Nm$, respectively. Notably, the wear volume of PTFE does not show a positive correlation with the friction coefficient.

2.3. Worn surface topography

With a surface sampling accuracy of 0.7 μm, a 200×200 node area in the centre of the worn surface is selected for quantitative analysis of surface topography. Typical surface topography at 20°C and 60°C are shown in Figure2.4. At 20°C, the worn surface exhibits anisotropic characteristics, with wear scars exhibiting clear directionality. Wear primarily occurs in a specific direction. At 60°C, the worn surface exhibits isotropic characteristics, caused by the formation of uniform and discrete adhesive nodes on the surface.



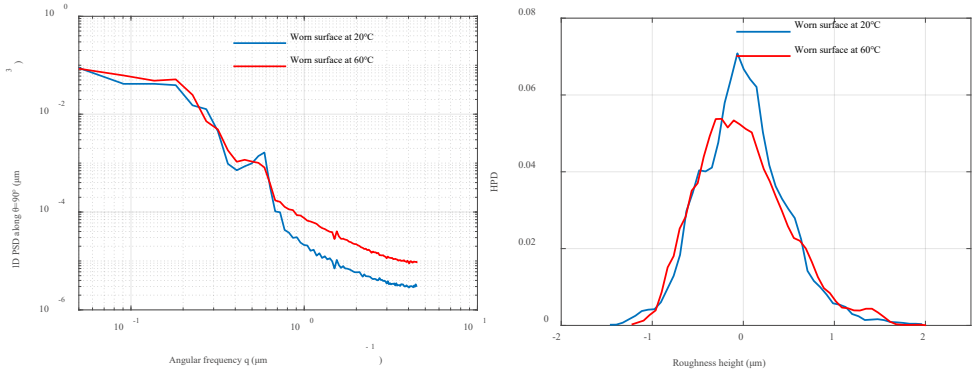
(a) 20°C, 30 mins (b) 60°C, 30 mins

Figure 2.4. Localized worn surface topography

1D PSD can be used to characterize the topography characteristics of rough surfaces, reveal microscopic changes in a single direction of the worn surface. By screening all frequency points projected to the specified direction angle θ at the spatial frequency q, and performing normalization processing, the 1D average PSD function is

$$C_{1D}(q,\theta)=\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}C_{2D}(q_x,q_y)\delta_d(q-q_\theta)dq_xdq_y}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(q-q_\theta)dq_xdq_y} \tag{2.1}$$

where δ is the Dirac function; q_θ is the specified direction projection length.



(a) 1D PSD along $\theta=90°$ (b) HPD

Figure 2.5. Worn surface statistical characteristics

Figure2.5 shows the 1D PSD along the friction direction (defined as $\theta=90°$) and the height probability distribution (HPD) of roughness at 20℃ and 60℃. The 1D PSD variation

of the worn surface reflects the transition between its anisotropic and isotropic characteristics. At 20°C, the worn surface exhibits strong anisotropy, with slight surface fluctuations along the friction direction, resulting in a low PSD of the high-frequency domain. At 60°C, the worn surface tends to be isotropic, with the energy distribution of the surface topography being relatively uniform in different directions, resulting in a higher overall PSD. The result is consistent with the observation of the worn surface topography, indicating that the wear mechanism significantly impacts the worn surface structure and its spectral characteristics.

The root mean square (RMS) roughness increases with temperature. The softening of PTFE at high temperatures intensifies wear, resulting in a rougher worn surface. The HPD of worn surfaces at both temperatures exhibits positive skewness, with numerous peaks formed by material spalling, exhibiting an asymmetric distribution. Higher temperatures exacerbate the unevenness of the worn surface height.

3. PREDICTIVE RECONSTRUCTION OF WORN SURFACE

3.1. Surface reconstruction

The surface topography reconstruction method proposed by Pérez-Ràfols is adopted, in which the PS and HPD are used as inputs to generate random surfaces that simultaneously satisfy both statistical characteristics [18]. First, two random rough surfaces with height sequences Z_h^0 and Z_s^0 are generated. The surface Z_h^0 has a height distribution consistent with the given HPD but does not meet the target PS requirement, whereas surface Z_s^0 satisfies the desired PS but deviates from the target HPD.

Both surfaces are then subjected to two-dimensional Fourier transforms, yielding their frequency-domain counterparts $\hat{Z}_h^0$ and $\hat{Z}_s^0$. Then, an iterative correction process is applied to gradually adjust the surfaces so that they meet both target distributions. This process alternates between two main steps:

- Adjusting the surface in the frequency domain to match the target PS.
- Correcting the surface in the spatial domain to match the target HPD.

Through iteration, both the PS and HPD of the reconstructed surface converge toward the specified target distributions, thereby realizing a random surface generation method that simultaneously satisfies the prescribed PS and HPD. The PS of Z_h^i is corrected and the Z_s^{i+1} that satisfies the target PS is characterized using the spectral representation method, that is

$$Z_s^{i+1} = FFT^{-1}\left[\sqrt{S} \times exp\left(i\,\phi_h^i\right)\right] \tag{3.1}$$

where, ϕ_h^i represents the phase in the frequency domain, which can be expressed in terms of the real and imaginary parts of $\hat{Z}_h^i$ obtained from the previous iteration as follows:

$$\phi_h^i = tan^{-1}\left(\frac{Im\left(\hat{z}_h^i\right)}{Re\left(\hat{z}_h^i\right)}\right) \tag{3.2}$$

At this stage, the PS of Z_s^{i+1} satisfies the target, but the HPD does not meet the requirements. By mapping the height distribution of Z_h^{I} onto Z_s^{i+1}, a new iteration of Z_h^{i+1}

that satisfies the target HPD can be obtained. Three types of errors are introduced to evaluate how well the generated surface matches the target surface in terms of PS and HPD [19].

i. Height probability distribution error err_h — quantifies the deviation between the actual height distribution of the generated surface and the target height distribution:

$$err_h = \sum_{i=1}^{N_b} |n_{ih} - n_{is}| \cdot \omega_b \tag{3.3}$$

where n_{ih} and n_{is} represent the frequency of occurrence of surface heights within a given height interval in the histograms of the target and generated surfaces, respectively; N_b is the total number of height intervals, and ω_b is the interval width.

ii. Power spectrum error (nonzero region) err_s — reflects the discrepancy between the nonzero power spectrum values of the generated surface and those of the target surface:

$$err_s = \sqrt{\frac{1}{N_s} \sum_{q_i=q_0}^{q_1} \left(\frac{|\hat{Z}_s(q_i)| - |\hat{Z}_h(q_i)|}{|\hat{Z}_s(q_i)|} \right)^2} \tag{3.4}$$

where $\hat{Z}_s(q_i)$ and $\hat{Z}_h(q_i)$ are the power spectrum values of the target surfaces at spatial frequency q_i; N_s denotes the number of nonzero spatial frequencies in the target power spectrum; q_0 and q_1 are the lower and upper cutoff spatial frequencies.

iii. Power spectrum error (zero region) err_{s0} — characterizes the performance of the generated surface in the frequency range where the target power spectrum is zero:

$$err_{s0} = \frac{\sqrt{\frac{1}{N_{s0}} \sum_{q_i<q_0 | q_i>q_1} |\hat{Z}_h(q_i)|^2}}{\frac{1}{N_s} \sum_{q_0 \le q_i \le q_1} |\hat{Z}_s(q_i)|} \tag{3.5}$$

where N_{s0} denotes the number of spatial frequencies for which the target power spectrum is zero. The iterative process continues until all three errors fall below the specified thresholds, at which point the generated surface is considered to meet the target requirements.

3.2. Topography characterization

(1) Power spectral density

The power spectral density ratio function corresponding to the 1D PSD frequency point $q_{1D}^i \in S_q$ is expressed as follows:

$$R_{1D}(q_{1D}^i) = \begin{cases} 1, & q_{1D}^i < q_s \\ \alpha_{psd}, & q_{1D}^i \ge q_s \end{cases} \tag{3.6}$$

where, α_{psd} is the power-spectrum scaling coefficient, and q_s is the adjustment frequency threshold. Using an interpolation function, the ratio function $R_{1D}(q_{1D}^i)$, which is discretely defined on the 1D PSD, is mapped into the two-dimensional (2D) frequency domain to obtain for each point $q_{2D}^{k,l}$ in the frequency coordinate system, its scaling factor [20]:

$$R_{2D}(q_{2D}^{k,l}) = R_{1D}(q_{1D}^i) + \frac{R_{1D}(q_{1D}^{i+1}) - R_{1D}(q_{1D}^i)}{q_{1D}^{i+1} - q_{1D}^i} \cdot (q_{2D}^{k,l} - q_{1D}^i) \tag{3.7}$$

where R_{2D} is the 2D PSD ratio function, and $q_{2D}^{k,l} \in \left[q_{1D}^{i}, q_{1D}^{i+1}\right]$. As worn surfaces typically exhibit pronounced anisotropy, the adjustment in the frequency domain is applied only to the frequency components along the target direction. Let $\tilde{\theta}$ denote the direction angle of an arbitrary frequency in the frequency domain, θ is the target direction, and θ_t is the directional tolerance. The directional mask matrix for the spatial frequency $q_{2D}^{k,l}$ in the 2D frequency domain is

$$Mar\left(q_x^k, q_y^l\right) = \begin{cases} 1, \left|\tilde{\theta}\left(q_x^k, q_y^l\right) - \theta\right| < \theta_t \\ 0, \left|\tilde{\theta}\left(q_x^k, q_y^l\right) - \theta\right| \geq \theta_t \end{cases} \tag{3.8}$$

The adjusted two-dimensional power spectral density function is

$$C_{2D}{}'\left(q_x^k, q_y^l\right) = \begin{cases} R_{2D}\left(q_{2D}^{k,l}\right) \cdot C_{2D}\left(q_x^k, q_y^l\right), & Mar\left(q_x^k, q_y^l\right) = 1 \\ C_{2D}\left(q_x^k, q_y^l\right), & Mar\left(q_x^k, q_y^l\right) = 0 \end{cases} \tag{3.9}$$

(2) Height probability density

Johnson transform is a distribution transformation method based on data statistical analysis, which can be used to convert between Gaussian sequences and non-Gaussian sequences [21]. The skewness and kurtosis of the input and output height sequences have the following relationship [22]:

$$Sk_0 = \frac{\left(\sum_{r=0}^{M-1}\sum_{s=0}^{N-1} f(r,s)^2\right)^{\frac{3}{2}}}{\sum_{r=0}^{M-1}\sum_{s=0}^{N-1} f(r,s)^3} Sk_\eta \tag{3.10}$$

$$Ku_0 = \frac{\left(\sum_{r=0}^{M-1}\sum_{s=0}^{N-1} f(r,s)^2\right)^2 Ku_\eta - 6\sum_{r=0}^{M-2}\sum_{p=r+1}^{M-1}\sum_{s=0}^{N-2}\sum_{q=s+1}^{N-1} f(r,s)^2 f(p,q)^2}{\sum_{r=0}^{M-1}\sum_{s=0}^{N-1} f(r,s)^4} \tag{3.11}$$

where Sk_0 and Ku_0 are the input skewness and kurtosis, and Sk_η and Ku_η are the given skewness and kurtosis. r, s, p, q is the index of the coordinate in space. The surface topography is regarded as a random process. By adjusting the power spectrum density and probability distribution characteristics, the generated surface topography has a specific skewness, kurtosis and spatial correlation. The autocorrelation function is a mathematical tool used to describe the correlation of random signals, often used in roughness modelling. The exponential form is

$$R(x,y) = S_q^2 \exp\left\{-2.3\left[(x/\beta_x)^2 + \left(y/\beta_y\right)^2\right]^{1/2}\right\} \tag{3.12}$$

where β_x and β_y are the autocorrelation lengths. By performing Fourier transform on the autocorrelation function, the output roughness height power spectral density function can be obtained:

$$S(k,l) = \frac{1}{MN}\sum_{r=0}^{M-1}\sum_{s=0}^{N-1} R(r,s)\exp\left(-\frac{j2\pi kr}{M}\right)\exp\left(-\frac{j2\pi ls}{N}\right) \tag{3.13}$$

where k,l is the frequency domain index. The two-dimensional digital filter function $f(r,s)$ can be obtained through the power spectrum density function:

$$f(r,s)=\frac{1}{MN}\sum_{r=0}^{M-1}\sum_{s=0}^{N-1}\sqrt{S(k,l)}\exp\left(-\frac{j2\pi kr}{M}\right)\exp\left(-\frac{j2\pi ls}{N}\right) \tag{3.14}$$

3.3. Predictive reconstruction

Numerical reconstruction of the worn surface relies on the PSD and HPD obtained from experiments as statistical characteristic parameters. However, due to limitations in the number of test surfaces and the scope of surface topography acquisition, it is difficult for the experimental worn surfaces to cover all possible topography characteristics. By adjusting the PSD and HPD extracted from existing test data, a reconstruction process independent of the specific measured surface can be achieved. It can be used to simulate and predict the topography of worn surfaces with arbitrary characteristics. Furthermore, the worn surface topography can be predictive based on operating conditions to simulate oil percolation. The predictive reconstruction process of the worn surface topography is shown in Figure3.1.

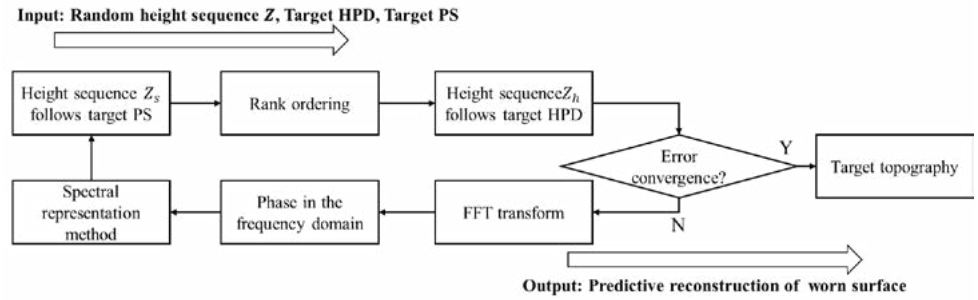
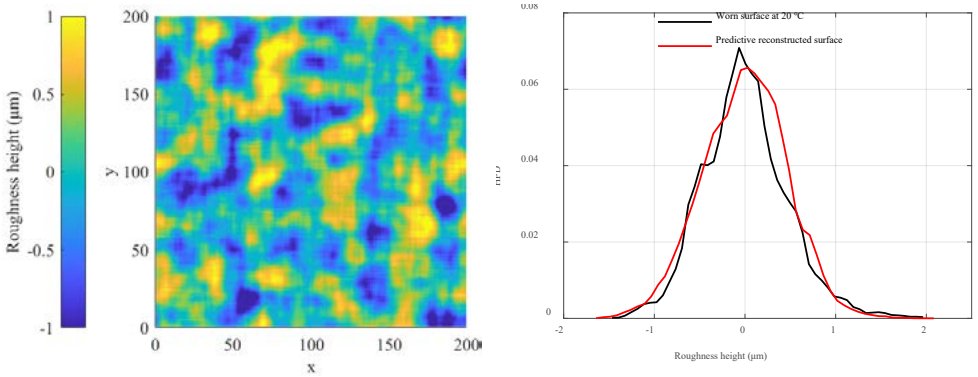
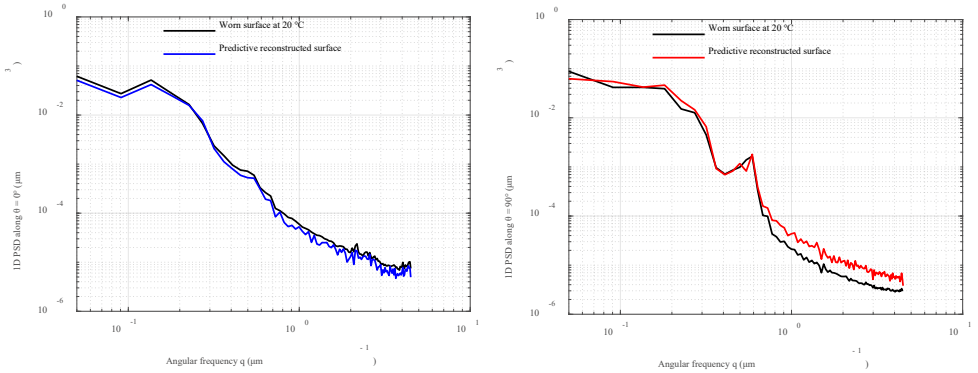


Figure 3.1. Flowchart for predictive reconstructed worn surface topography

To verify the effectiveness of the predictive reconstruction method, the worn surface at 20°C shown in Figure2.4 (a) is chosen as the research object. $\alpha_{psd}=1.8$, $q_s=0.6$ and $\theta_t=\pm5°$ are selected to adjust its PSD. The results are shown in Figure3.2.



(a) Reconstructed surface (b) HPD



(c) 1D PSD along $\theta = 0°$ (d) 1D PSD along $\theta = 90°$

Figure 3.2. Predictive reconstructed surface and its characteristics

Define $\theta = 90°$ as the seal's axial direction, i.e., the direction of reciprocating friction, and $\theta = 0°$ as the seal's circumferential direction, perpendicular to the reciprocating motion. By comparing the 1D PSD in these two directions, the anisotropic characteristics variation of the surface topography can be reflected.

The 1D PSD of the original worn surface in the high-frequency region of the friction direction is significantly lower than that in other directions, indicating that the surface exhibits significant anisotropy. During the prediction and reconstruction process, the 1D PSD in the high-frequency region is appropriately amplified to compensate for the surface anisotropy. The anisotropy of the predictive and reconstructed surface is weakened, tending towards an isotropic state, while the HPD of the roughness remains unchanged. The 1D PSD in the $\theta = 0°$ direction remains essentially the same before and after reconstruction. The surface energy increases in the $\theta = 90°$ direction and when $q > 0.6$, demonstrating that this method effectively adjusts the high-frequency characteristics in the target direction without affecting the roughness characteristics in other directions.

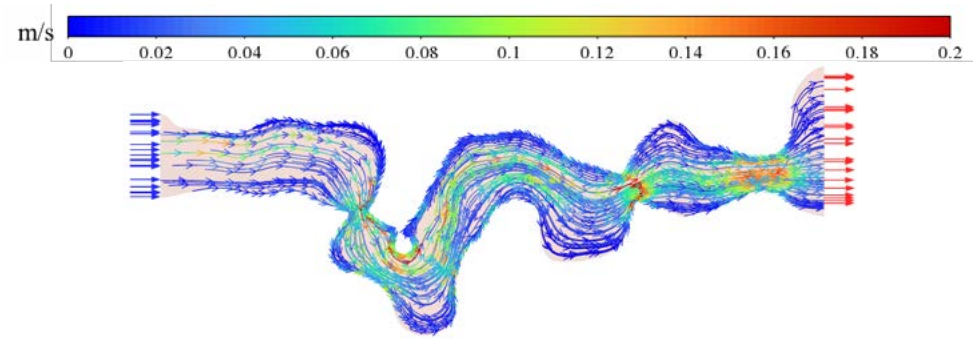
4. OIL PERCOLATION ANALYSIS

4.1. Actual worn surface

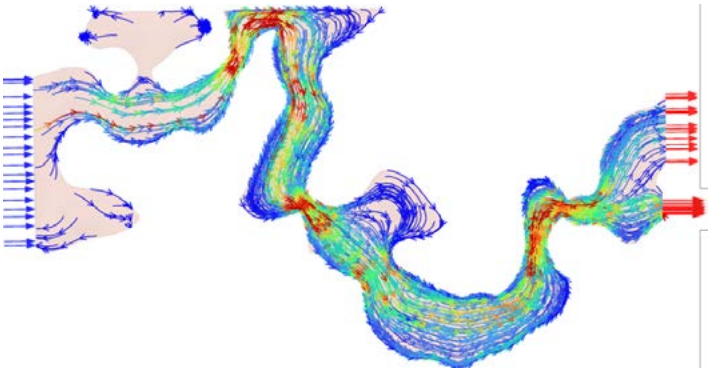
The paper adopts the oil percolation simulation method proposed in Ref [23] to conduct three-dimensional simulations of the percolation behaviour of worn surfaces obtained from tests at 20°C and 60°C. By extracting oil film thickness distribution data from the seal contact surface, an oil channel structure model of the oil percolation can be constructed, enabling the analysis of oil flow behaviour within the percolation channel. Figure4.1 shows the flow velocity distribution within the oil percolation channel on the sealing surface at different temperatures.

The oil leakage rates on the worn surface at 20°C and 60°C are $2.81 \times 10^{-5} cm^3/s$ and $3.62 \times 10^{-5} cm^3/s$. PTFE produces significant shear deformation and stress dispersion due to its excellent energy dissipation capabilities. This process suppresses the ability of abrasive particles to create deep scratches on the surface during ploughing effect. In contrast, adhesive wear primarily results from localized adhesion between the materials at the contacting surfaces. Under reciprocating shear loads, adhesive nodes fracture and material

delamination occur, and their formation is closely related to the adhesive strength of the contact interface. At high temperatures (e.g., 60°C), thermal softening of the material enhances adhesive strength, making it more likely to form deeper wear scars.



(a) Actual worn surfaces at 20°C



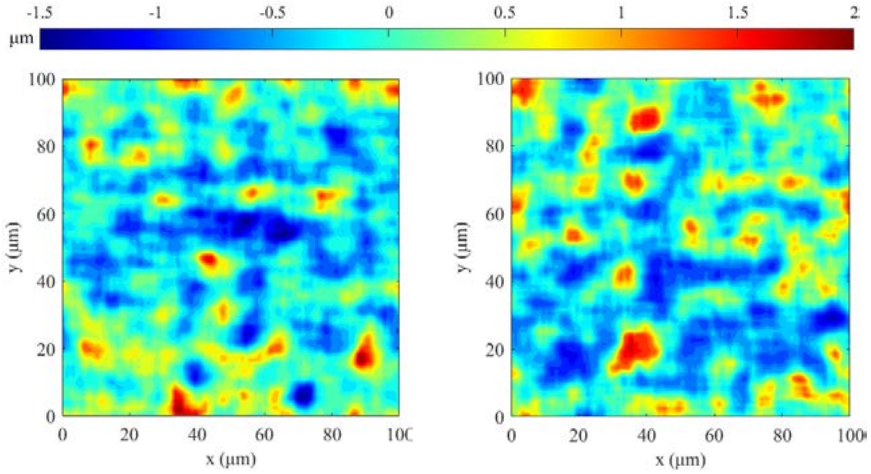
(b) Actual worn surfaces at 60°C

Figure 4.1. Oil percolation characteristics of actual worn surfaces

Therefore, the worn surface at 60°C is more susceptible to enhanced adhesive wear than that at 20°C. The oil flow channels exhibit variation in cross-sectional geometry, particularly in the critical constriction, where a larger channel area and higher oil flow rate are observed. This phenomenon indicates that high-temperature operating conditions not only alter the dominant wear mechanism but also significantly affect the lubrication and flow properties of the sealing surface, impacting the stability of sealing performance.

4.2. Predictive reconstructed worn surface

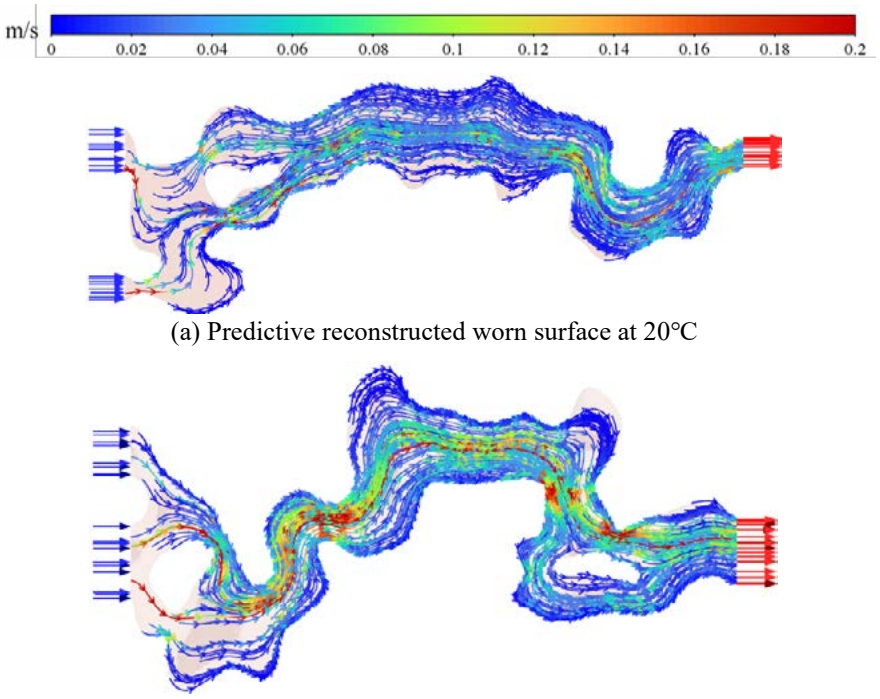
To investigate the influence of the reconstructed worn surfaces on the analysis results of percolation characteristics, analyses are conducted on these surfaces at 20°C and 60°C. The results are then compared with those from the test worn surfaces. The reconstructed worn surfaces for percolation analysis are shown in Figure4.2. The RMS roughness of the reconstructed surfaces is consistent with that of the actual worn surfaces. This consistency eliminates the interference of roughness differences on the percolation behaviour, thereby more accurately evaluating the influence of surface reconstruction on the percolation channel structure, velocity distribution, and other related factors.



(a) Reconstructed surface at 20℃ (b) Reconstructed surface at 60℃

Figure 4.2. Predictive reconstructed worn surfaces for percolation analysis

Figure4.3 shows the oil percolation characteristics on the predictive reconstructed worn surfaces at different temperatures.



(a) Predictive reconstructed worn surface at 20℃

(b) Predictive reconstructed worn surface at 60℃

Figure 4.3. Oil percolation characteristics of predictive reconstructed worn surfaces

At 20°C, a wide but shallow percolation channel is formed along the reciprocating motion direction on the reconstructed worn surface, and the oil flow velocity is relatively low. At

60°C, the width of the percolation channel at the critical constriction is narrower, but due to the increase in the wear depth of the sealing surface, a larger effective flow area is formed, and the local maximum flow velocity reached 0.181 m/s. Compared to the width, the height of the percolation channel directly determines the oil film thickness and has a greater impact on the overall flow capacity of the channel.

The numerical calculation of the net leakage rate of oil on the reconstructed worn surface is conducted, and it is compared with the analysis results of the leakage rate on the corresponding actual worn surface, as shown in Figure4.4.

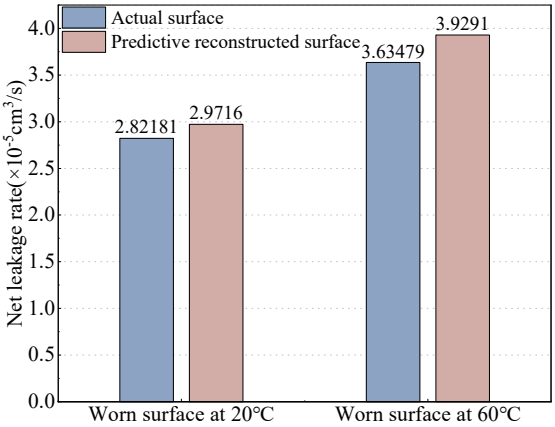


Figure 4.4. Comparison of net leakage rate results

Since the reconstructed surface maintains the same topography characteristic parameters as the actual surface, such as PSD, HPD, RMS roughness, etc., both show similar leakage characteristics. The leakage rate of the reconstructed surface is slightly higher than that of the actual surface. The error arises from the differences in the topography structure arrangement and topological form of the reconstructed surface compared to the actual surface. The topography affects the contact state and percolation channels, resulting in different leakage rates. However, the error is relatively small, and the reconstructed surface can more accurately reflect the change trend of the leakage rate under different temperatures. It still has good reliability in predicting oil leakage.

5. CONCLUSION

As the wear temperature increases, the surface damage pattern changes from continuous scratches along the reciprocating direction to discrete, localized adhesive failure nodes, indicating a shift from abrasive to adhesive wear. Material exfoliation depressions, resulting from shear-adhesive nodes, begin to appear on the worn surface. During the abrasive wear-dominated phase, the friction coefficient is high. In contrast, during the adhesive wear-dominated phase, the softening of the PTFE surface reduces shear strength and the friction coefficient. A method for predicting and reconstructing worn surface topography is developed. This method reconstructs a surface with identical PSD and HPD statistical characteristics from experimentally derived worn surface topography.

Furthermore, the anisotropic characteristics of the original surface are modified by adjusting its 1D PSD along a single direction. The HPD of the surface roughness can be modified using a Johnson transform. Comparison of oil percolation analysis results between the original and predictive reconstructed surfaces reveals that, due to differences in morphological arrangement and topology, the leakage rate calculations exhibit minimal error, accurately reflecting the leakage rate trends of the worn surfaces at different temperatures. At room temperature, grooves created by abrasive wear easily form percolation channels that penetrate the sealing surface. However, due to the shallow wear depth, leakage is low. At high temperatures, adhesive failure nodes cause significant material removal, forming large cross-sectional percolation channels that increase overall flow capacity and leakage.

6. REFERENCES

[1] G. K. Nikas, 'Eighty years of research on hydraulic reciprocating seals: Review of tribological studies and related topics since the 1930s', Proc. Inst. Mech. Eng., Part J, vol. 224, no. 1, pp. 1–23, Jan. 2010. doi: https://doi.org/10.1243/13506501jet607

[2] B. Wang, X. Meng, X. Peng, and Y. Chen, 'Experimental investigations on the effect of rod surface roughness on lubrication characteristics of a hydraulic O-ring seal', Tribol. Int., vol. 156, pp. 106791–106791, Apr. 2021. doi: https://doi.org/10.1016/j.triboint.2020.106791

[3] D. Liu, S. Wang, and C. Zhang, 'A multiscale wear simulation method for rotary lip seal under mixed lubricating conditions', Tribol. Int., vol. 121, pp. 190–203, Jan. 2018. doi: https://doi.org/10.1016/j.triboint.2018.01.007

[4] D. Zhu and Y.-Z. Hu, 'A computer program package for the prediction of EHL and mixed lubrication characteristics, friction, subsurface stresses and flash temperatures based on measured 3-D surface roughness," Tribol. Tran., vol. 44, no. 3, pp. 383–390, Jan. 2001. doi: https://doi.org/10.1080/10402000108982471

[5] A. P. Nyemeck, Noël Brunetiere, and B. Tournerie, 'A mixed thermoelastohydrodynamic lubrication analysis of mechanical face seals by a multiscale approach', Tribol. Trans., vol. 58, no. 5, pp. 836–848, Mar. 2015. doi: https://doi.org/10.1080/10402004.2015.1023407

[6] G. K. Nikas and R. S. Sayles, 'Study of leakage and friction of flexible seals for steady motion via a numerical approximation method', Tribol. Int., vol. 39, no. 9, pp. 921–936, Sep. 2006. doi: https://doi.org/10.1016/j.triboint.2005.09.003

[7] Y. Z. Hu and K. Tonder, 'Simulation of 3-D random rough surface by 2-D digital filter and fourier analysis', Int. J. Mach. Tools. Manuf., vol. 32, no. 1–2, pp. 83–90, Feb. 1992. doi: https://doi.org/10.1016/0890-6955(92)90064-N

[8] T. Yin, J. Fu, D. Wei, Z. Xie, E. Zhang, and B. Zhao, 'Thermal contact of roughness macro and micro model to reveal seal leakage mechanism under mixed lubrication', Tribol. Int., vol. 191, p. 109176, Mar. 2024. doi: https://doi.org/10.1016/j.triboint.2023.109176

[9] B. N. J. Persson, 'Elastoplastic Contact between Randomly Rough Surfaces', Phys. Rev. Lett., vol. 87, no. 11, Aug. 2001. doi: https://doi.org/10.1103/physrevlett.87.116101

[10] B. N. J. Persson, 'Theory of rubber friction and contact mechanics', J. Chem. Phys., vol. 115, no. 8, pp. 3840–3861, Aug. 2001. doi: https://doi.org/10.1063/1.1388626

[11] B. N. J. Persson, 'Contact mechanics for randomly rough surfaces', Surf. Sci. Rep., vol. 61, no. 4, pp. 201–227, Jun. 2006. doi: https://doi.org/10.1016/j.surfrep.2006.04.001

[12] C. Yang and B. N. J. Persson, 'Contact mechanics: contact area and interfacial separation from small contact to full contact', J. Phys.: Condens. Matter, vol. 20, no. 21, p. 215214, Apr. 2008. doi: https://doi.org/10.1088/0953-8984/20/21/215214

[13] B. N. J. Persson and C. Yang, 'Theory of the leak-rate of seals', J. Phys.: Condens. Matter, vol. 20, no. 31, p. 315011, Jul. 2008. doi: https://doi.org/10.1088/0953-8984/20/31/315011

[14] B. Lorenz and B. N. J. Persson, 'Leak rate of seals: Effective-medium theory and comparison with experiment', Eur. Phys. J. E, vol. 31, no. 2, pp. 159–167, Feb. 2010. doi: https://doi.org/10.1140/epje/i2010-10558-6

[15] F. Bottiglione, G. Carbone, and G. Mantriota, 'Fluid leakage in seals: An approach based on percolation theory', Tribol. Int., vol. 42, no. 5, pp. 731–737, May 2009. doi: https://doi.org/10.1016/j.triboint.2008.10.002

[16] M. Scaraggi and B. N. J. Persson, 'Time-Dependent Fluid Squeeze-Out Between Soft Elastic Solids with Randomly Rough Surfaces', Tribol. Lett., vol. 47, no. 3, pp. 409–416, Jun. 2012. doi: https://doi.org/10.1007/s11249-012-9996-6

[17] C. Peng, F. Fischer, K. Schmitz, and Hubertus Murrenhoff, 'Comparative analysis of leakage calculations for metallic seals of ball-seat valves using the multi-asperity model and the magnification-based model', Tribol. Int., vol. 163, pp. 107130–107130, Nov. 2021. doi: https://doi.org/10.1016/j.triboint.2021.107130

[18] F. Pérez-Ràfols and A. Almqvist, 'Generating randomly rough surfaces with given height probability distribution and power spectrum', Tribol. Int., vol. 131, pp. 591–604, Mar. 2019. doi: https://doi.org/10.1016/j.triboint.2018.11.020

[19] Y. Wang and Y. Liu, 'Preparation of measured engineering surfaces for modeling tribological systems, part I: Characterization and reconstruction', Tribol. Int., vol. 199, pp. 110026–110026, Jul. 2024. doi: https://doi.org/10.1016/j.triboint.2024.110026

[20] T. D. B. Jacobs, T. Junge, and L. Pastewka, 'Quantitative characterization of surface topography using spectral analysis', Surf. Topogr.: Metrol. Prop., vol. 5, no. 1, pp. 013001, Jan. 2017. doi: https://doi.org/10.1088/2051-672x/aa51f8

[21] N. L. Johnson, 'Systems of Frequency Curves Generated by Methods of Translation', Biometrika, vol. 36, no. 1/2, p. 149, Jun. 1949. doi: https://doi.org/10.2307/2332539

[22] Tatsunori Tomota, Yasuhiro Kondoh, and Toshihide Ohmori, 'Modeling Solid Contact between Smooth and Rough Surfaces with Non-Gaussian Distributions', Tribol. Trans., vol. 62, no. 4, pp. 580–591, Apr. 2019. doi: https://doi.org/10.1080/10402004.2019.1573341

[23] J. Wang, J. Li, N. Bauer, and K. Schmitz, 'Numerical study of leakage rate in reciprocating seal with non-Gaussian topography characteristics using percolation mechanism', Tribol. Int., vol. 199, p. 109955, Nov. 2024. doi: https://doi.org/10.1016/j.triboint.2024.109955

Biographies



Jiehao Wang received the bachelor's degree in mechanical design, manufacturing and automation from Shanghai University of Engineering Science in 2017, the master's degree in mechatronic engineering from University of Shanghai for Science and Technology in 2020. He is currently a Ph.D. candidate majoring in mechatronic engineering at Tongji University. His research areas include seals, friction and wear and surface topography.

18

Wear modeling and sensitivity analysis for seals in hydraulic cylinders.

Diego M. Chamorro, Gyan Wrat, Mohit Bhola, Henrik C. Pedersen and Jesper Liniger

dmcr@energy.aau.dk, gywr@energy.aau.dk, mobh@energy.aau.dk, hcp@energy.aau.dk, jel@energy.aau.dk.

Abstract.

Fluid power systems are widely used in demanding and harsh environments such as marine and offshore installations, where service, maintenance, and repair costs are extremely high. This includes usage in offshore wind turbines, which impose stringent and rigorous requirements on system lifetime, reliability, and availability. As a result, accurately predicting these complex systems' lifetime, degradation patterns, and optimal service intervals is becoming increasingly important for ensuring operational efficiency and minimizing levelized cost of energy. However, the persistent absence of reliable and validated methods to accurately predict component lifetimes under realistic and variable operating conditions means that such systems are typically designed, operated, and maintained based on historical failure data, empirical observations, and conservative safety assumptions.

Among the most critical and failure-prone components in these systems are hydraulic cylinders, particularly the hydraulic seals, which are highly prone to wear and degradation over time. Previous studies have effectively combined Finite Element Analysis (FEA) with advanced numerical simulation techniques to model and predict seal wear performance under a wide variety of operating conditions. Extending on these established approaches, this study develops and applies a comprehensive sensitivity analysis examining the correlation and interaction between key influencing factors such as ambient temperature, pressure, and duty cycle, to predict seal wear.

This research incorporates wear simulation modeling in which Ansys is used to model the contact and deformation behavior, and a numerical framework is developed in MATLAB to predict fluid and asperity contact pressures, lubricant film thickness, and wear. This approach integrates detailed surface contact modeling, material properties, and realistic boundary conditions. Through a series of carefully designed simulations, this study attempts to predict the wear behavior of hydraulic seals over time, offering valuable insights that can be used for improving component performance and enhancing the overall reliability and longevity of fluid power systems in demanding operational environments.

Keywords: 1 Sensitivity analysis, 2 Wear Prediction, 3 Finite element analysis, 4 Hydraulic cylinders, 5 Numerical simulations.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Reciprocating hydraulic seals serve as auxiliary components in hydraulic systems, yet they are also among the most critical parts because they prevent sealed fluid from leaking. If the seal operating mechanism is not well understood, it may be used incorrectly, potentially leading to seal failure and unpredictable consequences. White and Denny were the pioneers in analyzing the behavior and sealing efficiency of reciprocating seals between 1944 and 1947 [1] [2]. A further wave of experimental and theoretical analysis has been carried out since then.

According to experimental findings, reciprocating movement of the piston rod induces hydrodynamic forces. Two analytical methods are commonly used to study fluid film inverse hydrodynamic lubrication. The inverse hydrodynamic lubrication (IHL) approach simplifies the model by assuming full-film lubrication and exceeding the impact of fluid dynamic pressure on seal deformation, thereby limiting its effectiveness [3]. The elasto-hydrodynamic lubrication (EHL) method, however, incorporates a balance between static contact pressure and fluid pressure, which leads to a more realistic depiction of lubrication film behavior in a sealing context [4].

Among the type of seals that have been employed to analyse the lubrication characteristics, we can find O-ring seal, V- cup seals, and step seals. Previous studies on these kinds of seals have revealed that the lubrication in this interface is in the mixed lubrication state [5].

As a result of the interference fit between the seal and the inner wall of the cylinder, both surfaces initially possess fine asperities formed during the shaping and machining process. The reciprocating motion of the piston rod introduces frictional forces that not only smooth the cylinder wall but also cause migration and wear of the seal's surface profile. This wear leads to changes in the geometry of the sealing lip, which in turn, can significantly influence the integrity and performance of the sealing system [6] [7].

The most common kind of seal failure is wear, so it is crucial to accurately predict seal wear characteristics. For the prediction of wear in hydraulic seals, both experimental and simulation methods have been employed. Among the empirical work that has been analysed, we can find the studies of Ridgway et al. [8], which analysed the wear behaviour of the seal ring in slurry-free environment and reached the conclusion that non-uniform packing compression is not the sole factor influencing the lifetime and reliability of the gland seal. Yang et al. [9] analysed the friction behaviour of seals containing different PolyTetraFluoroEthylene (PTFE) content. Moreover, Lee et al. [10] suggested an experimental method for the accelerated wear of fluororubber (FKM) seals by introducing 0 – to 1 microns alumina particles mixed with the lubricant; nonetheless the outcome of this experiment indicated that the wear performance was similar to normal working conditions. Pan Deng et al. [11] evaluated the impact of different pressures to analyse their effect on the wear rate of PTFE piston seals. The outcome of this experiment indicated that the wear performance was similar to normal working conditions. Liu et al. [12] investigated how varying pressure levels influence the wear rate of PTFE piston seals. The outcome of these experiments indicate that wear was more pronounced at higher pressures. In addition to this, Gao et al. [13] analysed the wear behaviour of different coating seals. They found that the

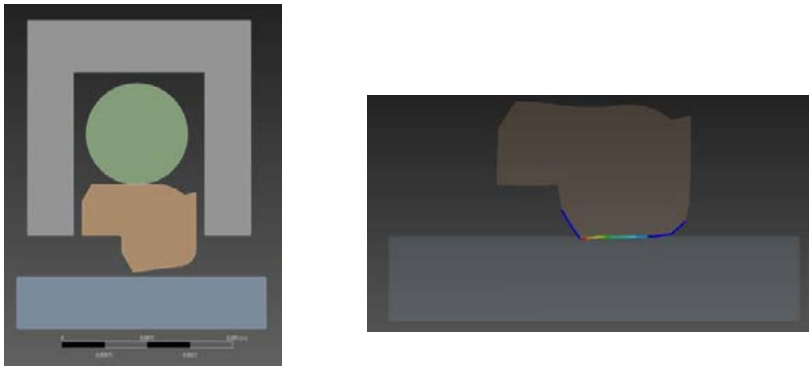
highest possible temperature and the increment of temperature related to the friction surface affect the wear resistance and coating durability. Furthermore, the relationship between leakage rate and skewness surface topography was explored by Ruhmani et al. [14]. Even though the tribological characteristics of a seal can be explored through experimental means, this entails manpower and resources. As a result, it has been explored to use numerical simulations to predict the wear behaviour. One of the first analytical models to predict wear on seals was implemented by Yang et al. [15]. In addition to this, Beksi et al. [16] implemented a birth-death technique on an FEM model to better study the seal wear in reciprocating seals. Moreover, other studies like the ones presented by Frolich et al. and Li et al. [17] [18] proposed methods to assess the effects of factors like friction and heat generation on wear. Nonetheless, these methods primarily examine the friction behaviour associated with the seals large–scale deformation under external compression, while neglecting the impact of microscopic factors at the lubrication interface on seal wear. The macro deformation of the seal generates heat, which in turn reduces the fluid's viscosity. Thus, the macro deformation mechanics are closely linked with the micro-hydrodynamics of the fluid film. Moreover, studies performed by Liu et al. [19] and Zhang et al. [20] employed A multiscale wear simulation approach for rotary lip seals operating under mixed lubrication conditions, considering the microscopic characteristics of the interacting surfaces. The results of these studies showed that decreasing the axial asperity density and increasing the circumferential asperity density can effectively reduce seal wear.

Even though other studies have modelled wear in hydraulic seals using various approaches, including the modelling of micro-elastohydrodynamic effects under different working conditions, no sensitivity analysis based on this approach has been conducted for hydraulic seals to examine how ambient temperature, sealing pressure, and speed during outstroke affect the wear. The purpose of this study is to describe wear using a micro-elasto hydrodynamic model that incorporates the effects of lubrication in a reciprocating seal. This is achieved by calculating the parameters present at the contact interface to account for lubrication effects and to improve the wear prediction using Archard’s wear equation.

2. SEAL DESCRIPTION

This study aims to analyse the contact between a step-seal made of an O-ring and a Turcon Ring as shown in Figure 1, Typically, the O-ring is manufactured from nitrile rubber, while the Turcon ring is made from a composite material based on polytetrafluoroethylene (PTFE). When the step seal is mounted on the rod, the sealing edge forms a steep contact pressure gradient on the high-pressure side and a gentler gradient on the low-pressure side. This distribution of pressure reduces fluid adhesion to the rod during the outstroke and helps return any remaining fluid film on the rod into the system.

The reciprocating rod step seal consists of an elastomeric O-ring, commonly manufactured from nitrile rubber (NBR), and a sealing ring made of a PTFE-based composite that is in direct contact with the rod surface. The O-ring provides elastic preload and accommodates installation and pressure-induced deformations, while the PTFE sealing ring is responsible for sealing performance due to its low friction coefficient and good wear resistance. This combination of materials enables the seal to operate reliably under mixed lubrication conditions, where both fluid film lubrication and asperity contact occur during reciprocating motion.



a) b)

Figure 1. a) Ansys FEA model of the Step-seal including the O-ring, the Turcon ring and the Housing, with the direction of the stroke, where the left side is pressurised and the right side is ambient pressure. b) Sealed pressure contact distribution from Ansys FEA model, red indicates the highest contact pressure.

3. MODEL OF MIXED LUBRICATION

3.1. Numerical Simulation

This study aims to characterize wear through a micro-elastohydrodynamic model that accounts for the influence of lubrication in reciprocating seals. The model calculates key parameters related to the lubricated contact, intending to determine the asperity contact pressure for use in the Archard's wear equation. On a microscale level, the lubrication model displays robust coupling among fluid mechanics, asperity contact mechanics, deformation mechanics, and thermo-mechanics effects. Therefore, an iterative computational procedure has been built and employed to manage the coupling relationship. These relationships are described by a set of governing equations (including the Reynolds equation), which are solved using the finite element method.

The employed multiscale wear simulation model integrates a microscale mixed Thermo-Elasto-Hydrodynamic (TEHD) lubrication model with a macroscale wear model to simulate the wear behaviour of reciprocating lip seals under mixed lubrication conditions. The microscale model is designed to resolve critical tribological parameters such as lubrication film thickness, hydrodynamic pressure, and asperity contact pressure. The governing equations form a coupled multiscale framework in which each equation represents a dominant physical mechanism governing reciprocating seal wear under mixed lubrication. The Reynolds equation describes the hydrodynamic behavior of the lubricant film, while asperity contact is modeled using the Greenwood–Williamson formulation, which represents solid–solid interactions through a statistical description of surface asperities. The model links microscale contact, hydrodynamic pressures, and thermal relations account for frictional heating and its influence on lubricant viscosity. The equations are implemented through an iterative numerical scheme in which hydrodynamic and asperity contact pressures are solved simultaneously and jointly determine the film thickness. Frictional heat

generation updates the lubricant viscosity and feeds back into the pressure solution until convergence is achieved at each time step. After convergence, the asperity contact pressure is incorporated into a modified Archard wear model to compute material loss. As a result, the model behavior emerges from the coupled evolution of lubrication, contact, thermal, and wear mechanisms rather than from any single equation in isolation.

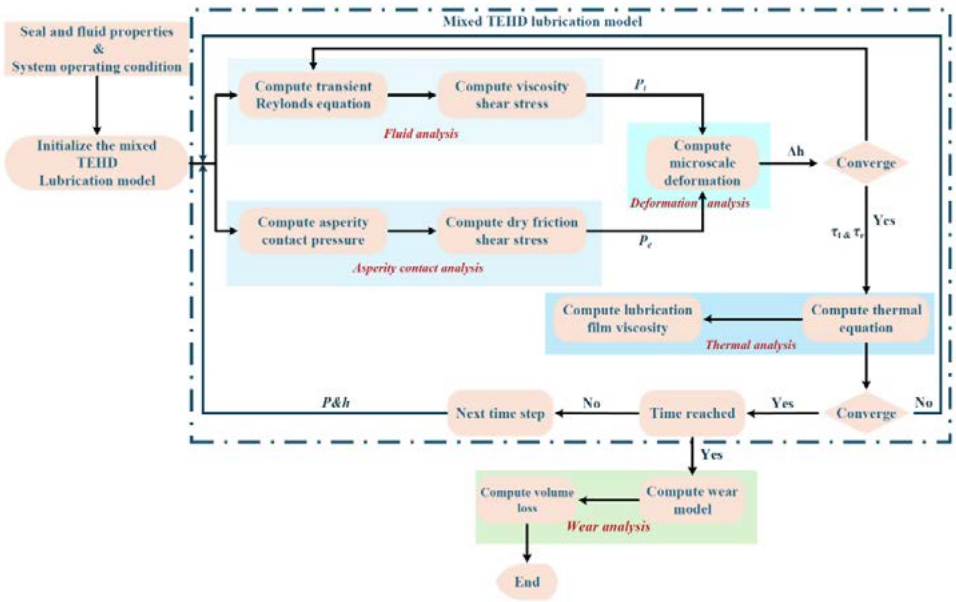


Figure 2. Schematic of the computational procedure TEHD: Thermo-elasto-hydrodynamic.

The thermal mechanics, fluid mechanics, and asperity contact mechanics, and deformation mechanics are recalculated until the lubrication film thickness, and the average temperature achieve convergence which has been explained in the following sections. Afterwards, the asperity contact pressure is used in the Archard's modified equation to calculate wear.

3.2. Mixed TEHD Lubrication Model

The multiscale wear simulation model used here incorporates a microscale mixed Thermo-Elasto-Hydrodynamic (TEHD) approach with a macroscale wear model to simulate the wear behavior of reciprocating lip seals under mixed lubrication conditions. The microscale model is designed to resolve critical tribological parameters such as lubrication film thickness, hydrodynamic pressure, and asperity contact pressure, employing the version of the Reynolds equation that considers cavitation given by Payvar et al. [21]:

$$\frac{\partial}{\partial x}\left(\phi_{xx}\frac{h^3}{\mu}\frac{\partial(F\Phi)}{\partial x}\right) 6u\frac{\partial}{\partial x}\{[1+(1-F)\Phi](h_T+\sigma\phi_{s.c.x})\} + 12\frac{\partial}{\partial t}\{[1+(1-F)\Phi](h_T+\sigma\phi_{s.c.x})\} \quad (1)$$

Where, h is the film thickness, μ is the fluid viscosity, σ is the root mean square roughness of the seal, and u is the rod speed. ϕ_{xx} is the pressure flow factor, related to the fluid

obstruction due to surface roughness, and $\phi_{s.c.x}$ is the shear flow factor. This represents the extra flow transport resulting from surface sliding. F is the cavitation index and Φ is the fluid pressure or the density of the cavitation region.

In the full film region: $\Phi > 0$, $F = 1$, $p_f = F\Phi,\ \rho = \rho_f$ (2)

In the cavitation region: $\Phi < 0$, $F = 0$, $p_f = 0,\ \rho = \rho_f(1 + \Phi)$ (3)

When solving the Reynolds equation, it is essential to define both the inlet and outlet boundary conditions. During the outstroke and instroke phases, the fluid pressure corresponds to the sealed pressure on the fluid side and the ambient pressure on the air side, respectively.

At fluid side: $\phi\ (x = 0) = p_s$ (4)

At air side: $\phi\ (x = L_x) = p_e$ (5)

where L_x is defined as the sealing region, p_s corresponds to the pressure within the sealed area, and p_e refers to the surrounding ambient pressure.

As described by Ran et al. [22] the film thickness can be expressed as:

$$h = h_{st} + \Delta h \quad (6)$$

where h_{st} denotes the static film thickness, and Δh represents the elastic deformation of the sealing lip in the normal direction, which can be determined using the influence coefficient method. Ran et al. [22] mentioned that the mean truncated film thickness is expressed as follows:

$$h_T = \int_{-h}^{\infty}(h + \delta)f(\delta)d\delta \quad (7)$$

Where δ represents the roughness amplitude of the seal surface measured relative to the mean level. $f(\delta)$ denotes the probability density function of δ. When the asperity heights of the seal surface follow a Gaussian distribution, the mean truncated film thickness can be expressed as follows:

$$h_T = \frac{h}{2} + \frac{h}{2}erf\left(\frac{h}{\sqrt{2}}\right) + \frac{1}{\sqrt{2\pi}}\, e^{-\frac{h^2}{2}} \quad (8)$$

Where erf represent the error function

The shear stress resulting from the fluid viscosity can then be determined by the following relationship, as described by [23]:

$$\tau_f = \frac{-\mu u}{h}(\phi_f - \phi_{fss}) - \phi_{fpp}\frac{h}{2}\frac{dp_f}{dx} \quad (9)$$

where ϕ_f, ϕ_{fss} and ϕ_{fpp} are shear stress flow factors.

Asperity contact analysis

When the elastic seal is compressed against the rod, contact occurs at multiple microscopic asperities within the nominal contact area. The distance or separation between the seal and rod surfaces greatly affects key sealing characteristics, including leakage and friction.

The asperity contacts P_c between the seal surface and the cylinder wall can be described by the Greenwood Williamson (G–W) surface contact model [24].

$$P_c = \frac{4}{3}\frac{1}{(1-\nu_s^2)}\sigma^{3/2}R^{1/2}\eta\frac{1}{\sqrt{2\pi}}\int_H^\infty (z-H)^{3/2}e^{-z^2/2}\,dz \quad (10)$$

$$P_a = \int_d^\infty (z-d)\;\phi(z)\,dz \quad (11)$$

Where η is the seal asperity density, R is the asperity radius, and σ is the RMS roughness of the seal surface

E represents the equivalent Young's modulus, which can be expressed as

$$E = \frac{E_{seal}}{1-\upsilon_{seal}^2} \quad (12)$$

Where E_{seal} and υ_{seal} are the Young's elastic modulus and Poisson's ratio of the seal, respectively.

The dry friction shear stress τ_c is also described by the following formula found in [22]

$$\tau_c = -\frac{u}{|u|}fp_c \quad (13)$$

For this case, f is the dry friction coefficient.

Thermal analysis

Since viscosity, the key parameter in the Reynolds equation, depends on temperature, its effect on sealing performance must be considered. Friction in the sealing area raises the surface temperature, and the resulting heat is mainly dissipated through conduction and convection. Part of the heat is conducted through the seal and rod, while a smaller portion is carried away by the lubricant. Wang et al. [25] introduced a modified quasi-3D energy equation that accounts for surface roughness to analyse temperature distribution T in the fluid film. Because the lubricating layer is only a few microns thick, the heat removed by the lubricant is negligible. Salant et al. [26] proposed a simplified model assuming all heat transfers to the rod, given its higher thermal conductivity compared to the seal. Where T_e is the ambient temperature L_x is the length of the surface contact, ρ_r is the density of the rod, c_r is the specific heat capacity of the rod, k_r is the thermal conductivity of the rod.

$$T - T_e = 1.07\frac{QL_x}{2k_r}\left[\frac{\rho_r c_r|u|L_x}{2k_r}\right]^{-1/2} \quad (14)$$

$$Q = (\tau_f + \tau_c)u \quad (15)$$

Because the viscosity of the lubricating film changes with both pressure and temperature, a viscosity–pressure–temperature relationship is used to calculate the film viscosity in the sealing region, which is expressed as [25]:

$$\mu = \mu_0 exp\left\{(ln\mu_0 + 9.67) \times \left[(1 + 5.1 \times 10^{-9} p_f)^{\zeta} \times \left(\frac{T - 138}{T_e - 138}\right)^{-\psi} - 1\right]\right\} \quad (16)$$

where μ_0 represents the viscosity under the atmospheric pressure p_e and ambient Temperature T_e. The dimensionless constants ζ and ψ can be defined as:

$$\zeta = a/[5.1 \times 10^{-9}(ln\,\mu_0 + 9.67)\,] \quad (17)$$

$$\psi = b(T_e - 138)/(ln\,\mu_0 + 9.67) \quad (18)$$

where a denotes the Barus viscosity–pressure coefficient (Pa^{-1}) and b represents the Reynolds viscosity–temperature coefficient (K^{-1}).

Deformation analysis

For this approach, an offline FEA model is first adopted for analysing deformation mechanics. The first is the macroscopic deformation used to calculate sealed pressure distribution and contact length; the other is the microscopic parameters such as the asperity contact pressure and the fluid pressure employed for the wear calculation.

The static contact pressure, P_{st}, is assumed to be the combined effect of the hydrodynamic pressure and the asperity contact pressure, expressed as $P_{st} = P_f + P_c$ (19).

Because the lubricating film is relatively thin, the microscale deformation of the sealing lip is assumed to exhibit linear elastic behaviour. By dividing the contact length into N axial nodes, the normal deformation of the seal lip at the *i*-th node caused by the film, (the change in film thickness described in eq. (2) Δh) can be calculated by the following expression as described by [22]:

$$\Delta h_i = \sum_{j=1}^{N} C_{ij}\,(p_f + p_c - p_{st})_j \quad (20)$$

Where C_{ij} is the deformation coefficient matrix obtained from FEA.

Wear analysis

Multiple theoretical and experimental models in previous work have been done regarding material wear, such as abrasive wear and adhesive wear.

Archard's wear model is one of the most commonly employed, and widely used for the calculation of wear on seals.

For this case, the wear volume loss is considered to be linear to the sliding distance, load, and the inverse of hardness of the softer material. The original equation is expressed as:

$$V = \frac{K}{H} WS \quad (21)$$

In this scenario V is the volume loss, K is the dimensionless Archard coefficient, W is the contact force, H is the surface hardness of the softer material, and S is the sliding distance.

The wear modulus then is

$$k = K/H \quad (22)$$

Then, the Archard's wear model may be written as.

$$V = kWS \quad (23)$$

When this is divided by the contact area A, the wear film thickness can be obtained by

$$h_W = kp_nS \quad (24)$$

in which case p_n is the normal contact pressure.

The lubricating conditions present in the rod seal are of mixed lubrication. For this micro hydrodynamic lubrication model a competition between the fluid and asperity contact is considered. If lubrication is not present in sealing zone, the normal contact pressure would be equal to the static pressure. Under mixed lubricating conditions, the sealing lip is supported by both the fluid dynamic pressure and asperity contact pressure. Wear in this case is generated through asperity contact; as a result, the wear model is redefined as

$$h_w = kp_cS \quad (25)$$

Where p_c is the asperity contact pressure obtained through the mixed lubrication model.

Table 1. Main simulation parameters

Parameter	Value	Parameter	Value
Ambient Temperature (T_e)	40°C	Compression rate of seal (ε)	8 %
Sealed Pressure (p_s)	0-20 MPa	Rod diameter (d)	25.4 mm
Ambient Pressure (p_e)	0.1 MPa	Initial sectional height of seal (l)	3.18 mm
Reference Viscosity (η_0)	0.04 Pa.s	Thermal conductivity rod (k_r)	25 W·m^{-1}·K^{-1}
Density of rod (ρ_r)	7700 kg·m^{-3}	Viscosity–temperature coefficient (β)	3.18 × 10^{-2} K^{-1}
Stroke length (S)	200 mm	Root mean square roughness of seal (σ)	0.6 μm
Speed of rod (u)	20–200 mm/s	Specific heat capacity of rod (c_r)	460 J·kg^{-1}·K^{-1}
Dry friction coefficient (f)	0.25	Viscosity–pressure coefficient (α)	2 × 10^{-8} Pa^{-1}
Young's modulus of seal (E_{seal})	380 MPa	Poisson's ratio of seal (ν seal)	0.4

Wear coefficient (k)	1.2×10^{-5} mm³/Nm		

Simulation values reported in Table 1 are based upon [19] and [22].

4 RESULTS AND DISCUSSIONS

The following results are presented showing how the asperity contact pressure varies as function of the sealed pressure at different temperatures. Figure 3 displays the curves of asperity contact pressure at different sealed pressures and temperatures.

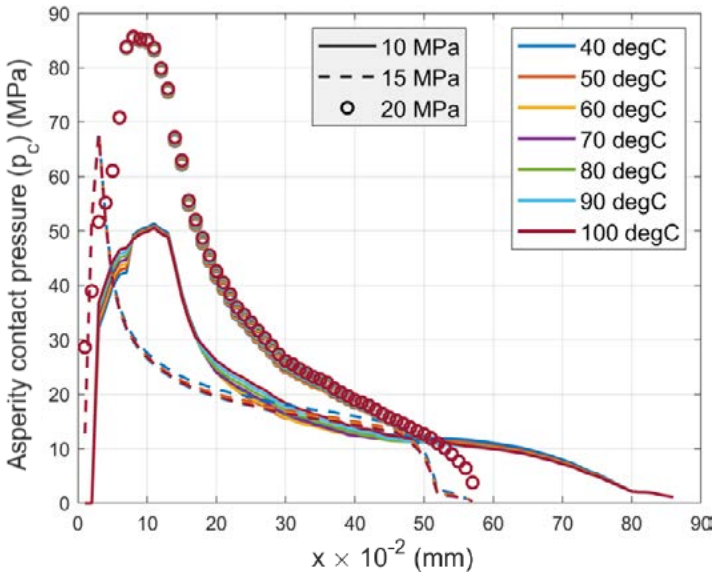


Figure 3. Curves of asperity contact pressures at various sealed pressures under varying ambient temperatures obtained from the computational procedure TEHD (thermal elasto-hydrodynamic).

Figure 3 displays the evolution of the asperity contact pressure along the sliding direction x for sealed pressures of 10, 15, and 20 MPa at ambient temperatures ranging from 40 °C to 100 °C. Generally, the asperity pressure displays a sharp initial peak near the contact entry, followed by a gradual decay towards the outlet region. This behaviour would reflect the localized concentration of contact stress at the starting edge, where the load is supported by asperities.

The decay of the asperity contact pressure follows a similar trend for all sealed pressures and temperatures, indicating that the contact mechanics remains dominated by surface topography rather than temperature-induced geometric changes. Compared with the lower sealed pressures, the overall asperity contact pressure magnitudes are higher as the sealed pressure increases, with maximum values of the asperity contact pressure approaching 65–75 MPa at a sealed pressure of 15 MPa and the asperity contact pressure would reach a maximum at 80–90 MPa at a sealed pressure of 20 MPa near the contact entrance.

Similar to the behaviour observed for each sealed pressure individually, all profiles nearly overlap through the entire contact domain, indicating that within this load range, the asperity deformation shows minor variation to ambient temperature variations. Nonetheless, when the ambient temperature rises, a slight increase in asperity contact pressure distribution can be seen.

Finally, the results indicate that the overall asperity contact pressure distributions augment when the sealed pressure increases. As the external load grows, the contact asperities experience augmented deformation, resulting in increased localized contact stresses at the edge of the contact zone.

4.1 Combined effect of sealed pressure, rod speed and ambient temperature.

The wear behaviour of the reciprocating hydraulic step rod seal and its lubrication performance are greatly influenced by the rod's operating outstroke speed. Therefore, the proposed method is used to analyse the seal wear and lubrication characteristics under various rod outstroke speeds, sealing pressures, and ambient temperatures. The results of this analysis may be seen in Figure 4. Showing the resulting wear as a function of sliding outstroke speed, temperature, and sealed pressure.

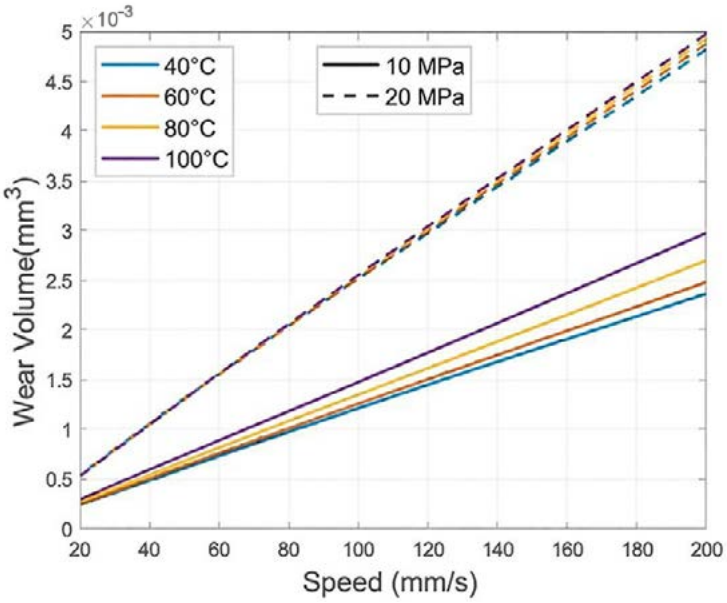


Figure 4. Wear volume per stroke of the rotary lip seal versus speed under different sealed pressures and ambient temperatures.

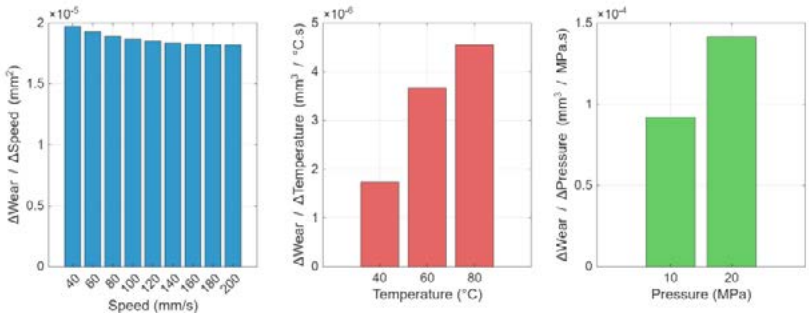


Figure 5. Sensitivity Analysis of Wear with each parameter.

The set of line plots shown in Figure 4 displays a dependence of wear volume on both parameters with different behaviour observed for each pressure condition. At both pressures, the wear volume increases with rising outstroke speed and temperature. This effect is more pronounced at 20 MPa, where wear volumes are higher all over the full range of test conditions. This suggests that the increased contact pressure would intensify the wear process, mainly due to augmented asperity deformation, and slightly worsened lubrication conditions at higher ambient temperature. However, the change in wear for both the 10 MPa and 20 MPa sealed pressures is more strongly influenced by outstroke speed than by the increase in ambient temperature. For the different speed cases analysed in this study we can observe that wear increases with rod outstroke speed. Having a maximum value of wear at a rod speed of 200 mm/s, and for the distinct sealed pressures the asperity contact profile has a low variation when the ambient temperature is changed for all the analysed sealed pressure cases. It is observed that wear is slightly higher at an increased ambient temperature of 100 degrees Celsius for both sealed pressures, with a lower variation between temperatures for the 20 MPa sealed pressure case. In Figure 5 it is observed that both speed and pressure have more or less the same influence, this confirms that controlling pressure and speed is key to reducing wear in hydraulic seals.

5 CONCLUSIONS

In this paper, a sensitivity analysis regarding accelerating wear factors, including sealed pressure, rod outstroke speed, and ambient temperature, is presented based on a wear simulation model which is developed for a reciprocating hydraulic step-rod seal under mixed lubricating conditions. From the model, it is observed that with increased sealed pressure the resulting calculated asperity contact pressure increases. In addition, ambient temperature rising from 40°C to 100°C causes a modest increase in the peak asperity contact pressure, despite the low overall variation. It is observed that with an increase in sealed pressure, the value of wear also increases, as a higher sealed pressure results in higher asperity contact pressure and increased contact between asperities. Moreover, it is observed for all sealed pressure scenarios that the calculated wear increases in an arguably mostly linear manner with the outstroke speed. The increased effect that the augmentation of both sealed pressure and operating outstroke speed had on the calculated wear could be explained by the elevated contact between asperities and the worsening lubricating conditions. It is observed that the highest accelerating factor out of the ones analysed in this study is the sealed pressure,

followed by operating outstroke rod speed, whereas the ambient temperature effect has a lower effect across all operating conditions.

In conclusion, this study presents a comprehensive investigation into the wear behavior of reciprocating hydraulic step-rod seals operating under mixed lubrication conditions, with a particular focus on the sensitivity of wear to key accelerating factors: sealed pressure, rod outstroke speed, and ambient temperature. Using a simulation model grounded in Archard's wear equation, the analysis reveals that sealed pressure is the most dominant contributor to wear progression. As sealed pressure increases, the asperity contact pressure rises significantly, leading to intensified surface interactions and greater material loss. Rod outstroke speed also plays a critical role, as higher speeds exacerbate dynamic contact conditions and reduce the effectiveness of lubrication, further accelerating wear. In addition, ambient temperature directly influences wear. This impact is attributed to the thermal exchange between the environment and the lubricated contact zone, with a direct effect on viscosity, confirming ambient temperature as a contributing factor in wear.

To complement the simulation-based insights, the study also emphasizes the importance of accelerated testing protocols for evaluating seal wear characteristics under controlled yet extreme operating conditions. These tests would simulate high-pressure, high-speed, and variable-temperature environments to induce wear over shortened timeframes, allowing for rapid assessment of seal durability and performance. Accelerated testing not only validates the simulation results but also provides empirical data that can be used to refine predictive models and improve material selection and seal design. Employing simulation approaches, this study offers a robust framework for understanding and managing wear in hydraulic sealing systems. The overarching vision is to enable predictive maintenance, reduce downtime, and enhance the reliability and operational lifespan of hydraulic components. Ultimately, controlling sealed pressure and rod speed emerges as the most effective strategy for minimizing wear. This dual approach simulation-driven analysis and accelerated testing would lay the foundation for smarter, more resilient hydraulic systems in demanding industrial applications.

6 REFERENCES

[1] C. M. White, D. F. Denny, “The sealing mechanism of flexible packings,” No. 3-47, HM Stationery Office, UK, 1948.

[2] G. K. Nikas, “Eighty years of research on hydraulic reciprocating seals: review of tribological studies and related topics since the 1930s,” Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol., vol. 224, no. 1, pp. 1–23, 2010. DOI: 10.1243/13506501JET607

[3] B. Kim, J. Suh, B. Lee, Y. Chun, G. Hong, J. Park, and Y. Yu, “Numerical analysis via mixed inverse hydrodynamic lubrication theory of reciprocating rubber seal considering the friction thermal effect,” Appl. Sci., vol. 13, no. 1, p. 153, 2023. DOI: 10.3390/app13010153.

[4] N. Dakov, C. Schüle, “Benefits and Applications of EHL Analysis in Sealing Technology on the Example of a Hydraulic Step-Seal,” Proc. 21st Int. Sealing Conf., pp. 635-645, Stuttgart, Germany, 2022. DOI: 10.61319/GTNCA8DN

[5] B. Wang, X. Li, X. Peng, Y. Li, Y. Chen, and J. Jin, “Thermoelastohydrodynamic mixed lubrication of combined rod seals operating at high pressures and speeds: mathematical modeling and numerical analysis,” J. Tribol., vol. 146, no. 4, p. 044104, 2024. DOI: 10.1115/1.4063267.

[6] L. Mazza and E. Goti, “Failure and damage of reciprocating lip seals for pneumatic cylinders in dry conditions,” Lubricants, vol. 12, no. 4, p. 119, 2024. DOI: 10.3390/lubricants12040119.

[7] B. Traber and O. Nahrwold, “Condition monitoring of reciprocating seals,” Proc. 21st Int. Sealing Conf. (ISC), Stuttgart, Germany, Oct. 12-13 2022. DOI: 10.61319/FNCU9FBU.

[8] N. Ridgway, C. B. Colby, and B. K. O’Neill, “Slurry pump gland seal wear,” Tribol. Int., vol. 42, no. 11-12, pp. 1715–1721, 2009. DOI: 10.1016/j.triboint.2009.04.047.

[9] X. B. Yang, X. Q. Jin, Z. M. Du, T. S. Cui, and S. K. Yang, “Frictional behaviour investigation on three types of PTFE composites under oil-free sliding conditions,” Ind. Lubr. Tribol., vol. 61, no. 5, pp. 254–260, 2009. DOI: 10.1108/00368790910976087.

[10] S. H. Lee, S. S. Yoo, D. E. Kim, B. S. Kang, and H. E. Kim, “Accelerated wear test of FKM elastomer for life prediction of seals,” Polym. Test., vol. 31, no. 8, pp. 993–1000, 2012. DOI: 10.1016/j.polymertesting.2012.07.013

[11] D. Pan, B. Fan, X. Qi, Y. Yang, and X. Hao, “Investigation of PTFE tribological properties using molecular dynamics simulation,” Tribol. Lett., vol. 67, no. 1, p. 28, 2019. DOI: 10.1007/s11249-019-1141-3.

[12] Y. Liu, D. Tang, J. Liu and J. Lumkes, “Wear behaviour of piston seals in flow meter of fuel dispenser under different pressure conditions,” *Flow Meas. Instrum.*, vol. 28, pp. 45–49, 2012.

[13] S. Gao, W. Xue, D. Duan, S. Li, and H. Zheng, “Effect of thermal–physical properties on the abradability of seal coating under high-speed rubbing condition,” Wear, vol. 394–395, pp. 20–29, 2018.

[14] E. Kozuch, P. Nomikos, R. Rahmani, N. Morris, and H. Rahnejat, “Effect of shaft surface roughness on the performance of radial lip seals,” Lubricants, vol. 6, no. 4, p. 99, 2018. DOI: 10.3390/lubricants6040099.

[15] L. Yang and T. Moan, “Numerical modeling of wear damage in seals of a wave energy converter with hydraulic power take-off under random loads,” Tribol. Trans., vol. 54, no. 1, pp. 44–56, 2010. DOI: 10.1080/10402004.2010.518340.

[16] L. Yan, L. Guan, D. Wang, and D. Xiang, “Application and prospect of wear simulation based on ABAQUS: A review,” Lubricants, vol. 12, no. 2, p. 57, 2024. DOI: 10.3390/lubricants12020057.

[17] D. Frölich, B. Magyar, and B. Sauer, “A comprehensive model of wear, friction and contact temperature in radial shaft seals,” Wear, vol. 311, no. 1-2, pp. 71–80, 2014. DOI: 10.1016/j.wear.2013.12.030.

[18] J. Fan, H. Ji, and Q. Wang, “Study of wear and friction heat generation models of finger seals based on energy theory of tribology system,” AIP Advances, vol. 12, no. 4, p. 045202, 2022. DOI: 10.1063/5.0080686

[19] D. Liu, S. Wang, C. Zhang, ‘A multiscale wear simulation method for rotary lip seal under mixed lubricating conditions’, Tribology International, vol. 121, pp. 190–203, 2018. DOI: 10.1016/j.triboint.2018.01.007.

[20] Z. Yi, Z. Yi, ‘Study on lubrication characteristics of rotary combination seals under stress relaxation’, Engineering Research Express, vol. 5, no. 2, p. 025074, 2023. DOI: 10.1088/2631-8695/acddb8.

[21] P. P. Payvar and R. F. Salant, “A computational method for cavitation in a wavy mechanical seal,” J. Tribol., vol. 114, pp. 199–204, 1992. DOI: 10.1115/1.2920861.

[22] H. Ran, S. Wang, and D. Liu, “A multiscale wear model for reciprocating rod stepseal under mixed lubricating conditions based on linear elasticity,” *Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol.*, vol. 235, no. 1, pp. 161–180, 2021. DOI: 10.1177/1350650120902602.

[23] N. Patir and H. S. Cheng, “An average flow model for determining effects of three-dimensional roughness on partial hydrodynamic lubrication,” J. Lubr. Technol., vol. 100, pp. 12–17, 1978. DOI: 10.1115/1.2920861.

[24] J. A. Greenwood and J. B. P. Williamson, “Contact of nominally flat surfaces,” Proc. R. Soc. Lond. A, vol. 295, pp. 300–319, 1966. DOI: 10.1098/rspa.1966.0242.

[25] B. Wang, X. Peng, and X. Meng, “A thermo-elastohydrodynamic lubrication model for hydraulic rod O-ring seals under mixed lubrication conditions,” Tribol. Int., vol. 129, pp. 442–458, 2019. DOI: 10.1016/j.triboint.2018.08.044.

[26] B. Yang and R. F. Salant, “Elastohydrodynamic lubrication simulation of O-ring and U-cup hydraulic seals,” Proc. IMechE, Part J: J. Eng. Tribol., vol. 225, pp. 603–610, 2011.

Biographies



Diego M. Chamorro is working as a Research Assistant in the Mechatronic Systems section at AAU Energy, where he investigates accelerated testing methodologies in fluid power systems. He holds a Master's degree in Mechanical Engineering with a focus on tribology from Luleå University of Technology in Luleå, Sweden. His research interests include mechatronic systems, tribology, and fluid power systems.



Gyan Wrat is a Post Doctoral Researcher at Aalborg University in the Energy Department, currently researching wind turbine technology. He holds a PhD and a Master's degree in Mechatronics from the Indian Institute of Technology, India. His research interests lie in the field of Mechatronics systems and Fluid Power Fault and Control.



Mohit Bhola is a Researcher at the Mechatronics Section, Energy Department at Aalborg University located in Aalborg, Denmark. He holds a PhD degree in Mechatronics from the Indian Institute of Technology (Indian School of Mines), Dhanbad, India. His current research interests are Fault Tolerant Control of Hydraulic Pitch Systems for Wind Turbines,

Digital Hydraulics, Hybrid Hydrostatic Transmission Systems, System Modeling, Control, Hardware in Loop (HIL) Simulations, and Automation.



Henrik C. Pedersen is a Professor and Vice-head at the Department of Energy at Aalborg University, Denmark. His field of work is in the analysis, design, optimization, and control of mechatronic systems in general and fluid power systems and components in particular. Special focus is on efficiency, reliability, condition monitoring, fault detection and diagnostics, and control of fluid power systems and components. Author of > 170 publications within these areas and project leader (P.I.) for several research projects within this and related areas.



Jesper Liniger is an Associate Professor at Department of Energy, Aalborg University, Denmark. His research interest includes fault analysis and fault detection in fluid power and perception and navigation in subsea robotics.

Novel ultrasonic position sensor for hydraulic clamping cylinders

Authors: Markus Fritton, Julia Wick, Albert Dorneich

Balluff GmbH, Zabergäustraße 8, 73765 Neuhausen a.d.F., Germany,
albert.dorneich@balluff.de

Abstract.

Piston position detection for hydraulic actuators is a common task in hydraulic systems. Sensors based on several sensing principles can be found on the market. Nevertheless, a robust and compact solution for short stroke cylinders was missing. To fill this gap, a novel ultrasonic position sensor was designed, built and tested for applications in hydraulic actuators.

Keywords. Clamping position sensing, condition monitoring, clamping cylinder, block cylinder, digitalization in hydraulics, ultrasonic sensing

1. INTRODUCTION

Modern industrial production is characterized by a high degree of automation and digitalization. Hydraulic clamping cylinders for automated clamping in machining centers make a significant contribution to efficient and high-quality manufacturing of metal parts such as gearboxes and engine blocks [1].
Until recently, there were no sensors available that could detect and monitor both the clamping status, and the clamping travel of the hydraulic actuators used. In this paper we describe the development of a novel position sensor based on ultrasonic sensing.
The presented sensor enables direct, quick and continuous position measurement in short-stroke hydraulic cylinders up to 80 mm. Additional data for condition monitoring of hydraulic components with quality control for workpieces can be provided. This generates reliable information on the clamping status or piston position and lets the machine operator detect critical deviations or changes in the clamping process or hydraulic system early on. Therefore, the sensor is ideally suited for applications such as swing, lever and vertical clamping cylinders, block cylinders, hydraulic directional control valves, hydraulic chucks and small hydraulic cylinders for precise positioning.

2. SENSING PRINCIPLE: OPPORTUNITIES AND LIMITATIONS

2.1. Comparison of sensing principles

The goal of this work was the development of a piston position sensor for applications with limited available space. Solutions which need additional space around the hydraulic cylinder

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

were therefore not preferred. Besides, a very compact design is required to avoid damage to the sensor in rugged environments. Since it is not desirable to use add-on parts to transfer the internal piston movement to an add-on sensor outside the cylinder, the measurement should be executed directly inside the cylinder.

Many sensing principles are available for position measurement such as [2]:

- Optoelectronic sensors (triangulation and time-of-flight) using optical lenses are contactless, precise and widely applied in open space but will fail in operation inside the cylinder because of the high-pressure and the opaque hydraulic oil.
- Magnetostrictive contactless sensors [3] provide a robust position detection inside the cylinder up to pressures of several 100 bars but have a poor ratio of the length of the sensor housing to the detection range for strokes < 100 mm.
- Draw wire sensors are prone to wear because of the mechanical movement of the wire.
- Oil flow sensors can provide an indirect indication of the piston position by tracking the oil volume. These sensors are often inserted into the supply line of more than one cylinder and can therefore enable overall monitoring of the status of the hydraulic system only.
- Microwave sensors may be suitable but are not yet available on the market [4].
- Alternately, clamping force sensors may be used for clamping monitoring [5].

As a result, ultrasonic time-of-flight-sensing was chosen as the favourite principle (Figure 1.1).

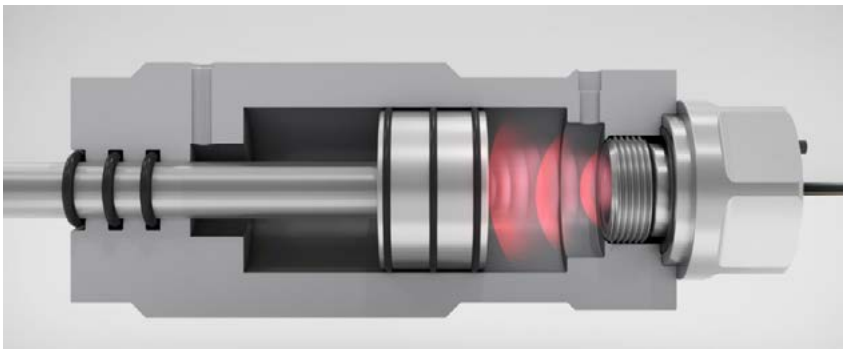


Figure 1.1: Hydraulic cylinder in cross section: rod, cylinder housing with two inlets for hydraulic oil, piston and sensor with cable are shown from left to right. The ultrasonic wave is displayed schematically in red.

2.2. Fundamentals of Ultrasound

Ultrasound is defined as sound in the frequency range of 20 kHz to 1 GHz which is not detectable by human hearing [6]. It can occur in solids, liquids and gases. If particles of a liquid are displaced from their initial position an acoustic wave will propagate through space and time due to the inertia and repulsive forces between the particles. Propagation of sound in matter is one of the fundamental phenomena in physics and may be exploited for technical applications such as sensing. It can be described by the pressure $p(\boldsymbol{x},t)$ and the sound particle velocity $v(\boldsymbol{x},t)$ and leads to a set of coupled second-order differential equations. In acoustics

as in other physical domains, phenomena such as superposition, interference, attenuation, reflection, refraction and diffraction must be considered.

In liquids, there are only longitudinal waves having a speed of propagation c given by (2.1) where E is the elasticity and ρ is the density:

$$c = \sqrt{\frac{E}{\rho}} \qquad (2.1)$$

Another important quantity in acoustic systems is the acoustic impedance Z given by (2.2).

$$Z = \rho \cdot c \qquad (2.2)$$

If an acoustic wave meets a boundary between two different materials having the impedance Z_1 and Z_2, the wave will be partly reflected and partly transmitted. If p_+ and v_+ are the pressure and the velocity of the incident wave and p_- and v_- of the reflected wave, respectively, the reflection factor r is defined by (2.3).

$$r = \frac{p_-}{p_+} = \frac{v_-}{v_+} \qquad (2.3)$$

To fulfil the continuity of pressure and velocity at the boundary, the reflection factor must follow (2.4).

$$r = \frac{Z_1 - Z_2}{Z_1 + Z_2} \qquad (2.4)$$

Table 1 shows the material parameters and impedance of relevant materials. The properties of ceramic are applied for the piezo element.

Property	Hydraulic oil	Steel	Ceramic
Speed of sound c [m/s]	1462	5920	10050
Density ρ [g/cm^3]	0.88	7.90	3
Impedance Z [Ns/m^3]	$1.29 \cdot 10^6$	$46.77 \cdot 10^6$	$30 \cdot 10^6$

Table 1: Material parameters and impedance of relevant materials

2.3. Ultrasonic Position Sensing

For ultrasonic position sensing, an ultrasonic wave is generated by a wave source such as a piezo element [2]. The emitted wave propagates through the hydraulic fluid inside the cylinder in the axial direction, is reflected by the piston, travels back to the sensor and is finally detected by the piezo element. It results from equation (2.4) and table 1 that approx. 95% of the amplitudes of the incident wave are reflected at the boundary between the oil and piston. The two-way time-of-flight τ_p is acquired by a time-to-digital converter (TDC, chapter 2.5). The piston position s_p is determined from τ_p using equation (2.5).

$$s_p = \frac{c(T,p) \cdot \tau_p}{2} \qquad (2.5)$$

The speed of sound $c(T,p)$ of the hydraulic fluid is dependent on the type, temperature T and the pressure p of the hydraulic fluid.

2.4. Temperature and pressure compensation

To compensate for the temperature and pressure, a second target – the reference target – is designed into the ultrasonic position sensor at a fixed position s_r. The reference target is designed in the shape of a three-armed star and has a fixed distance to the piezo element (Figure 2.1 at the left end of the sensor). The metal structure reflects only a fraction of the ultrasonic wave and allows most of the wave to propagate in the direction of the piston. The speed of sound can be calculated by the central processing unit (CPU, chapter 2.5) by using equation (2.6). The time-of-flight τ_r is obtained from the first echo by the TDC.

$$c(T,p) = \frac{2 \cdot s_r}{\tau_r} \quad (2.6)$$

Therefore, the desired quantity s_p can be continuously determined during the sensor operation by inserting *c(T, p)* into equation (2.5) without any a priori knowledge of the type, temperature and pressure of the hydraulic fluid.



Figure 2.1: Picture of the ultrasonic position sensor

2.5. Sensor design

Figure 2.2 shows the block diagram of the electronic circuitry of the ultrasonic position sensor. The position measurement is controlled by the CPU. An ultrasonic wave is generated by the piezo element which is electrically excited by the transmitter-electronics TX. The TDC starts its timer at the point in time of the wave transmission. After reflection of the wave by the reference target and the piston respectively, the wave is inversely transferred into an electrical signal by the piezo element. The received signal is amplified by the low noise amplifier LNA and the variable gain amplifier VGA. Finally, the signal is compared to a level by the comparator COMP. The VGA compensates for the attenuation of the ultrasonic wave which is dependent on the stroke of the cylinder, the hydraulic oil and the temperature. The output signal of COMP is used as a stop signal for the timer of the TDC. The TDC provides the time-of-flight resulting from the difference between the start and the stop time. After temperature compensation the piston position is calculated by the CPU. Finally, the output stage generates a standardized voltage or current signal proportional to the piston position. Alternatively, the piston position can be digitally transmitted to an IO-Link master [7], [8]. The power block converts the standardized supply voltage of 24 V into 3.3 V to feed the electronic circuitry and ensures electromagnetic compatibility (EMC).

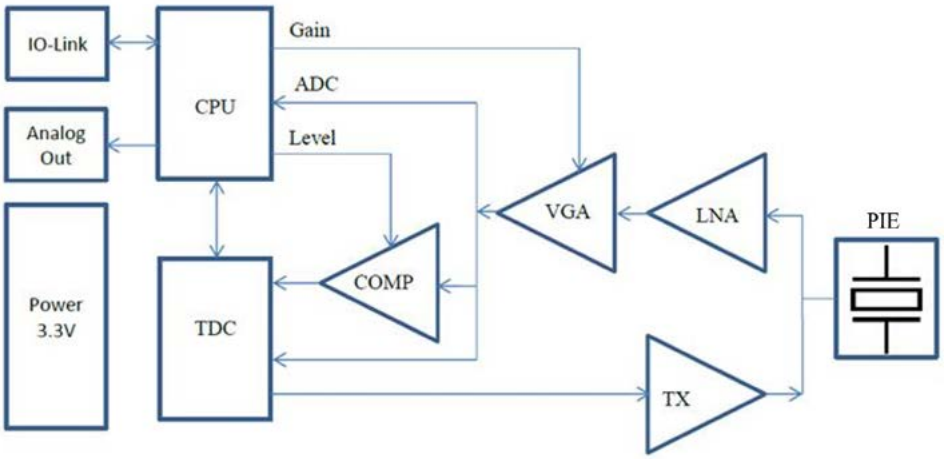


Figure 2.2: Block diagram of the electronic circuitry: CPU central processing unit, TDC Time -to-digital-Converter, COMP Comparator, VGA variable gain amplifier, LNA low noise amplifier, TX transmitter, PIE piezo element

Figure 2.3 illustrates the sensing principle for one measuring cycle. The emitted wave is visible between 0 and 5 µs. A first echo appears at approximately 12 µs resulting from the reference target (Figure 2.1). A second larger echo appears at about 67 µs which corresponds to the piston.

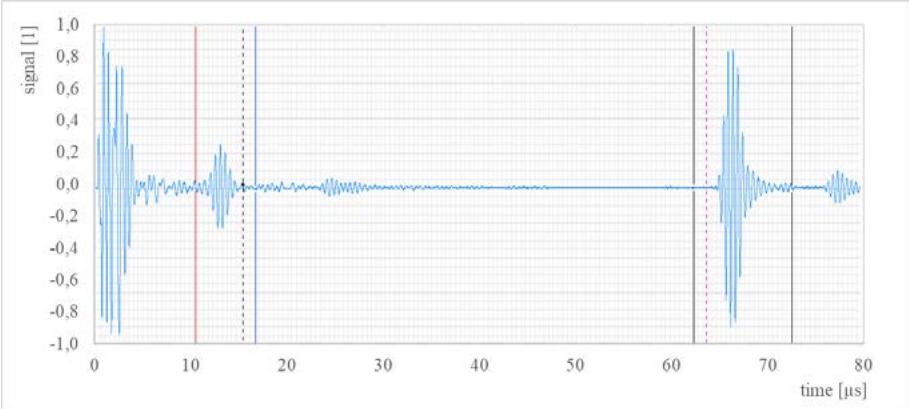


Figure 2.3: Sensor signal of one measuring cycle: x-axis time [s], y-axis normalized signal at the piezo element

Figure 2.4 demonstrates the sensor operation independently of the pressure. The plateaus at 27 mm and 6 mm correspond to the clamped and released state (“open position”) of the clamping device, respectively (blue curve). The transient phases of the piston in between these two states can be reliably monitored. The strong increase of the pressure (green curve) in the clamped position to 60 bars does not reduce the accuracy of the sensor output.

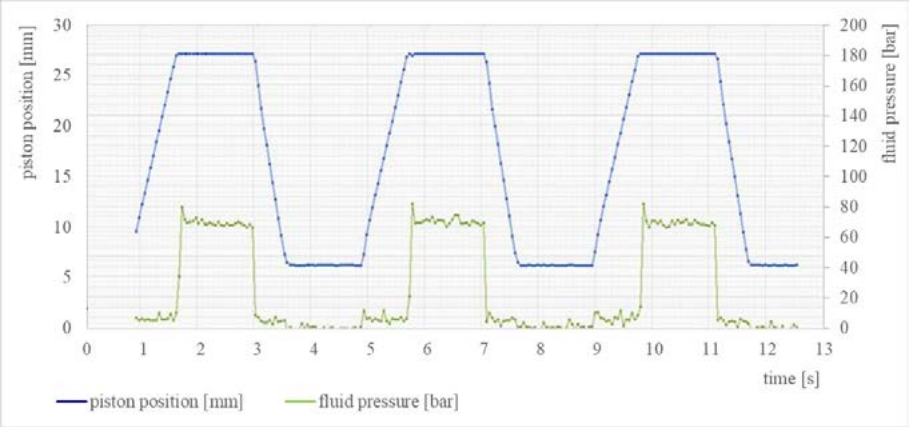


Figure 2.4: Sensor output: Piston position [mm] (blue curve) and fluid pressure [bar] (green curve) plotted during three clamping cycles as a function of time [s]

During the development of the sensor, a non-immersive solution was investigated. The sensor would not be wetted by oil. Instead, the sensor would be assembled behind the cylinder end wall (right side in Figure 1.1) outside the cylinder. This setup has some advantages for the cylinder manufacturer. Unfortunately, in this case about 95% of the wave amplitude is reflected at the boundary between the steel end wall and oil inside the cylinder (chapter 2.2). Consequently, the wave will travel back and forth inside the end wall and will superimpose on the piston echo. This will lead to a degradation in the quality of the measurement. Therefore, an immersive setup was finally chosen (Figure 1.1).

Additional measures have been taken to reduce the reflections at the boundary between the piezo element and the oil. However, a detailed description is beyond the scope of this publication.

2.6. Performance and limitation of the ultrasonic position sensor

The ultrasonic position sensor provides a selectable analogue voltage or current output (0…10V or 4…20 mA) and an IO-Link interface [7]. IO-Link is predestined as a digital, bidirectional, peer-to-peer communication system to support the configuration of the sensor and the transmission of sensor data [8] [9]. Table 2 lists the performance characteristics of the ultrasonic position sensor.

The limitation of the measurement range to currently < 100 mm is caused by attenuation of the ultrasonic wave, which leads to decreasing signals for long strokes. The maximum measuring rate is limited by the time-of-flight inside the cylinder. A new measurement cycle can be started after the arrival of the previous echo only considering the maximal measuring range. Noise will also limit the effect of the temperature and pressure compensation because short pressure peaks which appear in the hydraulic system must be acquired without filtering or averaging.

Performance characteristic	Value
Measuring range	0…80 mm
Pressure rating max.	350 bar
Burst pressure	455 bar
Dimension	∅ 30 x 30.9 mm
Sensor housing material	1.4104
Measuring rate max.	500 Hz (digital) 1000 Hz (analog)
Non-linearity max.	± 250 µm
Repeat accuracy	± 50 µm
Temperature drift max.	5 % full scale

Table 2: Performance characteristics of the ultrasonic position sensor [7]

The typical output characteristic at room temperature is shown in figure 2.5.

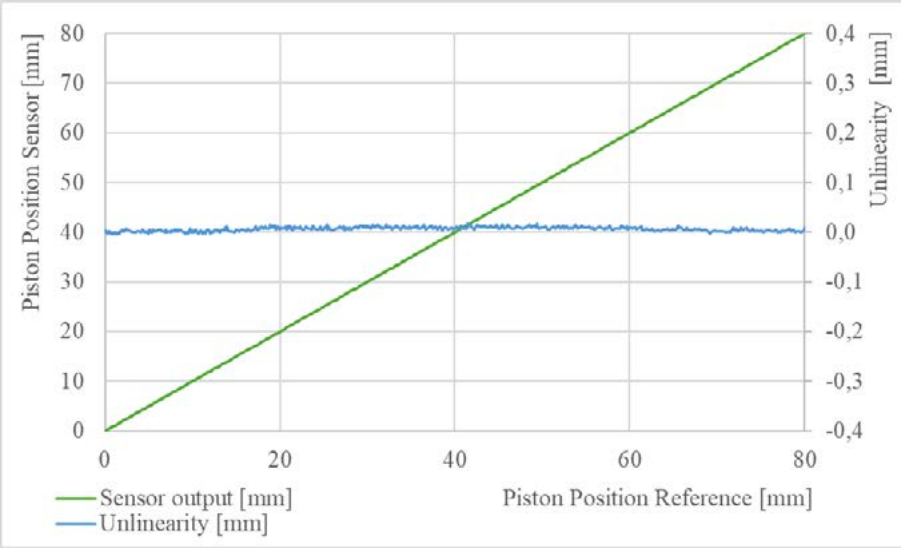


Figure 2.5: Output characteristic during travel over the full piston stroke of 80 mm. Piston position sensor [mm] (green curve) and unlinearity [mm] (blue curve)

3. APPLICATIONS FOR SWING CLAMPING CYLINDERS

Figure 3.1 represents a typical swing clamping cylinder. The hammer at the upper end of the piston rotates by 90° during its vertical movement to allow the workpiece to be placed in the “open position”. The ultrasonic position sensor provides information of the clamping status for high-efficiency PLC-controlled manufacturing in machining centers (figure 3.2).



Figure 3.1: Swing clamping cylinders: The ultrasonic position sensor is attached to the end face. The piston is equipped with a hammer to clamp the workpiece.

The automated manufacturing process consists of the main following steps:

1. Picking of the raw workpiece by the worker, a robot or an automated device from a tray or another container
2. Transport of the workpiece into the workspace and placement on the workbench of the machining center
3. Clamping of the workpiece by the clamping cylinder. The clamping cylinder should travel from its “open position” until “clamped with workpiece” which causes a mechanical stop.
4. Machining process such as milling, grinding or drilling. High forces and torques are applied to the workpiece during this step.
5. Releasing of the workpiece by the movement of the clamping cylinder to the “open position”
6. Picking of the processed workpiece by the worker, a robot or an automated device from the workbench
7. Transport of the workpiece out of the workspace and placement into a container of the processed workpieces

Several possible situations may occur after the clamping in step 3:

a. The workpiece is correctly inserted and clamped. The clamping position of the clamping cylinder (“clamped with workpiece”) is in the predefined range. Step 4 may be executed.
b. The workpiece is not placed at all due to faulty execution of the preceding steps. The clamping cylinder travelled too far to a position “clamped without workpiece”. The preceding steps 1 and 2 must be repeated.
c. The workpiece was placed at the wrong position and is consequently not clamped correctly
d. An incorrect workpiece was placed and is consequently not clamped correctly.
e. The workpiece is placed at the correct position but was not clamped.

Cases c, d and e may cause severe damage of the workpiece and the machining center in step 4 and leads to production downtime and high reject rates. By monitoring the entire clamping process, it is also possible to detect pressure fluctuations in the hydraulic unit, defects in the swivel mechanism and defective seals. Thus, the integration of the ultrasonic position sensor can help to prevent damage and downtime in the daily production process.



Figure 3.2: Machining center: Two swing clamp cylinders integrated in the workbench are visible. The cylinders are in the position “clamped with workpiece”

Polymer injection molding machines are another application for ultrasonic position sensors. Block cylinders are used in these machines to eject the molded part or to move inserts during the molding process. Ultrasonic position sensors can monitor this movement for process quality improvements and condition monitoring.

4. CONCLUSION

Ultrasonic sensing is a suitable principle for continuous, accurate and quick detection of piston position for short stroke cylinders. Modern electronic components, materials and compensation algorithms enable a compact and cost-efficient design for applications in machining centers and other automated machines. IO-Link makes it possible to achieve an excellent user experience during commissioning and configuration of the ultrasonic position sensor. The performance characteristics that are achieved have the potential to make this new type of sensor a game changer in position detection of short-stroke hydraulic cylinders and condition monitoring of the clamping process in factory automation.

5. REFERENCES

[1] H. Möhring, D. Biermann, F. Bleicher, S. Melkote, G. Kappmeyer, 'Fixtures and workpiece clamping systems in machining', CIRP Annals, Vol. 74, No. 2, pages 945-969, 2025.

[2] E. Hering, G. Schönfelder, 'Sensoren in Wissenschaft und Technik', Springer Vieweg, Wiesbaden, 2023.

[3] T. Burckhard, A. Feinäugle, S. Fericean, A. Forkl, 'Lineare Weg- und Abstandssensoren', Verlag Moderne Industrie, München, 2004.

[4] S. Fericean, A. Hiller-Brod, A. Dorneich, M. Fritton, 'Microwave displacement sensors for hydraulic devices', IEEE Sensors Journal, Vol. 13, No. 12, pages 4682-4689, December 2013.

[5] B. Denkena, B. Bergmann, C. Tiege, M. Schnittberger, S. Wörner, 'Zuverlässige Prognosen für den Zerspannprozess', NCFertigung, Vol. 41, No. 9, pages 80-83, 2020.

[6] R. Lerch, G. Sessler, D. Wolf, 'Technische Akustik', Springer Berlin, Heidelberg, 2009.

[7] Balluff, Ultrasonic sensors for short-stroke hydraulic cylinders, https://www.balluff.com/en-gb/products/areas/A0001/groups/G0104/products/F01408 , 2025

[8] What is IO-Link?, https://io-link.com/en/Technology/what_is_IO-Link.php, 2025.

[9] IEC 61131-9:2022, Programmable controllers - Part 9: Single-drop digital communication interface for small sensors and actuators (SDCI), 2022

Biographies



Markus Fritton completed his apprenticeship successfully in communication electronics with Balluff GmbH, Neuhausen, Germany, in 1993. He received the qualification as state certified technician in communication engineering from the DAA-Technikum, Würzburg, Germany, in 1998. He is currently working as a senior engineer at the Innovation Department of Balluff GmbH.



Julia Wick received the bachelor's degree in industrial engineering from DHBW Stuttgart in 2014. She is working as a Product Manager for fluid sensors at the Balluff GmbH in Neuhausen, Germany.



Albert Dorneich received the degree of Dipl.-Phys. in physics from University of Stuttgart in 1998. He is employed as a Technology Strategy Manager at the Innovation Department of Balluff GmbH in Neuhausen, Germany. He has been a member of the IEEE for 18 years and is currently a senior member. He serves as a member of the board of trustees of the Institut für Mikroelektronik Stuttgart (IMS Chips). His R&D activities include new and well-known sensor principles such as inductive, radar and ultrasonic as well as industrial RFID and wireless communication for use in production. He is the author of several reviewed conference and journal papers and holds various patents.

From Pressure Pulses to Noise Sources – NVH-Oriented Design of Hydraulic Systems

Benedikt Müller[1], Heiko Baum[1]

[1] *FLUIDON GmbH, benedikt.mueller@fluidon.com, heiko.baum@fluidon.com*

Abstract.

This paper shows how fluid sound propagation in complex fluidic systems can be described accurately and in a numerically efficient manner using 1D CFD. Key modeling components are presented, and examples illustrate how these can be used to create a complete model for analyzing the vibration and resonance behavior of the fluid.

Based on this 1D CFD, a simplified, one-sided fluid-structure coupling is presented. Mechanical alternating forces on pipes and components are derived from transient pressure and velocity profiles, whereby the influence of the structural response on the fluid is neglected. This assumption significantly reduces the complexity of the model and enables highly efficient, fast-computing simulations.

The approach allows early identification of NVH-relevant loads and specifically supports the design of robust and structure-friendly piping systems.

Keywords: 1D CFD, unilateral fluid-structure interaction, simulation, vibration fingerprint, resonance behavior, NVH-relevant loads, NVH optimization

1. INTRODUCTION

Despite technical advances in automation, the machine operator remains a central figure in agricultural and construction machinery. Complex operating environments, legal frameworks, and the need for situational decisions make complete autonomy unrealistic at this point. In many cases, semi-automated systems take over the execution, while humans retain overall control and responsibility. This is precisely why acoustic comfort in the workplace cabin is becoming increasingly important: noise and vibrations have a direct impact on concentration, fatigue, and long-term health – and are becoming a decisive quality factor in machine development. NVH optimization is thus not only a technical necessity, but also an ergonomic and economic one. At the same time, reducing noise emissions into the environment is becoming increasingly important. This is particularly relevant when construction machinery is used in sensitive areas such as inner-city areas or during night work, where low noise pollution is essential.

Improvements in noise reduction on the drive side – for example, through quieter engines or electrification – means that hydraulic noises, such as those caused by pressure ripple from piston pumps, are no longer masked as they used to be. These noises are now clearly audible. This development is exacerbated by ever-higher power densities and increasing system dynamics, which make resonances in the branched piping system more likely. The result is structural excitation and airborne noise emissions, which significantly influence the overall NVH impression of the vehicles.

2. HYDRAULIC PRESSURE RIPPLE AS A SOURCE OF NVH

Hydraulic pressure pulsations are primarily caused by time-varying volume flows. Typical excitation sources are displacement pumps with kinematically induced flow ripple, dynamic valve operations, and abrupt changes in flow boundary conditions, for example as a result of load changes or emergency shutdowns.

These excitations lead to the formation of pressure waves that propagate at the speed of sound in the fluid along the pipe system and are partially reflected at cross-sectional changes, branches, and open or closed boundary conditions. The physical principles of this wave propagation and the associated acoustic effects are described in detail in *technical acoustics* (see [1]).

Within this context, this article deliberately focuses on fluid acoustics rather than on the classic, component-oriented, concentrated parametric system simulation of fluid power systems. **Figure 2.1** illustrates the propagation and decay characteristics of pressure waves in a fluid power system with sufficient damping.

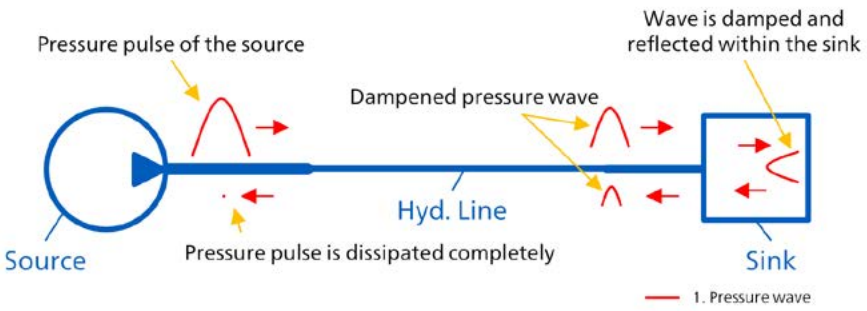


Figure 2.1: Pressure wave propagation in a system with high damping

In practice, the pressure ripple of the source – especially the pump – is often so pronounced that the introduced pressure wave is not completely dissipated. In addition, the effective system damping can be reduced, for example as a result high operating temperatures. Consequently, a relevant portion of the pressure wave remains in the pipe system and, after multiple reflections, reaches the excitation source again.

In the event of a resonance, this backward-travelling pressure wave hits a newly generated wave in-phase, causing the pressure pulsations to superimpose and reinforce each other (**Figure 2.2**).

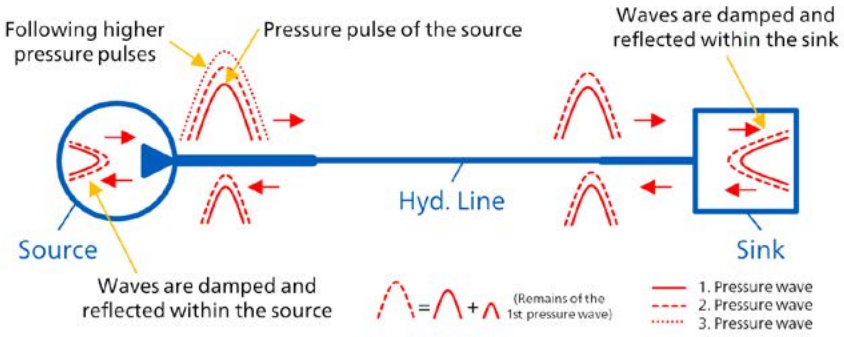


Figure 2.2: Pressure wave propagation in a system with insufficient damping

When excitation frequencies encounter the natural frequencies of the line system, resonant states arise with locally highly increased pressure amplitudes. From an NVH perspective, it is not only individual pressure peaks that are critical, but in particular, permanently acting, pulsation-induced alternating forces.

Two mechanisms of action are particularly important here:

- **Structural excitation** due to time-varying forces on pipe bends, elbows, branches, and fastening points
- **Sound emission**, either directly from vibrating pipes or indirectly via excited machine parts

While short-term pressure surges are often evaluated in terms of their maximum mechanical stress, pulsation-induced alternating forces lead to cyclic loads. These can cause material fatigue, loosening of connectors, or leaks in the long term, and have a significant impact on the overall NVH impression of modern machines. However, these effects are often underestimated in early development phases.

3. WHY 1D CFD?

A transient, one-dimensional simulation approach is used to analyze the fluid-acoustic effects described above. This approach enables a time-resolved simulation of the pressure and volume flow profiles along complex, branched pipe systems. A prerequisite for the validity of the one-dimensional analysis is that the pressure waves propagate as plane waves at the speed of sound in the relevant frequency range – an assumption that is well fulfilled for typical hydraulic pipe systems in mobile machines.

Compared to three-dimensional CFD approaches, 1D CFD offers significant advantages:

- Significantly reduced modeling effort
- Very short computing times
- High numerical stability for highly dynamic processes
- Possibility of extensive parameter and variant studies

Compared to concentrated parametric (0D) methods, 1D CFD enables a more realistic representation of wave propagation and resonance effects, while offering high numerical robustness and often lower computing costs.

Another significant advantage of the time domain approach is that it takes into account nonlinear effects such as those caused by real valve characteristics, pressure-dependent material properties, or transient boundary conditions.

4. COMPONENTS IN 1D CFD

With the *Piping Systems Library* [2], Fluidon has developed a model library that enables consistent mapping of hydraulic systems for fluidoacoustic analysis and NVH-oriented optimization. In addition to detailed pipe models, the library includes components for describing displacement units, valves, cylinders, and accumulators, thus enabling the consistent setup of acoustically relevant system models.

4.1. Pipes and Hoses

A central element in the acoustic simulation of hydraulic piping systems are pipes and hoses, whose behavior is decisive for the propagation of fluid sound in the system. The continuity equation and the Navier-Stokes equation are the basic fluid mechanics equations on which the calculation is based. A proven method for solving this partial differential equation system is the method of characteristics (MoC), which is characterized by high accuracy and low computational effort.

In the simulation of acoustic processes in pipes, it is necessary to take into account not only stationary friction but also dynamic friction, often referred to as frequency-dependent friction. The method developed by *Trikha* [3] implements this efficiently. *Theissen* [4] uses the method of characteristics with frequency-dependent friction to calculate oil hydraulic systems.

Figure 4.1 shows in the upper part that, in the case of slow oscillation, the flow profile is similar to that in the stationary case. However, if the frequency increases (lower part), history plays a major role in the shape of the flow profile, with higher shear velocities and thus increased friction occurring in the boundary layer.

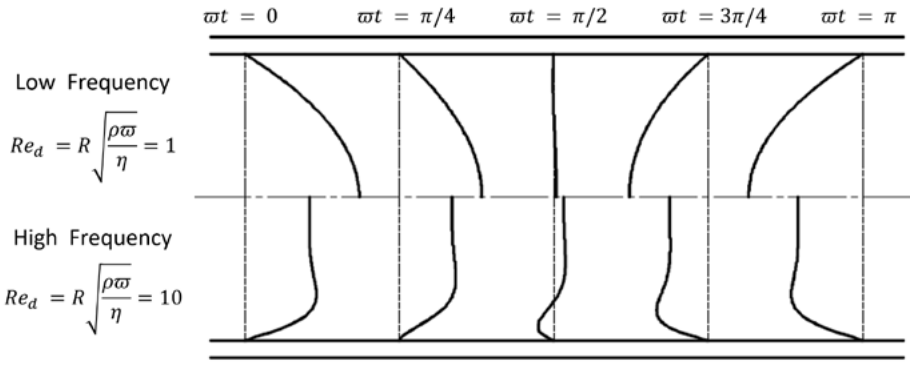


Figure 4.1: Flow profiles of oscillating pipe flows [5]

An important extension is the consideration of the viscoelastic behavior of hoses [6], [5], which is parameterized as a complex modulus of elasticity in the current implementation in DSHplus [7]. The viscoelastic component of the material behavior of hose material is represented by a Kelvin-Voigt model, which, connected in series with the linear elastic component, represents a so-called Zener body (**Figure 4.2**).

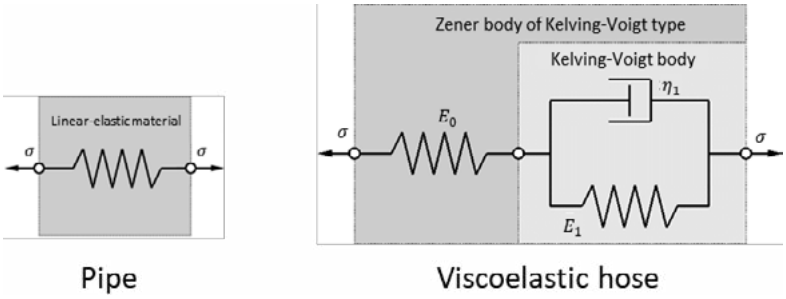


Figure 4.2: The viscoelastic component of the material behavior of hose material is represented by a Kelvin-Voigt model.

While the parameterization of a linear elastic steel pipe can be easily traced back to material and geometry values, measurements are still required for the exact parameterization of viscoelastic hose lines. The method for metrologically determining the four-pole transfer matrices in the frequency domain according to ISO 15086-3, which goes back to [8], has been used and further developed by Fluidon in numerous projects. Automated curve fitting methods then allow the models to be compared with the measurement data as a function of the operating parameters. **Figure 4.3** shows an example test setup for four-pole measurement. A comprehensive description of the measurement method, which originally comes from acoustics of exhaust systems for combustion engines, and the possible measurement setups, including a discussion of their specific advantages and disadvantages, can be found in *Tao et al.* [9].

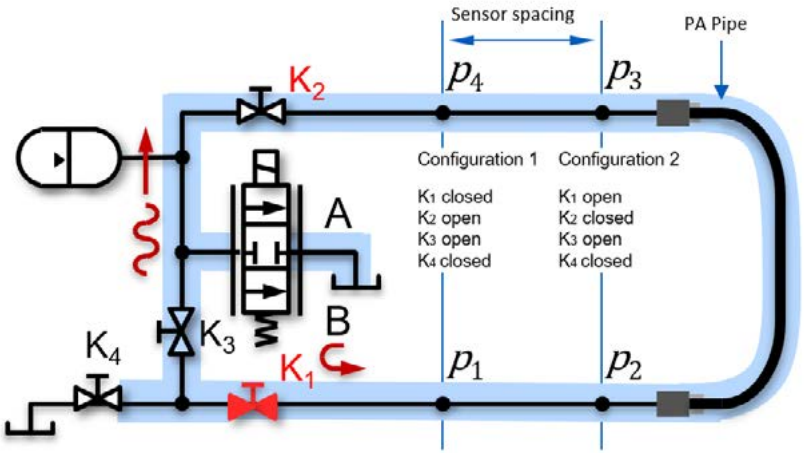


Figure 4.3: Test setup for four-pole measurement according to the secondary source method [10, p. 467]

Variable material data, thermohydraulic effects, cavitation, and gas transport are explicitly taken into account in the 1D CFD calculation and included in the solution of the conservation equations as time-dependent variables.

The time interpolation method proposed by *Savar et al.* [11] achieves greater flexibility in discretization with lower numerical damping, which is particularly advantageous when modeling complex systems. Further efficiency gains in the calculation of large models are achieved through parallel calculation of pipe sections.

4.2. Open / Closed End

In addition to the pipes, other components are required to model complete systems. The simplest boundary conditions are open and closed ends, at which either a pressure or a volume or mass flow is applied.

A pressure boundary condition can be used, for example, to represent a pressure source or a tank ($p \approx 0$). An applied volume flow can already represent a simple pump and forms the basis for more advanced models of displacement units.

If 1D CFD is coupled with concentrated parametric (0D) models, open or closed ends serve as defined interfaces to the 0D components.

4.3. Connectors

The algorithms of the method of characteristic and the resulting model structure do not allow two line elements to be connected directly. Instead, two or more lines are interconnected via connecting elements. With the help of the 2- to n-way connectors shown in **Figure 4.4**, up to 8 lines can be interconnected.

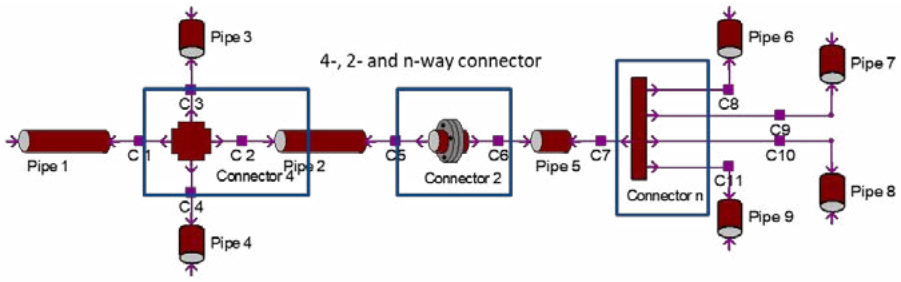


Figure 4.4: Connector types

4.4. Pumps and Motors

With the closed end, a volume flow can already be imposed on a system model, which can also be defined as variable over time, for example, according to a measured flow ripple or a pump start-up. In the case of a straight pipe closed at both ends, the pressure oscillation situation shown in **Figure 4.5** is obtained.

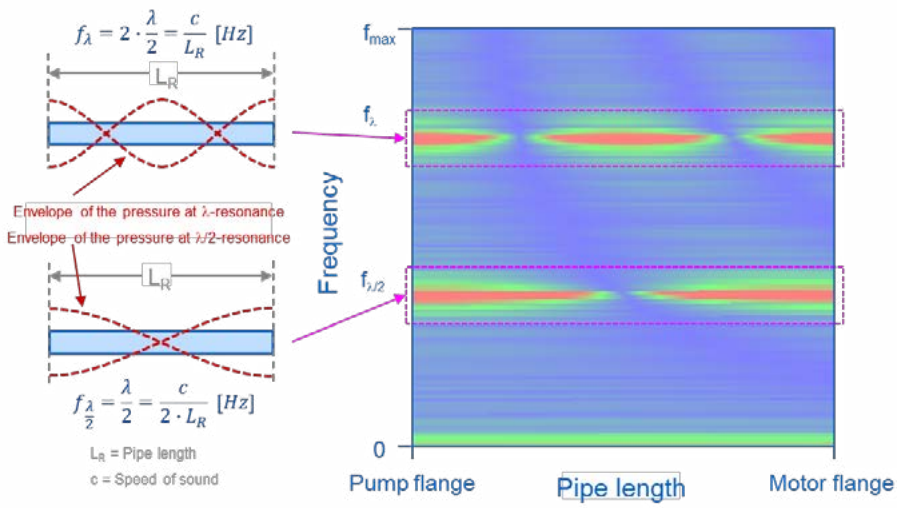


Figure 4.5: Resonances of a pipe with closed ends

A color-coded top view of the stored pressure values of the data matrix is displayed as a *pressure vector plot* [12]. The pipeline (abscissa) starts at the pump and ends at the consumer. The ordinate does not show the simulation time, but rather the excitation frequency. This is possible because the excitation frequency was changed linearly during the simulation, allowing a corresponding frequency to be assigned to each simulation time point. The horizontal lines indicate the first natural frequency ($\lambda/2$ resonance) and the second natural frequency (λ resonance). The pressure vector plot in Figure 4.5 clearly shows when the excitation frequency "meets" the natural frequencies of the pipeline at $f_{\lambda/2}$ and at f_λ.

Since the excitation frequency was continuously increased in the simulation, the resonances initially build up slowly, reach their maxima, and then decay again. In the top view, the shape of the envelope resembles the pressure waveform visualized in the left part of Figure 4.5 (but should not be confused with it), which makes it very easy to identify the resonances in the pressure vector plot.

However, this is often insufficient for the acoustic evaluation of hydraulic systems, as the internal structures of a displacement unit also contribute to the oscillation behavior. Acoustically, they act as a resonator in extension of the pressure line and thus change the resonance behavior of the overall system. **Figure 4.6** shows the pressure vibration situation in the pipeline plus displacement unit on both sides.

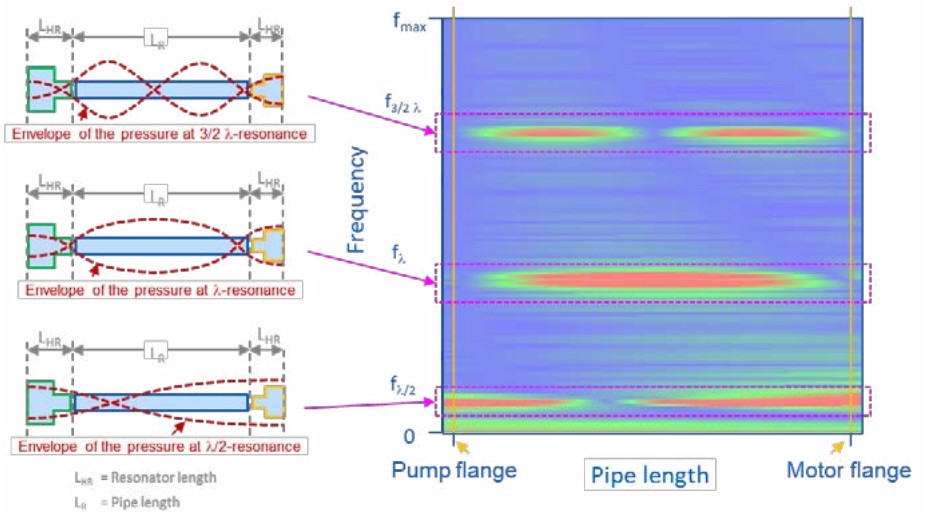


Figure 4.6: Resonances of the pump–pipeline–motor system

The resonator represents an equivalent model for the impedance of the displacement device. Until now, this had to be determined by measurement and is then represented in the system model by an equivalent model that is adapted as closely as possible to the measurement results using an optimization process.

In order to simultaneously impose a measured volume flow pulsation on the system model, Fluidon has developed the *hybrid pump model* [13], **Figure 4.7**. A modified Helmholtz resonator, which contains an additional throttle to increase damping, serves as the impedance replacement model. Experience has shown that it better suits for modeling the impedance of conventional hydraulic displacement units than the frequently used Norton model.

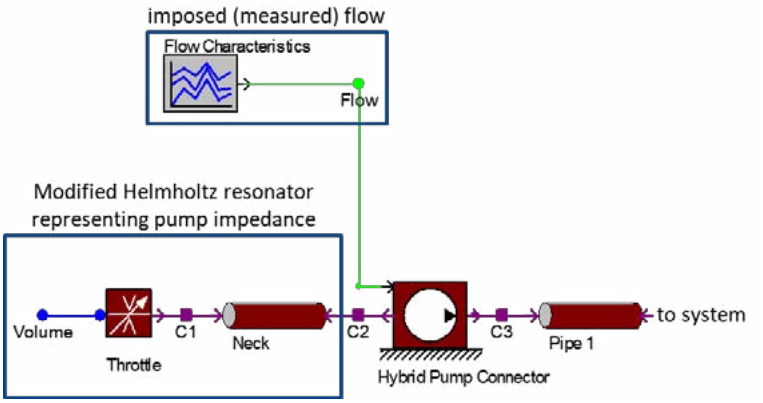


Figure 4.7: Hybrid pump model

Fluidon uses the secondary source method, which goes back to *Edge* and *Johnston* [14], to measure impedance. By using a frequency-variable excitation, the method was further developed in terms of results quality. It can be used on existing functional test benches with

relatively little effort [15]. If the pump impedance is known, the flow ripple can be calculated from a measurement on the same test bench setup with the secondary source switched off [16]. However, this requires good quality impedance measurement or the use of a good quality replacement model. A method that is independent of the pump impedance is measurement using a low-reflection line termination (RaLa) [4], [17].

4.5. Resistors and Valves

The simple resistance component (**Figure 4.8** left) is based on a calculation comparable to that for the double connector, which solves the characteristic equations and the continuity equation. While $p_1 = p_2$ applies to the connector

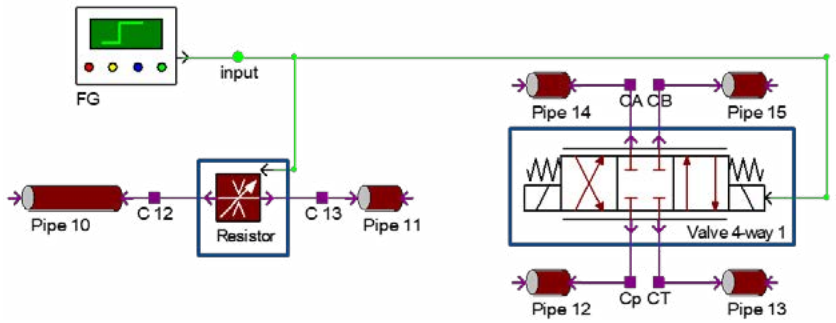


Figure 4.8: Resistance and valve component

In principle, valves can be modeled in 1D CFD as a combination of single hydraulic resistances with intermediate pipes and branches. Since a resistance cannot be directly coupled to a branch, but always requires an intermediate pipe, this approach leads to very extensive and confusing models, especially in the case of complex valve structures.

This is particularly true for valves with multiple effective control edges that are connected to each other via short channels, whose acoustic behavior is often negligible. The "valve" component was therefore introduced to efficiently map such components. It allows its connections to be flexibly linked to each other depending on a control signal. Optional characteristic maps for the effective opening cross-sections allow the mapping of almost any valve characteristics.

5. MODELING AND ANALYSIS OF COMPLEX SYSTEMS

The following section first describes the modeling of complex hydraulic pipe systems before presenting methods for analyzing vibration and resonance behavior based on this.

5.1. Model Creation of Complex Hydraulic Systems

Complex hydraulic systems are usually modeled on the basis of the circuit diagram and the associated CAD data. Steel pipes, valves, and manifold blocks can be mapped with sufficient accuracy in 1D CFD using CAD geometry and the corresponding material data. **Figure 5.1** shows an example of how the internal bores of a manifold block are taken into account in the model.

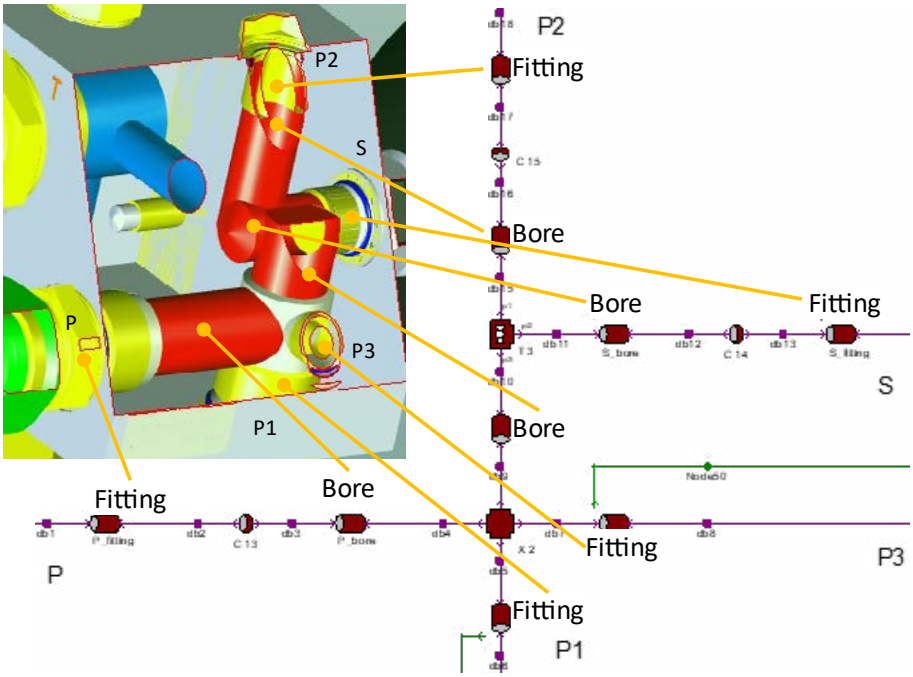


Figure 5.1: Depiction of a manifold block in the model

Hose crimp fittings are modeled as separate pipe sections. In many applications, return lines are not relevant to noise from an NVH perspective and can therefore often be neglected without significantly affecting the validity of the acoustic analysis.

As an example, the pressure-side model of the hydraulic system of a large mobile machine requires around 100 pipe components. The computing time for such a model on the *Fluidon Cube* platform [18] is around 9 hours (without parallelization).

5.2. Analysis of Vibration and Resonance Behavior

In many applications, the actual pulsation of the displacement units as excitation of the system is initially unknown, or it is useful to determine the system response independently of the excitation amplitude in order to obtain an amplitude-independent *vibration fingerprint* of the line system. For this purpose, a sine sweep with constant volume flow amplitude is applied to the system instead of the actual pulsation. The resulting system response is a direct measure of the admittance of the system. Admittance describes the frequency-dependent ratio of the system response to the applied excitation and is therefore a measure of the dynamic compliance of the fluidic system. It allows the identification of resonances and NVH-critical frequency ranges.

However, to analyze the vibration modes of a complex, branched piping system, it is not sufficient to evaluate the pressure curve at individual measuring points in the time or frequency domain. To systematically identify resonances and spatial vibration modes, Fluidon again uses the representation in the form of so-called *pressure vector plots* (**Figure 5.2**).

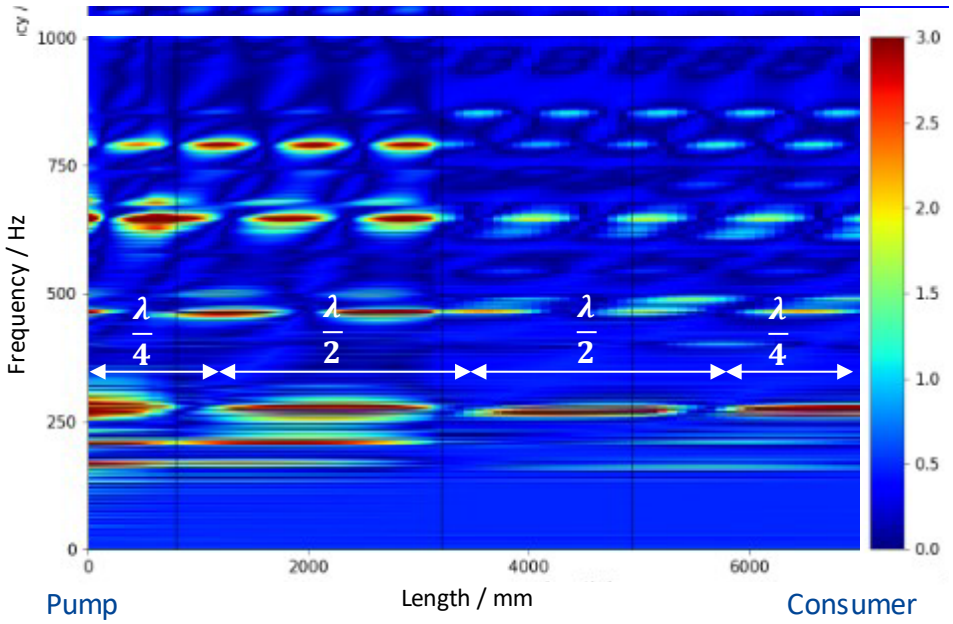


Figure 5.2: Pressure vector plot for analyzing resonances and vibration modes

As already introduced in Figure 4.5 and Figure 4.6, the x-axis describes the pipe length, for example from the pump outlet to a critical consumer. Vertical lines mark the transition between line sections, like a hose-pipe combination or a manifold block. The frequency is plotted on the y-axis, while the pressure admittance is represented as a color gradient. This visualization allows frequencies to be directly assigned to spatial vibration modes in the piping system.

The example shown clearly illustrates that a critical resonance vibration occurs at approximately 260 Hz, which is pronounced in the pipe branch shown over 3/2 wavelengths λ . To evaluate the overall system, such an analysis is performed for all relevant branches. The spatial characteristics of the critical vibration modes can be used to derive specific optimization measures, which are then evaluated using variant simulations.

Knowledge of the spatial vibration modes and dominant resonances in the pipe system forms the basis for the subsequent consideration of fluid-structure interaction. The pressure amplitudes can be used to directly calculate the dynamic forces acting and their transmission to pipes, fastenings, and adjacent structures.

6. FLUID-STRUCTURE INTERACTION

"A true, two-way fluid-structure interaction (FSI) exists when the flow is noticeably influenced by the structural change. However, the effect of the structural change on the flow is often so weak that it can be neglected. In this case, we refer to a one-way fluid-structure interaction." [19]

The article *Pressure Oscillation Analysis in Fuel Supply Systems* [20, pp. 337–358] describes how DSHplus takes bilateral FSI into account. However, the system of equations can be reduced so that only unilateral FSI is calculated. Based on the transient pressure and flow curves determined in the vibration fingerprint, the resulting hydrodynamic forces are calculated at selected points in the hydraulic system. This simplification significantly reduces the modeling effort and enables numerically very efficient simulations without significantly limiting the information content relevant to NVH issues. The aim is not to calculate deformations or stresses in detail, but to determine the time-dependent excitation forces that serve as input data for further structural dynamic or acoustic analyses.

The approach is particularly suitable for:

- Identifying NVH-critical line sections or system components
- Evaluating different plumbing concepts
- Early estimation of pulsation-induced continuous loads

To demonstrate the methodology, a practical pipe geometry is considered below.

6.1. Alternating Forces on a Double Pipe Bend

In addition to the acoustic pipe model, the spatial geometry of the piping system, as typically available in CAD planning, is required to calculate the fluid-dynamically induced forces. Based on the three-dimensional geometry, the piping system is divided into individual segments whose geometry is defined in the global coordinate system. **Figure 6.1** illustrates this transformation step from the CAD representation to the simulation model structure and shows an example of an assembly with several pipe bends and a connecting pipe section. Defining a boundary around this assembly makes it possible to specifically calculate the forces acting on the assembly.

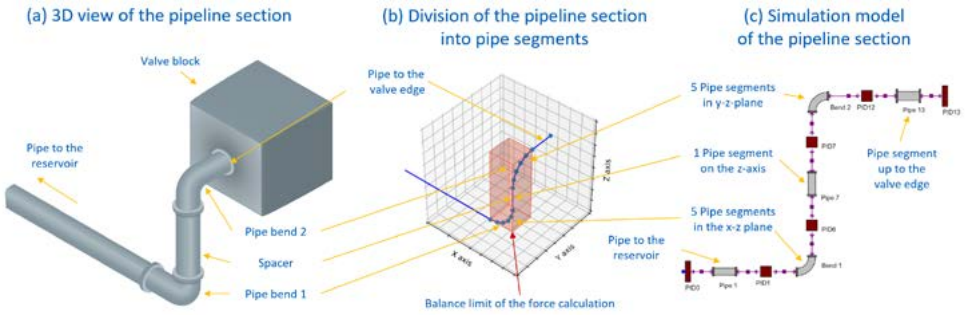


Figure 6.1: Preparation of the 3D representation for simulation

The forces are calculated by summing the time-dependent pressure and momentum components of all pipe segments within the balance boundary. Depending on the pipe routing and orientation in space, this results in force components in different spatial directions. The geometric arrangement of the example pipe already makes it clear in which directions forces can act preferentially, while other directions are in equilibrium in the static state. Changes in the operating state, for example due to pressure build-up or the start of flow, lead to corresponding changes in the distribution of forces on the assembly under consideration.

Figure 6.2 shows an example of such a simulation model and illustrates how the forces acting on a subassembly change over time. While pressure-related forces predominate in the static state, additional impulse forces occur when fluid flows through the pipe. Transient processes, such as the opening or closing of valves, lead to short-term changes in force, which are particularly noticeable in line bends and are often accompanied by visible or perceptible relative movements in real systems.

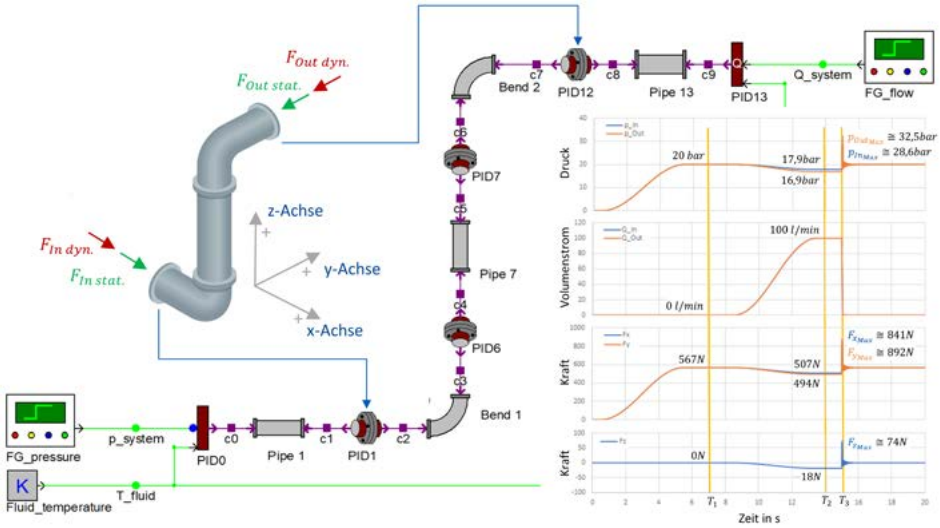


Figure 6.2: Simulation model of the pipe system

Similar to these temporary effects, permanent pressure and flow ripples induce dynamic forces, where resonance situations as described above lead to increased amplitudes.

Figure 6.3 shows the evaluation of a frequency-dependent excitation of the pipe system. With appropriate preparation of the simulation model, the force components shown can be calculated directly from the results of the vibration fingerprint in post-processing.

The illustration clearly shows that when certain excitation frequencies are reached, pronounced vibration modes of the fluid arise, which are directly reflected in increased dynamic forces on the assembly under consideration. The time- and frequency-dependent evaluation of the force components allows such resonance states to be clearly identified and spatially assigned.

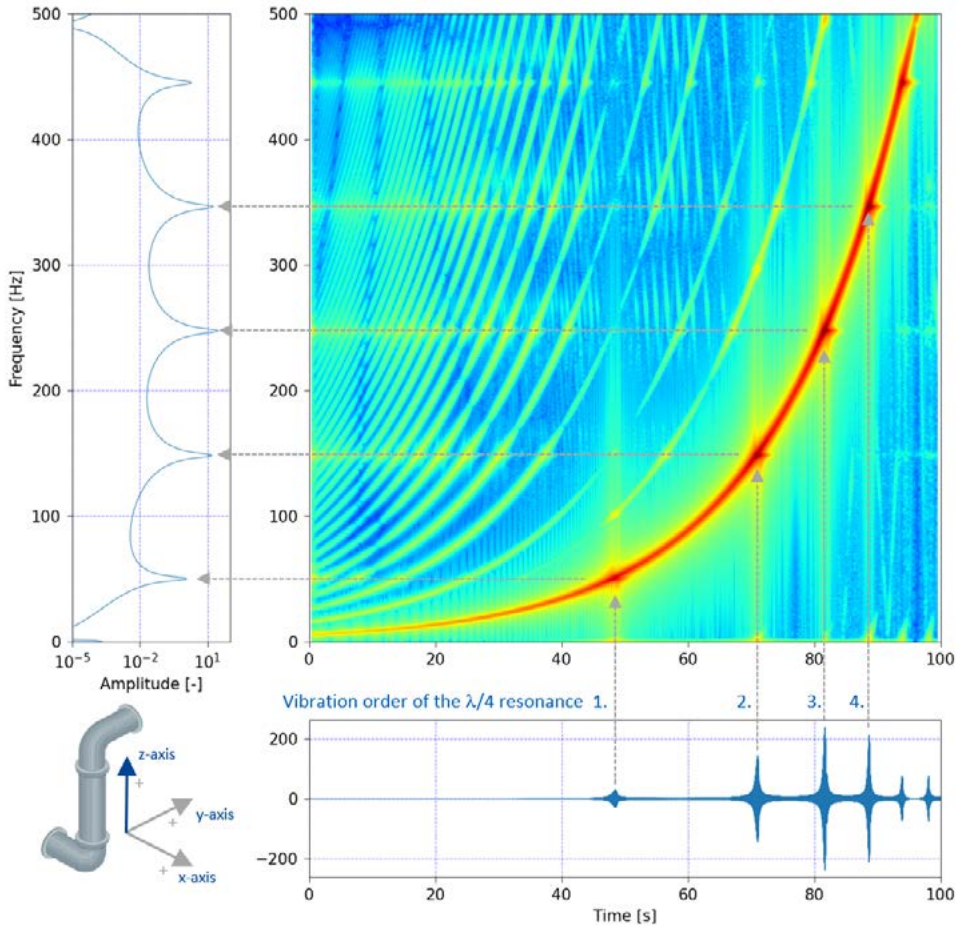


Figure 6.3: Analysis of the force component in the z-axis direction

The combination of pressure oscillation analysis and force calculation thus provides direct access to the structural mechanical evaluation of hydraulic systems. It enables dynamic alternating forces in relevant assemblies to be identified at an early stage and specifically incorporated into the design of line routing, mounting points, and structural measures. This approach is therefore a key tool for NVH-oriented system design.

7. SUMMARY AND NVH IMPLICATIONS

The results show that NVH problems in hydraulic systems are not exclusively attributable to rare extreme events such as pressure surges. Rather, flow ripple-induced continuous loads are a key cause of structural excitation and acoustic emissions and have a significant impact on the overall NVH impression of modern machines.

The numerically efficient simulation approach presented in this paper, based on transient 1D CFD in combination with simplified, unilateral fluid-structure coupling, makes it possible to identify and evaluate these effects early in the development process. NVH-critical excitation forces can be specifically localized and assigned to their cause in the hydraulic system.

Critical conditions can be avoided as early as the concept phase by making targeted adjustments to the line geometry, mounting concepts, or hydraulic boundary conditions. The approach thus represents a practical alternative to complex fully coupled FSI simulations and is particularly suitable for variant studies and the early design of structure-friendly and low-vibration line systems.

Overall, the methodology presented makes a significant contribution to the development of quiet, durable, and NVH-optimized hydraulic systems.

8. REFERENCES

[1] I. Veit, *Technische Akustik: Grundlagen der physikalischen, physiologischen und Elektroakustik*, 5., Durchges. Aufl. in Kamprath-Reihe. Würzburg: Vogel, 1996.

[2] "DSHplus - Piping system library 1D-CFD | FLUIDON." Accessed: Jan. 30, 2026. [Online]. Available: https://fluidon.com/tools/dshplus/1d-piping-simulation

[3] A. K. Trikha, "An Efficient Method for Simulating Frequency-Dependent Friction in Transient Liquid Flow," *Journal of Fluids Engineering*, vol. 97, no. 1, pp. 97–105, Mar. 1975, doi: 10.1115/1.3447224.

[4] H. Theissen, "Die Berücksichtigung instationärer Rohrströmung bei der Simulation hydraulischer Anlagen," Dissertation, RWTH Aachen, 1983.

[5] B. Müller, "Einsatz der Simulation zur Pulsations- und Geräuschminderung hydraulischer Anlagen," Dissertation, RWTH Aachen, 2002.

[6] K. Nagata and K. Takahashi, "A Simulation of the Resonator Hose Using Method of Characteristics," *Proceedings of the JFPS International Symposium on Fluid Power*, vol. 1996, no. 3, pp. 491–496, 1996, doi: 10.5739/isfp.1996.491.

[7] E. Pasquini, H. Baum, H. Murrenhoff, and K. Schmitz, "Die Schallgeschwindigkeit von Fluiden in nachgiebigen Leitungen – Teil II: Leitungen mit viskoelastischem Materialverhalten", [Online]. Available: https://www.yumpu.com/de/document/read/62639146/o-p-fluidtechnik-5-2019/36

[8] E. Kojima and K. A. Edge, "Experimental determination of hydraulic silencer transfer matrices and assessment of the method for use as a standard test procedure," presented at the 7th Bath International Fluid Power Workshop, 1994.

[9] Z. Tao and A. F. Seybert, "A Review of Current Techniques for Measuring Muffler Transmission Loss," SAE Technical Paper, 2003-01–1653, May 2003. doi: 10.4271/2003-01-1653.

[10] VDI Wissensforum GmbH, Ed., *SIMVEC Simulation und Erprobung in der Fahrzeugentwicklung 2016: Berechnung, Prüfstands- und Straßenversuch*, 1. Auflage. in VDI-Berichte, no. 2279. Düsseldorf: VDI Verlag, 2016. doi: 10.51202/9783181022795.

[11] M. Šavar, Z. Virag, and R. Korbar, "A variant of the method of characteristics," *Computer Assisted Mechanics and Engineering Sciences*, no. 9, Jan. 2002.

[12] H. Baum and B. Erzberger, “Berücksichtigung des akustischen Verhaltens komplexer Bauteile bei der simulativen Druckschwingungsanalyse,” in *SIMVEC – Simulation und Erprobung in der Fahrzeugentwicklung*, 1. Auflage., VDI Wissensforum GmbH, Ed., in VDI-Berichte, no. 2333. , Düsseldorf: VDI Verlag, 2018. doi: 10.51202/9783181023334.

[13] H. Baum, K. Becker, and A. Faßbender, “Hybrid Pump Model for 1D Hydraulic System Simulation,” presented at the 9th International Fluid Power Conference, Aachen, 2014.

[14] K. A. Edge and D. N. Johnston, “A new method for evaluating the fluid borne noise characteristics of positive displacement pumps,” presented at the Seventh International Fluid Power Symposium, Bath, 1986, pp. 253–260.

[15] B. Müller and H. Baum, “Impedance measurement in a hydrostatic drive,” presented at the 12th International Fluid Power Conference, Technische Universität Dresden, Jun. 2020, pp. 513–522. doi: 10.25368/2020.58.

[16] *B.S. ISO 10767-1:1996, Hydraulic fluid power - Determination of pressure ripple levels generated in systems and components, Part 1: Precision method for pumps*.

[17] H. Baum, “Wirkzusammenhänge begreifen: Teil IV – Die Förderstrompulsation von Verdrängerpumpen,” *CITplus*, vol. 23, no. 3, pp. 30–32, Mar. 2020, doi: 10.1002/citp.202000319.

[18] “FLUIDON | Cube.” Accessed: Jan. 03, 2026. [Online]. Available: https://cube-lab.app/

[19] “Fluid-Struktur-Kopplung,” *Wikipedia*. Mar. 13, 2023. Accessed: Feb. 04, 2026. [Online]. Available: https://de.wikipedia.org/w/index.php?title=Fluid-Struktur-Kopplung&oldid=231765849

[20] H. Baum and E. Pasquini, “Druckschwingungsanalyse in Kraftstoffversorgungsanlagen,” in *11. Tagung Einspritzung und Kraftstoffe 2018*, H. Tschöke and R. Marohn, Eds., Wiesbaden: Springer Fachmedien, 2019, pp. 337–358. doi: 10.1007/978-3-658-23181-1_17.

Biographies



Benedikt Müller studied mechanical engineering at RWTH Aachen University and earned his doctorate in 2002 with a thesis on NVH topics in fluid technology. He has many years of experience in mobile hydraulics, fluid acoustics, and the development and simulation of hydraulic systems.



Heiko Baum studied mechanical engineering at RWTH Aachen University and completed his doctorate in 2001. Since 2001, he has been the managing director of FLUIDON GmbH, where he focuses on model-based engineering, digital twins, and automated simulation workflows for hydraulic and pneumatic systems. His long-standing experience in the simulation and development of fluid power systems bridges industrial applications and research projects in the context of Industry 4.0 and the Asset Administration Shell.

Current developments in magnetostrictive position sensors

Olaf Kissing

Temposonics GmbH & Co. KG, olaf.kissing@temposonics.com

Abstract.

A magnetostrictive position sensor with a flexible sensing element simplifies the installation in a hydraulic cylinder. Such a sensor should achieve the same performance as a sensor with a rigid sensing element. In addition, a magnetostrictive position sensor should provide further information in addition to the position, as data can be recorded directly on site in the hydraulic cylinder.

Keywords. Magnetostriction, position sensor, flexible sensing element, hydraulic cylinder.

1. INTRODUCTION

Hydraulic cylinders are used for many purposed, such as to open and close lock gates (Fig. 1), feed rolls in a rolling mill, move the arm of an excavator, or orient the wheels of a vehicle according to the steering angle. All these movements require the position of the piston rod to be detected and transmitted to the control system. Magnetostrictive position sensors have proven themselves in applications such as these.



Figure 1.1: Hydraulic cylinder for moving a lock gate

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. WHAT IS MAGNETOSTRICTION?

A magnetostrictive position sensor uses the principle of magnetostriction to determine a position. Magnetostriction is a magnetic effect in which the mechanical dimensions of a ferromagnetic material such as nickel, iron, cobalt, or their alloys change under the influence of a magnetic field. This is known as the Joule effect. The reverse of this effect is also possible: The magnetic properties of a material, such as permeability, change when its mechanical dimensions are altered. This behavior is known as the Villari effect.

When a current flows through a wire made of magnetic material, this current generates a magnetic field that surrounds the wire. If this magnetic field encounters an external axial magnetic field, a torsional force is generated at this location, which causes a slight mechanical twist. The external axial magnetic field is generated by a permanent magnet that surrounds the wire or is located very close to it. This behavior of a wire made of magnetostrictive material is known as the Wiedemann effect.

These properties are used to determine a position with a magnetostrictive sensor.

3. WHAT IS A MAGNETOSTRICTIVE POSITION SENSOR?

Fig. 2 shows a magnetostrictive position sensor. This type of sensor measures the distance between the pickup system and the position magnet. To do this, a short current pulse is generated in the sensor electronics and applied to the waveguide. This current pulse travels through the waveguide and creates a radial magnetic field. The position magnet, which marks the position to be measured, generates an axial magnetic field. The momentary interaction of these two magnetic fields generates a torsional wave in the waveguide, which propagates as an ultrasonic wave at a speed of approximately 2800 m/s. This wave travels in both directions, toward the end of the waveguide and toward the pickup system. The wave traveling to the end of the waveguide is extinguished by damping elements. This eliminates disruptive reflections of this wave.

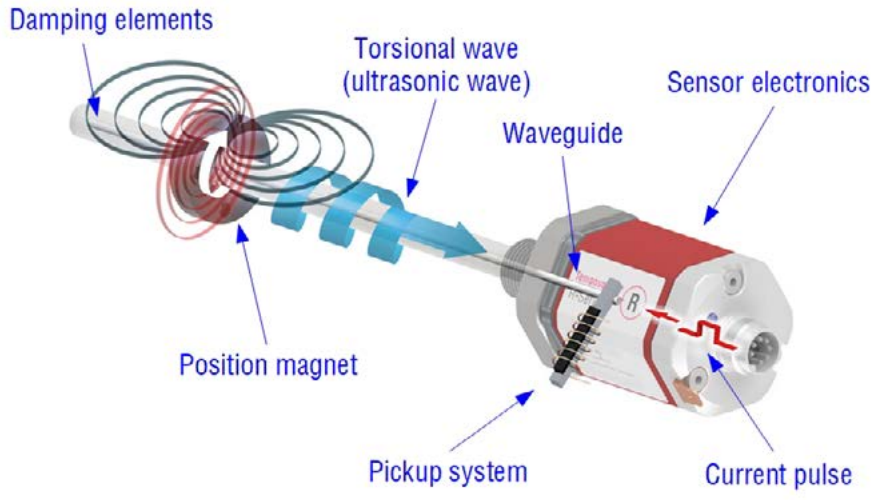


Figure 3.1: Design of a magnetostrictive position sensor

The wave traveling to the pickup system is converted into a voltage signal. When the current pulse is sent, a time measurement is started, which is stopped when the wave arrives at the pickup system. The speed at which the wave moves along the waveguide is called the gradient. The gradient of each position sensor is known, so that the position of the magnet on the waveguide can be determined based on the time measurement. The following applies:

Position = gradient × travel time

The main components of a magnetostrictive sensor are:

- Waveguide
- Position magnet
- Pickup system
- Sensor electronics

The waveguide can be described as the heart of a magnetostrictive position sensor. It is a thin wire with a diameter of 0.25…1 mm. The properties of the waveguide have a significant influence on the performance of the sensor. Important characteristics of a waveguide are:

- Magnetostriction coefficient: The coefficient indicates how far the magnetostrictive material expands when a certain field strength is applied.
- Wave attenuation: The amplitude of the wave is attenuated as the distance along the waveguide increases.
- Variation in propagation speed along the length: For accurate position measurement, it is necessary that the speed at which the wave travels along the waveguide is identical at every point along the waveguide.

These characteristics depend on the intrinsic properties of the material and the microstructure. They can be influenced by the material composition and manufacture of the waveguide. A nickel-iron alloy or similar is often used as the material for the waveguide. However, the exact composition of the material and the manufacturing process are a trade secret of the manufacturers, just like the recipe for Coca-Cola.

A permanent magnet is used as the position magnet. When using a magnetostrictive position sensor in a hydraulic cylinder, a rod-type sensor with a ring magnet is used in most cases, as shown in Fig. 2. The ring magnet surrounds the waveguide on all sides, ensuring a uniform magnetic field.

The pickup system detects the torsional wave and converts it into a voltage signal. The design and functionality of a pickup system are not relevant to this work. Interested readers can find more information in [Nyce, 2004].

The sensor electronics generate the current pulse that is applied to the waveguide and perform the time measurement. The time measurement signal is processed and made available at the output. The user can choose between different configurations such as analog, SSI, IO-Link, EtherCAT®, or PROFINET.

A magnetostrictive position sensor is an absolute measuring sensor that is characterized by high precision and repeatability. The measurement is contactless because the position magnet hovers above the rod. It is not necessary for the magnet to touch the rod. Such a sensor meets the requirements of both industrial and mobile hydraulic environments in terms of reliability and robustness in harsh environments.

4. INTEGRATION OF A MAGNETOSTRICTIVE POSITION SENSOR INTO A HYDRAULIC CYLINDER

A magnetostrictive position sensor is integrated into a hydraulic cylinder by inserting the sensor rod into the drilled piston rod (Fig. 3). The cylinder is sealed via the flange so that no hydraulic fluid can escape. The position magnet is mounted at the bottom of the piston rod so that the position of the piston can be detected at any time. The pressure-resistant sensor rod contains the sensing element with the waveguide. Only the flange with the sensor electronics housing is located outside the cylinder. This offers the advantage that no large attachments are required on the outside of the cylinder.

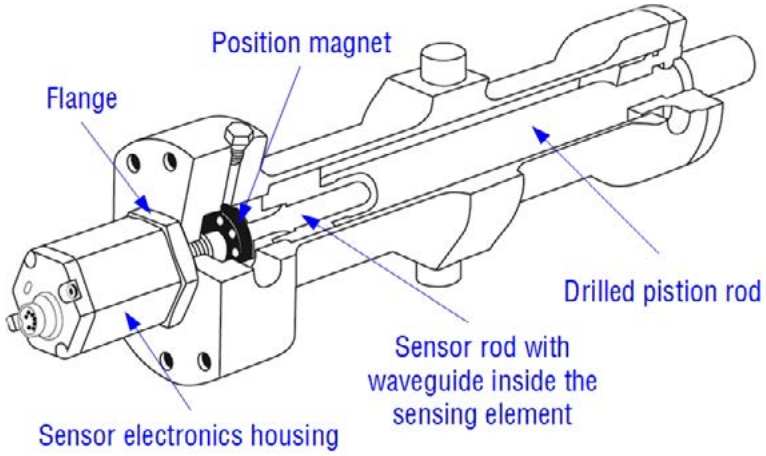


Figure 4.1: Integration of a magnetostrictive position sensor into a hydraulic cylinder

As described, a magnetostrictive position sensor measures the distance between the pickup system and the position magnet. The maximum distance that a sensor can measure is referred to as stroke length and can be selected in the range of 25…7620 mm for a rod sensor. When installing a magnetostrictive sensor in a hydraulic cylinder, the stroke length is usually slightly greater than the piston stroke of the cylinder. This ensures that the sensor can detect the entire travel path of the piston. The sensor rod usually has an outer diameter of 10 mm and a wall thickness of 1.5 mm. It protects the internal waveguide and withstands the hydraulic pressure inside the cylinder.

5. SENSOR WITH FLEXIBLE SENSING ELEMENT

As explained the stroke length of a magnetostrictive position sensor is slightly greater than the piston stroke of the hydraulic cylinder, and the sensor rod with the internal waveguide is rigid due to its compressive strength. Therefore, when installing the sensor, one piston stroke of space is required behind the cylinder. While this is usually possible when constructing the cylinder, there is often not enough space available after the cylinder has been integrated into the final application. If the sensor needs to be replaced, the lack of space behind the cylinder poses a challenge.

To reduce the space required for removal and reinstallation, the sensing element – essentially consisting of a waveguide and an enveloping protective tube – would have to be bent.

However, bending the waveguide impairs the propagation of the torsional wave. The more the waveguide is bent, the more the amplitude of the torsional wave is weakened. A weakening of the torsional wave makes detection at the pickup system more difficult and thus a reliable and precise position determination. If the waveguide is bent so much that it is plastically deformed, the microstructure of the waveguide is damaged. In this case, the propagation of the torsional wave is stopped at the damaged location. A position determination is no longer possible.

The following conditions therefore apply to a flexibly designed sensing element:

- Bending of the sensing element must not impair the function of the sensor: As described, bending of the waveguide weakens the signal of the torsional wave. The sensing element with waveguide must therefore be designed in such a way that this weakening does not affect reliable position determination.
- It must be possible to insert the flexible sensing element into the pressure-resistant sensor rod with an outer diameter of 10 mm and a wall thickness of 1.5 mm.
- The performance of the sensor with a flexible sensing element must be the same as that of a sensor with a rigid sensing element: In addition to reliable position detection, this applies in particular to linearity as well as shock and vibration resistance.

To meet these conditions, the interaction between the sensing element and sensor electronics must be adjusted. The choice of protective tube is also crucial: The protective tube must be stable enough to protect the internal waveguide, yet flexible enough to allow bending. After testing various tubes, a stainless-steel protective tube covered with a PU sheath was selected. The structure of this tube is shown in Fig. 4. It protects the internal waveguide from mechanical stress and, with the PU sheath, provides the necessary flexibility so that the waveguide can only be elastically deformed.



Figure 5.1: Schematic presentation of the structure of the protective tube in a bent state

As described, the amplitude of the torsional wave is attenuated when the waveguide is bent. With a suitable combination of sensing element and sensor electronics, reliable position measurement can be achieved even with a weaker torsional wave signal. Tests were conducted to determine the degree of bending of the flexible sensing element at which the position of the magnet on the measuring rod can still be reliably determined by adjusting the sensor electronics.

For example, a sensor with a flexible sensing element was wound in a curve of 100 mm and 165 mm and the linearity was measured (Fig. 5.2). Fig. 5.3 shows the linearity diagram of the sensor with flexible sensing element in a straight state after unwinding from a diameter of 165 mm. A laser interferometer is used as a reference to determine linearity. Despite the bending of the flexible sensing element, the internally set linearity limits of ± 0.1 mm are only minimally exceeded.



Figure 5.2: Sensor with flexible sensing element wound up in a diameter of 165 mm

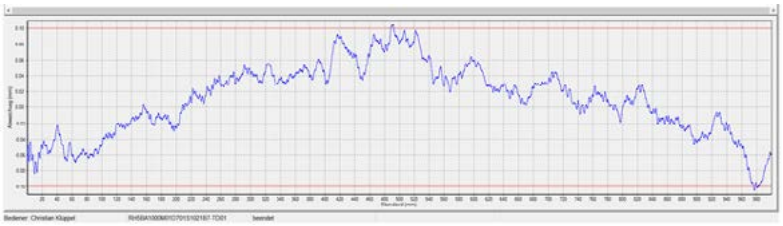


Figure 5.3: Linearity diagram of a sensor with flexible sensing element after unwinding from a diameter of 165 mm

Another test examined how different bending radii affect the amplitude of the torsional wave. To do this, the flexible part of such a sensor was pushed through devices with defined bending radii. The sensing element was pushed through the device up to the sensor electronics in order to assess how different bending radii dampen the torsional wave when it reaches the pickup system (Fig. 5.4). Table 5.1 shows how the signal strength changes as a percentage compared to the unbent state as the bending radius increases. By adjusting the sensor electronics, the signal strength decreases by only 3% even with an increasing bending radius. It should be noted that these tests were carried out under laboratory conditions. In harsh industrial environments, the permissible minimum bending radius must therefore be provided with a safety factor.

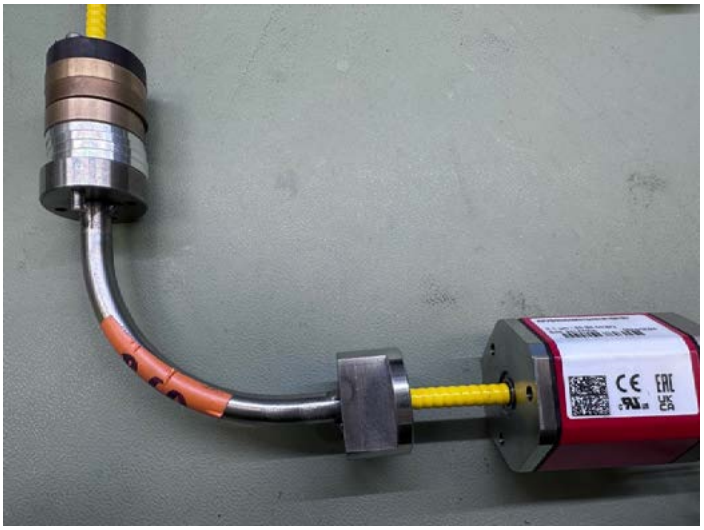


Figure 5.4: Sensor with flexible sensing element inserted into a device with a defined bending radius

Bending radius R	Amplitude of torsional wave
R = 0 mm (reference)	100 %
R = 80 mm	100 %
R = 75 mm	100 %
R = 70 mm	100 %
R = 65 mm	97 %
R = 60 mm	97 %
R = 55 mm	97 %
R = 50 mm	97 %

Table 5.1: Damping of the amplitude of the torsional wave depending on the bending radius

The tests show that a reliable magnetostrictive position sensor with a flexible sensing element can be constructed using a suitable combination of metal protective tube, waveguide, and sensor electronics. As shown in Fig. 5.5, the base unit of a magnetostrictive sensor, consisting of sensor electronics and sensing element, can be separated from the pressure-resistant sensor rod. While the pressure-resistant sensor rod remains in the cylinder, the base unit can be replaced. With a sensor with a flexible sensing element, only a space of approximately 150…200 mm behind the cylinder is required when installing and removing the base unit in the pressure-resistant sensor rod, regardless of the stroke length of the sensor. This means that a sensor with a flexible sensing element is a great simplification in

comparison to a sensor with a rigid sensing element regarding removal and reinstallation on site.

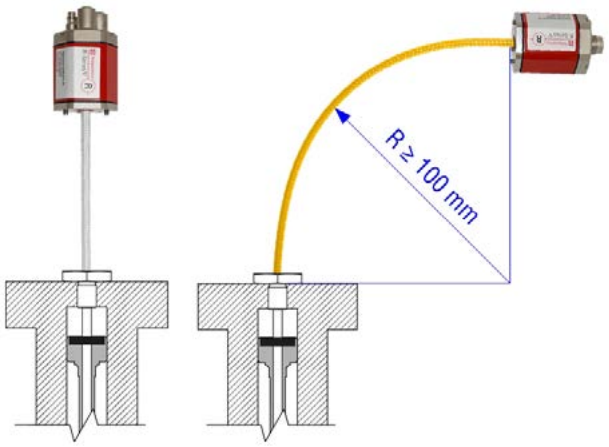


Figure 5.5: Installation of a base unit with rigid sensing element (left) and with flexible sensing element (right)

6. SENSOR WITH ADDITIONAL INFORMATION

In the context of Industry 4.0, a sensor should not only provide its actual data, but also transmit additional information [Anderl et al. 2015]. The actual data of a magnetostrictive sensor is the position of the piston in the cylinder and, if applicable, the velocity of the piston movement. In addition, such a sensor can transmit further information to the control system, such as:

- the total distance traveled by the piston,
- the number of cycles performed by the cylinder,
- the current, minimum, and maximum power supply to the sensor, or
- the quality of the measured sensor signal.

The total distance traveled by the piston and the number of cycles provide information about the operating life of the hydraulic cylinder and can be used to analyze maintenance intervals. The information on the sensor's power supply and the quality of the sensor signal describes the current status of the sensor. Users can use all this information to check both their system and the sensor during operation, thereby enabling predictive maintenance.

One aspect of development trends in sensor technology is the functional integration of different sensor components [Wertschützky 2018]. When integrated into a hydraulic cylinder, a magnetostrictive position sensor offers the advantage of being able to collect data “directly on site,” i.e., inside the cylinder. For example, the temperature of the hydraulic oil or the pressure inside the cylinder can be recorded by extending the magnetostrictive sensor. Currently, in order to record additional data, holes must be drilled and sealed in the cylinder, further sensors must be installed, and cables must be laid. Combining data recording in a single sensor reduces the cost of mechanical processing of the cylinder. In addition, by linking the data, conclusions can be drawn about the condition of the cylinder and thus the application. On the other hand, there are the costs of such an adaptation. In order for the

expansion of a magnetostrictive position sensor to be of interest to cylinder manufacturers and users, the costs must not exceed those of a separate sensor.

According to [Drese et al., 2025], sensors will become smarter in the future. Such smart sensors not only collect data, but also process it. In this context, a magnetostrictive sensor can be enabled to prepare the data currently collected for analysis and decision-making:

- By recording the piston travel, it is possible to process how often the piston moves in which area of the cylinder stroke.
- The number of cycles driven can be used to determine how smoothly or gently the piston moves through the cylinder (detection of stick-slip effects).

In this respect, this article is also a call to users, manufacturers, and scientists to work together to clarify what can and should be implemented and how.

7. REFERENCES

[Anderl et al. 2015] R. Anderl, A. Picard, Y. Wang, J. Fleischer, S. Dorsch, B. Klee, J. Bau-er: Leitfaden Industrie 4.0 - Orientierungshilfe zur Einführung in den Mittelstand, VDMA Verlag, Frankfurt 2015.

[Drese et al. 2025] K. Drese, J. Osten, T. Sauerwald, A. Schütze, M. Wurz, S. Zimmermann: Sensor Trends 2030, AMA Association for Sensors and Measurement and VDI/VDE-Gesellschaft Mess- und Automatisierungs-technik, 2025.

[Nyce 2004] D. S. Nyce: Linear Position Sensors Theory and Application, John Wiley & Sons, Inc., Canada 2004.

[Wertschützky 2018] R. Wertschützky: Sensor Technologien 2022, AMA Verband für Sensorik und Messtechnik e.V., Berlin 2018.

Biographies



Olaf Kissing received his diploma in industrial engineering from the University in Siegen in 2002 and his doctorate degree in engineering from the University of Siegen in 2010. He has been working as a product manager at Temposonics GmbH & Co. KG in the industrial sensors division for more than 10 years.

Additive Manufacturing in Hydraulics: Design Methodologies for Manifolds

Zita Tappeiner[1]*, Moritz Nieder[1], Robin Schulze Raestrup[1], Felix Spickenheier[1], Katharina Schmitz[1]

[1]*RWTH Aachen University, Institute for Fluid Power Drives and Systems (ifas), Campus-Boulevard 30, 52074 Aachen*

** Corresponding author: Tel.: +49 241 80 47729; zita.tappeiner@ifas.rwth-aachen.de*

Abstract.

Additive Manufacturing (AM) has established itself as a promising technology in the field of hydraulics, offering new opportunities for the design and optimization of components [1]. For converting classic designs into AM-compatible models, there are guidelines and methodologies available, including even specific approaches for manifolds [2]. However, the guidelines are usually very vague and provide limited concrete design specifications. In addition to this challenge, the available software solutions for the rapid creation of asymmetrical or bionic geometries are quite limited. Although there is a variety of software solutions on the market for designing AM components, their functionalities are not tailored to the requirements of AM. While conventional computer-aided design (CAD) is based on 2D sketches evolving into volumetric bodies, there is an increasing demand for advanced software solutions that facilitate freeform spatial design and provide constructive guidelines to enhance flow-optimized designs. This paper compares and evaluates various software and design methodologies to assess their suitability for additively manufactured hydraulic components.

In this paper we perform an analysis of the entire design process, including a comparison of pressure losses. The primary objective is to establish specific methodologies for the design of AM components. Results indicate significant shortcomings in conventional CAD software, which often limits designers' capabilities. To address these challenges, this study introduces an innovative design configurator implemented in the software Rhinoceros 3D® according to the approach of Biedermann [3]. This configurator facilitates the automated creation of manifolds by adjusting channel geometries to minimize pressure losses in hydraulic channels.

Keywords. Design, Additive Manufacturing, Manifold, Flow-Optimization.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

The use of AM compared to conventional production has increased in recent years. Consequently, the demand for integrated software solutions for product development is also increasing, particularly those that enable the automated creation of three-dimensional freeform surfaces.

To analyze which software solutions are currently available and how advanced their level of automation is, four different CAD software are compared in Table 1 based on various evaluation criteria. The programs considered are Siemens NX®, SolidWorks®, Rhinoceros 3D® (including Grasshopper®) and nTopology®. The evaluation criteria include a general definition, the capability for generative design, and automation possibilities. Due to the significant potential for material and weight reduction, the topology optimization capabilities of each program are also compared.

Table 1: Comparison of different software solutions for designing additively manufactured components

	Siemens NX	**SolidWorks**	**Rhino 3D + Grasshopper**	**nTopology**
Definition	Characterized by functionality, performance & stability [4]	Software known for conventional manufacturing [5]	Graphical programming approach [6]	Interesting for AM purposes [7]
Generative design	Direct Modeling [8]	Direct Modeling [8]	Based on NURBS [9]	Implicit modelling by utilizing functions [10]
Automation possibilities	Using external software [11]	Design automation of a cube [12]	Design automation of cable routing [6]	No current data found
Topology optimization	Yes [13]	Yes [14]	Yes [15]	Yes [10]

Siemens NX® was selected because it represents a widely used software solution in industrial product development and is characterized by extensive functionality, high performance, and stability [4]. SolidWorks®, in contrast, was originally designed for conventional manufacturing processes. Rhinoceros 3D®, including Grasshopper®, follows a graphical programming approach, whereas nTopology® is particularly suited for AM applications. It was developed within the DfAM framework and utilizes advanced design methods such as topology optimization and lattice structures, enabling increased design freedom [6]. In this paper, a comparison is drawn between classical CAD design (conventional design) using Siemens NX® and an automated design approach in Rhinoceros 3D®.

Design Automation (DA) refers to the computer-based automation of repeatable tasks in the design and design process. The advantages of DA are manifold: it reduces the time and effort required for the design of components through the automated creation of geometric bodies as well as their easy adaptation. This enables efficient design adjustments, facilitates the iterative design process, and minimizes human errors as well as the risk of faulty designs. [16], [17], [3].

For the design of valve block channels, which are represented by paths between valve cavities and connections, DA is presented through iterative path generation based on Position-Based Dynamics (PBD) simulation. PBD is already applied in areas such as special effects in films, physics engines in video games, and form-finding techniques. [3], [18]

2. METHODS

In order to compare conventional and AM-optimized design methods, both approaches must first be defined. To do this, the selected procedure is presented in each case. Conventional design refers to the approach in which the design is created manually in a CAD program. The alternative is a design configurator that automatically creates paths between the components. The methods presented in this chapter reflect the actual approach used in the design development and outline optional additional steps. These methods should not yet be regarded as general design methodologies. The results presented in this paper elucidate various aspects of the design process and highlight potential improvements in fluid-dynamic performance.

2.1. Conventional design process for product optimization

There are various methods for product development, such as the widely used VDI 2221 [19]. These methods provide a structure for the individual steps of product design. In practice, method-based approaches that simplify certain aspects of the methodology are often employed. The approach adopted in this paper is frequently used at the institute to implement application-oriented product development and is also employed in industry.

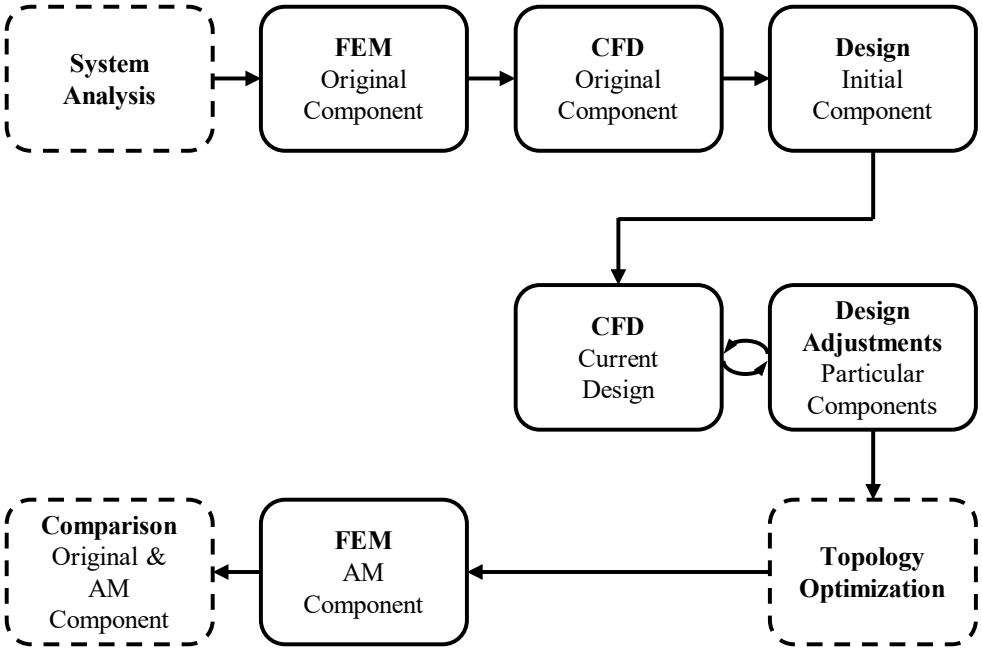


Figure 2.1. Methodological approach to the design of an AM manifold

The methodological approach follows the problem-oriented development process shown in Figure 2.1. This approach is based on established procedures for the structural optimization of hydraulic components and has been adapted to the specific requirements of a hydraulic manifold. It is divided into the necessary analysis steps and optimization processes, which lead, step by step, to a functional, production-ready optimization of the component. In addition, the process includes a comparison between the original and the new states of the manifold with respect to pressure drop, strength, and mass.

The steps shown with a dashed outline are not strictly required for the development of an AM manifold. In this paper, however, they are implemented to improve and validate the result. Topology optimization is generally advisable; since it was not implemented in the automated methodology used for comparison, it is presented here as optional. The mandatory steps for the (re)design of a manifold are the initial design, the iterative cycle of flow analysis and design adjustments, and strength verification. The intermediate topology optimization makes the FEM analysis a safety check; otherwise, it would be mandatory to verify the strength of the design.

The conventional design approach is applied to the manifold of a compact hydraulic power unit from a forklift. This manifold includes a high-pressure filter and two valves in the return line to the tank that controls the lowering of the fork. The primary functions implemented by the manifold are therefore “lifting” and “lowering” of the fork. The CFD simulation was performed in Ansys Fluent with an element size of 0,7 mm. During the lowering operation, a volumetric flow rate of 40 l/min was assumed. The fluid was modeled with a kinematic viscosity of 46 cSt and a density of 850 kg/m³. The thickness of the first prism layer was set

to 68 µm. The number of prism layers was limited to 10, with a growth rate of 1,2. The SST turbulence model was used.

2.2. Design Configurator

The procedure shown in Figure 2.2 represents an alternative design methodology. It is important to note that, within the scope of design development using the design configurator, no topology optimization has been implemented to date; instead, wall thicknesses are selected according to the system parameters, and the model is designed accordingly.

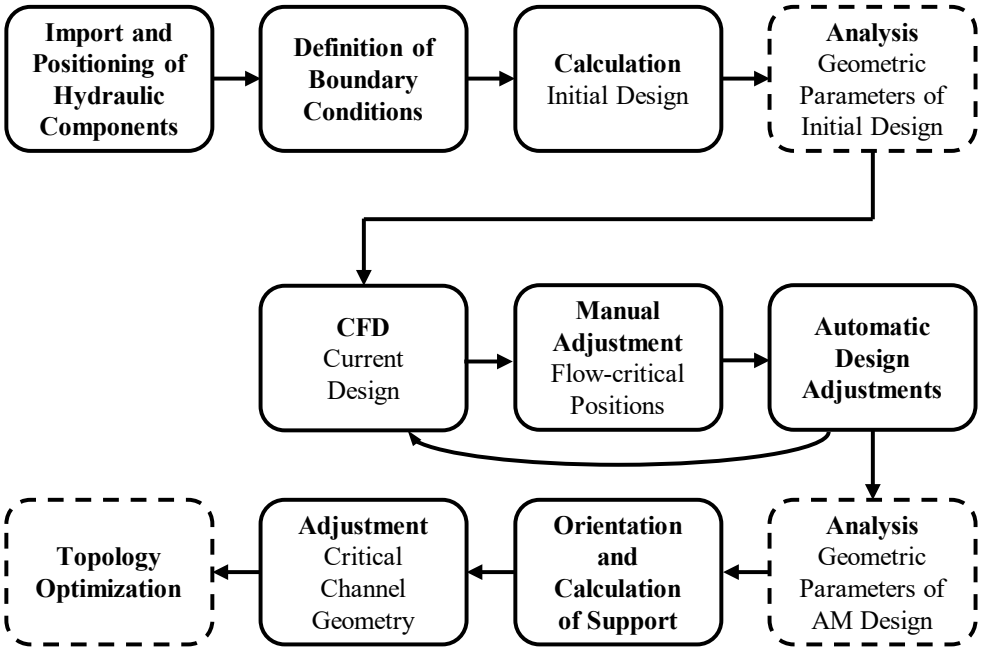


Figure 2.2. Methodological approach of the design configurator

The main differences to the conventional approach lie in the simplification of design adjustments using the configurator. Flow-critical areas can be modified in a targeted manner, and the rest of the model is automatically rebuilt. However, this carries the risk that manual channel adjustments may create new critical positions.

The design configurator used, which builds on the work of Manuel Biedermann, is implemented using Grasshopper®, a visual programming language and environment within the CAD software Rhinoceros 3D® [3], [20]. Version 8.1 of Rhino and version 1.0.0008 of Grasshopper are used. The path generation of the valve block channels is realized using a PBD algorithm through the Grasshopper plugin Kangaroo2 by Daniel Piker [18], [21]. The standard Kangaroo solver and its default settings are implemented. The implementation of the design configurator takes place on the valve block from Schmelzle, with the arrangement of the valve cavities based on Biedermann's layout [3], [22]. To implement the geometric boundary conditions, the blocks Anchor, Equal Angle, Coincident, Collider, Constant Tension, Direction, Dynamic Weight 1D, Length (Line), Load, and Rod are utilized in Grasshopper. With the help of these boundary conditions in the Kangaroo solver, various

geometric parameters of the design are modified. These include the channel length L [mm], the curvature radius R [mm], the curvature angle W [°], as well as the branching angles α, β, and γ.

In the first step, the hydraulic components, such as valve cavities and screw-in fittings, are imported and positioned within the build space of the valve block to be developed. During the second step, the channel routing is defined by initializing a path. This path is represented as an abstract line between the inlet(s) and outlet(s), with a distinction made between simple and branched paths. For simple channels with only one inlet and one outlet, a direct path is created, while for branched channels, branching points are used to account for multiple connections. Subsequently, the abstract line along the path is discretized with particles. Figure 2.3 shows the implementation of the first two steps in the design configurator.

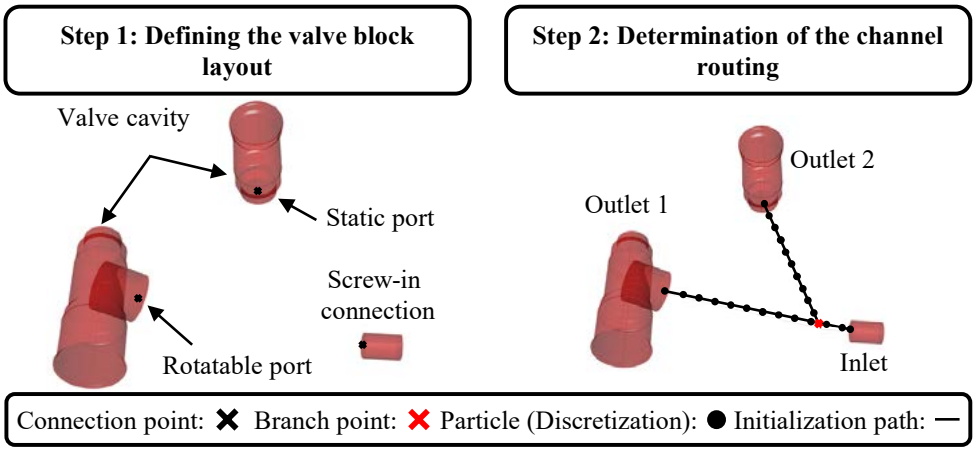


Figure 2.3. Defining the valve block layout and determining the channel routing

Within the third step, the boundary conditions of the paths or particles are defined, as they significantly determine the subsequent course of the channel path. The individual boundary conditions are weighted against each other, depending on the strength of this weighting, the influence of each boundary condition on the channel routing is determined. It may occur that two or more boundary conditions counteract each other, and a specific weighting must be provided for this purpose, as certain boundary conditions, such as collision avoidance, must be adhered to [6]. The boundary conditions are channel-specific.

In the fourth step, a suitable solver processes the defined boundary conditions using the PBD algorithm iteratively until either the convergence criterion is met or the maximum number of iterations is reached. Convergence is achieved when the threshold ϵ is reached. The threshold is calculated by dividing the square of the velocity of all path particles by the total number of particles N_{total} [3]. The solver gradually adjusts the position of the particles along the initial path in accordance with directions specified by the boundary conditions. The boundary conditions generate displacement vectors for each individual particle; the resulting displacement vector of a particle is obtained from the superposition of these various displacement vectors. Figure 2.4 shows the implementation of the boundary condition from step 3 in Grasshopper and step 4, illustrating both the initial and finished paths in the design configurator.

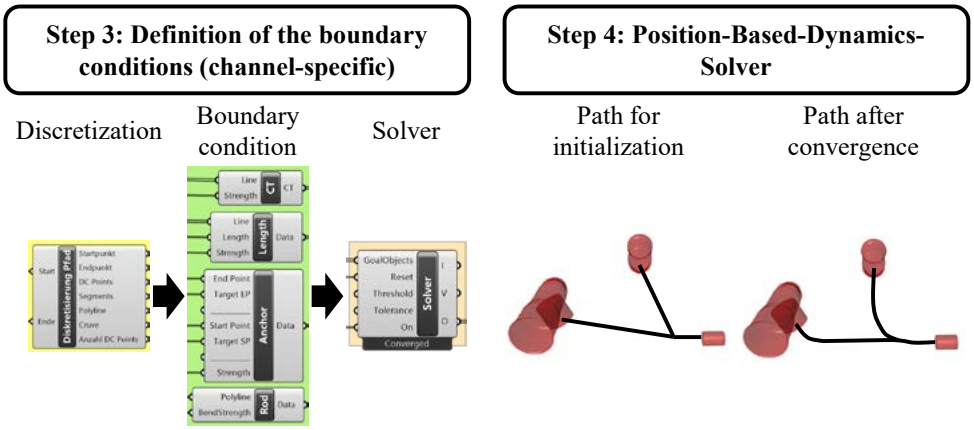


Figure 2.4. Definition of the boundary condition and Position-Based-Dynamic Solver

Finally, in the fifth step, the geometry of the valve block channel is constructed from the post-solver path by specifying the channel diameter and wall thickness. Figure 2.5 shows the CFD volume body and the AM channel geometry derived from the post-solver path.

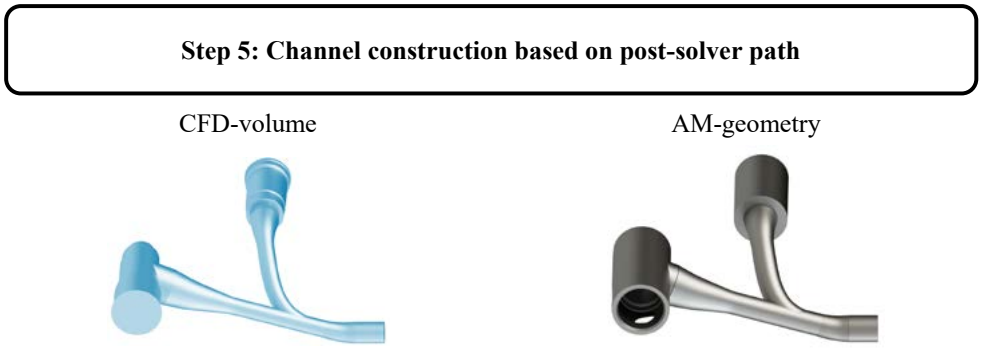


Figure 2.5. Finished valve block channel as CFD volume body and AM channel geometry

To illustrate the general applicability of the design methodology of the design configurator, various designs are created and examined. Based on this data set, a correlation analysis is conducted to determine guidelines for the design of the valve block channels for a specific operating point. To determine the Pearson correlation coefficient, the influences of the geometric parameters on the pressure loss are assessed using the correlation coefficient r. The correlation coefficient r is explained in detail in [24].

3. RESULTS

In this chapter, a fluid-dynamic comparison is made between the original forklift manifold and the component created using conventional design. This includes comparing the overall pressure drop across the component and presenting individual design adjustments at flow-critical locations. In addition, the results obtained with the design configurator are presented and compared in fluid-dynamic terms using a flow-path geometry.

Figure 3.1 shows a comparison between the original manifold and the component obtained through conventional design for AM. The component shown in a) is manufactured by cold extrusion followed by subsequent machining. Component b) is designed for AM, and the cavities likewise require finish machining. The individual optimizations that were carried out to achieve the final design are explained in detail. Compared to the original component, the new design saves 40 % in mass. When considering the change in pressure losses, a distinction must be made between two load cases: extension of the cylinder and retraction of the cylinder. During extension, 43 % pressure loss can be saved, and during retraction, 38 %.

a) b)

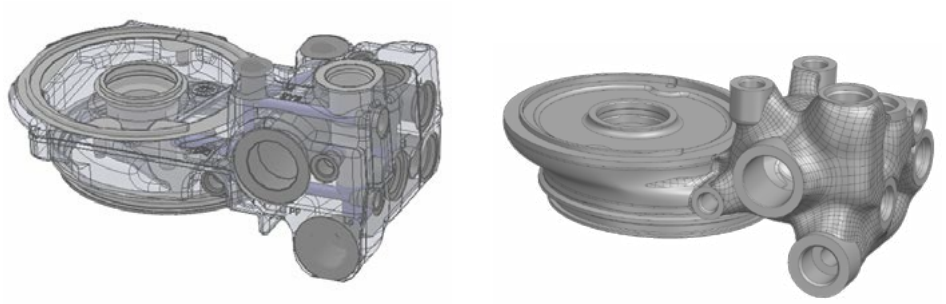


Figure 3.1. Results for the design of the valve block for a) conventional and b) additive manufacturing

The new layout of the hydraulic channels can be seen in Figure 3.2. Eliminating the 90° intersections when merging channels alone leads to a significant reduction in pressure losses.

a) b)

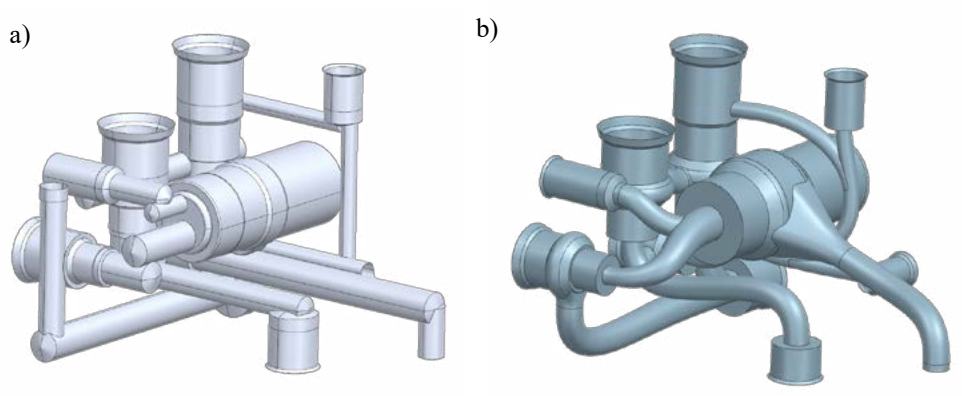


Figure 3.2. Channel route for the design of the valve block for a) conventional and b) additive manufacturing

Figure 3.3 shows the result of the CFD simulation for the old and new manifolds in the case of the cylinder lowering by the load mass. The 38 % reduction in pressure drop is evident from the color-coded contours in the CFD simulation.

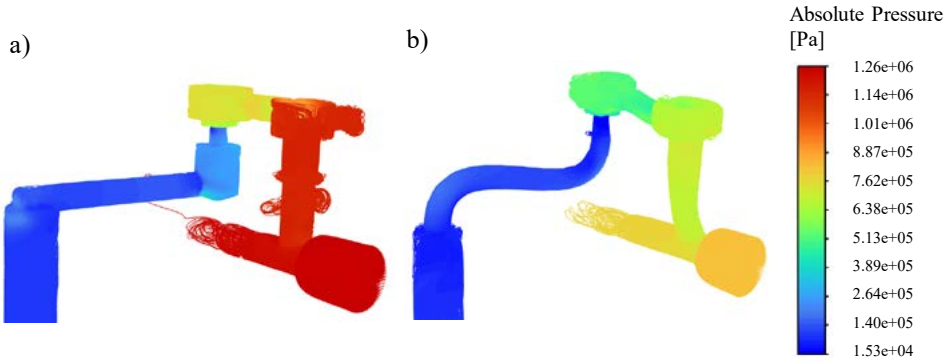


Figure 3.3. CFD-Simulation results of the manifold for a) conventional and b) additive manufacturing in the event of cylinder retraction (lowering)

The Figure 3.4 shows how the potential offered by the design freedom of AM can be exploited. Although the pressure loss is largely determined by the valve, the adapted valve outlet design and a flow-optimized connection can achieve an improvement of 35 %. The valve modification illustrates the approach used in conventional design. Components are optimized sequentially and verified in simulation. Flow paths are routed as NURBS curves between components without any optimization algorithm. At necessary locations, such as the connection to the filter, self-supporting channel cross-sections are implemented.

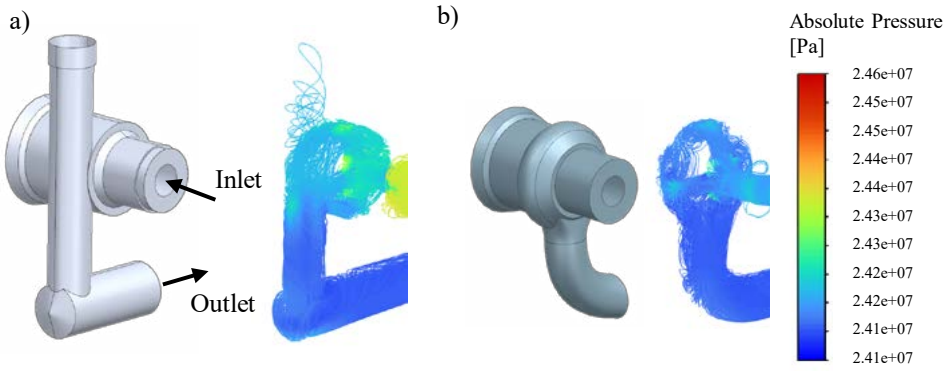


Figure 3.4. Valve design for a) conventional manufacturing and b) additive manufacturing

Conventional design highlights the substantial potential for efficiency gains; however, the design process itself is not particularly efficient. Once modifications are made, dependencies no longer hold, and repositioning components is feasible only with substantial effort. DA provides an alternative to this process. Thanks to the freely programmable and flexible extension of the design configurator, it is capable of creating a wide variety of channel designs and layouts. By adding or removing boundary conditions, as well as by inserting different components into the design space, customized valve manifold layouts and designs

are possible. Figure 3.5 illustrates this diversity by showing the possible modifications of a channel and the flexible adjustment of the valve manifold layout.

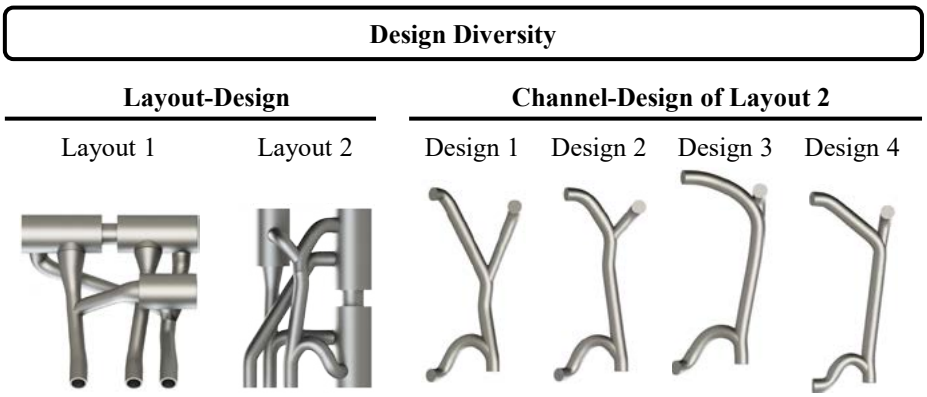


Figure 3.5. Realizable design variety using the design configurator

The design configurator enables the iterative optimization of channel flow due to its simple and rapid creation of various designs [23]. The workflow of this process is illustrated in Figure 3.6. Starting with the initial design, a CFD simulation is carried out and evaluated. In the evaluation, flow-critical positions are identified, and in the next step, the channel geometry is adjusted at these critical locations by modifying the boundary conditions in the design configurator. The resulting valve manifold layout is then generated, and the process is repeated until the predefined optimum is achieved.

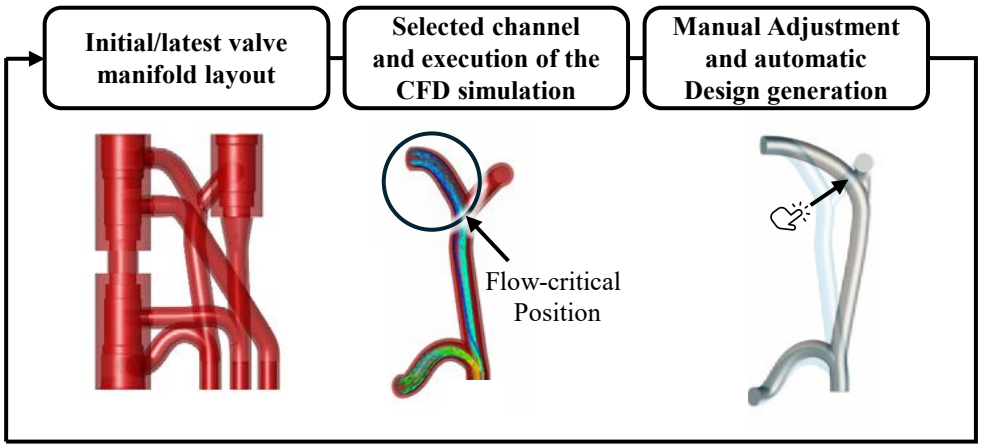


Figure 3.6. Workflow of flow optimization using the Design Configurator

The results of the case study for Layout 2 based on four channel designs are presented below. The case study examines flow optimization based on a defined operating point. The examined designs are iteratively created according to the workflow for flow optimization of the design configurator. To determine a local optimum, the pressure loss across the channel (Δp) is selected as an optimization parameter. Additionally, another parameter (the specific pressure loss) is introduced that describes the pressure loss across the channel divided by the channel length l ($\Delta p/l$) to determine the influence of geometry independent

from channel length. The pressure loss and the specific pressure loss of Designs 1 to 4 are shown in Figure 3.7.

The calculated results are determined using the simulation software ANSYS Fluent. The k-ω-(SST) solver is employed, and between 1000 and 3000 iterations are performed per design. HLP 46 hydraulic oil is used as the working medium for the simulation, with a density of 858 kg/m³ and a kinematic viscosity of 46 cSt. The mesh has the following characteristics: The minimum edge length is at least 0.5 mm, while the maximum edge length does not exceed 1 mm. The boundary layer consists of five layers, and the initial boundary layer has a thickness of approximately 66.71 µm with a growth rate of 1.2. The volume flow varies depending on the channel variant and ranges between 7.12 l/min and 125.61 l/min.

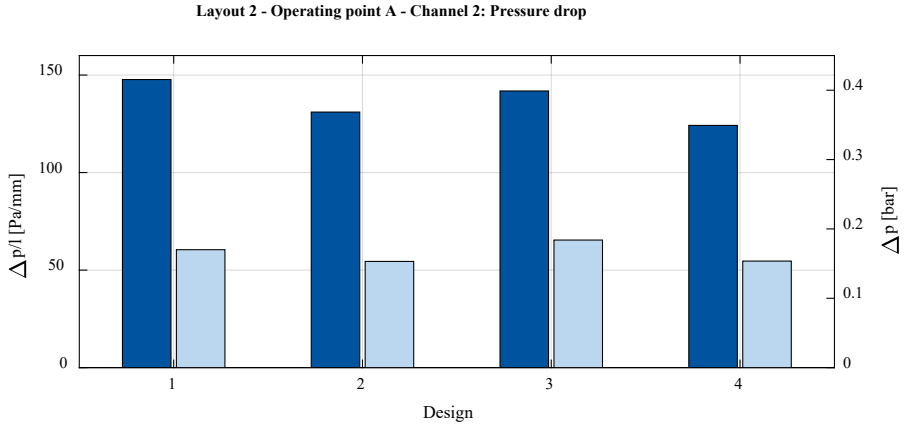


Figure 3.7. Pressure loss and pressure loss divided by channel length for designs 1 to 4

Design 2 has the lowest pressure loss at 0.153 bar. The specific pressure loss is 131.032 Pa/mm, making it the second lowest. Design 4 shows the lowest specific pressure loss with 124.187 Pa/mm; however, here the absolute pressure loss is at 0.154 bar, which is only slightly higher than that of Design 2. The reason for Design 4's poorer performance lies in its channel length; this measures 123.734 mm. The channel length of Design 2 is shorter and measures 116.896 mm. The shortest channel examined in this case study is the one of Design 1 with a length of 115.144 mm. For the case study, a correlation analysis is also conducted to investigate the correlation between various geometry parameters and the optimization parameter pressure loss. Conclusions regarding how to adjust the parameters cannot be drawn from the results due to the difficulty in classifying the geometry of the channels, as the geometric description depends on several individual parameters. The results of the correlation analysis indicate tendencies that certain parameters have a greater influence on the pressure loss of the channels, particularly the channel length should be mentioned.

For the investigation of operating point A, channel 2 from the Layout 1 and 2, the correlation coefficient is calculated to analyze the relationship between the number of occurring

curvature radii and angles with specific pressure loss (pressure loss divided by channel length), as well as channel length and absolute pressure loss. The correlation coefficient quantifies the potential influence of channel geometry on pressure loss. The significance is considered sufficient when a strong correlation exists with a correlation coefficient of at least ±0.5. The results from this analysis are presented in Figure 3.8, with the existing geometric parameters categorized into discrete ranges. This is due to the complex structure of the channels, which exhibits a variety of different curvatures with various parameters. Therefore, it is useful to categorize the geometric parameters into discrete ranges for a correlation analysis. The operating point considered has two inlets and one outlet. The volumetric flow rate at Inlet 1 is 12.88 l/min, and the volumetric flow rate at Inlet 2 is 7.12 l/min.

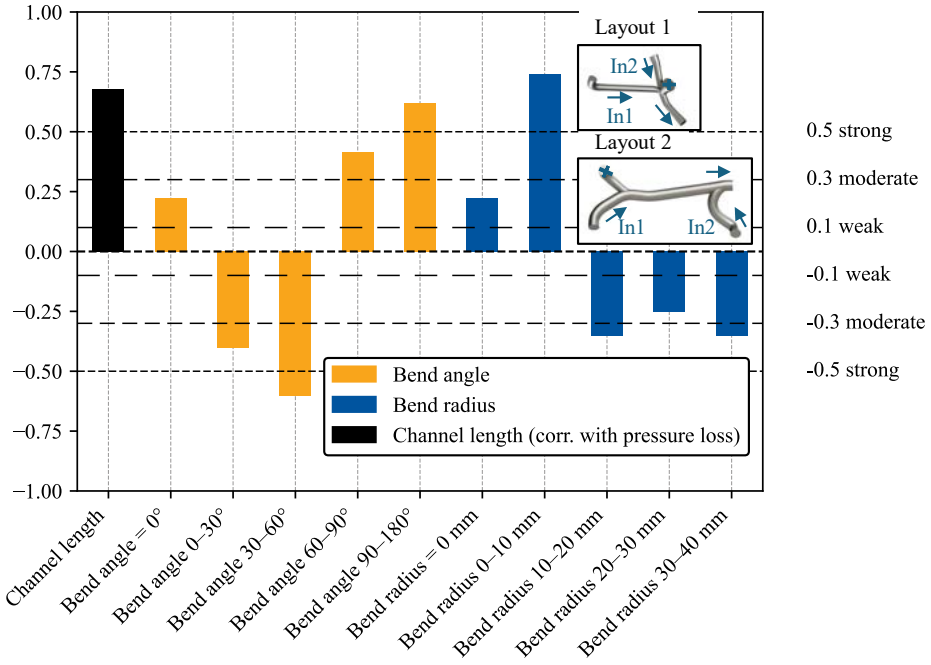


Figure 3.8. Correlation coefficients for operating point A – Channel 2

The analysis provides recommended design guidelines to achieve minimal pressure loss. The curvature angle shows two strong correlations. To reduce pressure loss, it is beneficial for the curvature angle to be in the range between 30° and 60°, where the correlation coefficient is strongly negative. Curvature angles from 90° to 180° should be avoided, as they exhibit a strong positive correlation; in this case, pressure loss increases with an increasing number of such curvature angles.

For curvature radii, a strongly positive correlation coefficient is observed in the range from more than 0 mm to 10 mm. Small radii should be avoided in design. The channel length shows a strongly positive correlation with pressure loss as well, leading to the conclusion that pressure loss increases with increasing channel length.

The design configurator was also used to automatically redesign the forklift manifold shown in Figure 3.1. The implementation process of this design configurator incorporated the use of parametric inputs. In the automated design process of the manifold, the flow-critical positions are identified and modified with minimal effort by utilizing the parametric input of the design configurator. With this a new manifold layout is devised by adjusting the position of the valve cavities or branches using parameters. By considering selected flow-paths in the CFD simulation and comparing them to the conventional valve block, the capabilities of AM are clearly visualized. Figure 3.9 shows the modification of a flow channel, which has led to a 44.6 % reduction in pressure loss.

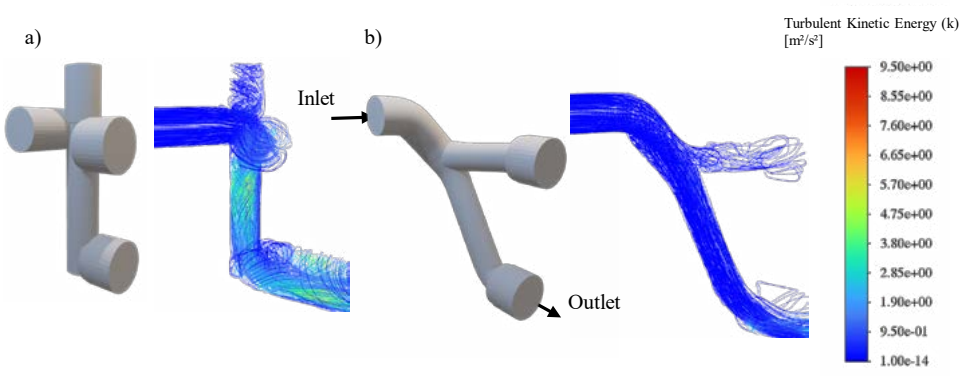


Figure 3.9. Channel optimization using design automation on the manifold

4. DISCUSSION

The analysis of pressure losses along one flow channel highlights the challenges of automated design optimization. Due to the high degree of design freedom, geometric characterization of the channel is very difficult, and losses cannot be attributed to individual parameters. An optimization focused solely on one geometric parameter, such as reducing channel length, does not lead to an optimal pressure-reduced flow in most cases within the valve block. Longer channels with geometries such as bends or branches often exhibit a lower specific pressure loss compared to shorter channels with the same geometries; however, this advantage is offset by extending the channel, resulting in a higher overall absolute pressure loss. For flow optimization of channel geometry, this means that a compromise must be found: between a preferably short channel length and bending as well as branching geometries that exhibit a lower specific pressure loss.

Although conventional design highlights the potential of the new AM design, its implementation is highly inflexible and time-consuming. A compromise would be to derive universally applicable rules that can be implemented in both design methodologies.

5. SUMMARY AND OUTLOOK

Based on the methods and results presented, two main conclusions can be drawn:

1. AM offers the opportunity to significantly reduce pressure drop in manifolds (up to a 43 % improvement).

2. DA enables the rapid generation of various manifold layouts that would be very time-consuming to achieve with conventional manufacturing.

These two findings reinforce both the use of AM in hydraulics and the development of automated design approaches. The next step toward establishing DA is the targeted adaptation of the algorithm for automated flow optimization. To this end, software solutions were already compared in the Introduction. In addition, simple design parameters must be defined so that knowledge-based engineering can be implemented. This includes, for example, defining suitable curvature radii, investigating channel intersections, and providing guidance on suitable self-supporting channel cross-sections. All of these goals are already being addressed within the HyRess research project to develop a guideline for designers. The resulting design guidelines are intended both to implement the constraints of AM and to achieve a flow-optimized component.

ACKNOWLEDGEMENTS

This research and development project is funded by the Federal Ministry for Economic Affairs and Energy (BMWE) within the Technologietransfer-Programm Leichtbau (TTP LB) and supervised by Projektträger Jülich.

6. REFERENCES

[1] Matthiesen, G., Merget, D. et al. 'Design and Experimental Investigation of an Additive Manufactured Compact Drive', International Fluid Power Conference, 2020.

[2] Rolinck, N., Schmitt, M. et al. 'Development Workflow for Manifolds and Fluid Components Based on Laser Powder Bed Fusion', Applied Sciences, Vol. 11, Nr. 16, pp. 7335, 2021. DOI: 10.3390/app11167335.

[3] Biedermann, M. 'Automated Design of Additive Manufactured Flow Components', 2022. DOI: 10.3929/ethz-b-000587405.

[4] Sadchikova, G. M. 'Application of NX Siemens PLM software in educational process in preparing students of engineering branch', AIP Conference Proceedings, 2017. DOI: 10.1063/1.4972454.

[5] O'Shea, B. 'Evaluation of Solidworks on how to integrate part preparation for Additive Manufacturing in the CAD modeling process', https://opus4.kobv.de/opus4-rhein-waal/frontdoor/index/index/docId/1382, 2022.

[6] Beutler, P., Berger, M. et al. 'Enhanced Design Automation for Hydraulic Manifolds Produced Using Additive Manufacturing', Procedia CIRP, Vol. 128, pp. 162–167, 2024. DOI: 10.1016/j.procir.2024.06.017.

[7] Mata, M., Pinto, M., Costa, J. 'Topological Optimization of a Metal Extruded Doorhandle using nTopology', U.Porto Journal of Engineering, Vol. 9, Nr. 1, pp. 42–54, 2023. DOI: 10.24840/2183-6493_009-001_001620.

[8] Junk, S., Burkart, L. 'Comparison of CAD systems for generative design for use with additive manufacturing', Procedia CIRP, Vol. 100, pp. 577–582, 2021. DOI: 10.1016/j.procir.2021.05.126.

[9] Guo, S. 'Modeling design of BH1500UU based on additive manufacturing and digital technology', 2020 International Conference on Innovation Design and Digital Technology (ICIDDT), pp. 544–547, 2020. DOI: 10.1109/ICIDDT52279.2020.00108.

[10] Dayanç, A., Canlidinç, M., Karakoç, F. 'An innovative methodology to design gyroid heat exchangers for metal additive manufacturing', International Journal of 3D Printing Technologies and Digital Industry, Vol. 7, Nr. 3, pp. 478–484, 2023. DOI: 10.46519/ij3dptdi.1365189.

[11] Zbiciak, M., Grabowik, C., Janik, W. 'An automation of design and modelling tasks in NX Siemens environment with original software - generator module', IOP Conference Series: Materials Science and Engineering, Vol. 95, pp. 12117, 2015. DOI: 10.1088/1757-899X/95/1/012117.

[12] Chowdhury, D. R. 'Design Automation Methodology of a Cube using Solidworks and VB.net', International Journal of Engineering Research & Technology (IJERT), Vol. 9, Nr. 6, 2020.

[13] Sedlacek, F., Kalina, T., Stepanek, M. 'Optimization of Components with Topology Optimization for Direct Additive Manufacturing by DLMS', Materials (Basel, Switzerland), Vol. 16, Nr. 15, 2023. DOI: 10.3390/ma16155422.

[14] Rosso, S., Savio, G. et al. 'Optimization Approaches in Design for Additive Manufacturing', Proceedings of the Design Society: International Conference on Engineering Design, Vol. 1, Nr. 1, pp. 809–818, 2019. DOI: 10.1017/dsi.2019.85.

[15] Lai, Y., Cai, Q. et al. 'A new evolutionary topology optimization method for truss structures towards practical design applications', Engineering Structures, Vol. 324, pp. 119326, 2025. DOI: 10.1016/j.engstruct.2024.119326.

[16] Johansson, J. 'Design Automation Systems for Production Preparation Applied on the Rotary Draw Bending Process', 2008.

[17] Wiberg, A. 'Towards Design Automation for Additive Manufacturing - A Multidisciplinary Optimization Approach', Linköping University Medical Dissertations Ser, 1st ed., Linkopings Universitet, https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=30401063, Linköping, 2019.

[18] Piker, D. 'Kangaroo: Form Finding with Computational Physics', Architectural Design, Vol. 83, Nr. 2, pp. 136–137, 2013. DOI: 10.1002/ad.1569.

[19] Verein Deutscher Ingenieure e.V. 'VDI 2221 Blatt 1: Entwicklung technischer Produkte und Systeme – Modell der Produktentwicklung', 2019.

[20] Robert McNeel & Associates: 'Rhino 3D Website', https://www.rhino3d.com/de/, 2025.

[21] Piker, D. 'Kangaroo 2.4.2 ReleaseNotes', https://github.com/Dan-Piker/Kangaroo-Documentation/blob/master/Kangaroo242ReleaseNotes.pdf, 2017.

[22] Schmelzle, J., Kline, E. V. et al. '(Re)Designing for Part Consolidation: Understanding the Challenges of Metal Additive Manufacturing', Journal of Mechanical Design, Vol. 137, Nr. 11, 2015. DOI: 10.1115/1.4031156.

[23] Pletzer-Zelgert, L. P., Dirks, S. et al. 'Comparative Evaluation of Optimization Algorithms for Automatic Build Orientation for Powder Bed Fusion of Metals Using a Laser Beam', 2024. DOI: 10.1007/978-3-031-42983-5_2.

[24] Siebertz, K., van Bebber, D., Hochkirchen, T. 'Statistische Versuchsplanung', Springer Berlin Heidelberg, Berlin, Heidelberg, 2017. DOI: 10.1007/978-3-662-55743-3.

Biographies



Zita Tappeiner received the bachelor's degree in mechanical engineering from RWTH Aachen University in 2020, and the master's degree in product development from RWTH Aachen University in 2021. She is currently working as group leader of the High-Performance Components research group at the RWTH Aachen University, Institute for Fluid Power Drives and Systems (ifas). Her research areas include hydraulic components, additive manufacturing, and flow optimization.



Moritz Nieder received the bachelor's degree in mechanical engineering from RWTH Aachen University in 2022 and the master's degree in production engineering from RWTH Aachen University in 2025. He is currently working as a research associate at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. His research areas

focus on high-performance components, specifically the development of ceramic spool valves.




Robin Schulze Raestrup and Felix Spickenheier completed their bachelor's theses at the Institute for Fluid Power Drives and Systems (ifas) at RWTH Aachen University. As part of their bachelor's theses, they dealt with various aspects of design and additive manufacturing and subsequently participated in the preparation of this paper. Felix Spickenheier is currently working as a student assistant at the institute. Both are now pursuing master's degrees.



Prof. Katharina Schmitz studied mechanical and chemical engineering at RWTH Aachen University and Carnegie Mellon University, Pittsburgh (USA) and graduated 2015 as Dr.-Ing. at RWTH Aachen University. Since 2018, she is full professor at RWTH Aachen University and director of the Institute for Fluid Power Drives and Systems (ifas). In addition, she is Vice Dean of the Faculty for Mechanical Engineering at RWTH Aachen, a position she holds since 2020. Prof. Schmitz’s awards and honors include several best paper awards and 2023 IMechE Joseph Bramah Medal award.

3D-Printed Polymer and Ceramic Spools as Alternatives to Steel for Hydraulic Valves

Jan Bartolj[a)], Ana Trajkovski[a)], Mitjan Kalin[a)], Katharina Schmitz[b)], Franc Majdič[a)]

[a)] *Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva ulica 6, 1000 Ljubljana, Slovenija*

[b)] *Institute for Fluid Power Drives and Systems (IFAS), RWTH Aachen University, 52074 Aachen, Germany*

jan.bartolj@fs.uni-lj.si

Abstract.

In this work, the possibility of replacing steel spools in proportional hydraulic valves with additively manufactured alternatives was investigated. Four materials were selected: E-Rigid Form, HTM 140, Aceaddity ceramic resin, and SuperHDT. Spools were printed with the same geometry as conventional steel parts, without any optimization, to allow for a direct comparison. The main focus was on the manufacturing process itself and the dimensional accuracy that could be achieved. For this purpose, the spools were measured in 3D and evaluated with respect to circularity and cylindricity, as these parameters are critical for proper valve function. Besides this, hydraulic tests with tap water [1] were also carried out in order to observe flow behaviour, pressure losses, and leakage. The results show that polymer and ceramic resins make it possible to manufacture functional prototypes that are lighter and corrosion resistant, but that dimensional stability and reliability still remain the main limitations.

Keywords: Additive manufacturing, Hydraulic valve spools, Polymer and ceramic materials, Geometric tolerancing (GD&T), Dimensional accuracy

1. INTRODUCTION

Spools are a crucial components in the hydraulic valves, being a movable part that is guiding flow of the fluid through the geometry of the valve. Valves come in various shapes and sizes, and their design us usually based on functionality as well as years of testing, optimisation and engineering practice [2]. Spool valves and so spools can be divided in hydraulics into conventional, proportional and servo valves. Each of those have their unique design based on their operation. In this study we will focus on proportional spools for valves that allow for continuous operation and management of flow, having in theory indefinite possible states between closed and fully opened. Proportional spools have specially made control notches

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

that are distinctive for this type of valves and that are working as nozzle, through which fluid us flowing and is being restricted, thus providing the ability for flow control.

Usually spools for hydraulic spool valves are of cylindrical shape because they need to fit into a hole that is drilled, grinded or honed by the rotating tool. Spools are conventionally made from steel, because this gives them excellent dimensional accuracy, wear resistance, long term reliability and manufacturability, because steel is very well-known engineering material.

However, manufacturing spools from steel also comes with some drawbacks. Machining spools and production of steel in general is highly energy intensive process. Also, if spools and other components were to be used in the water or wet environment, each component must be made of non-corrosive materials such as Stainless-steel or other alternatives. Third drawback is the weight of the steel spools since metals are in general denser compared to modern engineering polymers.

This all got us thinking if it is possible to create light and cheap spools that could withstand forces applied to them under hydraulic pressure and could also be manufactured with energy-less-intensive process thus having lesser impact on the environment.

Since there is a lot of different options to choose from in the engineering alternative materials we focused our direction into mainly polymers. We also wanted to be able to design and produce those vales by ourselves, so we used materials that could be Additively Manufactured (AM) on our 3D printers [3]. We have chosen four different engineering resins that have good thermal or other physical properties such as tensile strength or corrosion resistance. We have decided that this study was the first step towards better and optimized spools that could be used in hydraulics [4-7], but at this stage no geometry optimisation has been done.

Currently, polymers are used in hydraulics mainly for sealing applications and rarely as moving components, critical to the operation of the valves. So, study about feasibility in terms of material properties and dimensional accuracy of the AM manufactured spools, made from alternative materials, had to be made.

Since environmental awareness has been more and more important in recent times, we want to promote environmentally friendly or at least acceptable, hydraulic fluids [1]. In our case we will be using tap water as a hydraulic fluid, which will test our spools in tougher conditions, because water offers worse tribological protection and has tendency to corrode materials which will increase stress on materials. Water also has some positive effects that are combined with environmentally friendliness like non-flammability, availability in large quantities, …

Scope of this study is to manufacture spools originating from same geometry and create them using AM and alternative materials. We will be focusing on manufacturing feasibility, dimensional measurements and material evaluation while conducting the experiments. Goal is to determine whether the polymers are viable materials for spools in highly demanding environments like an inside of hydraulic valves [4].

2. MATERIALS AND METHODS

2.1. Spool Geometry and Design Constraints

As mentioned, we have used the same geometry that conventionally manufactured spool is based on. This is a geometry of quite standard CETOP 3 4/3 proportional directional spool, with outer nominal diameter of 12 mm. To allow the proportional flow control, there are control notches made, 3 per each chamber connection. Notches have width nominal dimension of 2.5 mm and have a depth 2.6 mm. Since this geometry of the spool is not adapted to the Additive Manufacturing (AM) we had to print the spools with a supporting structures that were removed after the printing process. Reference spool is made from Stainless-steel AISI 440-C, firstly turned to rough shape, surface heat treated and then fine grinded to the tolerance specified.

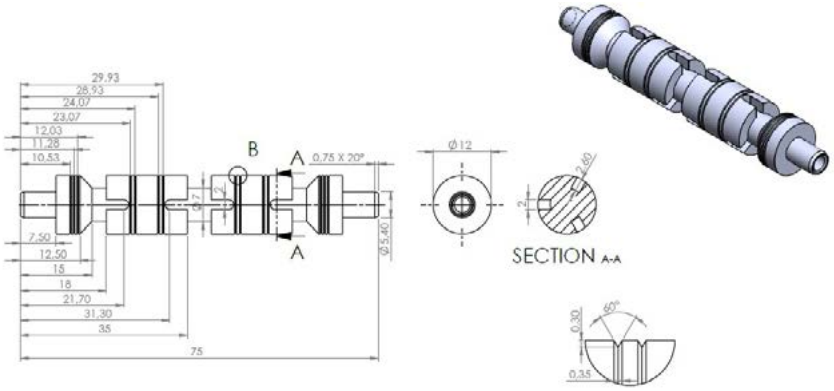


Figure 1: Detailed section from the manufacturing drawing of the spool for hydraulic proportional directional valves.

Usually, for practical application in water hydraulics it is mandatory for spools to achieve tight tolerances ranging from 0 to the negative value of 2 micrometres. This tolerance is crucial to ensure sufficient sealing between internal chambers of the valve by reducing the amount of leakage. Since we did not plan to do any postprocessing machining after 3D printing we did not expect to meet these criteria in full.

2.2. Used alternative materials

We have selected two polymer materials, one ceramic material and one hybrid/ceramic-reinforced resin. Another two polymer materials were from ETEC: High temperature mold material (HTM 140) with heat deflection temperature of 140 °C and Polyurethane-like resin named E-Rigid Form with very good tensile properties. Ceramic resin chosen was from Aceaddity and is named Ceramic Photopolymer 3D printing resin. According to manufacturer this resin contains 30 % ceramic particles. Lastly, we have chosen hybrid/ceramic-reinforced polymer resin SuperHDT that has excellent temperature resistance while still maintaining great load bearing ability. Most essential properties of each material can be found in the table below. In the results we will be presenting measured data with the short abbreviations of the names of each resin. Main advantages and material properties are shown in the Figure 2. Material properties are obtained from the manufacturers data and are dependant on the printing process.

Commercial name	E-Rigid Form	HTM 140	Ceramics 30 %	SuperHDT
Manufacturer	ETEC	ETEC	Aceaddity	3DMaterials
Material	Polyurethane-like polymeric resin	Polymeric resin	Polymeric resin, containing 30 % ceramic powder	Composite polymeric resin with nano-micro ceramic particles
Tensile strength [Mpa]	73	56	47.97 ± 10 %	39,5
Tensile modulus [Mpa]	3250	3350	819.52 ± 10 %	3
Elongation at break [%]	6	3,5	6.9 ± 10 %	<2
Temperature resistance [°C]	62.5	140	280	240
Essential properties	•Water resistance •High stiffness and dimensional accuracy. •Good for detailed geometries and tight tolerances. •Moderate resistance to water exposure; lower wear resistance compared to ceramics.	•Water resistance •Enhanced thermal stability. •Higher resistance to deformation under elevated temperature or prolonged stress. •Similar surface finish to E-Rigid Form, but slightly more brittle.	•Very high hardness and wear resistance. •Excellent corrosion resistance in water-based hydraulics. •Sintering shrinkage must be compensated in design. •Surface roughness and porosity depend heavily on post-processing.	•High density after sintering → improved mechanical strength. •Good dimensional stability and reduced brittleness. •Resistant to chemical attack and swelling in tap water. •Requires precise post-processing to meet hydraulic tolerances.

Figure 2: Alternative materials properties.

2.3. Additive Manufacturing process

Ceramics and SuperHDT were 3D printed using Stereolithography (SLA) process using desktop Anycubic Photon Mono M5 3D printer. SLA printing is a type of 3d printing process that uses photopolymeric resin and UV light to make printed parts. It has several crucial parameters, the most important being layer exposure time. This defines how long each layer is exposed to UV light from the illuminating display. Longer exposure time might make printed parts slightly larger, since more photopolymeric resin is hardening, but shorter times might create objects that are not fully hardened. Layer exposure time is mainly dependent on power of the light source and type of resin that is used for printing.

We used manufacturers recommendations to define parameters to manufacture and post-process the spools. Since recommended layer exposure time for SuperHDT was defined between 5.6 and 7 seconds for our printer, we have used both of these options. Similarly, when printing Ceramics, we used two different layer exposure times defined at 1.35 and 2 seconds.

HTM 140 and E-Rigid Form were manufactured by DLP process using Envisiontec (ETEC) D4K printer. Again, all manufacturers presets were used when defining printing and post-processing parameters. Post-processing in all cases was only cleaning the parts of any residual resin, supports removal and UV curing according to manufacturers specification. All spools that were compared during the research are shown on the Figure 3.



Figure 3: Spools made from alternative materials and standard spool made from AISI 440-C.

Additive manufacturing printers

ETEC D4K
The ETEC D4K is a professional DLP resin printer with a build volume of roughly 148 × 83 × 110 mm, prioritizing accuracy over size. It uses a 50 μm native XY voxel (with

enhanced modes down to ~25 µm) and typical 25–100 µm Z layers, enabling very consistent fine features. Recommended minimum wall thickness is ~0.5–1.0 mm, and dimensional accuracy is strong, with practical fit tolerances around ±0.05–0.15 mm for standard resins when well-calibrated.

Anycubic Photon Mono M5
The Anycubic Photon Mono M5 targets high detail at larger scale, offering a much bigger ~218 × 123 × 200 mm build volume. Its 12K LCD provides an ~19 µm XY pixel size, finer on paper than the D4K, with common 0.01–0.05 mm layers. Practical minimum wall thickness is ~0.5–1.0 mm, similar to DLP systems, while typical clearance fits require ~0.1–0.2 mm due to MSLA exposure and shrink effects.

2.4. Materials characterization

Dimensional properties

We have used different approaches to measuring the dimensional accuracy of 3D printed spools.

Firstly, we used Insize 3109-25A digital micrometre to measure the diameter of the spools. We took two representative spools of each material and took 6 measurements of the diameter per spool. After that, we have plotted all the results into a chart, where it is possible to compare the measured results for each resin.

Secondly measurements were performed on the Mahr MMQ 500 3D measuring machine. By using this, we wanted to measure the circularity and cylindricity. These two parameters are crucial for the geometry, so that spool fits properly into valve housing.

Cylindricity is a form of 3D geometric tolerance that controls how close the surface of the cylinders is to a perfect cylinder. Tolerance zone is the space between two coaxial cylinders within which the surface of a measured part must lie. It is basically a combination of circularity and straightness of the axis and taper.

Mass properties

After dimensional characterization we have also weighted the spools. We used laboratory Gram Lite EH scale to weight the spools and compare the results.

Tribological properties

For the tribological tests we prepared AM made steel discs made from Maraging Steel 1 (MS1) with a dimension of 25 X 25 X 6 mm representing the valve body. Steel discs were polished on automated preparation device RotoForce-3 to achieve surface roughness of 0.1 µm. Representing alternative materials, we made a pins with a dimensions of Φ 10 x 20 mm, with 12.5 mm radius on top.

Tribological tests were performed on a Camern-Plint TE 77 tribometer. In this test pin in sliding reciprocally on a printed disc with a frequency of 10 Hz. Stroke was 2.4 mm which is representative of a sliding distance inside valve housing. Normal load was 65 N. We used water as a lubricant. Testing period of each pin was 1 hour or 36 000 cycles. This was

determined in accordance with the normal load to show some kind of wear on each pin. 3 repetitions were done per each material pair. After the tests we evaluated the wear and coefficient of friction and averaged the results. Coefficient of friction was calculated from the ratio of normal and tangential force, that was dynamically measured during testing. Specific wear was calculated using the following equation (1).

$$Specific\ wear = \frac{\pi * d_{mid}^4}{64 * R} \quad (1)$$

Secondly, we observed the wear that was the result of the normal force and reciprocating motion between each material pair. Since we were using pure tap water as a lubricant, we were able to observe the rust that was forming on our MS1 discs.

Hydraulic properties

Since geometry of the spools was the same in all cases we did not expect any differences (improvements) in hydraulic measurable parameters. Mainly we wanted to determine if these materials are suitable for use in hydraulic environment. We used self-developed water hydraulic test rig capable of 11 kW hydraulic power or 350 bar at 30 l/min. Pump used is axial piston pump with 3 pistons. Spools were tested in the AM made valve housing which is connected to the also AM made mounting plate. AM valve has been optimised in terms of complex internal geometry to impose as low as possible resistance to the fluid flow [8]. AM mounting plate has been optimised as well, but mainly in terms of transducer placement, because we can measure the pressure extremely closely to the valve itself, thus reducing the uncertainty in the measurement coming from other sources.

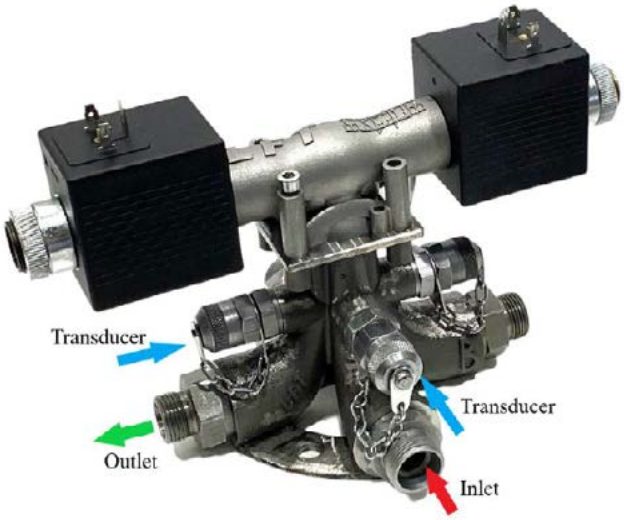


Figure 4: Valve in the testing assembly with magnets, mounting plate (AM) and hydraulic couplings.

Simple hydraulic scheme is shown on the Figure 5. As mentioned before, we have used 11 kW water hydraulic test rig. Each spool was tested separately to test its ability to withstand

the load in form of hydraulic pressure. We were measuring hydraulic pressure on the input and at the output from the valve. Mainly we wanted to determine if alternative materials are suitable for usage in real hydraulic applications. So, leakage test was performed first, followed by pressure drop characteristics dependent on the volumetric flow. Lastly pressure was increased in the pressure chamber till the breakage point of the spool itself. Leakage test was performed with spool in the middle neutral position. After applying the hydraulic pressure, we have measured the amount of leaked fluid on the neighboring A and B channels. For the pressure drop test we have gradually applied volumetric flow, and measured the pressure drop between the pressure (P) and working (A) port. Spool opening was kept constant at predetermined displacements, spaced by 0,25 mm. Since internal structure of the valve is symmetrical, we have switched position of the spool to the other side, just to confirm that both values were comparable.

Hydraulic tests were performed only using Ceramics and SuperHDT spools. This was due to the unavailability of the E-Rigid Form and HTM-140 resins during the stage of printing actual spools for usage in hydraulic valve.

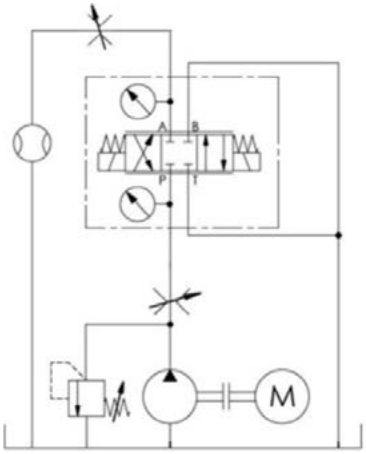


Figure 5: Simple hydraulic scheme for testing purposes.

3. RESULTS

3.1. Dimensional properties

Figure 6 is showing the measured results for each spool made from different materials. For the SuperHDT and Ceramics there are two sets of measurements, because these materials were printed at two different times of exposure per layer. It is evident that SuperHDT at both settings as well as E-Rigid Form and HTM 140 were printed with dimensions that were smaller than nominal dimension. Comparing these results to steel we can also see that distribution of measurement is quite large. Since steel spools were grinded to achieve the tolerance, the distribution of measurements is expectedly low. Largest distribution could be seen for the E-Rigid Form, but this is also the closest by the mean value to the nominal. HTM 140 was also relatively close comparing to the others, deviating from the nominal

dimension by approximately 80 µm. It is interesting to note that both printed Ceramic spools were measured above the nominal dimension. We can also clearly see the increase in the dimension for the Ceramic spool where we used higher exposure time. Sam principle can be observed for longer time of the exposure at SuperHDT.

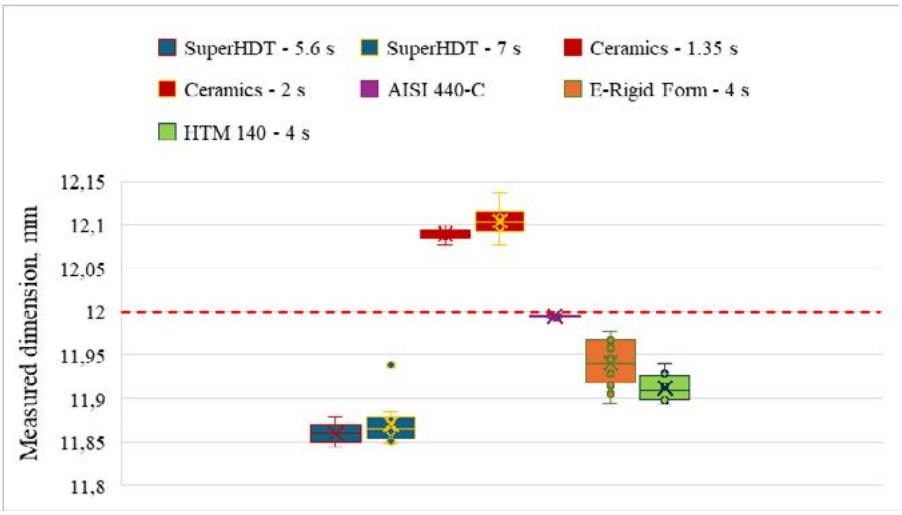


Figure 6: Measurements taken by micrometre plotted and compared according to 12 mm nominal dimension.

Cylindricity

In our case we measured each spool to determine the deviation from the perfect cylinder shape. Expected, from the previous circularity measurements, was that Ceramics printed at both exposures would perform the worst, which is actually the case here. Ceramics printed at 1.35 s exposure has a cylindricity 83.8 µm and printed at 2 s 112.3 µm. Again, we can see that elongation of exposure time made a result worse. HTM 140 resin preformed moderate with a cylindricity of 44.7 µm and E-Rigid Form 50.5 µm. Interestingly SuperHDT resin has a poor result at 5.6 s exposure time of 70.7 µm, but this result is greatly improved for increased exposure time at 7 s and was measured at 23.2 µm. Still the best performance was measured for the grinded Stainless-steel AISI 440-C spool at 2.4 µm.

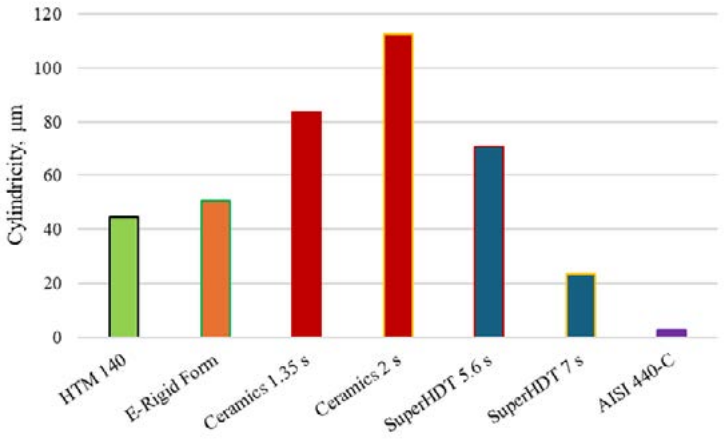


Figure 7: 3D measurement of cylindricity for each alternative material and standard AISI 440-C stainless steel.

3.2. Mass properties

Polymer spools weighted in at 6.21 g (E-Rigid Form) and 6.32 (HTM 140). Spool made from 30 % ceramics is weighing 7.01 g and partially ceramic spool (SuperHDT) 6.45 g. Our reference spool made from Stainless-steel has a measured weight of 44.82 g. In conclusion, all spools made of alternative materials are significantly lighter compared to the reference that is commonly used.

3.3. Tribological properties

Coefficient of friction

Firstly, we measured the coefficient of friction. This result is calculated based on the normal load and the measured value of tangential force needed to ensure reciprocating motion of the pin. Looking at the results of the measured coefficients of friction, we can observe how the values are changing per increasing cycle count. We can see the expected rising at the beginning where the load was applied and then all values star to converge towards certain values. Only E-Rigid Form is showing in this case that it would probably need more time and cycles to end the fluctuation in the results. No measurable differences between base and elongated exposure times were detected during testing for both Ceramics and SuperHDT, so we only included samples made with manufacturers recommended exposure times.

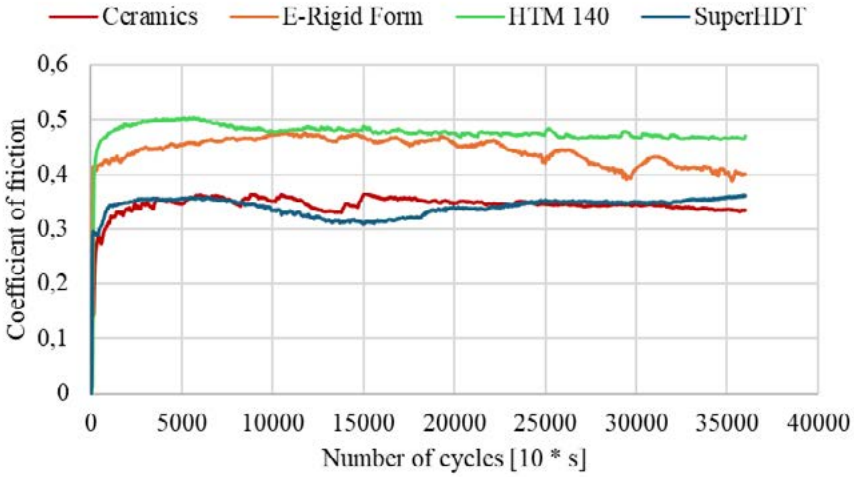


Figure 8: Coefficient of friction measured per 36000 cycles from the pin-on-disc tribological test.

Lowest averaged coefficient of the friction was achieved by the SuperHDT pin with a value of 0.342. Both polymeric pins did perform slightly worse with E-Rigid Form measuring 0.423 and HTM 140 measured 0.471. Here it is important to mention that only Ceramics had a noticeable sound when tribological test was performed. For comparison we have added the measured value of tribological pair AISI 440-C and MS1. It is very interesting that both ceramic and ceramic-powder-filled resin perform slightly better compared to steel-on-steel contact when lubricated with water.

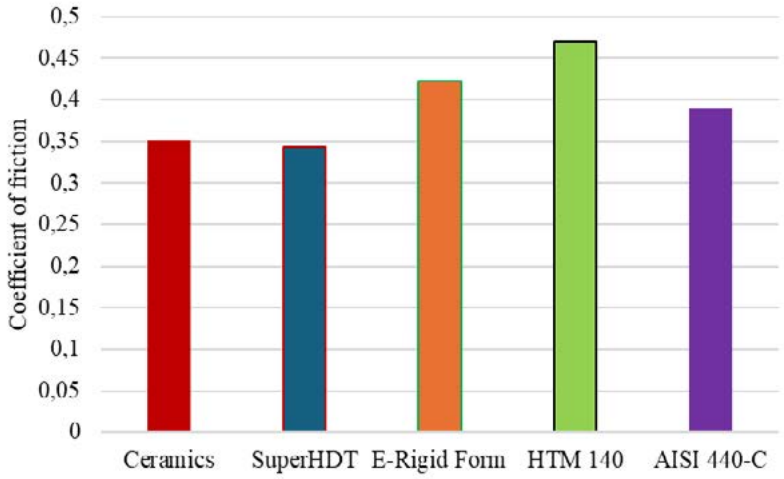


Figure 9: averaged value calculated of friction coefficient compared between alternative materials and standard AISI 440-C stainless steel.

Specific wear

It is evident that the smallest wear trace was left in the case of the Ceramics, because having the highest surface strength. Expectedly this also led to the smallest calculated specific wear of 0.223 mm^3/Nm. Impressively, HTM 140 also achieved quite low specific wear at 0.798 mm^3/Nm. Looking at the abrasion mark we can see that material was performing as a homogenous solid, without any material being compressed under the load force. Material degradation can be spotted in the case of E-Rigid Form. In the middle section of the pin abrasion mark it is evident that material did not withstand the load applied. Because of this, the highest Specific wear was calculated for this material at 1.986 mm^3/Nm. SuperHDT material had an abrasion circle on the pin with a diameter of approximately 4 mm with calculated specific wear 1.06 mm^3/Nm.

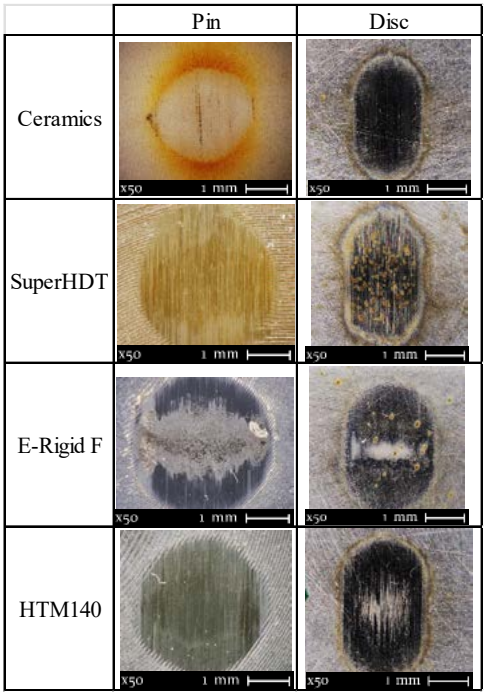


Figure 10: Digital microscope images of wear tracs on Pins from alternative materials and discs from MS1 steel.

Figure 11 is representing the measured specific wear values for each material.

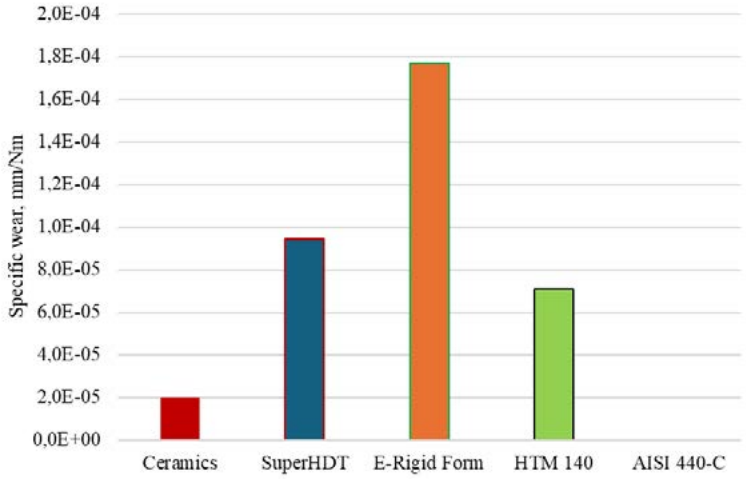


Figure 11: Specific wear for each material after tribological pin-on-disc test.

3.4. Hydraulic properties

Firstly, leakage of the hydraulic fluid was measured to determine the ability of the spools to sufficiently seal the internal chambers of the valve between themselves. Since circularity and cylindricity GD&T were not achieved, we did not expect the sealing and consequently to be sufficient and so that leakage to be quite high. This is confirmed on the Figure 12 where we can see that leakage measured was about 6 times higher in the case of two ceramic materials compared to AISI 440-C grinded spool. This was expected to have great consequences on the following results.

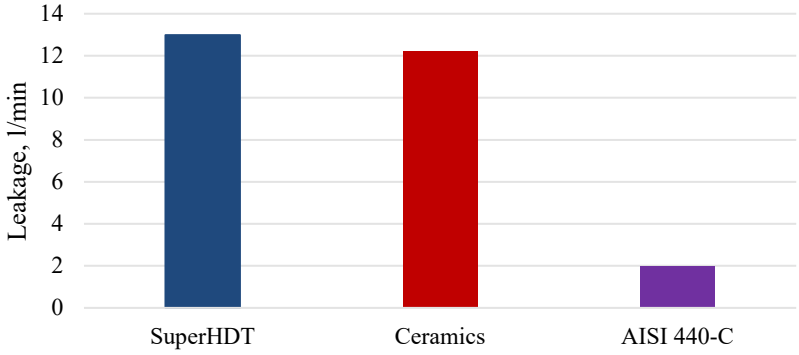


Figure 12: Measured leakage of hydraulic fluid between internal chambers of the valve.

Secondly, we measured the spool and valve characteristics during hydraulic operation. On the Figure 13, the results are shown for the 30 % valve opening. Effect of the internal leakage are quite obvious in this case. Since all measurements were supposed to show similar characteristics, due to leakage, the difference is quite noticeable. It is worth mentioning that results are shown at quite low pressure. This was done to prevent the spools from breaking under load prematurely. After determining the pressure drop per volumetric flow

characteristics, we have used 15 % valve opening at high pressure to determine how high hydraulic pressure the spools could withstand. SuperHDT spool did brake at 8.5 MPa of pressure, while spool made out of 30 % ceramics did not break at 7.5 MPa but the leakage was so high that we could not increase hydraulic pressure any longer due to limitations of our hydraulic aggregate. But Ceramic spool did brake once 10 MPa of hydraulic pressure was applied as an impulse.

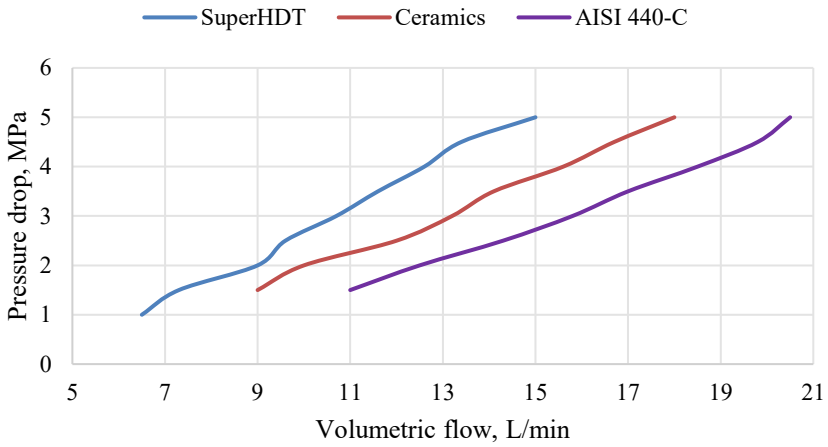


Figure 13: Measure performance of alternative spools compared to standard AISI 440-C steel spool.

Spools broken during actual hydraulic testing can be seen on the figure 14 below. Interesting, while also expected is, that both spools on the figure below did snap at the same location. This is obviously the weakest point of the spool geometry.



Figure 14: Spools from alternative materials, that were broken during actual hydraulic testing.

4. CONCLUSIONS

When comparing the additively manufactured spools with conventionally grinded steel spools, it is clear that the printed parts do not reach the dimensional tolerances required for efficient use in hydraulics. This was confirmed both by manual measurements and 3D evaluations, with circularity and cylindricity values outside the acceptable range. Some form of post-processing therefore appears unavoidable. On the other hand, the use of alternative materials results in spools that are about seven times lighter than steel, which can be advantageous in many applications. From a tribological point of view, the coefficient of friction is comparable, with ceramics and ceramic-filled resins even showing better performance than steel in certain cases. The weak point remains wear, which is significantly higher for all alternatives.

We have proven that simple spool geometry that was 3D printed could not withstand higher pressures that are usually applied to the hydraulic components. Both tested spools did brake at the pressures equal or less than 10 MPa. So, if there was an application of lower pressure hydraulics these spools might be feasible. For the lower pressure and power, we have proven that characteristics are relatively similar, but still the shape that is not reaching the defined GD&T tolerances, is mainly held back by the amount of leakage between the internal chambers of the valve.

For the further development, we are suggesting the improvement of manufacturing process in a way to create spools that are achieving or are at least closer to the tolerances that are required. Also, it would make sense to use materials that could withstand higher loads.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Solidworld d.o.o. 3D advanced solutions for their support in manufacturing the spool prototypes and testing pins.

Special thanks are also due to Poclain Hydraulics d.o.o. for conducting the 3D measurements of the manufactured spools, which were essential for determining the geometric specifications presented in this study.

5. REFERENCES

[1] Krawczyk J, Sobczyk A, Styczek J, Walczak P. Tests of New Methods of Manufacturing Elements for Water Hydraulics. Materials Research Proceedings 5, 2018, p. 200–5. https://doi.org/10.21741/9781945291814-35.

[2] Amiad Water Systems Eliptix R-30 & R-31 Hydraulic Control Valve n.d. https://au.amiad.com/products/eliptix-r-30-r31-hydraulic-control-valve-irr/ (accessed August 28, 2025).

[3] Ngo TD, Kashani A, Imbalzano G, Nguyen KTQ, Hui D. Additive manufacturing (3D printing): A review of materials, methods, applications and challenges. Compos B Eng 2018; 143:172–96. https://doi.org/10.1016/j.compositesb.2018.02.012.

[4] Klimek A, Kluczyński J, Łuszczek J. Performance Analysis of Additively Manufactured Hydraulic Check Valves with Different Postprocessing. Materials 2023; 16:7302. https://doi.org/10.3390/ma16237302.

[5] Peng T, Wang Y, Zhu Y, Yang Y, Yang Y, Tang R. Life cycle assessment of selective-laser-meltingproduced hydraulic valve body with integrated design and manufacturing optimization: A cradle-to-gate study. Addit Manuf 2020; 36:101530. https://doi.org/10.1016/j.addma.2020.101530.

[6] Arena M, Ambrogiani P, Raiola V, Bocchetto F, Tirelli T, Castaldo M. Design and Qualification of an Additively Manufactured Manifold for Aircraft Landing Gears Applications. Aerospace 2023;10:69. https://doi.org/10.3390/aerospace10010069.

[7] Hofmann U, Fankhauser M, Willen S, Inniger D, Klahn C, Löffel K, et al. Design of an additively manufactured hydraulic directional spool valve: an industrial case study. Virtual Phys Prototyp 2023;18. https://doi.org/10.1080/17452759.2022.2129699.

[8] Bartolj J, Trajkovski A, Hočevar M, Schmitz K, Majdič F. SLM 3D printed proportional directional water hydraulic valve, tested with optimized mounting plate. In: Ericson L, editor. Proc. 19th Scandinavian International Conference on Fluid Power (SICFP'25). River Publishers; 2025. p. [insert page numbers if available]. https://doi.org/10.13052/rp-9788743808251A02

Biography



Jan Bartolj is a PhD student and researcher at the Faculty of Mechanical Engineering, University of Ljubljana, specializing in additive manufacturing and its applications in the hydraulic industry. His research focuses on the design, optimization, and numerical simulation of hydraulic components, with a strong emphasis on innovative 3D printing solutions. He has collaborated on multiple industrial and academic projects, authored several scientific publications, and was

awarded the Golden Diploma by the Slovenian Fluid Power Society in 2023. Driven by the ambition to become a global leader in this field, he is dedicated to advancing the hydraulic industry through modern manufacturing technologies.

Hydraulic Substitute Base for Demolition Hammers

Johannes Klotz[1]*, Sascha Hasenoehrl[2], Prof. Dr.-Ing. Sven Matthiesen[2], Prof. Dr.-Ing. Marcus Geimer[1]

[1]*Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT)*

[2]*Institute of Product Engineering (IPEK), Karlsruhe Institute of Technology (KIT)*

**Corresponding author: Johannes Klotz (johannes.klotz@kit.edu)*

Abstract.

Demolition hammers are currently tested on mineral materials or on mechanical substitute base systems such as the steel ball energy absorber Dynaload. These bases are either destroyed during testing or not designed to replicate different materials. This work investigates a hydraulic substitute base as an equivalently effective, repeatable, adjustable, and low-wear alternative. Various arrangement concepts are evaluated. The selected concept combines a tool of a demolition hammer and a cylinder piston into a single impact body. The design approach uses a co-simulation that couples an FEM model of the impact mechanism with a hydraulic base model. Parameter studies show that by tuning hydraulic capacitance, resistance, and inductance, the tool displacement curves and dissipated energy can be targeted to approximate real materials. Therefore, this paper shows how a hydraulic substitute base could be adjusted. For concrete-like behavior (penetration depth ≈ 0.4 mm, energy absorption ≈ 83%), the simulation yields a required volume of 0.3 ml with mineral oil and an orifice cross-section of about 1.1 mm^2. With the identified adjustment ranges for capacitance and resistance, a hydraulic substitute base for a handheld rotary hammer with 4.2 J impact energy can be designed. The methodology can be transferred to other impact demolition hammers, such as hydraulic breakers, enabling reproducible testing.

Keywords. hydraulic substitute base, co-Simulation, rotary hammer testing.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Demolition hammers are used to break up mineral materials or building materials. Their core is an impact mechanism that acts on the chisel at a frequency in the single-digit or double-digit Hz range. The resulting impact force crushes the material beneath the chisel tip. Prototype testing remains indispensable for developing such machines because model assumptions and simplifications in simulations leave residual uncertainties. However, tests on mineral materials are problematic. These materials are destroyed during the test, limiting long-duration testing. In addition, mineral materials are structurally inhomogeneous, which reduces the repeatability of test results. Especially for the tests required by EU directives to assess vibration emissions, high repeatability is essential [1]. Consequently, "artificial" test bases - also called substitute bases - are used [2]. They should replicate mineral materials in an equivalently effective way with low scatter. Existing solutions, such as the Dynaload system [3] specified in DIN EN ISO 28927-10 [2] for hand-held demolition hammers without rotary motion, are predominantly mechanical and only partially meet testing needs. For example, the mechanical substitute base Dynaload used for vibration tests is not designed to represent different materials via adjustable parameters [3].

This paper addresses the research gap stemming from the lack of an adjustable substitute base. It investigates alternatives to purely mechanical systems, with a focus on an adjustable hydraulic substitute base. Due to the low compressibility of hydraulic fluids, high impact forces can be damped while chisel displacements remain small. Hydraulic components also allow targeted adjustment of stiffness and damping [6]. The paper shows, using a rotary hammer as an example, how such a substitute base should be designed in order to combine equivalent effectiveness, repeatability, adjustability, and wear resistance.

2. STATE OF THE ART IN TECHNOLOGY AND RESEARCH

2.1. *Substitute Bases*

For experimental evaluation of demolition, substitute bases are used to reproduce the behavior of mineral materials and provide a reproducible test environment. In practice, predominantly mechanical substitute base systems are employed [2], [3].

A well-established representative is the Dynaload system, developed by the PNEUROP Tools Committee [3]. Dynaload is a standardized test device for evaluating vibration, noise, and performance of hand-held percussive tools according to DIN EN ISO 28927-10 [2]. The system consists of a metallic cylinder filled with steel balls, as shown in Figure 1. The tool transfers the impact wave to this column of balls via a metallic piston plate. Part of the impact energy is dissipated upon impact through elastic deformations and relative motions of the balls. About 15% to 20% of the impact energy is reflected back to the tool, creating a loading situation similar to working on a hard mineral material. The system is available in several sizes (cylinder diameters 20 mm, 40 mm, 60 mm) and can be fitted with various tool - piston combinations. These variants extend test-stand capabilities (e.g., rotary hammers can also be tested in hammer-drilling mode where the tool additionally rotates). PNEUROP test series showed that Dynaload provides high repeatability and is insensitive

to small manufacturing tolerances or wear; repeatability remained stable throughout the test day and from one day to another. [3]

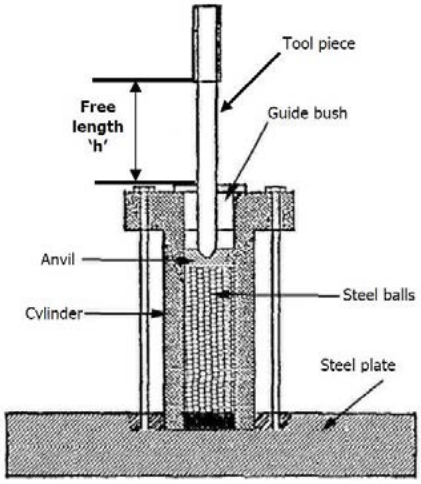


Figure 1: Mechanical steel ball energy absorber Dynaload [3]

For hydraulic breakers, the literature presents a test bench for determining impact energy as described in [5]. A key element is a shock- and vibration-absorbing base realized as a multilayer combination of sand, crushed stone, wood, and rubber, as shown in Figure 1. This design serves to decouple environmental influences and ensure repeatable measurements. [4]

The base is not intended to be adjustable. Moreover, it is reasonable to expect that the layered materials - especially crushed-stone layers - densify over time, changing their mechanical properties and affecting repeatability.

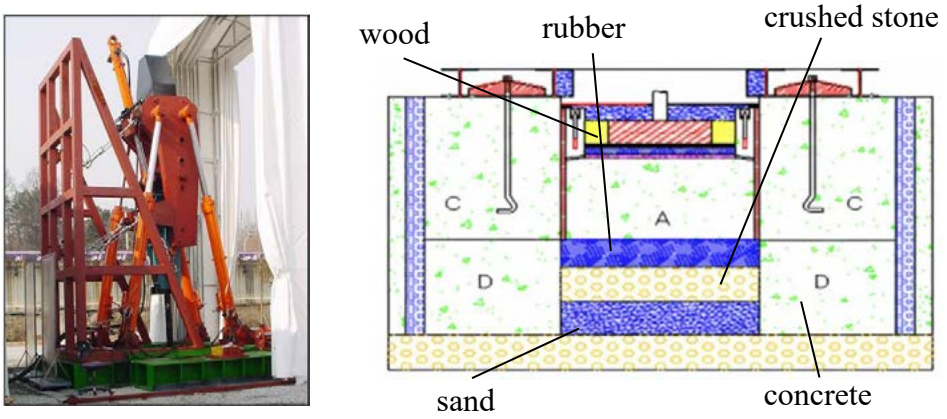


Figure 2: (left) test bench for hydraulic breakers; (right) impact absorbing base [4]. Text boxes were added by the authors.

In summary, mechanical substitute bases like Dynaload are suitable for standardized comparative measurements, but they cannot be tuned to different material properties and therefore cannot provide equivalently effective replication of different mineral materials. An advanced concept - such as the hydraulic substitute base presented by Klotz et al. offers the possibility to adjust both stiffness and damping by varying oil volume and orifice cross-section, thereby reproducing the behavior of different materials in a repeatable manner [6].

2.2. *Simulation of Hydraulic Impact Damping*

Hasenöhrl et al. investigate the shock-induced propagation of strain waves in a rotary hammer, taking into account the geometry of the impact body, using FEM simulation [7].

Klotz et al. describe a simulation environment enabling the determination of system parameters required to develop a hydraulic substitute base. The described co-simulation couples a structural FEM model of the impact mechanism with a hydraulic model of the substitute base, thus realistically representing the dynamic interaction between tool and base. In the FEM model, relevant impact events in the rotary hammer and at the hammer - tool interface are represented, modeling flying piston, firing pin, and tool as deformable bodies. The FEM model outputs the time-dependent displacement of the hydraulically pressurized piston area, which is passed to the hydraulic model via a co-simulation engine. The hydraulic model describes the substitute base as a hydraulic system comprising capacitance, and resistance. From piston displacement, the model computes the cylinder pressure history and the resulting reaction force, which is fed back to the FEM model. This bidirectional coupling captures energy flows between the impact mechanism and the substitute base comprehensively. By varying hydraulic parameters - especially the enclosed oil volume and the orifice cross-section - the influence on stiffness, damping, and energy absorption can be systematically investigated. Low hydraulic capacitance leads to a stiffer base with smaller penetration depth and higher peak pressures, whereas a larger orifice increases damping and thus energy absorption. In simulation, hydraulic damping could be varied between nearly 0% and 84% of the initial impact energy of the modeled rotary hammer. In this way, the interaction of hydraulic components can be tuned so that the substitute base reproduces the behavior of different mineral materials in an equivalently effective manner [6].

This simulation-driven approach allows different design variants to be evaluated before building hardware. The co-simulation thus serves as a tool for designing and parameterizing the hydraulic substitute base by revealing the relationships between geometric, hydraulic, and dynamic quantities and enabling virtual tuning to different base types.

3. CONCEPT SELECTION

3.1. *Requirements for a Substitute Base for Demolition Hammers*

For developing and testing demolition hammers, a test base is required that reproduces the behavior of mineral materials in an equivalently effective way. Because natural rock or concrete - due to inhomogeneity, anisotropy, and irreversible destruction under impact loading do not allow repeatable results, a substitute base is needed that can emulate the governing physical characteristics of the real material. From the authors' perspective, such

a base must satisfy four key requirements: equivalent effectiveness, repeatability, adjustability, and wear resistance.

Equivalent effectiveness means that the dynamic response of the substitute base - i.e., the time histories of penetration depth, contact force, and dissipated energy - matches those of a real mineral material as closely as possible. Only if the energy uptake and rebound of the base reflect typical building materials or rock tool loads can be represented realistically.

Repeatability describes the ability to reproduce identical impact events multiple times with minimal deviation, enabling comparisons of demolition hammers tested at different times and places.

Adjustability is required to emulate different base types such as concrete, granite, or limestone with varying stiffness and damping. An adjustable base allows systematic variation of material parameters and investigation of substrate influence on impact-mechanism behavior.

Finally, the substitute base must have high wear resistance to provide stable characteristics over many tests even at high impact frequencies and energies, enabling cost-effective testing.

3.2. *Evaluation of possible System Concepts for Substitute Bases*

Various technical concepts can in principle meet the requirements of equivalent effectiveness, repeatability, adjustability, and wear resistance. The following assesses mechanical, pneumatic, mechatronic, and hydraulic approaches for their suitability from the authors' perspective.

Mechanical systems combine springs, dampers, and mass elements. Very stiff bases can be realized, but their characteristics are only limitedly variable. Adjusting stiffness or damping typically requires part changes, increasing test effort. If damping relies on relative motion of components, wear and long-term stability can be problematic.

Pneumatic systems provide a soft, strongly damping characteristic due to the high compressibility of air. System parameters can be readily adjusted via pressure, volume, or orifice size. However, they are only of limited suitability for hard bases such as concrete or granite because the achievable stiffness is far below that of real mineral materials.

Electronic/mechatronic linear actuators or other controlled actuators could in principle emulate the interaction between tool and base. In practice, the high impact forces and very short strokes characteristic of demolition hammers impose stringent demands on structural stiffness and dynamics. Electric linear motors lack the required structural stiffness to absorb or dissipate these forces equivalently. Mechatronic concepts are therefore presently of limited suitability for reproducing the physical boundary conditions of a real mineral material.

Hydraulic systems combine high power density with low compressibility of the working fluid. As a result, they can represent both high impact forces and small deformations as observed in hard mineral materials. By adjusting enclosed oil volume and flow resistance, stiffness and damping can be tuned. At the same time, dissipative energy conversion in the fluid provides high repeatability with low wear. The simple adjustability of hydraulic components enables flexible adaptation to different material characteristics, making the system particularly suitable to meet the requirements of equivalent effectiveness, repeatability, adjustability, and wear resistance.

In summary, pneumatic and electric systems can reproduce the stiffness of mineral materials only to a limited extent. Mechanical systems have been developed but are impractical for adjusting damping characteristics and are prone to wear. Hydraulic systems therefore represent the most promising approach to achieve an equivalently effective, adjustable, reproducible, and low-wear substitute base. The following describes hydraulic concepts.

3.3. *Evaluation of possible Arrangement Concepts for Hydraulic Substitute Bases*

Two arrangement concepts were considered and are illustrated using a rotary hammer as an example. The first concept, as shown in Figure 3, is straightforward: the real mineral material is replaced by a hydraulic system consisting of a cylinder and other components. A commercial chisel is clamped in the rotary hammer as the tool. The hammer impacts a hydraulic cylinder via the chisel; the chisel and the cylinder piston are two separate bodies. Advantages include use of tools with original geometry, thus accounting for geometry effects on stress-wave propagation. Tools can be easily exchanged, allowing chisels of different length, diameter, and tip geometry. Another advantage is that the classical coefficient of restitution k can be used as a comparison metric, calculated from the velocities of the colliding bodies:

$$k = \frac{c_2 - c_1}{v_1 - v_2} \text{ [8]} \qquad (3.1)$$

where $v_1 - v_2$ is the relative velocity of chisel and base (i.e., cylinder piston) before impact, and $c_2 - c_1$ is the velocity difference after impact. A disadvantage is the wear at the chisel tip and on the hydraulic piston caused by the high-frequency impacts.

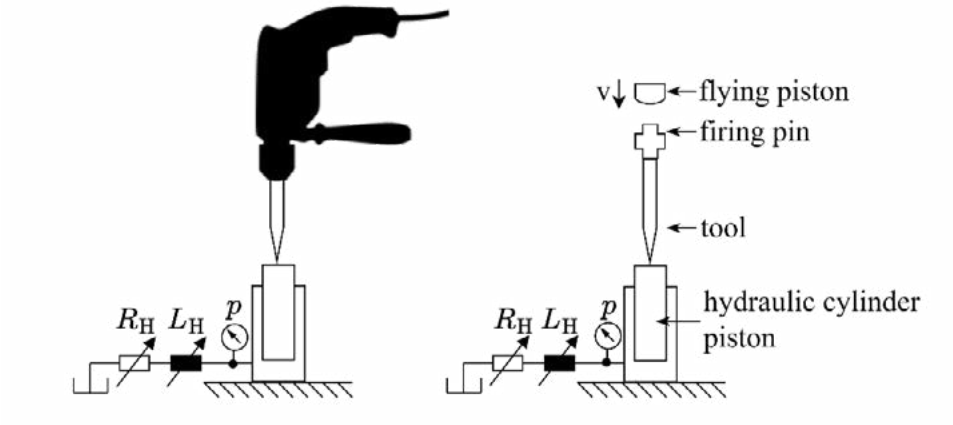


Figure 3: Arrangement concept 1: (left) The rotary hammer strikes a hydraulic cylinder piston with a standard tool; (right) Schematic representation of the relevant impact bodies.

Figure 4 shows the second arrangement concept. It differs from Concept 1 in a way that the tool acts as the cylinder piston. Therefore, no additional contact partner like the hydraulic cylinder piston in Concept 1 is needed. This arrangement has one fewer impact body than Concept 1. A drawback is that a coefficient of restitution according to Eq. (3.1) between tool and base cannot be computed, since the base in this concept is the pressure fluid and its "velocity" is hard to define. However, dissipated energy can be used as a comparison metric alongside penetration depth - specifically the ratio to the initial kinetic energy $\frac{E_{\mathrm{diss}}}{0{,}5 \cdot m_{\mathrm{fp}} \cdot \left(v_{\mathrm{fp}}(t_0)\right)^2}$. Concept 2 requires a custom-built cylinder that allows low-friction movement of the piston in the vertical direction and has a DSH-plus mount at the top. Cylindrical grinding can be used to produce a precisely fitting gap seal that is both low-friction and low-leakage. This eliminates the need for a friction-prone and flexible elastomer seal. The tool can be supported radially at the top and bottom to absorb radial forces transmitted by the rotary hammer.

Concept 2 is selected because of the advantage that the energy of the chisel is directly transferred to the hydraulic system, and a defined part of this energy is reflected back to the chisel. There is no additional contact partner. Due to the larger number of bodies between the flying piston and the hydraulics in Concept 1, more reflections of the incoming stress wave occur before they reach the hydraulics. With Concept 2, the impact characteristics can be influenced more directly by varying hydraulic parameters.

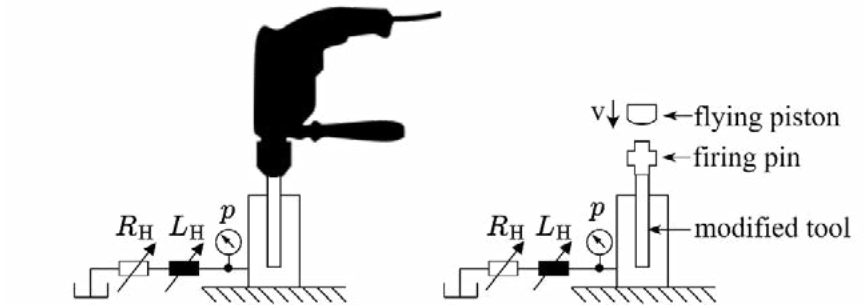


Figure 4: Arrangement concept 2: (left) The rotary hammer tool also functions as a cylinder piston; (right) Schematic representation of the relevant impact bodies of the rotary hammer.

4. HYDRAULIC SUBSTITUTE BASE

4.1. *Components of a Hydraulic Substitute Base*

A hydraulic substitute base for demolition hammers reproduces the mechanical interaction between tool and mineral material via a hydraulic system. The goal is to replicate the forces, displacements, and energy flows occurring during the impact process in an equivalently effective manner. The hydraulic substitute base can be described as a combination of hydraulic capacitance, hydraulic resistance, and hydraulic inductance. These elements determine the system's stiffness, damping, and inertia characteristics. [6] Their interplay allows the complex behavior of mineral materials to be replicated in an equivalently effective way; by varying these parameters, both stiffness and dissipated energy can be precisely adjusted. A hydraulic substitute base thus provides an adjustable, reproducible, and low-wear foundation for studying and developing percussive tools.

4.1.1. Hydraulic Capacitance

The hydraulic capacitance with elastically expanding vessel walls C'_{H} measures the compliance of a pressure chamber considering both fluid compressibility and wall elasticity: $C'_{\mathrm{H}} = -\frac{\mathrm{d}V}{\mathrm{d}p}$ [9]. Where $\mathrm{d}V$ is the volume change resulting from a pressure change $\mathrm{d}p$. It holds that $C'_{\mathrm{H}} = \frac{V}{K'}$ [9]. Where V is the chamber volume and K' the effective bulk modulus including both fluid compressibility K and wall elasticity [9]. Because the impact acts on a fluid in a hydraulic substitute base, at least one hydraulic capacitance is present. Low hydraulic capacitance leads to high stiffness and thus small tool displacements at high impact forces - behavior typical of hard rock. Larger capacitance produces a softer characteristic with greater penetration and lower peak pressures.

Figure 5 shows the penetration characteristic of a flat chisel clamped in a rotary hammer impacting concrete. The median penetration is 0.4 mm, with the maximum depth reached after 0.17 ms. Scatter during the intrusion phase is low. Using the co-simulation described in [6], a required hydraulic capacitance of $C'_{\mathrm{H}} = 4.4 \cdot 10^{-16}\ \frac{\mathrm{m^4 \cdot s^2}}{\mathrm{kg}}$ was determined to

achieve the same penetration depth without damping. In the simulation, the elasticity of the walls and possible gas components in the liquid were neglected: $K' = K$.

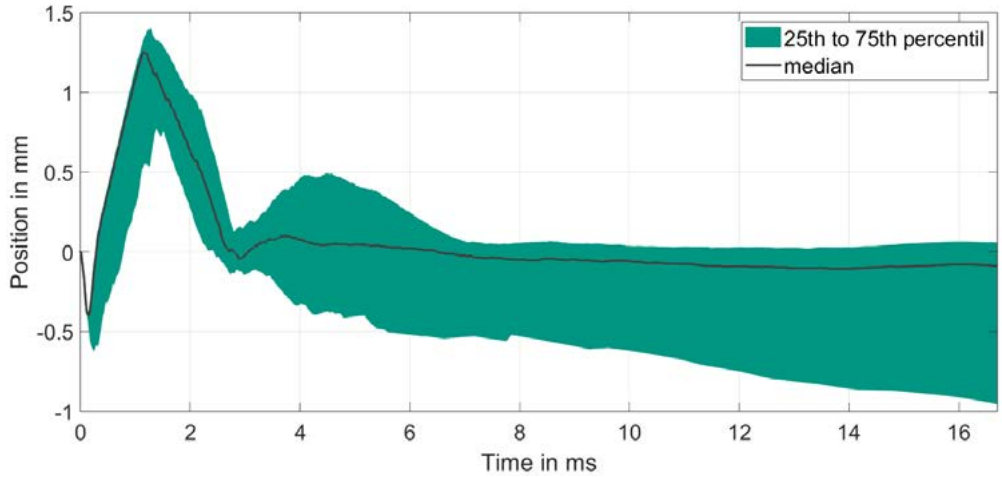


Figure 5: Displacement profile of 172 impacts of a rotary hammer with 4.2 J impact energy (Bosch GBH 4-32) on concrete C50/60 with 150 N feed force, recorded with a vibrometer.

The hydraulic capacitance can be practically varied via the adjustable oil volume to emulate different material stiffnesses. For mineral oils ($K \approx 1.6$ GPa), the volume required to achieve the penetration depth obtained in Figure 5 without damping is 0.7 ml. Softer materials require a larger volume. Harder materials require a smaller volume, which can be challenging to realize from a design standpoint.

If such small volumes are not feasible, a pressure fluid with a higher bulk modulus can be used to achieve the same capacitance at a larger fluid volume. Table 1 lists alternatives to mineral oil with significantly higher bulk modulus and thus higher system stiffness. At the same time, viscosity must be considered due to the precharge system and flow resistances. Very stiff media such as glycerol are highly viscous at room temperature, but blending with water allows adjustment to mineral-oil-like viscosity levels [15]. Replacing mineral oil with an ester fluid (HEES/HFDU, $K \approx 1.9$ GPa), allows the fluid volume to be increased by $\sim 20\%$ for the same capacitance without changing viscosity grade. For maximum stiffness, glycerol–water mixtures are an option. With about 10 wt% water, viscosity can be reduced to mineral-oil-like values ($\nu_{40°C} = 49\,{}^{\mathrm{mm}^2}/_{\mathrm{s}}$), while largely retaining the high bulk modulus ($K_{40°C} = \rho_{40°C} \cdot \left(u_{\mathrm{sound},40°C}\right)^2 = 1{,}222.1\,{}^{\mathrm{kg}}/_{\mathrm{m}^3} \cdot (1{,}862.9\,{}^{\mathrm{m}}/_{\mathrm{s}})^2 = 4.24$ GPa) [15], [16]. Ionic liquids offer high stiffness and good tribological properties at viscosities similar to mineral oil and are suitable as hydraulic fluids [10]. The bulk moduli in Table 1 refer to the properties of pure liquid. If the liquid contains free or dissolved air, the compression module of the fluid decreases and the hydraulic capacity increases. Figure 6 shows the influence of air on the compression modulus of a mineral oil. At low pressures, the influence of air is greater than at higher pressures. To avoid free air, the system must be

carefully vented. It is important to ensure during the design phase that no air pockets can form.

Table 1: Alternatives to mineral oil with higher stiffness

Fluid family	**Bulk modulus**	**Viscosity**
Mineral oil (HLP 46)	$K_{0\text{bar}} = 1.64$ GPa [10] $K_{200\text{bar}} = 1.86$ GPa [10]	$\nu_{40°\text{C}} = 47.07\ {}^{\text{mm}^2}/_{\text{s}}$ [10] $\nu_{100°\text{C}} = 7.36\ {}^{\text{mm}^2}/_{\text{s}}$ [10]
HEES (BIO-FR 46, ester fluid)	$K_{210\text{bar}} = 1.87$ GPa [11]	$\nu_{40°\text{C}} = 49.7\ {}^{\text{mm}^2}/_{\text{s}}$ [11] $\nu_{100°\text{C}} = 9.7\ {}^{\text{mm}^2}/_{\text{s}}$ [11]
HFDU (between VG 46 and VG 68)	$K_{210\text{bar}} = 1.87$ GPa [12]	$\nu_{40°\text{C}} = 55\ {}^{\text{mm}^2}/_{\text{s}}$ [12] $\nu_{100°\text{C}} = 12\ {}^{\text{mm}^2}/_{\text{s}}$ [12]
HFC (VG 46)	$K_{140\text{bar}} = 1.55$ GPa [13] $K_{200\text{bar}} = 1.9$ GPa [13]	$\nu_{40°\text{C}} = 46\ {}^{\text{mm}^2}/_{\text{s}}$ [14]
Glycerin (≥ 99 %)	$K_{20°\text{C}} = \rho_{20°\text{C}} \cdot \left(u_{\text{sound},20°\text{C}}\right)^2 =$ $1{,}261.3\ {}^{\text{kg}}/_{\text{m}^3} \cdot$ $(1{,}923\ {}^{\text{m}}/_{\text{s}})^2 = 4.66$ GPa [15]	$\nu_{40°\text{C}} = 227.39\ {}^{\text{mm}^2}/_{\text{s}}$ [15] $\nu_{100°\text{C}} = 12.07\ {}^{\text{mm}^2}/_{\text{s}}$ [15]
Ionic liquids (EMIM-EtSO4)	$K_{0\text{bar}} = 3.1$ GPa [10] $K_{200\text{bar}} = 3.32$ GPa [10]	$\nu_{40°\text{C}} = 39.44\ {}^{\text{mm}^2}/_{\text{s}}$ [10] $\nu_{100°\text{C}} = 7.66\ {}^{\text{mm}^2}/_{\text{s}}$ [10]

Another way to reduce penetration depth is to increase the precharge pressure. Simulations have shown that with a precharge of 350 bar, a penetration depth of 0.4 mm can be realized with a volume of 1.5 ml. A positive side effect of increasing the preload pressure is the reduced effect of dissolved air, as shown in Figure 6.

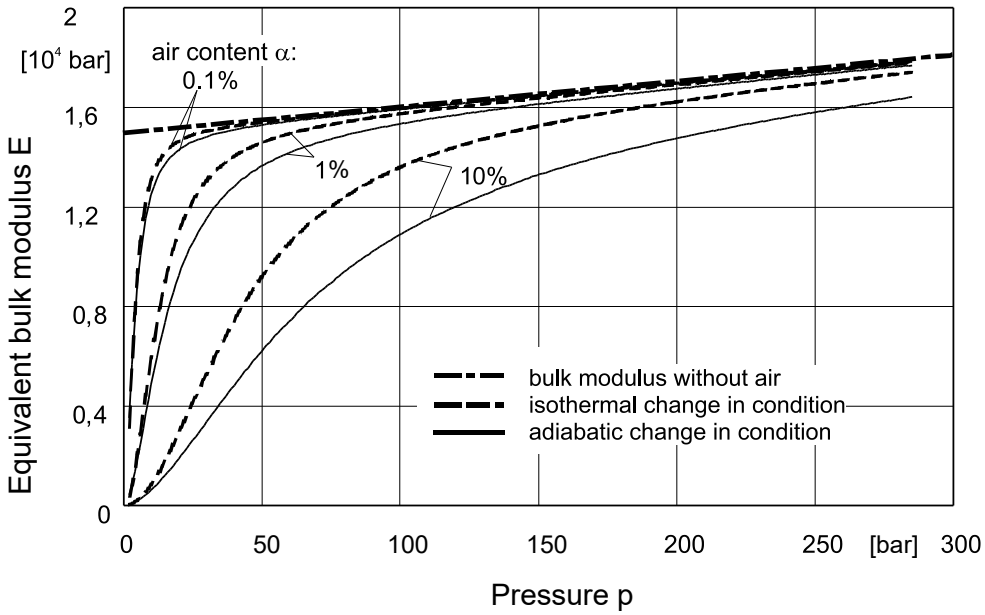


Figure 6: Equivalent bulk modulus of mineral oil-air mixture [17]

Figure 7 shows the measured penetration characteristic in concrete and the simulated penetration characteristic in the hydraulic substitute base for different parameter combinations. In the simulation, the elasticity of the walls and possible gas components in the liquid were neglected: $K' = K$.

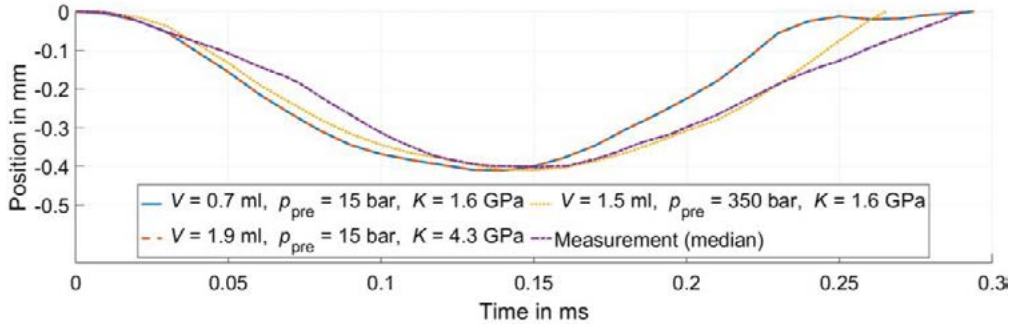


Figure 7: Measured and simulated profiles with identical penetration depth into the base. The cross-section of the orifice in the simulation is set to zero. The hydraulic resistance is therefore infinite. There is also no hydraulic inductance.

4.1.2. Hydraulic Resistance

Hydraulic resistance represents dissipative losses in the system and determines damping of the pressure history. It can be realized via orifices or throttles whose cross-section influences the outflow and thus pressure history. Compared with throttles, orifices show viscosity-

/temperature-independent flow behavior assumed turbulent flow occurs. As the orifice cross-section increases, energy absorption increases. Targeted adjustment of hydraulic resistance therefore allows tuning of the damping characteristic to different rock types. According to the guideline on the Dynaload steel ball energy absorber cited in DIN EN ISO 28927-10, about 15%– 20% of the impact energy in real materials is reflected back to the tool (rotary hammer) [3]; the remainder is dissipated in the base. Simulations have shown that, for 17% of the impact energy to be reflected back to the tool at a penetration depth of 0.4 mm, an orifice cross-section of 1.1 mm^2 and an oil volume of 0.3 ml are required.

4.1.3. Hydraulic Inductance

Hydraulic inductance describes the inertia of the fluid flow and is directly related to the moving fluid mass. It is determined by the geometry and length of the line as well as the fluid density. High hydraulic inductance affects pressure build-up in the system and causes a phase shift between force and displacement. This allows influencing the impact dynamics, particularly the temporal evolution of energy transfer between tool and substitute base. Due to the required low hydraulic capacitance, it is possible only to a limited extent to integrate inductances intentionally between capacitance and resistance. The additional line volume would simultaneously increase capacitance. A long pipeline with a small cross-section would also exhibit significant temperature dependence due to viscosity.

5. SUMMARY, DISCUSSION AND OUTLOOK

The paper demonstrates the advantages of a hydraulic substitute base for demolition hammers as an adjustable, low-wear, and highly reproducible alternative to conventional mechanical test bases. A coupled FEM–hydraulic co-simulation is used to identify the hydraulic capacitance and resistance required to replicate the penetration depth and energy absorption of concrete C50/60. This approach enables a material-realistic and transferable testing methodology for the development and assessment of demolition hammers.

Based on the requirements and mechanisms derived in this work, a prototype hydraulic substitute base will be designed, built, and instrumented. The challenge here is to achieve high stiffness. When using conventional pressure fluids in the design, the oil volume must be limited to 1.5 ml. If this is not possible for design reasons, a combination of alternative pressure fluids and a pressure preload can be used to increase the permissible volume. At the same time, care must be taken to ensure that the wall elements are constructed stiffly. One way to ensure high stiffness would be to manufacture the cylinder bore, as well as the bores for the pressure sensor and aperture, from a solid steel block. This eliminates the need for voluminous and elastic pipes or hoses. Consequently, the volume can be kept to a minimum. Keeping the pressure fluid chamber free of air is also challenging. The goal is to experimentally confirm the displacement and energy curves determined in co-simulation and to verify the identified adjustment ranges. If validation succeeds, the setup can serve as an adjustable, repeatable, and low-wear alternative to the Dynaload steel ball energy absorber and extend its scope of use, especially for material-specific studies. In addition, development of a larger hydraulic test rig is planned to enable testing of hydraulic breakers.

6. ACKNOWLEDGEMENTS

The project underlying this publication was funded by the Deutsche Forschungsgemeinschaft (DFG) under project ID 517363628. Responsibility for the content of this publication rests with the authors. The authors would like to express their sincere thanks for the support.

7. AUTHOR CONTRIBUTIONS

Conceptualization: Johannes Klotz
Methodology: Johannes Klotz
Software: Johannes Klotz
Validation: Johannes Klotz and Sascha Hasenoehrl
Formal analysis: Johannes Klotz and Sascha Hasenoehrl
Investigation: Johannes Klotz and Sascha Hasenoehrl
Resources: Marcus Geimer and Sven Matthiesen
Data curation: Johannes Klotz and Sascha Hasenoehrl
Writing—original draft preparation: Johannes Klotz
Writing—review and editing: Johannes Klotz, Sascha Hasenoehrl and Marcus Geimer
Visualization: Johannes Klotz
Project administration: Johannes Klotz
Supervision: Marcus Geimer and Sven Matthiesen
Funding acquisition: Marcus Geimer and Sven Matthiesen

8. REFERENCES

[1] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast) (Text with EEA relevance), Vol. 157, 2006. Accessed: May 16, 2025. [Online]. Available: http://data.europa.eu/eli/dir/2006/42/oj/deu

[2] DIN, DIN EN ISO 28927-10:2011-07—Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 10: Percussive drills, hammers and breakers, Jul. 2011.

[3] DYNALOAD - design, construction, use and maintenance - power tools, PNEUROP Tools Committee. Accessed: 10. Juli 2025. [Online]. Available: https://doczz.net/doc/4598702/dynaload---design--construction--use-and-maintenance

[4] J.-W. Park and H.-E. Kim, „DEVELOPMENT OF THE TEST SYSTEM FOR MEASURING THE IMPACT ENERGY OF A HYDRAULIC BREAKER“, Proceedings of the JFPS International Symposium on Fluid Power, Bd. 2005, Nr. 6, S. 75–79, 2005, doi: 10.5739/isfp.2005.75.

[5] CIMA (Construction Industry Manufacturers Association), Measuring Guide for Tool Energy Rating for Hydraulic Breakers, Milwaukee, USA, 1996

[6] J. Klotz, S. Hasenöhrl, S. Matthiesen, and M. Geimer, „Co-Simulation eines hydraulischen Ersatzuntergrunds zur wirkungsäquivalenten Abbildung mineralischer Materialien“, DS 140: Proceedings of the 36th Symposium Design for X (DFX2025), S. 1–10, 2025, doi: 10.35199/dfx2025.03.

[7] S. Hasenöhrl, J. Klotz, S. Matthiesen and M. Geimer „Simulation of Strain Waves in Impact Driven Tools: Towards Virtual Testing of an Adjustable Substitute Workpiece", Karlsruher Institut für Technologie (KIT), Karlsruhe, 2025. [Online]. Available: https://www.nafems.org/publications/resource_center/nwc25-0007167-paper/?srsltid=AfmBOorrZFj66XBld-MB5_pgRSfTrcltWKkF9cP6PZ6xG2zyzQ-Vp9Z5

[8] A. Böge and W. Böge, Technische Mechanik: Statik – Reibung – Dynamik – Festigkeitslehre – Fluidmechanik, 35th ed. 2024. Wiesbaden: Springer Vieweg, 2024. doi: 10.1007/978-3-658-44424-2.

[9] D. Findeisen and S. Helduser, Ölhydraulik: Handbuch der hydraulischen Antriebe und Steuerungen. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015. doi: 10.1007/978-3-642-54909-0.

[10] M. Kambic, R. Kalb, T. Tasner, and D. Lovrec, „High Bulk Modulus of Ionic Liquid and Effects on Performance of Hydraulic System", *ScientificWorldJournal*, Bd. 2014, S. 504762, Jan. 2014, doi: 10.1155/2014/504762.

[11] „Hydraulic-BIO-FR-46-68". Accessed: Oct. 21, 2025. [Online]. Available: https://www.anglomoil.com/wp-content/uploads/2022/08/Hydraulic-BIO-FR-46-68.pdf

[12] „Houghton - Technical Data Sheet". Accessed: Oct. 21, 2025. [Online]. Available: https://fireresistantfluids.com/wp-content/uploads/tds_quintolubric_855_EN.pdf

[13] „Danfoss technical data of Oils". Accessed: Oct. 21, 2025. [Online]. Available: https://www.btpco.com/download/training/hydraulic%20oil/Danfoss%20oil%20Book/Danfoss%20techniacal%20data%20of%20Oils.pdf

[14] „TDS_HOUGHTO-SAFE C646_EN". Accessed: Oct. 21, 2025. [Online]. Available: https://home.quakerhoughton.com/wp-content/hub/Files/TDS_HOUGHTO-SAFE%20C646_EN.pdf

[15] „Physical_properties_of_glycerine_and_its_solutions". Oct. 21, 2025. [Online]. Available: https://www.cleaninginstitute.org/sites/default/files/research-pdfs/Physical_properties_of_glycerine_and_its_solutions.pdf

[16] L. Negadi u. *a.*, „Effect of temperature on density, sound velocity, and their derived properties for the binary systems glycerol with water or alcohols", *The Journal of Chemical Thermodynamics*, Bd. 109, S. 124–136, Juni 2017, doi: 10.1016/j.jct.2017.01.011.

[17] K. Schmitz and H. Murrenhoff, *Fundamentals of Fluid Power*, vol. 7. Düren: Shaker Verlag, 2018. [Online]. Available: https://www.shaker.de/de/site/content/shop/index.asp?lang=de&ID=8&ISBN=978-3-8440-6306-6

Biographies



Johannes Klotz received his B. Sc. in Mechanical Engineering in 2021 and his M. Sc. in Mechanical Engineering in 2023, both from Karlsruhe Institute of Technology (KIT), Germany. He is currently working as a Research Associate at the Institute of Mobile Machines (Mobima) at the Karlsruhe Institute of Technology (KIT). His research focuses on the development of a hydraulic replacement base for demolition hammers and simulation of hydraulic systems.



Sascha Hasenöhrl received his B. Sc. in Mechatronics in 2020 and his M. Sc. in Mechatronics in 2023, both from Karlsruhe Institute of Technology (KIT), Germany. He is currently working as a Research Associate at the Institute of Product Engineering (IPEK) at the Karlsruhe Institute of Technology (KIT). His research focuses on the development of a hydraulic replacement base for demolition hammers.



Prof. Dr.-Ing. Sven Matthiesen is head of the Institute for Product Engineering (IPEK), Karlsruhe Institute of Technology (KIT), Chair of Device Design and Machine Elements. He is a peer reviewer for the DFG, a member of several scientific societies, former chairman of WiGeP - Scientific Society for Product Development, author of over 200 scientific publications, and inventor of 18 patents from his time at HILTI Corporation and his current research on power tool received the Engineer (Dipl.-Ing.) degree in mechanical engineering and the Ph.D. degree from University Karlsruhe (TH) respectively. He was with HILTI, Liechtenstein, from 2003 to 2010, as a Design Engineer for power tools and the Head of development for screw technology. He has a professorship for power tools and machine elements, since 2010. His research interests include design methods to support engineers, power tools, and human–machine interaction.



Prof. Dr.-Ing. Marcus Geimer received the Diploma degree in mechanical engineering from the RWTH Aachen University, Germany, in 1990, and the Ph.D. degree from the Institute of Hydraulics and Pneumatics, today named the Institute for Fluid Power Drives and Systems, RWTH Aachen University. He started his industrial career in 1995 in the field of construction at the company Krupp Berco Bautechnik GmbH, Germany, where he was the Leader of the research group for hydraulic breakers. In 2000, he was with Bucher Hydraulics GmbH, Germany, where he led the construction and customer development for mobile hydraulics. Since 2005, he has been a Full Professor and the Director of the Institute of Mobile Machines (Mobima), Karlsruhe Institute of Technology (KIT), Germany. His research and teaching focuses on automation systems, drive concepts, hydraulics and simulation. In addition, he is Faculty Director of Mechatronics.

Hydraulic Artificial Muscles for Exoskeleton Actuation Solutions

Mathias Niebergall, Jonas Pfister, Roman Strobel, Jonathan Scheible

Mathias.Niebergall@thu.de, *Jonas.Pfister@thu.de* , *Roman.Strobel@thu.de* , *Jonathan.Scheible@thu.de*

Abstract

Exoskeletons are designed to enhance the human physical capabilities and to reduce physical strain for various applications in terms of e. g. medical rehabilitation, every day assistance or physical working performance. The actuation of active exoskeletons by means of hydraulic artificial muscles (McKibben principle) are convincing solutions due to their light weight, compactness, exceptional power capacity, remarkable resiliency, and low investment costs. [1] The artificial muscle presented comprises an inner elastomeric hose surrounded by a textile, braided and reinforced by aramid fibers. The muscle is activated by hydraulic fluid supply that cause a radial expansion of the inner pressurized hose and a corresponding axial contraction of the muscle as actuation effect [2], [3], [4], [5], [6], [7], [8]. This paper presents a mathematical-physical model of the hydraulic artificial muscle based on conservation of mass and conservation of energy. Prototypes of hydraulic artificial muscles have been developed for selected drive functions and experimentally tested.
Based on [9] a hydraulic muscle with a length (without fittings) of 200 mm and with an inner diameter (elastomer tube) of 6 mm and an outer diameter of ca. 10 mm (elastomer tube + fiber fabric) has been experimentally evaluated. Relating to the initial length of the hydraulic muscle a contraction of up to 32 % is reachable. With a pressure level of 10 MPa an advanced force density per mass of 110 kN/kg of the muscle is now accessible.
Due to its flexible structure, a hydraulic muscle is much more compliant and adaptable than a rigid hydraulic cylinder. This makes the muscle design ideal for wearable actuation systems. Additionally, the muscle is hermetically sealed, due to its closed design.
An exoskeleton test setup has been developed for experimental investigations. The paper finally discus hydraulic muscles for exoskeleton drive functions.

Keywords Hydraulic Artificial Muscle, Mathematical-Physical Modeling, High Pressure Actuation, Exoskeleton Drive

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

New innovative actuator systems for exoskeletons with hydraulic artificial muscles are very attractive solutions due to their remarkable compactness, lightweight, robustness, and low investment costs. Various prototypes of hydraulic artificial muscles have been developed and built. A sectional view of a typical prototype is shown in Figure 1.1. The muscle has an inner elastomeric hose, surrounded by an aramid fiber textile sleeve, at both ends equipped with a typical hydraulic hose fitting. The fitting composed of a swage nipple and a swage ferrule, as well as a screw coupling with an O-ring sealing element.

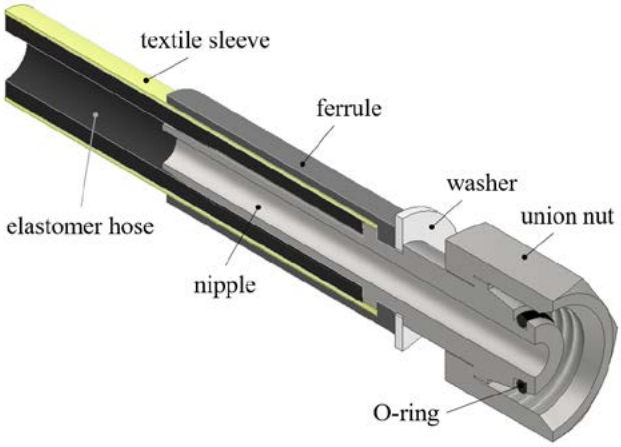


Figure 1.1 Sectional view of the hydraulic artificial muscle [9]

This publication focuses on the following topics:

- introduction of a quasi-static theoretical model of the hydraulic muscle (chapter 2)
- description of the experimental test set up (section 3.1)
- experimental investigations on hydraulic artificial muscles with three muscle length and two muscle diameters for the verification of the model equations (section 3.2)
- experimental high-pressure test (10 MPa) to determine the high force density of a hydraulic artificial muscle (chapter 4)
- an exoskeleton application solution for hydraulic artificial muscles (chapter 5)

The contraction of the hydraulic artificial muscle is illustrated schematically in Figure 1.2. Fluid supply activates the initially non-pressurized muscle with the radius R_0, the initial length L_0 and the fiber angle α_0 (compare Figure 1.3). The pressurized hose surface leads to radial hydraulic forces with increasing circumferential stress in the surrounding textile reinforcement sleeve. With radial displacement of the inner pressurized bladder over the

whole length of the muscle, the fiber angle and the radius increases up to α_1 and R_1 respectively. The corresponding axial contraction of the hydraulic artificial muscle to the length L_1 is the actuation effect.

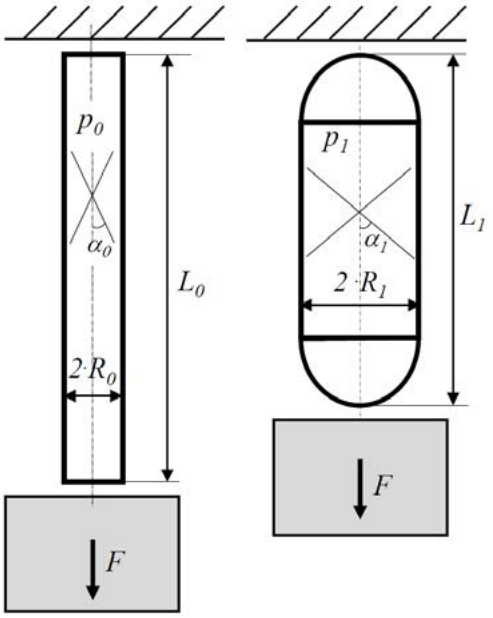


Figure 1.2 Sketch of the simplified non-pressurized muscle (left) and the pressurized activated muscle (right)

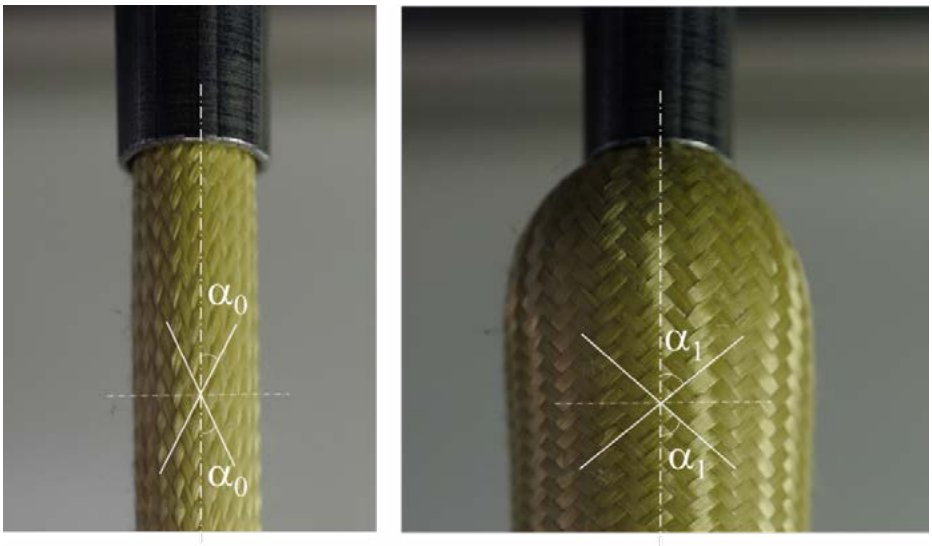


Figure 1.3 Fiber angle of a non-pressurized muscle α_0 and of a pressurized muscle α_1 [9]

2. STATIC MODEL OF HYDRAULIC ARTIFICIAL MUSCLES

In the last decades fluidic (pneumatic or hydraulic) artificial muscles have been investigated and discussed in several studies (e. g. Schulte (1961) [2], Chou, Hannaford (1996) [3], Tondu and Lopez (2000) [4], Tiwari, Meller et al. (2012) [5], Pan et al. (2020) [6], Mori, Suzumori (2009) [7], Zhang et al. (2022) [8]).

The physical approaches for fluidic artificial muscles are typically based on energy conservation, virtual work principles, and on a simplified cylindrical geometry. In general, the fluidic work is done by increasing the volume of the muscle, and leads finally to the mechanical work, done by the muscle as it contracts and changes length. [5]

For the below-described mathematical-physical modeling of the hydraulic artificial muscle, following assumptions are taken into account:

- quasi-static consideration of the actuation process
- compressible muscle including elastomer hose, textile sleeve and fluid
- homogeneously distributed volume flow q
- constant change in constant homogeneously distributed density of the fluid ρ
- cylindrical tube shape (the volume around the end cap sphere (compare Figures 1.2 and 1.3) will be later subtracted)

Initially an artificial muscle approach to mass conservation has been derived and subsequently the relationship to energy conservation is discussed.

Conservation of mass.

The continuity equation (2.1) has here the following components:

- the mass flow supplied to the hydraulic muscle (inflow),
- the mass flow of displacement of the muscle (deformation), and
- the mass flow of elasticity behaviour of the muscle (compression),

$$\underbrace{\rho \cdot q}_{inflow} + \underbrace{\rho \cdot \frac{d}{dt} \cdot \int_V dV}_{deformation} + \underbrace{\rho \cdot \frac{1}{\kappa} \cdot \dot{p} \cdot \int_V dV}_{compression} = 0 \tag{2.1}$$

with the mass density of the fluid ρ, the flow rate into the muscle q, the change of the muscle volume dV, the compression number of the muscle κ, and the change of pressure dp/dt. The compression number κ indicates the pressure required for the elastic muscle deformation.

The balance of volume flow, in case of equal density of the fluid, is shown in equation (2.2), which takes into account the inflow rates (+) and outflow rates (-) via the signs.

$$q - \frac{d}{dt} \cdot \int_V dV - \frac{1}{\kappa} \cdot \dot{p} \cdot \int_V dV = 0 \tag{2.2}$$

Integration over time from the initial state at t_0 to the final contraction state at t_1 yield to the equation of the volumes (2.3).

$$\int_{t_0}^{t_1} q\ dt - \int_V dV - \frac{p}{\kappa} \cdot \int_V dV = 0 \tag{2.3}$$

The change of the muscle volume dV can be described by infinitesimal steps of the two main influencing variables: the radius of the muscle R and the length of the muscle L (compare Figure 1.2 and equation (2.4)).

$$dV = \left(\frac{\partial V}{\partial R}\right)_{L_0} \cdot dR + \left(\frac{\partial V}{\partial L}\right)_{R_1} \cdot dL \tag{2.4}$$

The partial derivatives of (2.5a) are (2.5b) and (2.5c).

$$V = R^2 \cdot \pi \cdot L \tag{2.5a}$$

$$\left(\frac{\partial V}{\partial R}\right) = 2 \cdot R \cdot \pi \cdot L \quad \text{(2.5b)}, \qquad \left(\frac{\partial V}{\partial L}\right) = R^2 \cdot \pi \tag{2.5c}$$

Integration of (2.4) leads to the deformation relationship (2.6) with respect to the state limits (R_1, L_1) of the actuated muscle (compare Figure 1.2).

$$V = \int dV = \int_{R_0}^{R_1} \left(\frac{\partial V}{\partial R}\right)_{L_0} \cdot dR + \int_{L_0}^{L_1} \left(\frac{\partial V}{\partial L}\right)_{R_1} \cdot dL \tag{2.6}$$

The solution is the volume of muscle deformation (2.7). Here the term V_1 represents the radial expanded and contracted volume, and the term V_0 the initial volume of the muscle.

$$V = \underbrace{R_1^2 \cdot \pi \cdot L_1}_{V_1} - \underbrace{R_0^2 \cdot \pi \cdot L_0}_{V_0} \tag{2.7}$$

Based on (2.3) and (2.7) equation (2.8) describes the proportions of the change in volume of the artificial muscle.

$$\int_{t_0}^{t_1} q\ dt - \left(R_1^2 \cdot \pi \cdot L_1 - R_0^2 \cdot \pi \cdot L_0\right) - \chi \cdot \left(R_1^2 \cdot \pi \cdot L_1 - R_0^2 \cdot \pi \cdot L_0\right) = 0 \tag{2.8}$$

with the compression factor $\chi = p / \kappa$ as dimensionless parameter that describes the average compliance during the muscle actuation.

Conservation of energy.

The energy conservation of hydraulic artificial muscles (2.9) follows from the inclusion of pressure in equation (2.8).

$$\underbrace{\int_{t_0}^{t_1} q \cdot p \, dt}_{supply} - \underbrace{\left(R_1^2 \cdot \pi \cdot L_1 - R_0^2 \cdot \pi \cdot L_0\right) \cdot p}_{deformation\,/\,actuation} - \underbrace{\chi \cdot \left(R_1^2 \cdot \pi \cdot L_1 - R_0^2 \cdot \pi \cdot L_0\right) \cdot p}_{elastic\ deformation} = 0 \quad (2.9)$$

In the literature (e. g. [2], [3], [4], [5], [6], [7], [8]), deformation energy is equated with the mechanical work of displacement in the form (2.10a), that lead here to equation (2.10b).

$$-\int_{L_0}^{L_1} F \, dL = \int_{V_0}^{V_1} p \, dV \quad (2.10a), \qquad -\int_{L_0}^{L_1} F \, dL = p \cdot (R_1^2 \cdot \pi \cdot L_1 - R_0^2 \cdot \pi \cdot L_0) \quad (2.10b)$$

The second term of the displacement work in equation (2.11) replaces the muscle deformation term in (2.9). Thus, the conservation of energy approach has the following form:

$$\int_{t_0}^{t_1} q \cdot p \, dt - \int_{L_1}^{L_0} F \, dL - \chi \cdot \left(R_1^2 \cdot \pi \cdot L_1 - R_0^2 \cdot \pi \cdot L_0\right) \cdot p = 0 \quad (2.11)$$

Due to the rounding of the end caps after the muscle contraction (compare Figure 1.2), equation (2.11) can be corrected by taking the V_R volume (2.12) into account.

$$V_R = \underbrace{R_1^2 \cdot \pi \cdot 2 \cdot R_1}_{cylinder} - \underbrace{\frac{4}{3} \cdot R_1^3 \cdot \pi}_{sphere} \quad (2.12)$$

Then the conservation of energy approach for hydraulic artificial muscles has finally the following form (2.13):

$$\underbrace{\int_{t_0}^{t_1} q \cdot p \, dt}_{supply\,(E_s)} - \underbrace{\int_{L_1}^{L_0} F \, dL}_{actuation\,(W)} - \underbrace{\chi \cdot \left(R_1^2 \cdot \pi \cdot L_1 - V_R - R_0^2 \cdot \pi \cdot L_0\right) \cdot p}_{elastic\ deformation\,(E_e)} = 0 \quad (2.13)$$

The energy conservation approach (2.13) contains the hydraulic energy supply E_s to the muscle, the mechanical work W done and the energy share of the elastic compression / energy storage (and where appropriate energy losses) E_e of the muscle. For the verification of the energy balance (2.13) the mechanical work W, the energy supplied E_s and the Energy of the elastic compression E_e are experimentally determined and presented in the following chapter 3.

The compression factor $\chi = p_m / \kappa$ is a dimensionless parameter that describes the average

compliance during muscle deformation, by means of

- the average measured pressure p_m during the muscle contraction, and
- the compression number κ, that indicates the average pressure required for muscle contraction, dependent on the muscle geometry.

Here the compression factor χ was determined using equation (2.13), rearranged according to the compression factor when all other data were available as measured data (compare section 3.2). Due to the experimental determination of χ, it also includes the energy losses (in case of relevance) during the contraction of the artificial muscle.

3. EXPERIMENTAL ANALYSIS OF HYDRAULIC ARTIFICIAL MUSCLES

The experimental investigations of the artificial muscles and the verification of the model equations (2.13) have been carried out with an appropriate test setup. The experimental equipment is presented in section 3.1 and the test results are discussed in section 3.2.

3.1. Description of the experimental equipment

The test setup for hydraulic artificial muscles is shown schematically in Figure 3.1 and the test chamber of the test setup is shown in Figure 3.2 as photography.
The hydraulic artificial muscle (1) is suspended in the test chamber and loaded with a defined weight (8). The muscle is activated by a volume flow of a pressure fluid (mineral oil) from a hydraulic power unit (2). The muscle contraction is controlled by means of a 4/3-way valve.

The functional tests of the artificial muscle are monitored by measurement technology using the following components:

- pressure sensors (5) for monitoring the pressure load of the muscle,
- flow rate sensor (3) for measuring the volume flow into the muscle,
- force sensor (4) for monitoring the quasi-static load,
- laser distance sensor (7) for detecting the muscle contraction,
- displacement sensors (6) for the measurement of the muscle diameter,
- camera (9) for visual recording the changing fiber angle of the textile shell, and
- amplifier (10) for recording and logging the measurement data.

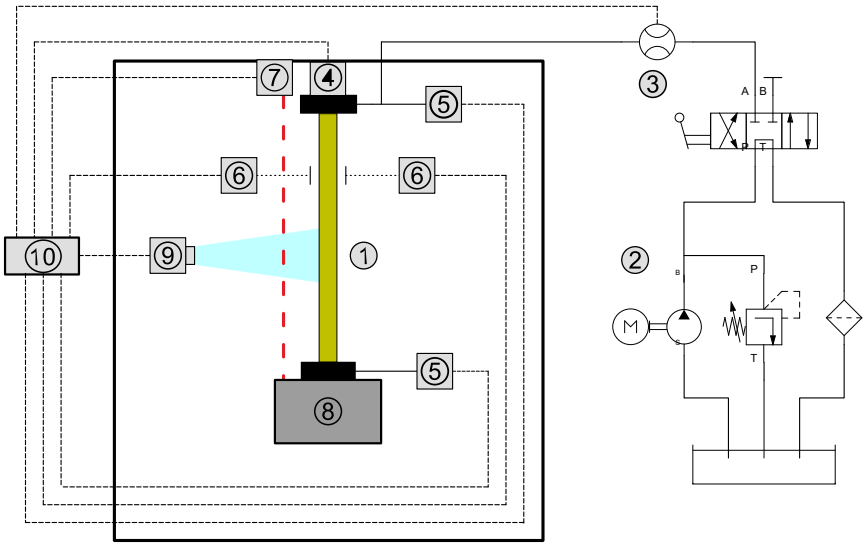


Figure 3.1 Schematic representation of the test setup with hydraulic test environment (dashed lines denote electrical signal lines)

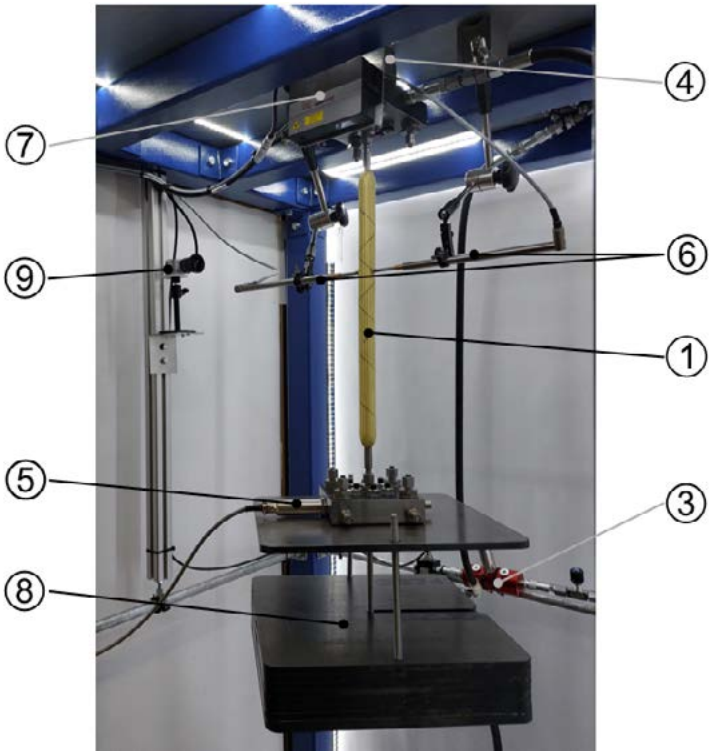


Figure 3.2 Photo of the test setup for experimental investigations of hydraulic artificial muscles (here exemplarily a pressurized muscle (1) with a length of 0.6 m and 100 kg load)

3.2. Experimental investigations of the hydraulic artificial muscle

Non-pressurized hydraulic artificial muscles with three different initial lengths L_0 of 0.3 m, 0.45 m and 0.6 m and with two different initial outer diameters D_0 of 8 mm and 10 mm have been configured for muscle tests. The artificial muscles have been activated by a volume flow q with an increase of the corresponding pressure p up to 5 MPa. The actuation effect is the contraction of the artificial muscles caused by a quasi-static load of 310.8 N for the D08-muscle (D_0 = 8 mm, α_0 = 24°) or 993.6 N for the D10-muscle (D_0 = 10 mm, α_0 = 17°) respectively. The pressurized artificial muscles have an expansion of the outer diameter to D_1 (α_1 = 44° for the D08-muscle, α_1 = 48° for the D10-muscle) and they are shortened to the respective length L_1. Relating to the initial length L_0 of the non-pressurized muscles, and the measured length changes (L_0 - L_1), the contractions ε can be calculated with equation (3.1).

$$\varepsilon = \frac{L_0 - L_1}{L_0} \cdot 100\% \tag{3.1}$$

The experimental initial data (L_0, $D_0 = 2\,R_0$ and F), the measured test results in the contracted state of the muscle (L_1, R_1), and the calculated test results from measured data (ε, κ, χ, E_s, W, E_e) are summarized in Table 3.1. The contraction ε of the artificial muscle show a dependence on the muscle diameter and on the load force. The compression number κ for the contraction of the artificial muscle show a dependence to the muscle length.

The associated measurement results are shown exemplary for the D10-muscles with their three initial lengths (0.3 / 0.45 / 0.6 m) in Figure 3.3. The calculated energy shares, based on equation (2.13), associated to the measurement results, are shown in Figure 3.4. The energy supplied for muscle contraction E_s corresponds to the sum of mechanical work W and elastic energy E_e, due to the compression number κ (or factor χ).

Table 3.1 Measurement data and calculation results of the initial state ($_0$) and the contracted state ($_1$) of the tested hydraulic artificial muscles

	D08 - muscle			D10 - muscle			unit
initial muscle length L_0	0.3	0.45	0.6	0.3	0.45	0.6	m
initial outer diameter D_0	8	8	8	10	10	10	mm
load force F	310.8	310.8	310.8	993.6	993.6	993.6	N
muscle length L_1 (5 MPa)	0.2405	0.3582	0.4804	0.2211	0.3280	0.4443	m
inner radius R_1 (5 MPa)	5.1	5.3	5.2	11.6	11.8	11.6	mm
contraction ε (5 MPa)	20.9	20.9	21.3	27.0	27.6	26.8	%
compression number κ	0.80	1.44	2.09	4.20	4.84	5.30	MPa
compression factor χ	1.56	1.00	0.78	0.342	0.307	0.301	-
supplied energy E_s	53.59	70.33	86.34	121.21	182.52	241.48	J
mechanical Work W	19.75	29.47	40.29	81.40	124.23	161.66	J
elastic energy E_e	33.84	40.85	46.05	39.82	58.29	79.82	J

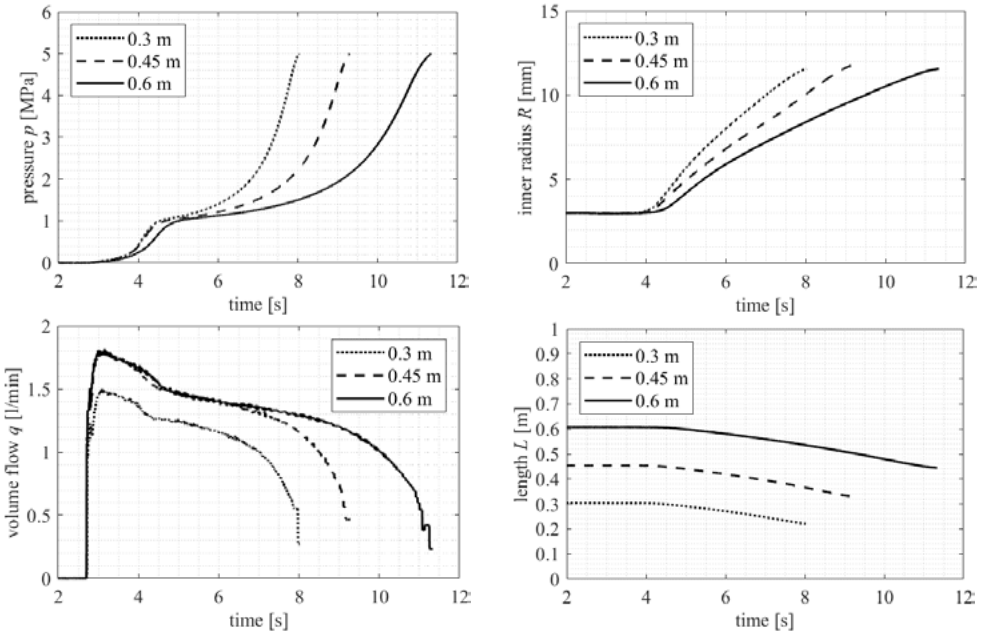


Figure 3.3 Measured test results of pressure p, flow rate q, muscle diameter D and muscle length L for the hydraulic artificial D10-muscles of the initial lengths L_0 = 0.3 / 0.45 / 0.6 m

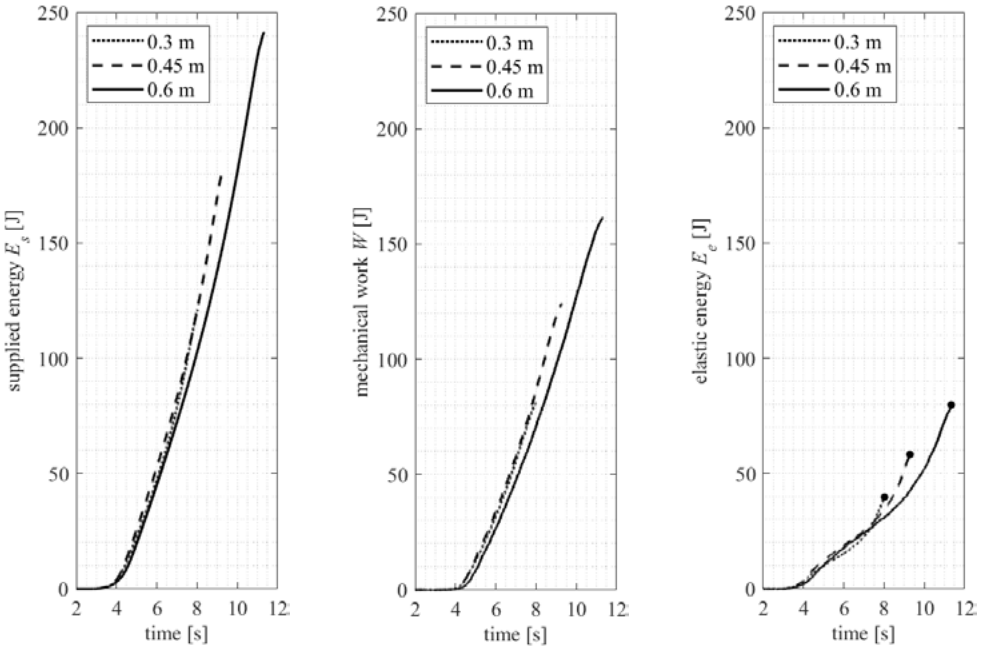


Figure 3.4 Energy conservation shares (compare equation (2.13)) for the hydraulic artificial D10-muscles of different length based on the measured data of Figure 3.3

4. ARTIFICIAL MUSCLES WITH HIGH FORCE DENSITY

For the investigation of the maximum force performance as well as the force density of the artificial muscle a hydraulic test setup has been developed, that is shown in Figure 4.1 with the measuring equipment. Here the muscle has to be clamped within the test setup to prevent its axial contraction during the recording of the blocked force. The muscle has an initial length of L_0 = 200 mm, a diameter D_0 = 10 mm, and a mass including fittings and filled with fluid of m = 0,120 kg. The measured force characteristic curve dependent on pressure of the blocked muscle is shown in Figure 4.2. Because of the clamping the blocked force rises with increasing pressure up to a high pressure of 10 MPa and a force level of F_t = 14200 N. Calculated with (4.1), the force density per mass of the muscle is δ_F = 110 kN/kg.

$$\delta_F = \frac{F_t}{m} \tag{4.1}$$

The improvement in power density is based on a significantly higher pressure level (10 MPa) compared to 5.6 MPa in [9] due to improved form-fit connection between muscle and fitting.

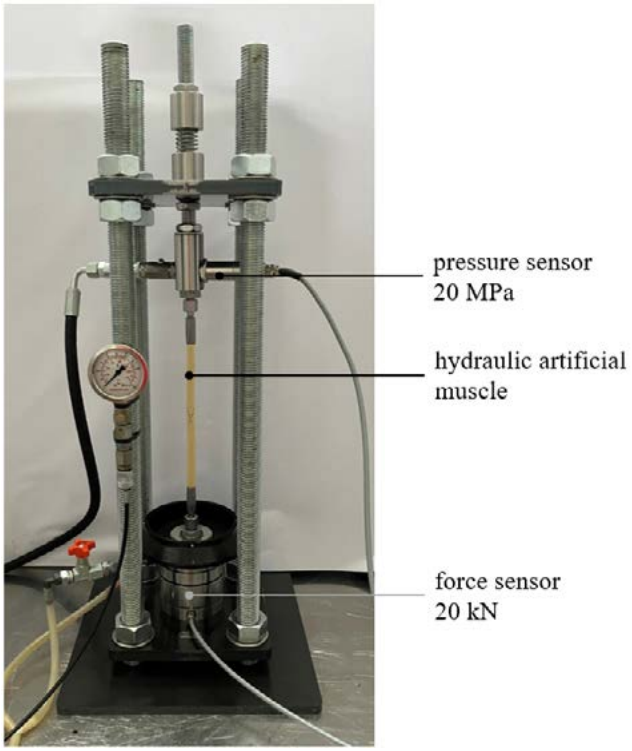


Figure 4.1 Test setup for determining the force density of hydraulic artificial muscles

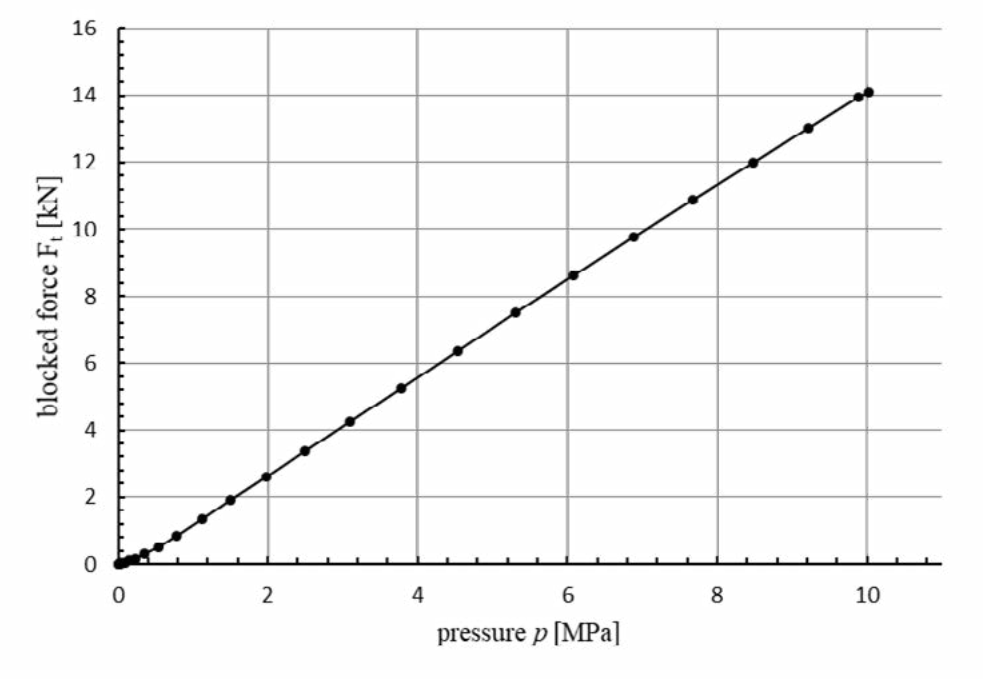


Figure 4.2 Measured blocked force of the clamped muscle in Figure 4.1

5. HYDRAULIC EXOSKELETON APPLICATIONS

Hydraulic artificial muscles are much more compliant and adaptable, due to their high force density advantages as well as because of their flexible textile braided structure compared to rigid electromechanical drives or conventional hydraulic cylinders. The muscles are hermetically sealed, due to the closed design using elastomeric hoses. For these reasons, the hydraulic artificial muscles are ideally suited for wearable actuation systems. For experimental investigations, an exoskeleton test setup has been built that enables the testing of selected exoskeleton drive functions as shown in Figure 5.1.

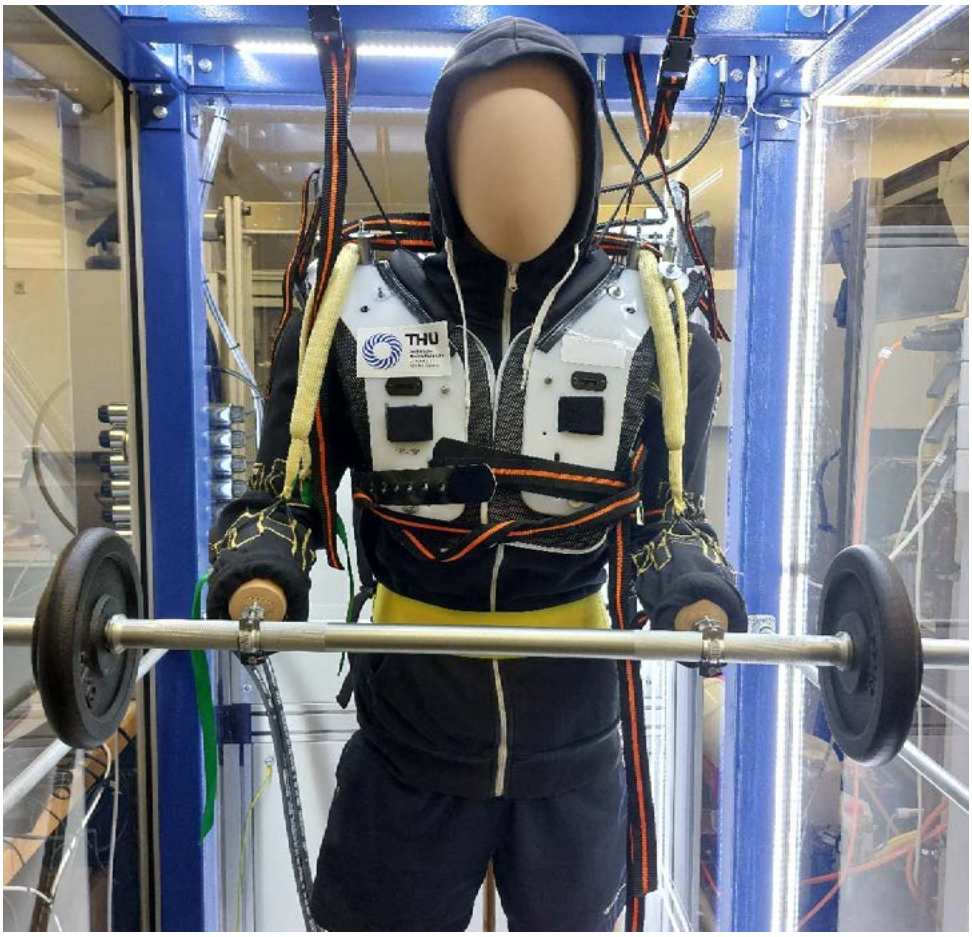


Figure 5.1 Exoskeleton test setup for power support for the upper body (here 20 kg)

6. CONCLUSION

The actuation of active exoskeletons by means of hydraulic artificial muscles are convincing solutions, due to their light weight, compactness, exceptional power capacity, remarkable resiliency, and low investment costs.

A mathematical-physical model of hydraulic artificial muscles is initially presented, based on an approach of conservation of mass and conservation of energy (chapter 2). For the test of the model equations, hydraulic artificial muscles with three different initial length and two initial diameters have been prepared and experimentally investigated. The experimental results confirm the model equations, taking into account adjusted compression factors that describe the elasticity or the energy storage capacity of the muscle (chapter 3).

With a further experiment the maximum blocked force of a hydraulic artificial muscle (length 200 mm, diameter 10 mm) under high pressure requirements of 10 MPa has been determined. It results in a measured blocked force of 14200 N with a very competitive force density per mass of δ_F = 110 kN/kg (muscle mass m = 0,120 kg) (chapter 4).

An exoskeleton test setup has been built for experimental investigations that enables the testing for selected exoskeleton drive functions. These investigations demonstrate that hydraulic artificial muscles are much more compliant and adaptable, due to their high force density advantages as well as due to their flexible textile braided structure compared to rigid electromechanical solutions or conventional hydraulic cylinders. The closed muscle design using elastomeric hoses enables a hermetically sealing.

In summary, the hydraulic artificial muscles are ideally suited for wearable exoskeleton actuation systems (chapter 5).

Acknowledgement

The authors would like to thank the foundation council of the Hydraulics Center of Excellence of the Ulm University of Applied Sciences for the friendly support of the project.

7. REFERENCES

[1] D. M. G. Preethichandra, L. Piyathilaka et al., Passive and Active Exoskeleton Solutions: Sensors, Actuators, Applications, and Recent Trends, in: Sensors 2024, 24, 7095, 2024

[2] H.F. Schulte, The characteristics of the McKibben artificial muscle, The Application of External Power in Prosthetics and Orthotics, Washington, DC: Nat. Acad. Sci.-Nat. Res. Council, 1961

[3] C.-P. Chou, B. Hannaford, Measurement and Modeling of McKibben Pneumatic Artificial Muscles, IEEE Transactions on Robotics and Automation, Vol. 12, No. 1, 1996

[4] B. Tondu, P. Lopez, Modeling and Control of McKibben Artificial Muscle Robot Actuators, IEEE Control Systems Magazine, 20 (2): 15-38, 2000

[5] R. Tiwari, M. A. Meller, Hydraulic Artificial Muscles, Journal of Intelligent Material Systems and Structures, 23(3) 301-312, 2012

[6] M. Pan, Zhe Hao, Chenggang Yuan, Andrew Plummer, Development and Control of Smart Pneumatic McKibben Muscles for Soft Robots, in: 12th International Fluid Power Conference, 2020

[7] M. Mori, K. Suzumori, S. Seita, M. Takahashi, T. Hooya, K. Kusumoto, Development of Very High Force Hydraulic McKibben Artificial Muscle and Its Application to Shape-Adaptable Power Hand, Proceedings of the IEEE International Conference on Robotics and Biomimetics, 2009

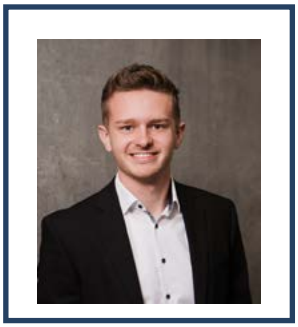
[8] Z. Zhang, J. Che, P. Liu, Y. Jia, Y. Gong, A Modified Physical Model of High-Strength Water Hydraulic Artificial Muscles Considering the Effects of Geometry and Material Properties, Journal of Mechanical Engineering Science, Vol. 236 (2), 2022

[9] M. Niebergall, R. Strobel, Development of a hydraulic artificial muscle with high force density, in: 14th International Fluid Power Conference, Dresden 2024, pp. 94 – 103, 2024

Biographies



Mathias Niebergall received the master's degree in 1990 and his doctorate in mechanical Engineering in 1996 from the University of Kassel. After working in industry with Bosch and Bosch Rexroth, he is since 2014 Professor at the Ulm University of Applied Science (THU) at the Department of Mechanical and Automotive Engineering. His research focuses on hydraulic drive systems.



Jonas Pfister holds a bachelor's degree in automotive Engineering and is currently studying for a master's degree in Systems Engineering at Ulm University of Applied Science. In addition to his studies, he works at the Hydraulics Center of Excellence at the THU.



Roman Strobel received the master's degree in Systems Engineering at Ulm University of Applied Science. In addition to his employment in industry, he works at the Hydraulics Center of Excellence at the THU.

Jonathan Scheible holds a bachelor's degree in automotive engineering. In addition to his employment in industry, he works at the Hydraulics Center of Excellence at the THU.

New hydrogen fuel lines from carbon steel tubes and their qualification for high-pressure hydrogen applications

Christian Lessmeier, Christian Möller, Lars Danhäuser, Helwig Brabander

BENTELER Steel/Tube GmbH & Co. KG, Paderborn

christian.lessmeier@benteler.com

Abstract

Recent advancements in steel tube technology have led to the development of carbon steel tubes specifically designed for hydrogen fuel lines in mobile machinery or in automotive applications. These developments address the material and safety challenges associated with hydrogen as an energy carrier on the pathway to climate neutrality. Especially the development and proof of operational safety for carbon steel tubes made from new materials for high-pressure hydrogen is challenging.

This paper provides insights into qualification process of the BENTELER HYRESIST® Mobility tubes for high pressure hydrogen applications. These fluid line tubes developed by BENTELER Steel/Tube from a ferritic-pearlitic carbon steel are an economic alternative to austenitic stainless-steel tubes, which are the current state of the art for hydrogen fluid lines. The main content of this paper will be the properties of the new steel grade, the hydrogen testing and its results under high-pressure hydrogen conditions of this steel in general and as a hydrogen fluid line, as well in the context of the current hydrogen fluid lines regulations.

The general hydrogen compatibility of the new steel material was proven by a burst disk test with pressurized hydrogen. Additionally, tubes made from this steel underwent a testing procedure where the tube is subjected to a cyclic loading under high-pressure hydrogen at different temperatures, followed by a hydraulic burst test to proof lifetime safety. These dynamic investigations were carried out using processed tubes in combination with fittings as a prototype hydrogen fluid line. The methodology successfully demonstrated the general hydrogen compatibility of the BENTELER HYRESIST® Mobility tubes. Finally, the paper will conclude with an evaluation of this innovative solution for high pressure hydrogen lines in relation to the current state of the art.

Keywords. BENTELER HYRESIST® Mobility, hydrogen fluid lines, seamless cold-drawn precision tubes, carbon steel, high-pressure hydrogen, burst disk test, pressure cycling test.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

restricted

1. INTRODUCTION

The practical evaluation and the prediction of hydrogen embrittlement in steels are despite numerous scientific studies a persistent problem. There is no generally accepted load-independent material criterion for the hydrogen sensitivity of steels under pressurised hydrogen. The reduction of area or elongation at fracture in the quasi-static tensile test is often used as a technological comparison value. It is often neglected that actual loads of real components under hydrogen typically occur by multiaxial loading and under dynamical load conditions. Occasionally, notched specimens are used, although the test result is critically dependent on the notch sharpness.

Fracture mechanics tests, such as determining the crack propagation rate as a function of the stress intensity factor, are effective methods for the investigation of hydrogen embrittlement resistance. However, these tests are not suitable for small and thin-walled tubes because the extraction of standard CT-samples is not possible because of the specimen size.

For the newly developed high-pressure tubes BENTELER HYRESIST® Mobility by BENTELER Steel/Tube a different testing approach was chosen. The tubes for hydrogen application made from a specially developed ferritic-pearlitic carbon steel are tested by an approach that is presented below. This approach is focusing at 70 MPa (700 bar), which is a pressure stage that is especially needed for automotive or other mobile machinery applications. It uses the burst disk test according to DIN EN ISO 11114-4 and a dynamical cycling internal pressure test, which is linked to a subsequent hydraulic burst-test. Moreover, the experimental method and its results are compared to the current automotive standards.

2. HYDROGEN COMPATIBILITY TESTS USING THE BURST DISC TEST

The evaluation of the hydrogen embrittlement resistance of steels can be carried out using the burst disc pressure test developed by Fidelle et al. [1-5], in accordance with DIN EN ISO 11114-4 [7]. In this test, a disc-shaped specimen is clamped and subjected to continuously increasing gas pressure until bursting or crack initiation occurs (Fig. 2.1). The embrittlement effect of hydrogen (H_2) is quantified by comparing the burst pressures under hydrogen (P_{H2}) and helium (P_{HE}), with helium serving as an inert reference gas. The ratio P_{HE}/P_{H2} is used as a measure of the material's susceptibility to embrittlement. It should be noted that this ratio depends on the pressure increase rate, which must be kept constant during each test.

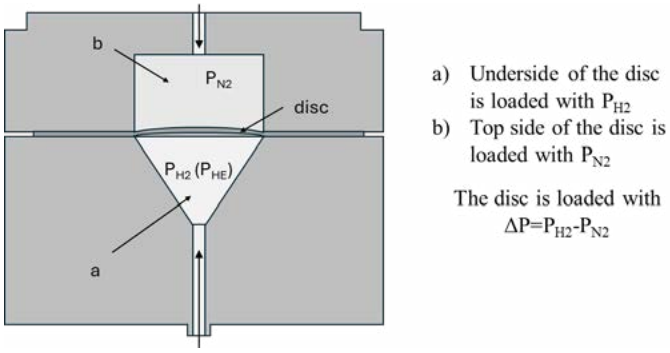


Figure 2.1. Principle of the burst disk testing according to DIN EN ISO 11114-4 [7]

restricted

Based on extensive measurements on steels that have proven successful in practice or have failed under hydrogen exposure, Fidelle identified a burst pressure ratio of $P_{HE}/P_{H2} = 2$ as a critical value, which was adopted by DIN EN ISO 11114-4. The ratio of P_{HE}/P_{H2} is also known as hydrogen embrittlement index (HEI). For materials with values of the HEI < 2, no component failure due to hydrogen embrittlement is expected.

A key advantage of this approach is that the strength and the ductility of the material are reflected in the test result. The loading conditions are particularly favourable for the effect of pressurized hydrogen and resemble those in pressure vessels or hydrogen-carrying fuel lines.

2.1. Experimental boundary conditions

According to DIN EN ISO 11114-4 (method A), the test disc must meet specific properties and dimensions. It must be flat and the surface is produced either by grinding or machining to ensure an appropriate surface finish. The dimensions are as follows: The diameter measures 58 mm, the thickness 0,75 mm and the flatness deviations is less than 1/10 mm with a surface roughness Ra < 1 µm.

If an evaluation of the steel at an operating pressure exceeding 30 MPa is required, a disc thickness greater than 0,75 mm shall be used. Due to the hydrogen pressure of 70 MPa in automotive fuel lines, the test discs had thicknesses ranging from 1,49 to 1,52 mm. Tests were conducted using hydrogen (purity > 99.9995 mol-%) and helium (both gases with H_2O < 3 µl/l), with pressure increase rates evenly distributed between 0,3 and 40 MPa/min. For complete material evaluation, 15 tests (six with helium and nine with hydrogen) as specified in DIN EN ISO 11114-4, method A, are considered sufficient. For disc thicknesses above 0,75 mm, the burst pressure Pr´ must be systematically corrected to account for deviations from the test of a disc with standard wall thickness. The corrected burst pressure Pr´ is determined by the following equation (2.1):

$$Pr' = \frac{Pr \cdot 0{,}75}{e_m}, \text{where } e_m \text{ is the mean disc thickness} \qquad (2.1)$$

Corrected burst pressures are shown as a function of the pressure ramp rate. The pressure ramp rate is calculated by dividing the actual burst pressure by the test time and is expressed in MPa/min. For each hydrogen test, the ratio Pr'_{He}/ Pr'_{H2} is determined, where Pr'_{He} is the theoretical helium burst pressure associated with the same pressure increase rate as the hydrogen test (obtained by regression from the corrected helium burst pressure), and Pr'_{H2} is the corrected hydrogen burst pressure. The ratios Pr'_{He}/ Pr'_{H2} shall be represented in relation to the pressure increase rate. The hydrogen embrittlement index (HEI) of a material is the maximum of these ratios. A material with an index not exceeding 2,0 is considered hydrogen compatible.

2.2. Specimen preparation

Since it is not possible to extract a disc with a diameter of 58 mm directly from a small thin-walled tube, a forged specimen of the same material was first cut out from a continuously cast billet. To eliminate the dendritic cast structure, the material was forged at about 1100 °C into a round bar of 58 mm diameter. Following the cutting of 2 mm thick discs from the round bar, both sides were grinded to achieve thicknesses of 1,49-1,52 mm and fulfill the surface roughness requirements according to DIN EN ISO 11114-4. Through a subsequent

restricted

heat treatment, a ferritic–pearlitic structure was established, comparable to the microstructure of a cold-drawn tube with a small wall (fig. 2.2). For both microstructural conditions, a grain size class of 10 was identified, together with comparable hardness values of approximately 125 HV1, which correspond to an estimated tensile strength of about 400 MPa. Consequently, the burst disc specimens exhibit microstructural characteristics and mechanical properties analogous to those of hydrogen fuel lines manufactured from the same material.

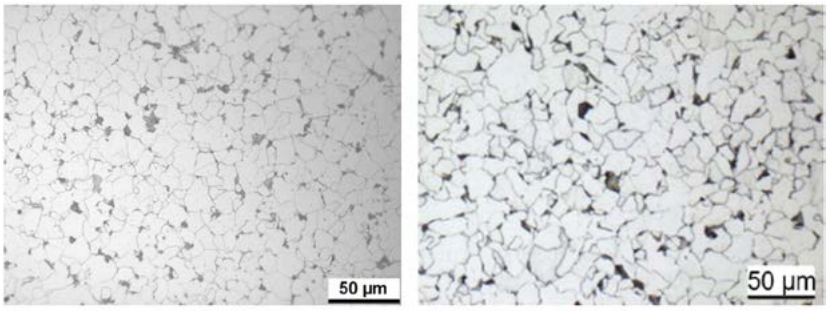


Figure 2.2. Comparison of microstructure between tube (left) and disk (right)

2.3. Results

The burst disc pressure tests were carried out in accordance with DIN EN ISO 11114-4 (Method A). Fig. 2.3 shows the corrected hydrogen and helium burst pressures as a function of pressure rise rate, as well as the resulting burst pressure ratios. All burst pressure ratios are significantly below the critical value of 2,0. According to DIN EN ISO 11114-4 (Method A), the hydrogen embrittlement index (HEI) of a material corresponds to the maximum value of this ratio, which in this case is 1.16. Based on this HEI, the examined tube material can be considered compatible with hydrogen at a pressure of 70 MPa.

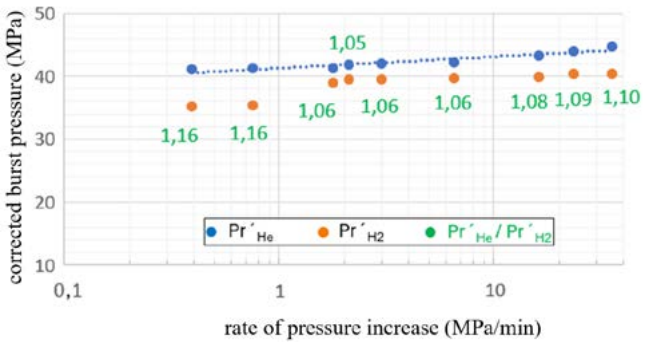


Figure 2.3. Corrected hydrogen and helium burst pressures as a function of pressure rise rate

restricted

3. CYCLIC PRESSURE TESTING OF HIGH-PRESSURE TUBES

To evaluate the potential sensitivity of steels to hydrogen in practical applications, it is essential to characterize their cyclic behaviour. The aim is to prevent the failure of components due to real world conditions, which could otherwise have severe consequences. To guarantee the durability of steel components exposed to hydrogen, fatigue tests under maximum loading conditions and varying temperature conditions are recommended.

The tested components consist of the hydrogen-carrying tube (seamless cold-drawn, outer diameter: 6.35 mm, wall thickness: 1.675 mm) with an outside corrosion protection layer and with tube fittings at both ends. The corrosion protection coating was realized as BENTELER ZISTA®SEAL and BENTELER ZISTAPLEX®. ZISTA®SEAL is a flexible chromiumVI-free zinc coated tube with a thin sealer as a top coat, offering good corrosion protection, even after tube processing. ZISTAPLEX® is a tube with a flexible duplex coating consisting of chromium (VI)-free zinc coating with an additional red coloured organic coating, offering high corrosion protection, even after tube processing. The tube fittings also play a key role in the functional and practical design of hydrogen fluid systems, as they must ensure gas-tightness under high hydrogen pressure and stability against mechanical impacts. Moreover, a 316L stainless steel tube of the same size with identical tube fittings was used as a reference.

3.1. Mechanical and microstructural characterisation of the tube

The mechanical properties of the high-pressure hydrogen tubes made of BENTELER carbon steel are listed in Table 3.1. The seamless cold-drawn and normalized BENTELER HYRESIST® Mobility (BHM) series-tube has a ferritic-pearlitic structure with a grain size class of 10 (see fig. 2.2 above).

Table 3.1. Mechanical characteristics of the high-pressure hydrogen tubes (BHM1) with outer diameter: 6.35 mm and wall thickness: 1.675 mm

	Yield Strength $Rp_{0.2}$ (MPa)	Ultimate Tensile Strength R_m (MPa)	Elongation A (%)
BENTELER HYRESIST® Mobility 1	≥ 250	≥ 370	≥ 35

3.2. Experimental Procedure and Results

For cyclic tests with hydrogen internal pressure, respectively three high-pressure tubes with a length of 100 mm and with fittings on both tube ends (Fig. 3.1) were subjected to cyclic ageing with up to 50,000 pressure cycles at three different temperatures of +20, -40 and 60 °C. In one cycle of 20 to 30 seconds, the hydrogen pressure in the tube was increased from 0 to 87.5 MPa and then decreased back to 0 MPa. The maximum value of 87.5. MPa is equivalent to a nominal working pressure of 70 MPa (700 bar) multiplied by a safety

restricted

factor of 1.25, according to EC79 or ISO 19887. The pressure test gas was hydrogen with a purity of >99.999 mol-% (hydrogen 5.0).

Following this cyclic ageing, hydraulic pressure tests were conducted until failure of the tubes. The determined burst pressure was compared with that of a non-preloaded reference tube to assess the effects of cyclic ageing under a pressurised hydrogen atmosphere. This hydraulic burst pressure tests were carried out at 20 °C and the results are shown in Table 3.2. No significant difference between the burst pressures of the tube samples preloaded with hydrogen and the reference samples (without hydrogen contact) can be observed. This proves the excellent hydrogen resistance of the BHM1 steel tube.

Table 3.2. Results of cyclic internal pressure tests and burst pressure tests

Temperature (°C)	**Pressure cycles**	**Burst pressure (MPa)**
BHM1 – *Reference without hydrogen*		
20	0 (reference)	330,0
BHM1 – cyclic loading with hydrogen		
20	10.000	339,7
-40	10.000	332,5
-60	50.000	329,3
Stainless steel (1.4404) – reference without hydrogen		
20	0 (reference)	350,0
Stainless steel (1.4404) - cyclic loading with hydrogen		
20	10.000	338,5
-40	10.000	343,3

Having a closer look at the results, it is remarkable that the BHM1 tube has a slight increase in burst pressure under the influence of hydrogen at 20 and -40 °C. At -60°C, the burst pressure is reduced by only 0.2%. While the burst pressures of the austenitic stainless-steel tube decrease by 3.3% under the same conditions (20 and -40°C, no testing at -60°C). Summarized, the BHM1 high-pressure tube even shows a very slightly higher resistance to high-pressure hydrogen embrittlement than the stainless-steel tube made of 316L (1.4404).

restricted

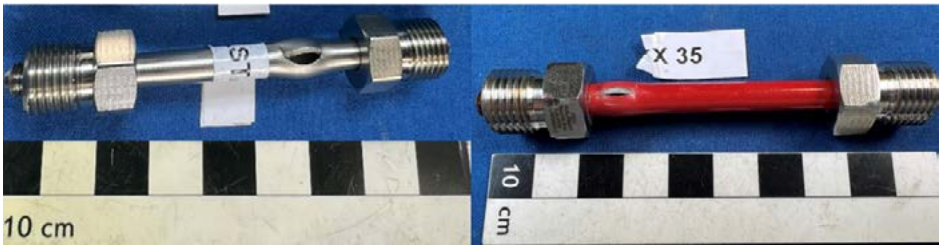


Figure 3.1. Test components after the hydraulic burst pressure test, 316L stainless steel (left) and BHM1-tube (right), each with tube fittings on both ends

4. TESTING APPROACHES AND STANDARDS

To provide a comprehensive context for the present work, the following paragraphs introduce an alternative testing approach which is under development for small sized tubes (4.1). Furthermore, the current standards for automotive applications are summarized (4.2).

4.1. Perspective on hydrogen service life testing of tubes and tube materials

As an alternative to the component testing described above, research is currently working on further efficient qualification procedures to assess the service life of tubes made of new materials under the influence of pressurized hydrogen.

In the project *hydrogen lightweight tube* (funded by the German Federal Ministry for Economic Affairs and Energy (BMWi), funding code 03EI3070A) the use of longitudinally welded lightweight tubes for hydrogen use is qualified. The tubes are manufactured from strip steel, which was developed based on innovative material concepts. Regarding hydrogen qualification, the project is characterised by the development and testing of efficient test methods for hydrogen compatibility, especially for tubes with small outer diameters and thin walls.

Since the influence of hydrogen can lead to a significant reduction in fatigue life in the high-cycle fatigue range [6], an evaluation of cyclically stressed tube will be performed by life-time experiments, commonly referred to as “Wöhler”- tests. The required lifetime of the tube to be qualified in this project is below 10^6 cycles and falls within the time-dependent fatigue region of the “Wöhler”-curve. For this purpose, a test method is being developed with which the necessary material tests (low cycle fatigue) can be carried out quickly and cost-effectively in a specially adapted autoclave under pressurized hydrogen.

A second testing method using the C-ring technique, which enables easy sample production from thin sheets and tubes, is tested for the fatigue strength characterisation of products and materials under pressurized hydrogen. The project includes adapting this solution to the specific tube geometry and extending it for fatigue tests under pressurized hydrogen. This offers an additional quite simple method to characterize the fundamental hydrogen compatibility under pressurized hydrogen loading. This innovation enables targeted, geometry-specific material qualification and, in the future, cost-effective further developments and implementation into simple quality assurance measures.

restricted

4.2. Automotive standards for hydrogen compatibility testing of components

For the North American market, the applicable standard is ANSI HGV 3.1 – "Fuel system components for compressed hydrogen gas powered vehicles". In Europe, until July 2022, EC/EG 79/2009 was applied for the type approval of hydrogen-powered vehicles, together with the associated regulation EU 406/2010, which specified a detailed test sequence for high-pressure lines consisting of four tests. This sequence included an initial corrosion test, a cyclic gas pressure endurance test, a hydraulic pressure cycle test, and a final external leakage test. These tests were widely used and can still be considered state of the art. [8, 9]

The current applicable European regulation for type approval of hydrogen vehicles primarily refers to UN Regulation No. 134 (UN GTR 134), which, however, is vague regarding the test requirements for high-pressure hydrogen fluid lines. The incomplete description of the state of the art in the European regulatory framework has been addressed by the international standard ISO 19887-1: *Gaseous Hydrogen — Fuel system components for hydrogen-fuelled vehicles - Part 1: Land vehicles*. It has been developed based on the ANSI HGV 3.1 and is being published in October 2024 for hydrogen land vehicles. It's translated European derivatives might follow. Hydrogen rail vehicles are excluded in Part 1, but Part 2 of ISO 19887 for rail vehicles is under development. [8,10]

The qualification framework defined in HGV 3.1 and ISO 19887-1, contains a durability assessment of hydrogen fluid line assemblies by a multi-stage verification regime that integrates endurance (continuous operation), overpressure resilience, and post-cycling leak integrity evaluation.

The endurance test reproduces the thermomechanical and hydrogen-assisted degradation processes expected in operational vehicle environments, thereby serving as an accelerated surrogate for long-term service exposure. Most significant characteristic of this continuous operation test is the completion of the endurance cycling up to a maximum pressure of 875 bar of firstly hydrostatic pressure and secondly hydrogen pressure cycles for each 50 000 cycles at different temperatures (-40 °C, room temperature and +85 °C). The maximum pressure is set according to the specific pressure class of the application as defined in the normative pressure classification tables of both standards, starting from H25 (250 bar) up to H70 (700 bar). Maximum test pressure is calculated with a factor of 1.25 to 875 bar for a nominal working pressure (NWP) of 700 bar.

After the endurance test the assembly undergoes a hydrostatic strength test using an overpressure that exceeds the nominal working pressure by at least a factor of 2.5, enabling the detection of subcritical defects, microstructural degradation, and plastic deformation phenomena that may have accumulated during cyclic loading. The final stage comprises a leakage assessment performed at defined pressure conditions to evaluate the integrity of all sealing interfaces, welded joints, and mechanical connections. The compliance with the mentioned acceptance criteria in combination with additional tests like corrosion resistance, automotive fluid exposure and a bending test (for the proof of material ductility) ensures that the fluid line assemblies retain its structural reliability, hydrogen compatibility, and leak-tightness throughout the simulated design service life.

The requirements for high-pressure hydrogen fluid lines of ISO 19887-1 are aligned with ANSI HGV 3.1, except that the material selection for high-pressure components is not limited to non-corrosive austenitic stainless steels. This opens the possibility for the series

restricted

use and qualification of material innovations such as e.g. the hydrogen fluid lines from carbon steel tubes, which is presented in the previous chapters of this article. [8, 10]

5. CONCLUSION

In summary, it can be stated that the presented test series successfully demonstrates the hydrogen compatibility of the high-pressure tubes BENTELER HYRESISIT Mobility. This enables the qualification of components made from these tubes, e.g. as a hydrogen fluid line assembly. According to current regulatory requirements, several standards govern the commercialization of hydrogen vehicles, covering both fuel cell and hydrogen combustion technologies.

In these standards, in addition to demonstrating the basic material compatibility under hydrogen exposure, component tests for fuel lines are required. These tests essentially replicate the expected operating conditions of the components over their service life. Both verifications were provided in a modified form within the experiments presented.

The technological neutrality regarding the material selection for hydrogen fuel lines of ISO 19887-1 enables the possible future use of innovations such as the high-pressure fuel line tubes (BENTELER HYRESIST® Mobility). These tubes are made of carbon steel specifically developed for the application in hydrogen fuel systems, other hydrogen mobile machinery or similar stationary applications. Altogether this innovation is supporting the cost-efficient development of hydrogen-powered vehicles and further hydrogen applications.

6. REFERENCES

[1] J. P. Fidelle, R. Broudeur, C. Roux, Effect of Hydrogen on Behavior of Materials, AIME Conf. Jackson Lake, USA. publ. AIME (1976) 507.
[2] J. P. Fidelle, R. Bernardi, R. Broudeur, C. Roux, M. Rapin, ASTM STP 543 (1974) 221.
[3] J. P. Fidelle, R. Broudeur, C. Pirrovani, C. Roux, ASTM STP 543 (1974) 34.
[4] J. P. Fidelle, L. R. Allemand, J. M. Deloron, C. Roux, MCchanique Electricitt, No. 229 (1969) 1.
[5] Colloque sur “Les Essais de Disque sous Pression”, Ivry, 29. 5. 1975, organist par INSTRUMENTS SA et le Commisariat l'Energie Atomique
[6] Michler, T. et al., Int. J. Fatigue 96 (2017) 67
[7] DIN EN ISO 11114-4:2017 - Ortsbewegliche Gasflaschen - Verträglichkeit von Werkstoffen für Gasflaschen und Ventile mit den in Berührung kommenden Gasen - Teil 4: Prüfverfahren zur Auswahl von Stählen, die gegen Wasserstoffversprödung unempfindlich sind.
[8] ANSI/CS HGV 3.1 - Fuel system components for compressed hydrogen gas powered vehicles.
[9] EC/EG 79/2009 - Verordnung Nr. 79/2009 des Europäischen Parlaments und des Rates über die Typgenehmigung von wasserstoffbetriebenen Kraftfahrzeugen mit der zugehörigen Durchführungsverordnung EU 406/2010.
[10] ISO 19887-1 - Gaseous Hydrogen — Fuel system components for hydrogen-fuelled vehicles - Part 1: Land vehicles.

restricted

Biographies



Dr. Christian Lessmeier received a university degree in Mechanical Engineering from the University of Paderborn in 2010. From 2010 to 2017, he served as a research associate at the same University, contributing to teaching and research in machine elements, drive technology, and mechanical design, including a research stay at the Sino-German Institute of Mechatronics in Qingdao. He completed his PHD (Dr.-Ing.) degree in Mechanical Engineering focusing on Condition Monitoring in 2017. Since then, he has been working as a Project Engineer on several projects and as Project Manager for hydrogen tube technologies at BENTELER Steel/Tube, focusing on advanced engineering and materials development. Recently his focus changed to innovations, and he is currently working as a Technology and Innovation Manager in the department Strategy/Communication & Innovation at BENTELER Steel/Tube GmbH & Co. KG.

restricted

From Turbulence to Tranquillity: A Multidisciplinary Blueprint for Optimising Liquid Cooling in Precision Machinery

Enrico Pasquini[1], Jonas Wack[1], Afshin Sattarifar[1] & Peter Gehr[2]

[1]AVL Deutschland GmbH, Junkers-Ring 6, 85098 Großmehring, enrico.pasquini@avl.com

[2]AVL Software and Functions GmbH, Im Gewerbepark B29, 93059 Regensburg

Abstract

High-power precision optical systems are essential for cutting-edge technology – from laser additive manufacturing and gravitational-wave interferometry to directed-energy drone defence. The encountered intense thermal loads typically require liquid cooling, even though pressure ripples and flow-induced vibrations in the coolant can undermine mechanical precision and degrade optical accuracy.

This paper introduces a multidisciplinary engineering methodology for systematically analysing and mitigating vibration issues in liquid cooling systems for highly mechanically sensitive applications. The approach integrates 1D acoustic simulation, 3D vibro-acoustic analysis, 3D computational fluid dynamics (CFD), 1D/3D thermal simulation, mechanical design, materials science, and experimental validation into a unified workflow. Its performance is demonstrated on a representative cooling circuit which, in its original state, generates unacceptably high oscillating flow forces that impair the optical performance of the overall system.

The first step in the methodology is the development of a system-level, frequency-domain 1D model to determine the baseline oscillating pressure, flow rate and force distribution throughout the cooling circuit and identify the driving mechanism of the disturbances. To cover complex components (e.g., hoses or pumps), surrogate models derived from 3D simulation or from direct experimental measurements are integrated into the 1D framework. Based on the initial results for the forces and the associated system sensitivities, mitigation measures are derived which enable a reduction of the disturbing forces by more than 90 %.

By systematically integrating multiple engineering disciplines into a unified framework, this approach represents a significant advancement over traditional, isolated analysis techniques. The ability to iteratively refine models across multiple domains ensures a robust and optimised solution for mitigating flow-related disturbances. The methodology is particularly valuable for applications requiring ultra-high precision, where even minor perturbations can severely impact system performance.

Keywords: Duct acoustics, precision engineering, condition number, multidomain optimisation, surrogate modelling.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

High-precision, high-power optical machinery is indispensable in the modern world. It enables ultra-accurate medical diagnostics, such as optical coherence tomography for retinal imaging, and is vital to cutting-edge, optics-based manufacturing methods (laser-based manufacturing, additive manufacturing, optical lithography). It underpins major scientific breakthroughs, such as the use of laser interferometers to detect gravitational waves, and plays a critical role in defence as well, for instance by countering drone attacks with directed energy weapons. The high thermal loads commonly encountered in such machinery typically require liquid cooling to prevent degradation in optical performance and structural failure.

However, in high-precision applications, liquid cooling is not free of cost: The presence of a coolant enables *hydraulic communication* even between components of the cooling circuit that are far apart from each other. As a result, mechanical or flow disturbances – for example, originating at the pump – can propagate to sensitive components such as a directly cooled laser beam source and adversely affect their optical performance. The issue is illustrated on the example of a cooling circuit for a high-precision optical manufacturing system.

1.1. Cooling System Under Consideration

In its baseline configuration, the cooling circuit comprises an external gear pump (displacement volume per rotation of 2.85 cm³ and $z = 9$ teeth per gear), two microchannel heat exchangers (see Figure 6 for dimensions), a return tank, interconnected piping and flexible PVC hoses. A CAD model of the duct system is shown in Figure 1:

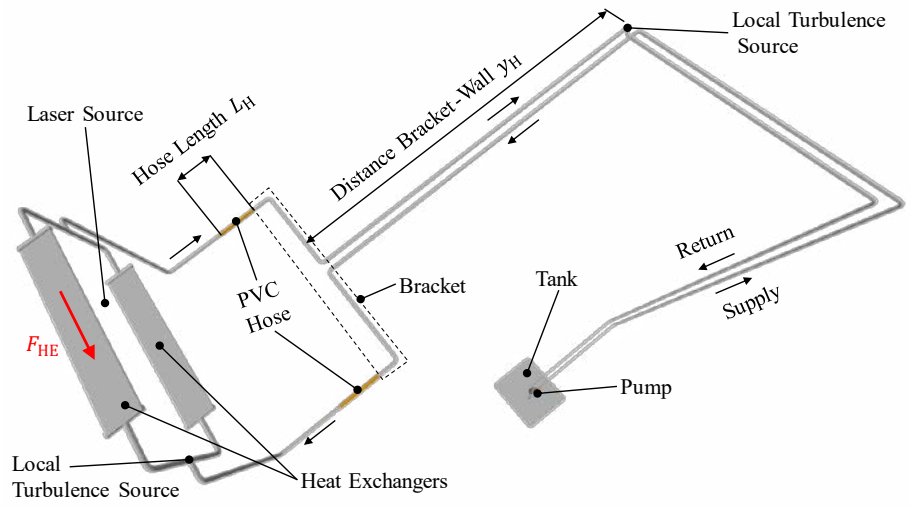


Figure 1: Cooling system under consideration.

The highly mechanically sensitive laser source is directly mounted between both heat exchangers. To be able to adjust the laser beam's orientation during operation, the beam

source – and thus its piping – must be suspended in a movable manner. This is achieved by two hoses within the routing.

1.2. Precision Losses and their Mitigation

During initial commissioning of a prototype of the system, it was observed that the operation of the cooling system was associated with unacceptably high optical precision losses due to unbalanced forces acting on the laser beam source, see Figure 1.

In order to effectively suppress the disturbing oscillatory forces, suitable – preferably passive – mitigation measures are to be taken. Because the mechanical design of the rest of the machine is already comparatively mature, major structural changes on the laser source and its support and actuator system are not possible anymore. Therefore, a direct reduction of the hydraulic excitation is necessary. Due to space constraints, the use of bulky resonators and silencers is not possible, leaving the adjustment of the position and length of the already existing hoses as the only viable option.

Before any mitigation strategy can be designed, the operating modes of the system need to be identified. The duct circuit acts as an oscillatory system which consists of fluid (coolant) and solid (duct walls, valve housing) elements. Both the fluid and the solid elements possess compliance and inertia. Thus, the unsteady forces and moments exerted by the duct system on the laser source result from the response of this oscillatory system to external as well as internal stimuli. In many applications, the inertia and most of the compliances of the solid parts can be neglected (weak fluid-structure interaction), such that the oscillatory forces acting on the duct structure are primarily a result of the pressure fluctuations at directional changes (e.g. bends) or area changes (e.g. conical sections) of the duct system.

In this case, the estimation of the oscillatory forces and moments reduces to determining the unsteady pressure and flow field in the duct system – a field commonly referred to as *duct acoustics*.

2. FUNDAMENTALS OF DUCT ACOUSTICS

In order to investigate the flow and pressure oscillations in a duct system by numerical analysis, a suitable model depth must be selected for each component. Depending on the properties of the individual component, either a 0D, 1D or 3D model must be chosen.

2.1. 0D vs. 1D vs. 3D Models

0D models are most suitable if the internal degrees of freedom of a component are not of relevance to its transmission behaviour. This is the case if the travel time L/c of a pressure wave of speed c through a component of characteristic length L is substantially smaller than the period $1/f$ of the frequency of interest, i.e. $fLc^{-1} \ll 1$. A typical example would be the model of a small return tank, where the pressure can be assumed to be uniform throughout the component and internal wave propagation is not of concern. Also, for a sufficiently short pipe section, compressibility effects can be ignored, such that the flow rate is constant along the section length. If the product fLc^{-1} is not small with respect to unity, wave phenomena within the component need to be considered. If the wave propagation primarily occurs along a single dimension (e.g. along the longitudinal axis of a slender pipe section), 1D models

suffice. If even this simplification does not apply and substantial changes of state variables might occur along more than one spatial dimension, one is required to resort to 2D or even 3D models. For most components encountered in duct acoustics, 1D models suffice.

2.2. Frequency vs. Time Domain Analysis

Irrespective of the chosen model depth, a choice must be made whether the system analysis is to be carried out in the time or frequency domain. Table 1 provides an overview over various aspects that drive the decision between time and frequency domain analysis.

Table 1: Comparative capability analysis of time vs. frequency domain analysis.

Aspects	Time Domain	Frequency Domain
Analysis of Transients (Valve closure & pump ramp-up)	✓	✗ (Only linearised)
Consideration of system nonlinearities (Cavitation, large amplitudes)	✓	✗ (Only linearised)
Determination of Steady-state modes, resonances & transfer functions	✓ (Slow)	✓
Acoustic optimisation (Silencer design, shifting of resonances)	✓ (Very slow)	✓
Efficient for large range of frequencies	✗ (Large amounts of data, extremely slow)	✓

For the cooling circuit under consideration, only steady-state oscillations are of interest and the amplitudes are small, such that frequency-domain simulation can demonstrate its advantages. If the analysis is limited to the frequency domain, the computationally efficient *transfer matrix method* can be employed.

2.3. Transfer Matrix Method

In the case of a 1D hydraulic system model, both pressure p and flow rate Q are required to uniquely describe the flow state at a given cross-section. Thus, to characterise the relation between the states at the inlet and outlet of any 1D component in the frequency domain, the concept of a transfer function has to be generalised to a transfer matrix $\boldsymbol{T}$. This generally frequency-dependent and complex-valued function relates the phasor vectors at the inlet with those at the outlet:

$$\begin{pmatrix} p^*_{\mathrm{In}} \\ Q^*_{\mathrm{In}} \end{pmatrix} = \boldsymbol{T} \begin{pmatrix} p^*_{\mathrm{Out}} \\ Q^*_{\mathrm{Out}} \end{pmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{pmatrix} p^*_{\mathrm{Out}} \\ Q^*_{\mathrm{Out}} \end{pmatrix} \tag{2.1}$$

For example, the transfer matrix $\boldsymbol{T}_{\mathrm{Pipe}}$ of a straight circular pipe of length L and cross-sectional area A, filled with an inviscid fluid of sonic speed c, is given by:

$$\boldsymbol{T}_{\text{Pipe}} = \begin{bmatrix} \cosh\left(\frac{\mathrm{i}\omega L}{c}\right) & \frac{\rho c}{A}\sinh\left(\frac{\mathrm{i}\omega L}{c}\right) \\ \frac{A}{\rho c}\sinh\left(\frac{\mathrm{i}\omega L}{c}\right) & \cosh\left(\frac{\mathrm{i}\omega L}{c}\right) \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{\omega L}{c}\right) & \frac{\mathrm{i}\rho c}{A}\sin\left(\frac{\omega L}{c}\right) \\ \frac{\mathrm{i}A}{\rho c}\sin\left(\frac{\omega L}{c}\right) & \cos\left(\frac{\omega L}{c}\right) \end{bmatrix} \tag{2.2}$$

The transfer matrix of a series connection of two components A and B is readily obtained by matrix multiplication of the respective matrices T_{A} and T_{B}. Further details regarding the interconnection and usage of transfer matrices can be found in the literature [1][2]. By combining multiple matrix elements, a system model as described in section 3 can be set-up. For components whose transfer matrix cannot be determined by analytical estimates, experimental methods have to be employed.

2.4. TMM Test Bench

For the experimental determination of the transfer matrix of a specific component, a so-called transfer matrix method (TMM) test bench as shown in Figure 2 is used [3].

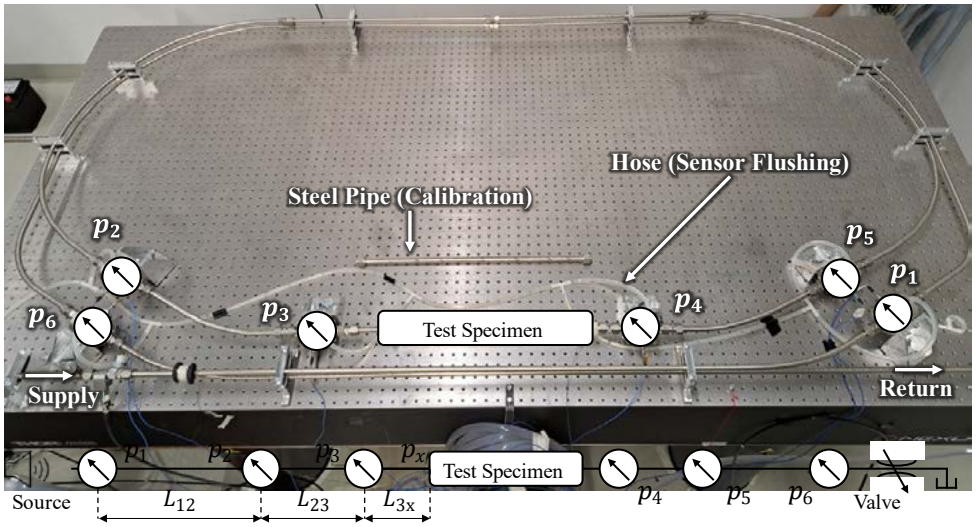


Figure 2: AVL Transfer Matrix Method Test Bench.

The test bench consists of up to six unsteady ("dynamic") pressure transducers, a load valve and a suitable acoustic source. The source as well as the valve must be located either before the first sensor or after the last sensor. The acoustic source excites the system and creates a wave pattern that is characteristic for the system under investigation.

2.4.1. Transfer Matrix Determination

To determine the transfer matrix $\boldsymbol{T}$ of a component experimentally, the phasors of pressure and flow rate at both the component inlet and outlet must be known to solve equation 2.1 for the matrix elements T_{ij}. Since there are only two equations for four unknowns T_{ij}, two

distinctly different measurements have to be carried out to create four independent equations. This is typically achieved by adjusting the load valve to force a different wave pattern. For the specific case of a symmetric transfer behaviour of the investigated component, only one measurement is required.

Compared to measuring unsteady pressures, non-invasive measurement of unsteady flow rates is technically challenging. To overcome this, the TMM replaces the direct measurement of a single flow rate with the measurement of two unsteady pressures by utilising a rearranged form of equation 2.1:

$$p_{\mathrm{In}}^{*} = p_{\mathrm{Out}}^{*} \cdot T_{11} + Q_{\mathrm{Out}}^{*} \cdot T_{12} \rightarrow Q_{\mathrm{Out}}^{*} = \frac{p_{\mathrm{In}}^{*}}{T_{12}} - p_{\mathrm{Out}}^{*} \frac{T_{11}}{T_{12}}. \tag{2.3}$$

Thus, if two pressures and the TFM elements of the components between the pressure sensors are known, a flow rate can be determined. Because their TFM elements can be calculated with high accuracy by equation 2.2 (or more advanced models), steel pipes are used between the sensors. In the actual measurement, three instead of two sensors are used for state determination in order to expand the frequency range that can be analysed. Based on the three pressures per side and known transfer behaviour of the steel pipes between the sensors, the states at the inlet and the outlet of the test specimen can be estimated by solving the following equation system for the unknown phasors p_x^* and Q_x^* at a location x (only shown for sensors 1, 2 and 3, see Figure 2):

$$\underbrace{\begin{pmatrix} p_1^* \\ p_2^* \\ p_3^* \end{pmatrix}}_{\vec{b}} = \begin{bmatrix} \cosh\left(\frac{\mathrm{i}\omega L_{1x}}{c}\right) & \sinh\left(\frac{\mathrm{i}\omega L_{1x}}{c}\right) \\ \cosh\left(\frac{\mathrm{i}\omega L_{2x}}{c}\right) & \sinh\left(\frac{\mathrm{i}\omega L_{2x}}{c}\right) \\ \cosh\left(\frac{\mathrm{i}\omega L_{3x}}{c}\right) & \sinh\left(\frac{\mathrm{i}\omega L_{3x}}{c}\right) \end{bmatrix} \underbrace{\begin{pmatrix} p_x^* \\ \frac{\rho c}{A} Q_x^* \end{pmatrix}}_{\vec{x}} = \boldsymbol{A} \begin{pmatrix} p_x^* \\ Z_{\mathrm{C}} Q_x^* \end{pmatrix} \tag{2.4}$$

The determination of the states – and thus the transfer matrix of the specimen – becomes challenging if the signal-to-noise ratio assumes small values. Disturbing noise levels are particularly relevant in the low-frequency range. The accuracy of experimental results in this part of the spectrum can be improved by optimising the sensor spacing.

2.4.2. Numerical Conditioning and Sensor Spacing

When solving a linear equation system $\vec{b} = \boldsymbol{A}\vec{x}$ like equation 2.4 for $\vec{x}$, the sensitivity to errors (i.e. noise) in $\vec{b}$ can be quantified through the so-called *condition number* $\kappa(\boldsymbol{A})$. It is defined as the ratio of the largest to the smallest singular value of $\boldsymbol{A}$. Large condition numbers imply that small errors in the measured pressures $\vec{b}$ can lead to large errors in the estimated acoustic states $\vec{x}$, whereas condition numbers close to unity indicate a well-conditioned system with minimal noise amplification. Since – for a given (angular) frequency ω and sonic speed c – the sensor spacings are the only variables affecting the conditioning of matrix $\boldsymbol{A}$, the condition number κ can serve as a design metric for sensor placement: by selecting appropriate sensor spacings, the measurement matrix can be made robust for the frequencies of interest, improving the reliability of reconstructed acoustic quantities.

Analytically, it can be shown that κ is independent of the absolute distance L_{3x} of the sensor array relative to the specimen – only the relative sensor spacings L_{12} and L_{23} affect numerical conditioning, as shown in Figure 3:

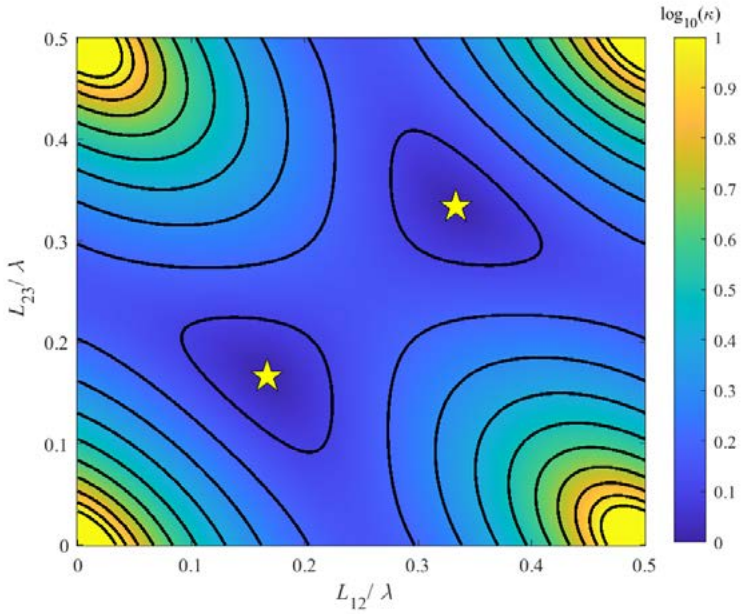


Figure 3: "Heatmap" of the condition number κ for a TMM three-sensor array.

The theoretical minimum occurs when the three measured pressure phasors are spaced 120° apart. Under this condition, $\kappa = 1$, which corresponds to perfect numerical conditioning. To achieve this, the relative sensor spacings must be chosen to:

$$\frac{L_{12}}{\lambda} = \frac{L_{23}}{\lambda} = \frac{1}{6}. \tag{2.5}$$

In the equation above, $\lambda = c/f = 2\pi c/\omega$ refers to the wavelength. In practice, selecting sensor spacings close to $\lambda/6$ for a frequency of interest ensures low condition numbers and reliable results. Conversely, sensor spacings approaching 0 or multiple integers of $\lambda/2$ lead to near-linear dependence of the measurement rows and result in severe noise amplification. Physically, this occurs because the equation system "expects" practically identical pressures which – due to noise – will not be recorded in practice.

3. System-Level Model

Based on the fundamentals presented in section 2, a high-fidelity augmented 1D system model of the cooling circuit can be set up for system analysis and optimisation. Depending on the complexity of the respective component, either analytical or surrogate models from various sources are chosen to represent its passive properties in the system model, as it is shown in Figure 4:

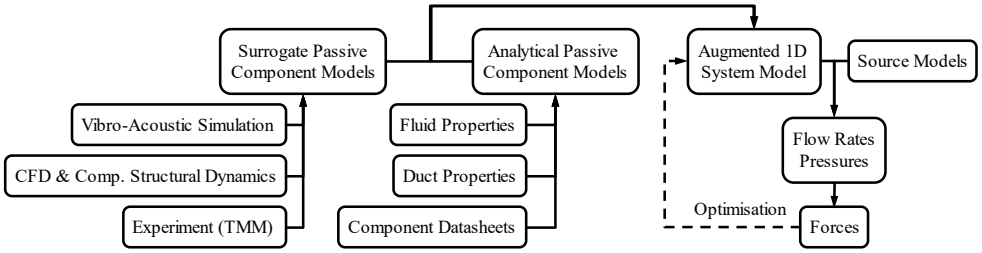


Figure 4: Data sources for the multidisciplinary augmented 1D system-level model.

3.1. Analytical Component Models

The geometry of most components of a duct system can be described by bulbous (i.e. acoustically "short", $fLc^{-1} \ll 1$) cuboids and slender cylinders. The acoustic behaviour of such components (e.g. tanks and pipes) can be estimated with sufficient confidence based on analytical relations alone. To parameterise the respective component models, the properties listed in Table 2 are required:

Table 2: Required properties for parameterisation of analytical component models.

Category	Property
Fluid properties	Mass density ρ
	Sonic speed c or bulk modulus K
	Viscosity η or kinematic viscosity ν
Duct material properties	Young's modulus E (potentially frequency-dependent and complex)
	Poisson's ratio μ
Duct geometry	Section length L
	Inner diameter d (or equivalent for non-circular sections)
	Wall thickness e
Operational conditions	Average flow rate $\overline{Q}$ (to assess flow regime)
	Average fluid temperature $\overline{T}$ (for fluid properties)
	Mean pressure level $\overline{p}$ (for fluid properties)

If the fluid is not a pure substance or a ubiquitously used, well-documented mixture, one must resort to e.g. manufacturer's data sheets to obtain the relationships between the thermodynamic state of the fluid and the respective fluid property.

3.2. Surrogate Component Models

In contrast to acoustically short or slender, tubular geometries, the transmission behaviour of components with complex internal geometries usually cannot be calculated purely by theory. In this case, surrogate component models have to be utilised. For the present system, the pump, the PVC hoses as well as the heat exchanger must be represented through surrogates.

3.2.1. Pump Surrogate Model

Like other positive displacement machinery, the pump's acoustic behaviour can be characterised through its source impedance (passive property) Z_P and its source flow ripple Q_P^* (active property, see section 3.3.1) according to a NORTON equivalent circuit [4]. Ideally, both these properties are determined experimentally through a TMM test setup as outlined in the standard ISO 10767-1:2015 [5]. If such measurements are not available (e.g. during the conceptual phase), reasonable estimates can often be made through semi-analytical approaches as shown below.

The source impedance Z_P of positive displacement pumps – as seen from the duct system – generally exhibits the characteristics of a simple hydraulic spring, damper and mass system:

$$Z_\mathrm{P} = \frac{C_\mathrm{P}}{\mathrm{i}\omega} + D_\mathrm{P} + \mathrm{i}\omega M_\mathrm{P} \tag{3.1}$$

The model parameters C_P, D_P and M_P are directly related to the characteristics of the fluid enclosed between the gears and the pump flange. Due to the comparatively small size of the pump, the impedance will be governed by the hydraulic stiffness C_P in the frequency range of interest. This stiffness is typically substantially smaller than what would be expected from an estimate based on the enclosed fluid volume alone since the additional compliances of sealings and the pump casing are present. STULEMEIJER reports on the *effective fluid volume* of pumps with a known displacement volume per revolution [6]. This data can be used to construct a semi-empirical correlation to estimate the effective fluid volume of the pump in question and thus its hydraulic capacity. The dissipative and inertial contributions can be neglected here due to the pump's size and the frequencies of interest.

3.2.2. PVC Hose

The surrogate component model for the PVC hoses can be derived in three different ways: 1D theory (i.e. analytically), 3D vibro-acoustic analysis, or experimentally on the TMM test bench. The first two methods require knowledge of the geometry and the material behaviour of the hose, which is determined by a dynamic mechanical thermal analysis (DMTA). In contrast, the experiment on the TMM test bench does not require prior knowledge of the material behaviour of the hose or the geometry of the test specimen.

The results of the three different methods, represented as a mass-spring-mass model are shown in Figure 5. Differences between 1D theory and 3D vibro-acoustic simulation can be explained by the fact that 1D theory cannot capture 3D effects (e.g. ballooning at the transition from steel pipe to PVC hose). The experimental data from the TMM test bench is in good agreement with the other two methods. The stiffness C from vibro-acoustic

simulation aligns better with the measurement since the stiffness-enhancing clamping effect at both ends of the hose are taken into account. If at least another measurement with a different hose length is performed, a length-dependent surrogate model for the PVC hose can be derived.

The significant deviations between simulation and measurement at low frequencies can be explained by the increasingly bad condition number κ (frequency range $f < 10$ Hz), while the low sensor coherence γ^2 – indicative of e.g. a low signal-to-noise ratio – explains the discrepancies at high frequencies (frequency range $f \geq 200$ Hz).

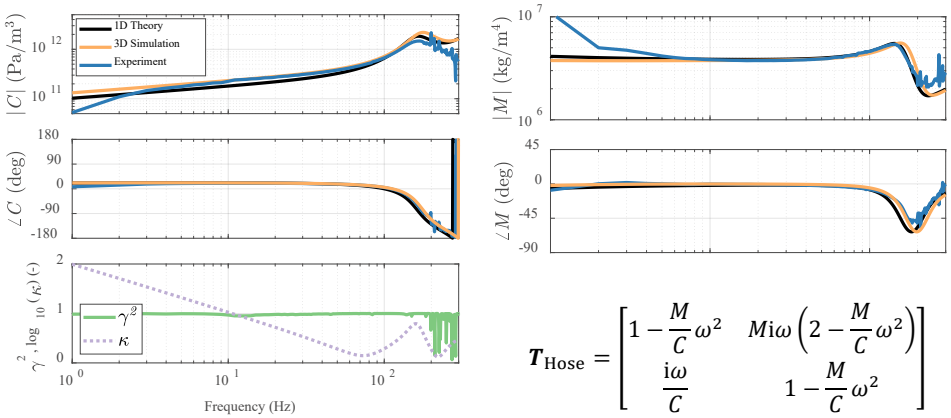


Figure 5: Comparison of various surrogate component models of the PVC hose.

3.2.3. Heat Exchanger Surrogate Model

As shown in Figure 6, the array of parallel narrow channels in the heat exchangers is connected via manifolds to the piping of the cooling circuit. Due to its highly branched and complex internal structure, a direct representation through a fully parametric TFM would be impractical and computationally inefficient. For this reason, the assembly of two heat exchangers and their respective manifolds is modelled through a surrogate model that extends from the "Inlet" to the "Outlet" port marked in Figure 6:

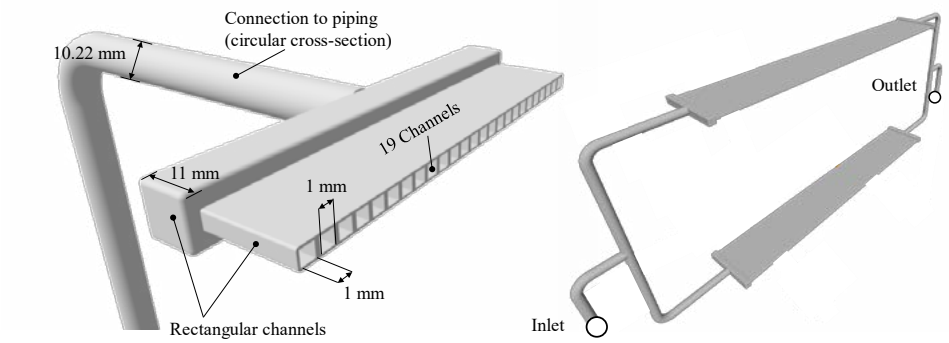


Figure 6: Heat exchanger cross-section with relevant dimensions.

The surrogate model is derived from a 3D vibro-acoustic simulation performed using commercially available FE software. Based on the calculated flow rates and pressures at the inlet and outlet of the heat exchanger assembly, the TFM elements of the surrogate model are determined using the same method as for experimental data (see section 2.4.1.).

3.3. Source Models

To obtain the pressures, flow rates and flow forces, the augmented 1D system model needs to be excited by source models. As outlined, several types of sources may excite liquid columns to oscillate. When modelling acoustic sources, a distinction is made between those of deterministic and those of stochastic nature. A typical example of a deterministic acoustic source is the flow ripple generated by the coolant pump.

3.3.1. Pump Flow Ripple

The coolant pump imposes a flow ripple phasor Q_{P}^{*} on the system boundary. For positive displacement pumps, the kinematic source flow ripple can often be reasonably approximated by a rectified sine wave with the amplitude Q_{P}' in the time domain. The sine amplitude is directly connected to the so-called degree of non-uniformity $\delta_{\mathrm{P}} = Q_{\mathrm{P}}'/\overline{Q}_{\mathrm{P}}$ commonly used in the pump industry [2]. The degree of non-uniformity δ_{P} can be reasonably approximated based on the number of pumping elements z (teeth per gear, pistons, etc.) by $\delta_{\mathrm{P}} = 2/z$ for $z \gg 1$. For the coolant pump at hand, $\delta_{\mathrm{P}} = 2/9 \approx 0.22$. Fourier decomposition of the assumed $Q_{\mathrm{P}}(t)$ shows that spectral contributions to the flow ripple amplitude spectrum (AS) only occur at integer multiples n of the fundamental pumping frequency $f_{\mathrm{P}} = f_{\mathrm{S}} \cdot z$ (with f_{S} denoting the pump shaft frequency), and that their amplitudes decay with the square of the order n as follows:

$$|Q_{\mathrm{P}}^{*}(n \cdot f_{\mathrm{P}})| = \frac{2}{\pi} \frac{\overline{Q}_{\mathrm{P}} \delta_{\mathrm{P}}}{4n^2 - 1} \tag{3.2}$$

3.3.2. Local Turbulence Sources

Even for nominally laminar flow throughout the system, strong deflections within a component may induce local turbulence (LT). These turbulent regions act as acoustic sources. In the investigated system, local turbulence occurs at the double bend and within the piping of the heat exchangers as depicted in Figure 1. The strength of local turbulence sources must be determined through CFD analysis or experimentally.

A common feature of sources due to local turbulence is their decay towards higher frequencies. Due to the randomness of the turbulent fluctuations, the corresponding sources show a stochastic behaviour, which is why they are characterised by a power spectral density (PSD) rather than an amplitude spectrum (AS). It is important to note that a PSD represents how the power of a signal is distributed across the frequency range. However, it *does not* represent a worst-case scenario that – for instance – can directly serve to calculate the maximum occurring pressure at a given location. Furthermore, it is not straightforward to combine deterministic and stochastic sources due to the different means of representation (amplitude and phase vs. PSD).

3.4. Pressures, Flow Rates & Forces

For the current model, one deterministic (pump) and two stochastic (local turbulence at the double bend and within the exchangers) sources are considered. Combining both deterministic and stochastic sources poses a particular challenge due to their different representations: Since a PSD does not contain any phase information anymore, a simple superposition of all sources is not guaranteed to provide the worst-case pressure and force amplitudes: Depending on the phase angle between the sources, superimposing them might result in constructive or destructive interference. To circumvent this issue, the relative phase angles of the stochastic sources with respect to the deterministic one are systematically varied which results in an envelope for the desired output quantities rather than single values.

This method is applied to the present system as well. The aim of the optimisation process is to minimise the axial reaction force amplitude $|F_{\mathrm{HE}}|$ at the heat exchangers which occurs primarily due to the pressure difference $\Delta p_{\mathrm{HE}} = p_{\mathrm{Inlet}} - p_{\mathrm{Outlet}}$ across them (see Figure 1 and Figure 6). Thus, the magnitude of the pressure difference $|\Delta p_{\mathrm{HE}}|$ at the heat exchanger is the hydraulic quantity of primary interest for the optimisation process. It is plotted as a function of the frequency in Figure 7. The dotted red curve shows the load case if only the pump source is active. The blue envelope band is obtained by adding the sources from local turbulence (heat exchangers and double bend) and varying their phase angles.

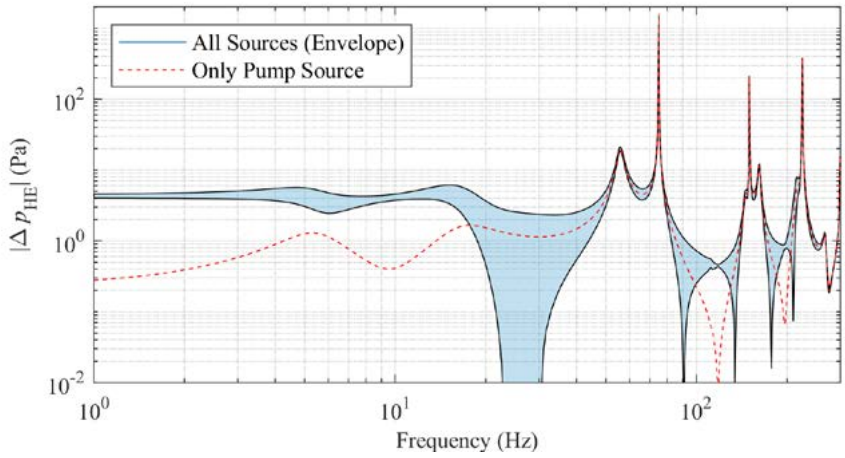


Figure 7: Baseline heat exchanger pressure difference $|\Delta p_{\mathrm{HE}}|$ as a function of the frequency.

As expected, the low-frequency range (below 20 Hz) is governed by the stochastic sources since these frequencies are well below the fundamental pumping frequency $f_{\mathrm{P}} = 75$ Hz. The distinct peaks due to pump excitation at 75 Hz and its harmonics (150 Hz, 225 Hz, …) are the most dominant contributions within the considered frequency range. When comparing the spectrum of the pressure difference (i.e. the forces on the laser source) with the characteristics of the control loop responsible for the laser beam orientation, the second-order peak at 150 Hz can be identified as the main driver of the precision loss of the prototype. Therefore, the goal of the following system optimisation will be the reduction of this peak.

4. OPTIMISATION WORKFLOW

Once the dominant force contributions and their causative sources are identified, targeted mitigation strategies can be derived. These measures are first evaluated virtually through surrogate-model enhanced 1D simulation. The innovative aspect of the proposed workflow is the ability to conduct rapid design iterations using a reduced-order system model while retaining the fidelity of full 3D simulations where necessary.

4.1. Hose Optimisation

A hose provides frequency-dependent compliance and damping through its viscoelastic material properties. Strategically positioning and sizing such a hose can attenuate pressure differences across sensitive components, thereby reducing the reaction force transmitted to the heat exchangers. This beneficial effect is to be maximised such that only minimal vibration affects the sensitive laser source. The optimisation problem possesses the following degrees of freedom:

- Hose length L_H.
- Hose position y_H, defined as the lateral distance of the bracket from the wall. The position affects local boundary conditions (e.g., support stiffness, damping) reflected in the section's transfer matrix.

The constraints are defined as follows:

- Box constraint on length of hose: 50 mm $\leq L_\mathrm{H} \leq$ 350 mm.
- Box constraint on position: 100 mm $\leq y_\mathrm{H} \leq$ 1900 mm.

The length of the viscoelastic hose is limited to prevent unwanted fluid-structure interaction. The limits for y_H are prescribed by layout rules and clearance envelopes (e.g., walls, frames, neighbouring equipment). In addition, the piping layout should be designed to route the flow directly to the heat exchanger, avoiding any detours or counterflow sections. Therefore, the axial constraint reads $L_\mathrm{H} + y_\mathrm{H} \leq 2100$ mm.

The optimisation problem of minimising the force magnitude at the heat exchangers is then given by $|F_\mathrm{HE}(f, L_\mathrm{H}, y_\mathrm{H})| \propto |\Delta p_\mathrm{HE}(f, L_\mathrm{H}, y_\mathrm{H})|$ and subject to the box constraints on L_H, y_H, and the non-linear routing constraint. For the optimisation, a global *surrogate-based optimisation* with *adaptive sampling* is employed. Each evaluation of $|\Delta p_\mathrm{HE}(f, L_\mathrm{H}, y_\mathrm{H})|$ requires a frequency-domain simulation using the full transfer-matrix chain, which is computationally expensive and involves nonlinear constraint interactions (via position-dependent boundary conditions). The chosen approach proceeds as follows:

1. Evaluation of objective function: A set of quasi-random points is generated, and the objective function is evaluated at these points.
2. Interpolation for surrogate: A radial basis function (RBF) is used to interpolate the objective function to set up the surrogate model [7]. The cheap-to-evaluate surrogate approximates the expensive objective function.
3. After building the surrogate, the algorithm generates a large set of candidate points across the domain. Each candidate is scored using a merit function that balances

the predicted objective value from the surrogate and the distance from existing sample points. The candidate minimising the merit function is selected for true evaluation [8][9], and the surrogate is updated with this new data point.

The algorithm is repeated until a certain budget of evaluations is exhausted. This particular method is used since it is well suited for problems involving expensive evaluations, nonlinear constraints and unknown gradients.

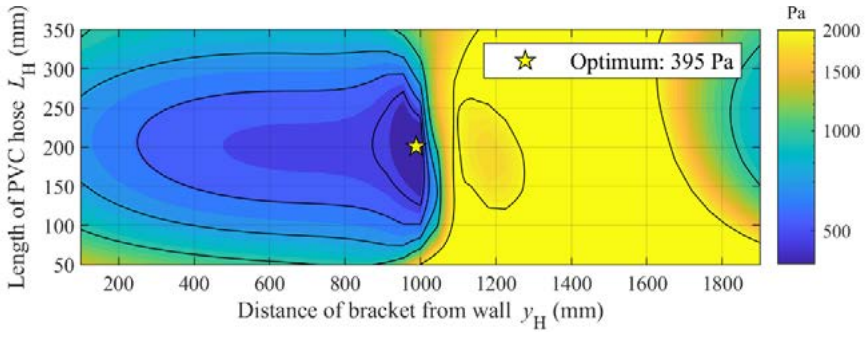


Figure 8: Contour plot of $|\Delta p_{\mathrm{HE}}|$ as a function of y_{H} and L_{H}, optimum marked with star.

The optimisation results are depicted in Figure 8. The minimal pressure difference (and thus force) is obtained for a hose length $y_{\mathrm{H}} = 200$ mm and a wall distance of $L_{\mathrm{H}} = 1000$ mm for the bracket. With the optimisation, a setup with only 3.8 % of the maximum pressure difference – and thus force – on the heat exchangers (baseline pressure difference $|\Delta p_{\mathrm{HE}}| = 10400$ Pa) could be identified.

Intuitively, one would expect the hose-related pressure suppression to monotonically increase with the hose length and provide minimal forces when the suppression is directly *before* the evaluated points. However, since the hoses not only act as dissipative but also as reflective elements, the maximal suppression is reached at different location and length.

Following virtual testing, the most promising mitigation measure undergoes detailed mechanical design. A detailed 3D vibro-acoustic analysis of the final hose design is then carried out, enabling an assessment of the stress distribution and – in a subsequent analysis – fatigue life under realistic operating conditions which are obtained from the augmented 1D system model. This phase ensures mechanical integrity and durability, considering factors such as vibrational loads and dynamic stress concentrations.

4.2. Engineering Framework for Development, Validation and Industrialisation

In addition to the mitigation measures described above, the development of other suppression devices such as chamber-based or membrane-type silencers, might be necessary for other systems to address acoustic, vibrational, or flow-induced disturbances. In such cases, a systematic and multidisciplinary development process is recommended, encompassing mechanical design, material science, and the application of appropriate computer-aided engineering (CAE) methods, including coupled thermal, thermo-

mechanical, and structural analyses, as well as reliability & lifetime assessment and industrialisation. Furthermore, the resulting components may need to be qualified and integrated for operation under non-standard environmental boundary conditions, including high- and low-temperature regimes, cleanroom environments, or vacuum conditions.

5. CONCLUSION & OUTLOOK

In this paper, a comprehensive workflow for identifying and mitigating flow-related disturbances in cooling circuits for high-precision applications via multidisciplinary simulation was presented. By employing surrogate models, the high fidelity of 3D simulations and experimental investigations could be carried over to a multidisciplinary augmented 1D system model. By optimising the length and location of PVC hoses, a substantial reduction of performance-degrading force disturbances could be achieved. Future research may focus on further refining surrogate modelling techniques and exploring adaptive or actively actuated mitigation strategies to dynamically counteract varying operational conditions.

6. REFERENCES

[1] Wylie, E. B., & Streeter, V. L. (1978). *Fluid transients*. New York.

[2] Theissen, H. (1984). *Die Berücksichtigung instationärer Rohrströmung bei der Simulation hydraulischer Anlagen.* (Doctoral dissertation, Rheinisch-Westfälische Technische Hochschule Aachen, Germany).

[3] Kojima, E., & Edge, K. A. (1994). Experimental determination of hydraulic silencer transfer matrices and assessment of the method for use as a standard test procedure. In *Innovations in Fluid Power: 7th Bath International Fluid Power Workshop.*

[4] Johnston, N., Dumnov, D., & Melling, J. (2023). Modelling of Pumps and Motors as Source Flow Ripple and Source Impedance. In *Fluid Power Systems Technology* (Vol. 87431, p. V001T01A008). American Society of Mechanical Engineers.

[5] International Organization for Standardization (2015). *Hydraulic fluid power — Determination of pressure ripple levels generated in systems and components — Part 1: Method for determining source flow ripple and source impedance of pumps*. (ISO Standard No. 10767-1:2015).

[6] Stulemeijer, I. P. J. M. (1981). *Beschreibung der Druckschwankungen in hydrostatischen Anlagen und ihre Auswirkungen auf den Luftschall*. (Doctoral dissertation, Technische Hogeschool Eindhoven, The Netherlands)

[7] Gutmann, H. M. (2001). A radial basis function method for global optimization. Journal of global optimization, 19(3), 201-227.

[8] Audet, C., & Dennis Jr, J. E. (2002). *Analysis of generalized pattern searches. SIAM Journal on optimization*, 13(3), 889-903.

[9] Abramson, M. A., Audet, C., Dennis Jr, J. E., & Digabel, S. L. (2009). *OrthoMADS: A deterministic MADS instance with orthogonal directions*. SIAM Journal on Optimization, 20(2), 948-966.

Biographies






Enrico Pasquini received the bachelor's degree in mechanical engineering from FH Aachen University of Applied Sciences in 2010, the master's degree in aerospace engineering in 2014 and a PhD in mechanical engineering from RWTH Aachen University in 2020, respectively. He is currently working as a technical lead at the Department of Precision Engineering of AVL Deutschland GmbH. His research areas include duct acoustics and mathematical modelling.

Jonas Wack received the master's degree in mechanical engineering from Technical University of Munich in 2013 and a PhD in engineering from University of Stuttgart in 2020. He is currently working as a development engineer at the Department of Precision Engineering of AVL Deutschland GmbH. His research areas include duct acoustics and experimental methods.

Afshin Sattarifar received the bachelor's degree in mechanical engineering from University of Guilan/Iran in 2012, the master's degree in computational engineering in 2016 and a PhD in mechanical engineering in 2024 from Ruhr University Bochum. He is currently working as a development engineer at the Department of Precision Engineering of AVL Deutschland GmbH. His research areas include duct acoustics and structural health monitoring.

Peter Gehr received the master's degree in mechanical engineering from Regensburg University of Applied Sciences in 2020. He is currently working as a development engineer at AVL Software and Functions GmbH. His research areas include mathematical modelling and control engineering.

Research of the cavitation in the rear brake chamber of hydraulic rock drill

Weihang Xi[1], Xiang Tian[2], Lulu Gao[1,*], Fei Ma[1]

Authors affiliation, [1] University of Science and Technology Beijing

[2] Siton Machinery Manufacturing Co., Ltd

**gaolulu@ustb.edu.cn*

Abstract: The hydraulic rock drill is a type of hydraulic impactor that relies on hydraulic power to deliver periodic impact forces externally. As a core piece of equipment in mining operations, the service life of its key internal moving components and supporting seals directly determines the overall operational efficiency of the equipment. This paper focuses on the cavitation erosion issue affecting the rear guide sleeve, which is a primary support component in the rock drill. Utilizing a rock drill impact test bench, experiments were conducted. The results indicate that the negative pressure zone generated during the outward movement of the impact piston from the rear braking chamber is the main cause of cavitation near the rear guide sleeve. By comparing the motion of the impact piston within the braking chamber and the corresponding pressure variations under different clearance conditions, it was found that the formation of the negative pressure zone primarily depends on the penetration depth of the piston into the braking chamber and the maximum pressure within the chamber. Furthermore, by comparing the variations in the impact performance of the rock drill under different clearance conditions, it is proposed that the optimal clearance for this type of rock drill should be within the range of 0.04 to 0.06 mm.

Keywords: Hydraulic rock drill, Cavitation, Clearance

1 Introduction

Hydraulic rock drills serve as critical equipment in modern mining and tunnel excavation [1,2], as shown in Fig 1.1. The stability and continuity of a rock drill's operation directly determine the overall working efficiency of mining equipment. Therefore, the service life of its core internal components is critical, primarily including the impact piston, guide sleeve, and sealing elements.

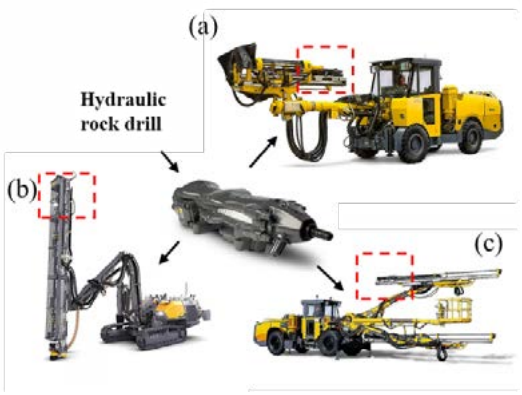


Fig 1.1 Mining machinery equipped with hydraulic rock drill
(a) Underground drilling rig, (b) Surface drilling rig, (c) Drill jumbo

Scholars have conducted in-depth research on the motion of the impact piston through mechanism analysis and hydraulic simulation techniques. Gorodilov [3] proposed a segmented modeling approach based on the assumption of constant inlet pressure and neglecting the impact and spool switching processes. To couple the motion of piston and spool, Hu [4] provided an AMESim-based model and based on this model Yang [5] added a plunger pump model to analyze the effect of inlet flow fluctuations on impact performance. Besides, Guo [6] employed the bond graph method to model the coupled motion, achieving lower complexity compared to AMESim.

Based on the motion analysis of the impact piston, the correlation between piston motion and internal cavitation has received some scholarly attention. For instance, Tian [7] and Jia [8] through experimental analysis, established a relationship between the service life of the front water seal and the impact piston's stroke in a rock drill, finding that seal life decreases

as the stroke increases. Separately, Geng [9,10] in a study on the cavitation mechanism in a dual-cushion rock drill, demonstrated that the piston's kinematic parameters upon entering the front cushioning chamber are a critical factor in the formation of cavitation in the front chamber.

While existing research has extensively addressed the front guide sleeve and seals, the rear guide sleeve has received limited attention. To bridge this gap, this study is dedicated to investigating the rear-side problem. By integrating laser displacement and pressure sensors, we capture the key parameters of piston motion and pressure dynamics at the rear cushioning chamber to elucidate the mechanism of cavitation erosion and develop a targeted solution.

2 Problem statement

2.1 Structure and working principle

Dual-control hydraulic rock drills (DHRD) have become the primary type of rock drill equipment in current mining machinery due to their high energy efficiency. Their basic structural composition is illustrated in Fig 2.1. The main body of DHRD contains impact piston (1), guide sleeve (2), and cylinder sleeve (3). It is used to perform the shock motion. The motion is controlled by the spool of direction valve (5), installed on the valve seat (4). There are four chambers formed by part 1 and 3, i.e., front chamber (FC), piston signal chamber (PSC), tank chamber (TC), and rear chamber (RC). Also, there is a signal chamber for direction valve (VSC). The pressure oil is input to power the whole system from the inlet port of direction valve.

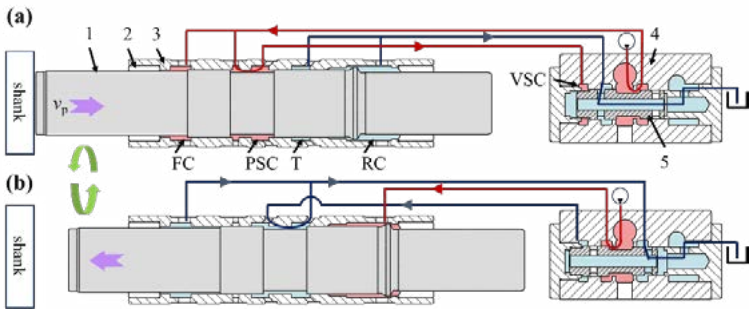


Fig 2.1 Working principle of DHRD

(a) Return stroke state, (b) Impact stroke state

The impact motion of DHRD contains two typical states, i.e., return stroke state (RSS) and impact stroke state (ISS), and the switching of these states above are shown in Fig 2.2. The working principle can be explained as follows: DHRD switches between RSS and ISS repeatedly by the coupling motion of part 1 and 5. The operating state during RSS and ISS can further be divided into acceleration and braking states according to the direction of movement of part 1.

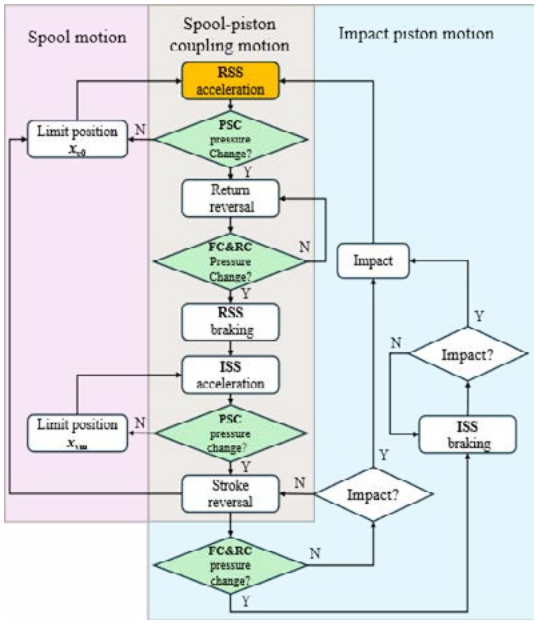


Fig 2.2 Flowchart of working principle of DHRD

2.2 Cavitation in the rear braking chamber

As illustrated in Fig 2.3, when the impact piston moves past a specific position in the cylinder, a sealed oil chamber forms between the piston, the cylinder, and the guide sleeve, and this chamber is the rear braking chamber. The chamber serves as an anti-impact mechanism to protect the piston from striking the rear guide sleeve in the period between RSS braking and ISS acceleration.

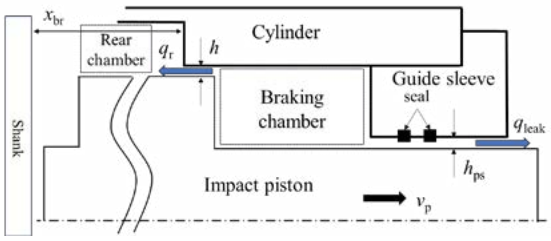


Fig 2.3 Structure of braking chamber

Damage progression in the rear guide sleeves of the same rock drill model reveals a clear time-dependent progression, as shown in Fig 2.4.

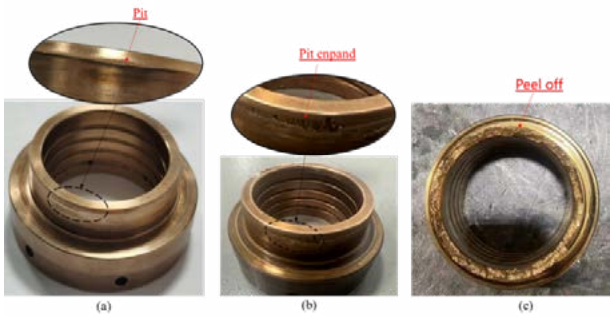


Fig 2.4 Cavitation phenomenon in different stages

(a) Early stage (less than 50 impact hours) (b) Intermediate stage (between 50 and 100 impact hours) (c) Late stage (over 100 impact hours)

Damage to the rear guide sleeve leads to hydraulic oil contamination. The resulting metallic debris poses a significant threat to both the rear seal and the piston surface, substantially reducing the service life of these components. Hence, determining how to optimize the rear braking chamber's structure to mitigate cavitation-induced damage to the rear guide sleeve is a critical issue for improving the service life of rock drills.

3 Method and materials

3.1 Hydraulic rock drill test bench

As shown in Fig 3.1, a hydraulic rock drill impact test bench (HRDITB) was established to acquire the pressure within the rear braking chamber and the motion of the impact piston.

The HRDITB is equipped with (A) an impacted system serving as the load, (B) pressure sensors mounted on the rock drill to record pressure variations in the rear braking chamber, and (C) a laser displacement sensor to track piston movement.

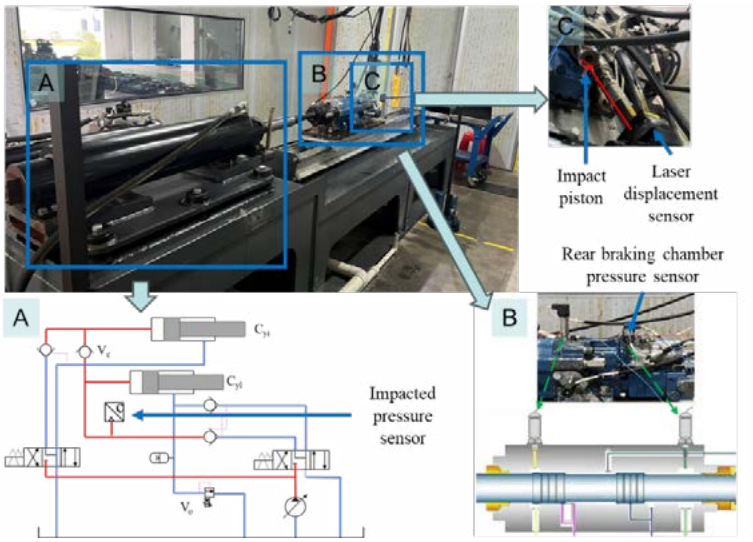


Fig 3.1 Hydraulic rock drill impact test bench

The load from the impacted system is generated by a hydraulic circuit where the impact cylinder (Cyi) is connected to a loading cylinder (Cyl). The pressure in the rodless chamber of Cyl is controlled by an overflow valve (Vo), thereby establishing a stable pressure differential across the intervening check valve. This differential pressure is the definitive load provided by the system. The sensors parameters are shown in Table. 3.1.

Sensor	Model	Range	Accuracy	Sample Rate
Laser displacement sensor	TS803-140/100	140~240 mm	± 0.05% FS	8192Hz
Pressure sensor	YN-63-IV-400	0~400 MPa	± 0.5% FS	8192Hz

Table. 3.1 Sensors used in the HRDITB

3.2 Motion and pressure variation characteristics of the impact piston in the rear braking chamber

Based on the mechanical structure, the position at which the impact piston enters the rear braking chamber is identified as x_{bl} = 58.5 mm. As shown in, the laser displacement data

confirm that the piston travels beyond this point. A calculation of the piston's velocity reveals that it decelerates abruptly within the braking chamber. Thus, the braking phase is defined by the piston's velocity.

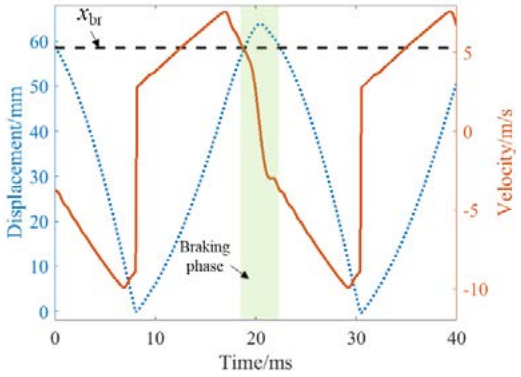


Fig 3.2 Braking phase in the motion of the impact piston

Based on Fig 3.2, the motion of the impact piston within the braking chamber (the braking phase) can be defined. To further analyze the relationship between piston motion and pressure changes during this process, the braking phase is subdivided according to the relative pressure relationship between the braking chamber and the rear chamber. The correlation between the piston's penetration distance into the braking chamber (x_{pb}), as well as the piston velocity, with the chamber pressure are separately shown in Fig 3.3.

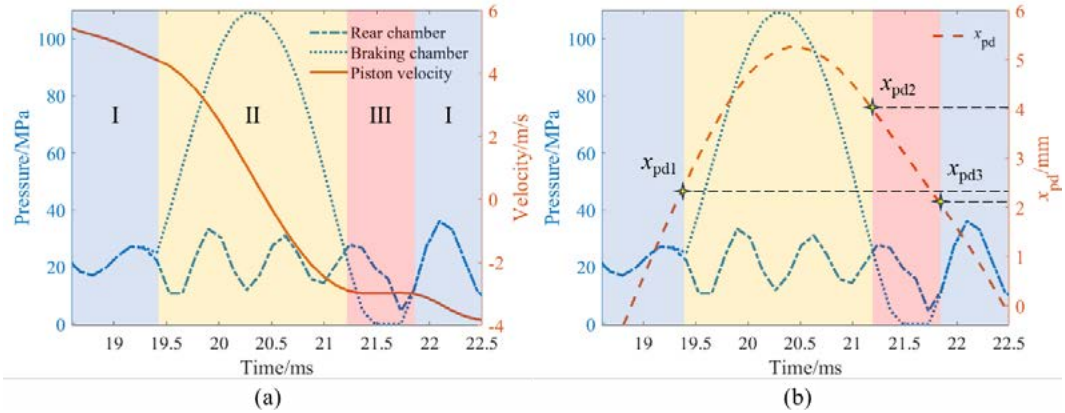


Fig 3.3 Motion and pressure variation in the rear braking chamber

As shown in Fig 3.3, the braking phase can be further divided into three states:

In part I, the braking chamber pressure equals the rear chamber pressure, the piston has entered the rear braking chamber less than 2 mm. During this state, the braking chamber does not form a sealed cavity. For most of the time, the pressure in the braking chamber equals that in the rear chamber. However, as the impact piston approaches the transition point x_{pd1} or x_{pd3}, a minor pressure difference arises between the braking chamber and the rear chamber due to localized throttling effects.

Then in part II, the braking chamber pressure exceeds the rear chamber pressure, the piston has entered the rear braking chamber between 2~5.3 mm. During this phase, the braking chamber is nearly fully sealed and functions approximately as a hydraulic spring on the impact piston, resulting in a linear change in piston velocity.

Part III occurs only when the impact piston exits the braking chamber, typically at a position about 4 mm within it. At this point, the braking chamber remains sealed, and the lack of fluid replenishment causes a sharp drop of the pressure in braking chamber as the piston moves. Consequently, the braking chamber no longer provides the outward thrust on the piston. Since the resistance from the front chamber pressure is relatively low, the piston exhibits a state of uniform motion during this phase.

Vapor cavitation is defined as the phase change resulting from a rapid drop in local pressure below the fluid's saturated vapor pressure. Hence the rapid pressure-drop in part II and the negative pressure in part III are the core elements of the appearance of the cavitation. Therefore, this study employs an experimental approach to investigate the influence of the clearance size on the pressure dynamics within the braking chamber, aiming to identify the optimal clearance value for mitigating cavitation in the current design.

4 Result analysis and discussion

Based on the analysis of the impact piston's motion within the braking chamber and the corresponding pressure variation in the chamber, it can be concluded that cavitation erosion on the rear guide sleeve is associated with the local negative pressure generated during the part III of the piston braking phase, as well as the piston's movement. During this state, the sealing length formed between the piston and the cylinder is insufficient to fully isolate the

braking chamber. As a result, high-pressure oil from the rear chamber flows into the braking chamber through this restricted passage via throttling. This flow, combined with the pressure drop induced by the piston's motion, ultimately leads to the occurrence of local cavitation.

To investigate the influence of different clearances on the pressure variation in the braking chamber and the motion of the piston within it, this study designed three sets of experiments with varying clearances for comparative analysis. In the experiment, the clearance between the braking chamber and the impact piston was controlled by adjusting the diameter tolerance of the cylinder's braking chamber. ISO VG 46 hydraulic oil was used in the tests, and the oil temperature was monitored via a temperature sensor installed in the tank. All experiments were conducted with an initial oil temperature of 60°C. Given the short duration of each test, it was assumed that the internal oil temperature of the rock drill remained consistent with the oil temperature in the tank.

The corresponding peak pressure (p_{brm}) in the braking chamber, the maximum penetration distance of the piston into the braking chamber (x_{pdm}), and time from piston entry into to exit from the braking chamber (t_{br}) for each clearance h are presented in Tab 4.1.

h /mm	p_{brm}/MPa	x_{brm}/mm	t_{br}/ms
0.02	114.1	1.6	1.7
0.04	120.5	5.25	3.8
0.083	108.6	5.17	3.3

Tab 4.1 Peak braking chamber pressure and critical impact piston motion parameters under different clearances

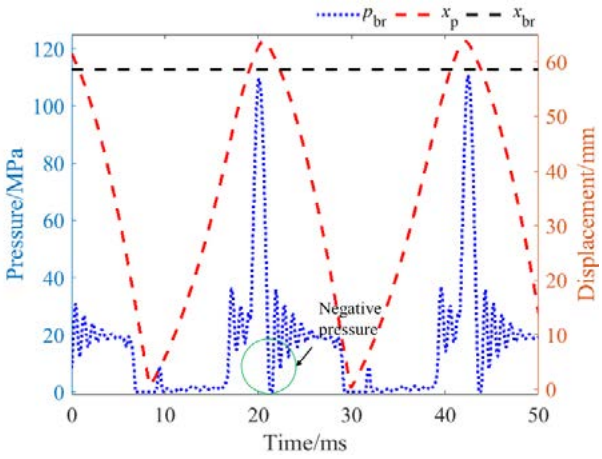


Fig 4.1 Clearance h = 0.083 mm, pressure and impact piston motion variations in the braking chamber

As shown in Fig 4.1, the original structural clearance h = 0.083 mm. Under this clearance, a high-pressure zone exceeding 100 MPa is formed in the braking chamber due to compression by the penetrating impact piston. Since the sealing length formed between the piston and the cylinder at this stage exceeds 5 mm, insufficient oil replenishment occurs in the braking chamber during the piston's outward movement, resulting in a negative pressure zone. This zone corresponds to the part III in Fig 3.3 and is the primary cause of cavitation in the braking chamber region.

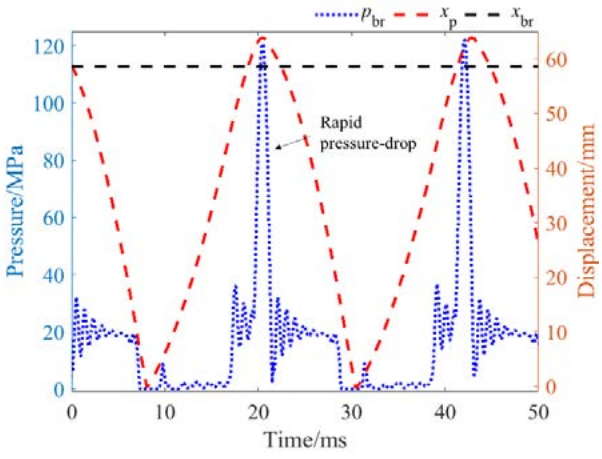


Fig 4.2 Clearance h = 0.04 mm, pressure and impact piston motion variations in the

braking chamber

As shown in Fig 4.2, the reduction in clearance accelerates the high-pressure buildup process in the

rear braking chamber, increase

ng the maximum peak pressure by approximately 12 MPa. However, the change in x_{brm} is minimal, so the pressure loss in the braking chamber caused by the outward motion of the impact piston remains essentially the same as when the h is 0.083 mm. Nonetheless, due to the increased p_{brm}, it does not enter a negative pressure state.

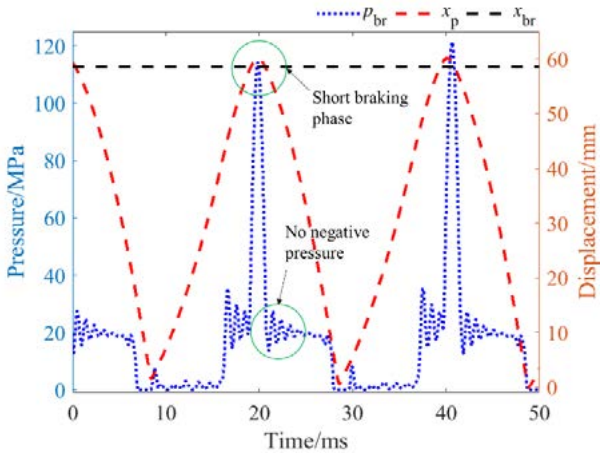


Fig 4.3 Clearance h = 0.02 mm, pressure and impact piston motion variations in the braking chamber

As shown in Fig 4.3, further reducing the clearance to 0.02 mm significantly decreases the sealing length required for the braking chamber to build high pressure. In this case, x_{brm} is only 1.6 mm, which is far smaller than the value under larger clearance conditions. Evidently, the increase in p_{brm} and the shortening of x_{brm} resulting from the reduced clearance jointly diminish the likelihood of the braking chamber entering a negative pressure range. Consequently, under this clearance condition, the rear braking chamber will be free from cavitation risk. However, as the tolerance for the guide sleeve is 0.015 mm, if the clearance h is set too close to this tolerance, it may impair the guiding function of the guide sleeve and significantly increase the probability of wear between the cylinder and the piston. Therefore,

the actual selected clearance value should be greater than twice this tolerance.

Based on the above analysis, it can be determined that the ideal clearance value should fall within the range of 0.03 to 0.083 mm. To explore the optimal parameters within this range, the impact performance of each group of rock drills was compared, and the results are presented in Tab 4.2 and Fig 4.4.

h	Stroke/mm	Impact energy/J	Impact frequency/Hz	Impact power/kW
0.04	60.1	361	49.4	17.8
0.06	63.5	327	46.8	15.3
0.083	63.6	331	46.3	15.3

Tab 4.2 Hydraulic rock drill impact performance under different clearances

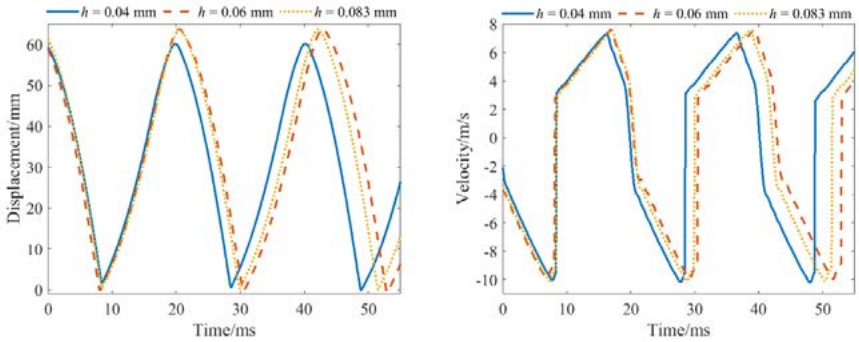


Fig 4.4 Hydraulic rock drill impact performance under different clearances

Based on Fig 4.1~Fig 4.4 and Tab 4.2, when the clearance value is between 0.06 and 0.08 mm, reducing the clearance only increases p_{brm} and lowers the risk of the braking chamber entering a negative pressure state during the piston's outward motion. However, when the clearance is reduced to 0.04 mm, this reduction significantly affects the sealing length required for the rear braking chamber to close, shortening x_{brm} and thereby reducing the total travel of the impact piston and increasing the impact frequency. At the same time, due to the reduction in energy loss within the braking chamber, the impact energy per piston stroke is also enhanced, leading to an overall improvement in the impact power of the rock drill.

5 Conclusion

This study, addressing the cavitation erosion issue in the rear braking chamber of a rock drill's cylinder, systematically investigated the influence of different clearance fits on the chamber's pressure dynamics in response to piston motion. The following principal conclusions were drawn:

(1) Cavitation in the braking chamber of a rock drill primarily occurs during the outward motion of the impact piston from the chamber. When the piston is approximately 4 mm away from exit the braking chamber, the braking chamber and the rear chamber gradually become connected. Due to the extremely low pressure in the braking chamber at this stage, the high-pressure oil from the rear chamber rushes into the braking chamber at high speed, leading to cavitation in the oil. This, in turn, causes vapor erosion on the adjacent guide sleeve.

(2) The clearance primarily influences the x_{brm} by affecting the rate of pressure buildup within the chamber. When the x_{brm} is less than 2 mm, the pressure drop in the braking chamber induced by the piston during the impact stroke is insufficient to create a negative pressure zone, thus preventing oil cavitation. However, if the clearance approaches the tolerance of the guide sleeve, it can impair the sleeve’s guiding function, leading to wear between the piston and the cylinder and reducing the piston's service life. Therefore, based on a comprehensive analysis of the experimental results, the optimal clearance for the model studied in this paper is approximately 0.04 mm.

6 Reference

[1] X. Sun, J. Cheng, X. Zhang, Z. Wan, Y. Lian, K. Liu, Mechanism of Drilling Parameter Distortion Induced by Drill Rod Buckling While MWD Application, Rock Mechanics and Rock Engineering 58 (8) (2025) 9459–9477, https://doi.org/10.1007/s00603-025-04579-5.

[2] H. Sun, M. Zhu, Y. Dai, X. Li, Y. Rui, W. Zhu, H. Zhu, Digital twin of advanced geological conditions in rock tunnel construction, Automation in Construction 178 (2025) 106407, https://doi.org/10.1016/j.autcon.2025.106407.

[3] L. Gorodilov, Mathematical models of hydraulic percussion systems, Journal of mining science 41 (5) (2005) 475–489, https://doi.org/10.1007/s10913-006-0010-2.

[4] Q. Hu, C. Yang, H. Zheng, M. Chen, Dynamic simulation and test research of impact performance of hydraulic rock drill with no constant-pressurized chamber, Automation in Construction 37 (2014) 211–216, https://doi.org/10.1016/j.autcon.2013.08.011.

[5] S.-y. Yang, Y.-b. Ou, Y. Guo, X.-m. Wu, Analysis and optimization of the working parameters of the impact mechanism of hydraulic rock drill based on a numerical simulation, International Journal of Precision Engineering and Manufacturing 18 (7) (2017) 971–977, https://doi.org/10.1007/s12541-017-0114-4.

[6] G. Yong, L. Deshun, Y. Shuyi, L. Xuejun, C. Jianwen, Hydraulic–mechanical coupling modeling by bond graph for impact system of a high frequency rock drill drifter with sleeve distributor, Automation in Construction 63 (2016) 88–99, https://doi.org/10.1016/j.autcon.2015.12.012.

[7] X. Tian, W. Xi, F. Ma, Z. Zhang, S. Wang, X. Geng, Experimental study on impact coupling characteristics of high-pressure accumulator for unconstant-pressurized chamber rock drills, Advances in Mechanical Engineering 16 (10) (2024) 16878132241289264, https://doi.org/10.1177/16878132241289264.

[8] C. Jia, A. Wang, L. Zong, L. Yu, Structural optimization and performance improvement of rock drill seals based on orthogonal test, Journal of Engineering and Applied Science 71 (1) (2024) 14, https://doi.org/10.1186/s44147-023-00354-x.

[9] X. Geng, W. Ma, F. Ma, Z. Zhou, Y. Liu, X. Tian, Effects of Double Damper System's Floating Characteristic on Percussion Performance of Hydraulic Rock Drill, IOP Conference Series Materials Science and Engineering 394 (2018), https://doi.org/10.1088/1757-899X/394/3/032060.

[10] X. Geng, W. Ma, F. Ma, Z. Zhou, Y. Liu, X. Tian, Cavitation erosion of the damping piston in double damping system of hydraulic rock drill and its influencing factors, AIP Advances 8 (11) (2018), https://doi.org/10.1063/1.5058162.