TECHNISCHE UNIVERSITÄT WIEN

DISSERTATION

Spline-based Methods for Fluid-Structure-Contact Interaction in Orthogonal Cutting

carried out for the purpose of obtaining the degree of Doctor technicae (Dr. techn.), submitted at TU Wien, Faculty of Mechanical and Industrial Engineering, by

Jaewook Lee
Matrikelnummer: 11945925

under the supervision of

Univ.Prof. Dr.-Ing. Stefanie Elgeti
Institute of Lightweight Design and Structural Biomechanics, E317

reviewed by

Prof. Marek Behr, Ph.D.
Chair for Computational Analysis of Technical Systems
RWTH Aachen, Germany

Prof. Dr. sc. habil. Julia Kowalski
Chair of Methods for Model-based Development in Computational Engineering
RWTH Aachen, Germany

Preface

This project has received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) — Projektnummer 439919057

Affidavit

I declare in lieu of oath, that I wrote this thesis and carried out the associated research myself, using only the literature cited in this volume. If text passages from sources are used literally, they are marked as such.

I confirm that this work is original and has not been submitted for examination elsewhere, nor is it currently under consideration for a thesis elsewhere.

I acknowledge that the submitted work will be checked electronically-technically using suitable and state-of-the-art means (plagiarism detection software). On the one hand, this ensures that the submitted work was prepared according to the high-quality standards within the applicable rules to ensure good scientific practice "Code of Conduct" at the TU Wien. On the other hand, a comparison with other student theses avoids violations of my personal copyright.

Vienna, November 2025

Jaewook Lee

Summary

In high-performance machining, process efficiency and product quality are dependent on tribological interactions at the tool-chip interface, known as the secondary shear zone. While cooling lubricants are employed to mitigate the extreme thermo-mechanical loads in this region, the precise physical mechanisms governing their behavior remain obscured from direct experimental observation. Furthermore, current industrial simulation tools rely on empirical, "black-box" friction models that lack the physical fidelity required for predictive process optimization. This dissertation addresses this fundamental gap by developing a novel, high-fidelity numerical framework for modeling the Fluid-Structure-Contact Interaction at the microscopic level and exploring approaches for reduced order modeling in this context.

This work begins by identifying the specific modeling requirements through experimental investigation. Orthogonal cutting tests reveal that while lubrication significantly reduces loads at low speeds, this effect diminishes at high speeds, highlighting the limitations of current experimental and numerical capabilities. To resolve this, a spline-based numerical framework is proposed and developed. This framework utilizes Isogeometric Analysis for the thermo-elasto-plastic solid and standard Finite Elements for the incompressible fluid, coupled via a partitioned Robin-Neumann scheme to robustly simulate the interaction, including the fully enclosed fluid pockets.

The fully coupled, high-fidelity model is employed to investigate the governing tribological mechanisms at the contact interface. Specifically, this investigation aims to verify the hypothesis that friction in the secondary shear zone is dominated by the mechanical interlocking of microscopic surface topographies. However, comparisons with experimental data reveal fundamental discrepancies regarding the influence of contact pressure and relative velocity. This outcome challenges the initial assumption of mechanical dominance, demonstrating that mechanical interaction alone is insufficient to explain the observed behavior and indicating that there are other physical mechanisms that play significant roles.

Finally, to bridge the gap between high-fidelity simulation and practical engineering application, this work develops two reduced model strategies. First, a scalar surrogate model based on Gaussian Process Regression is implemented, achieving a computational speed-up of over 10^7 for multi-scale coupling. Second, a methodology for generating full-field reduced models is established. Leveraging deep learning architectures developed in a parallel project, it is demonstrated that Deep Neural Networks can effectively learn compact latent-space representations of continuous functions defined in a spatial domain, offering a promising direction for future real-time, full-field "digital twins".

Kurzfassung

In der Hochleistungszerspanung werden Prozesseffizienz und Produktqualität maßgeblich von den komplexen tribologischen Wechselwirkungen in der sekundären Scherzone bestimmt. Obwohl Kühlschmierstoffe eingesetzt werden, um die extremen thermomechanischen Belastungen in diesem Bereich zu senken, entziehen sich die zugrundeliegenden physikalischen Wirkmechanismen der direkten experimentellen Beobachtung. Industrielle Simulationswerkzeuge greifen daher meist auf empirische „Black-Box"-Modelle zurück, denen jedoch die nötige physikalische Tiefe für eine prädiktive Prozessoptimierung fehlt. Die vorliegende Arbeit begegnet diesem Defizit durch die Entwicklung eines neuartigen, hochauflösenden numerischen Frameworks zur mikroskopischen Modellierung der Fluid-Struktur-Kontakt-Interaktion sowie durch die Erarbeitung effizienter reduzierter Modelle.

Ausgehend von experimentellen Untersuchungen werden zunächst die spezifischen Anforderungen an die Modellierung definiert. Orthogonalschnittversuche zeigen, dass die lastreduzierende Wirkung von Kühlschmierstoffen bei hohen Schnittgeschwindigkeiten abnimmt, was die Grenzen aktueller experimenteller und numerischer Methoden verdeutlicht. Zur Auflösung dieser Phänomene wird ein spline-basiertes numerisches Framework entwickelt. Dieses kombiniert die Isogeometrische Analyse für den thermo-elasto-plastischen Festkörper mit der Finite-Elemente-Methode für das inkompressible Fluid. Eine partitionierte Robin-Neumann-Kopplung gewährleistet dabei die robuste Simulation der Interaktion, insbesondere innerhalb vollständig eingeschlossener Fluidtaschen.

Das Framework wird anschließend genutzt, um die dominierenden Reibungsmechanismen zu untersuchen. Im Fokus steht die Überprüfung der Hypothese, dass Reibung primär durch die mechanische Verzahnung mikroskopischer Oberflächentopographien entsteht. Der Abgleich mit experimentellen Daten offenbart jedoch fundamentale Diskrepanzen hinsichtlich des Einflusses von Kontaktdruck und Relativgeschwindigkeit. Dieses Ergebnis widerlegt die Annahme einer rein mechanischen Dominanz und indiziert, dass weitere physikalische Effekte eine tragende Rolle spielen.

Um die Lücke zwischen hochauflösender Simulation und praktischer Anwendbarkeit zu schließen, werden zwei Strategien für reduzierte Modelle vorgestellt. Erstens ermöglicht ein skalares Ersatzmodell auf Basis von Gauß-Prozess-Regression eine Beschleunigung um den Faktor 10^7 für Multiskalen-Kopplungen. Zweitens wird ein methodischer Ansatz für reduzierte Vollfeld-Modelle etabliert. Unter Rückgriff auf Deep-Learning-Architekturen eines Parallelprojekts wird demonstriert, dass Tiefe Neuronale Netze in der Lage sind, komplexe kontinuierliche Feldgrößen effizient in kompakten Latent-Spaces zu repräsentieren – ein vielversprechender Schritt in Richtung echtzeitfähiger digitaler Zwillinge.

Contents

List of Figures

Acronyms

ALE Arbitrary Lagrangian-Eulerian. XV, 10

CFD Computational Fluid Dynamics. 64

DNN Deep Neural Network. 5, 12, 15, 45

EMUM Elastic Mesh Update Method. 10, 15

FEM Finite Element Method. 10, 14, 25

FFD Free-Form Deformation. 15

FSCI Fluid-Structure-Contact Interaction. IX, 3, 5, 6, 11, 13, 14, 15, 16, 25, 32, 63, 64

FSI Fluid-Structure Interaction. 14, 15

GPR Gaussian Process Regression. 5, 11, 15, 32, 63

IGA Isogeometric Analysis. 7, 9, 14, 25, 32

MQL Minimum Quantity Lubrication. 2, 63

NPP Nearest Point Projection. 9, 14, 25

PSZ Primary Shear Zone. 2

SSZ Secondary Shear Zone. 2, 13

Notation

Greek Symbols

α^R	Robin boundary condition coefficient
$\boldsymbol{\sigma}^f$	Fluid Cauchy stress tensor
$\boldsymbol{\xi}$	Parameter coordinate vector $(\xi_1, \ldots, \xi_D)$
χ	Reference coordinate (ALE formulation of Navier-Stokes equation)
$\dot{\bar{\epsilon}}^{pl}$	Equivalent plastic strain rate
$\dot{\epsilon}_0$	Reference strain rate
η	Dynamic fluid viscosity
Γ	Boundary of a domain (Γ_D: Dirichlet, Γ_N: Neumann, Γ_R: Robin)
$\hat{\theta}$	Non-dimensional temperature (Johnson-Cook hardening law)
μ	Scalar friction coefficient
Ω	Computational domain (Ω^s: Solid, Ω^f: Fluid)
ρ^s, ρ^f	Density of solid and fluid, respectively
σ_y	Yield stress
θ_{melt}	Melting temperature
$\theta_{transition}$	Transition temperature
Ξ_d	Knot vector associated with parametric direction d
ξ_d	Parameter coordinate in direction d

Latin Symbols

$\mathbf{b}^s, \mathbf{b}^f$	Body force density vector for solid and fluid
$\mathbf{C_i}$	Control point coefficient associated with multi-index $\mathbf{i}$
$\mathbf{I}$	Identity tensor
$\mathbf{i}$	Multi-index $\mathbf{i} = (i_1, \ldots, i_D)$
$\mathbf{n}$	Outward unit normal vector
$\mathbf{P}$	First Piola-Kirchhoff stress tensor
$\mathbf{u}$	Solid displacement vector
$\mathbf{v}$	Fluid velocity vector
$\mathbf{v}_\#$	Mesh velocity vector (ALE formulation)
$\mathcal{N}_\mathbf{i}$	Tensor-product B-spline basis function
$\mathcal{S}$	Spline mapping function
A, B, n, C, m	Material parameters for the Johnson-Cook hardening law
D	Dimension of the parameter space ($D = 2$ for surfaces, $D = 1$ for curves)
g_n	Normal gap function (contact)
k_j	j-th knot value in a knot vector
M	Dimension of the physical (target) space ($M = 2$ in this work)
$N_{i,p}$	Univariate B-spline basis function of index i and degree p
p	Fluid pressure
p_d	Polynomial degree in parametric direction d
t	Time

Publications and Copyrights

Results shown in this thesis have been published in the following articles:

- J Saelzer, Y Alammari, A Zabel, D Biermann, J Lee, and S Elgeti. "Characterisation and modelling of friction depending on the tool topography and the intermediate medium". en. In: *Procedia CIRP* 102 (2021), pp. 435–440. ISSN: 22128271. DOI: 10.1016/j.procir.2021.09.074
- Andreas Zabel, Jannis Saelzer, Gerhard Poll, Florian Pape, Haichao Liu, Berend Denkena, Lars Ellersiek, Stefanie Elgeti, and Jaewook Lee. "Effizientes Schmieren der Spanbildungszone/Controlled use of cooling lubricants for tribological optimization of machining processes – Efficient lubrication of the chip formation zone". In: *wt Werkstattstechnik online* 112.01–02 (2022), pp. 44–49. ISSN: 1436-4980. DOI: 10.37544/1436-4980-2022-01-02-48
- J Lee, S Elgeti, J Saelzer, and A Zabel. "Development and Validation of a Micro-Tribology Model for the Chip Formation Zone". en. In: *Procedia CIRP* 117 (2023), pp. 317–322. ISSN: 22128271. DOI: 10.1016/j.procir.2023.03.054
- Jaewook Lee, Jannis Saelzer, Jacques Zwar, Florian Zwicke, Felipe Gonzalez, Thomas Spenke, Norbert Hosters, Gero Polus, Andreas Zabel, and Stefanie Elgeti. "Spline-Based Framework for Microscopic Contact Zone Modeling in Lubricated Orthogonal Cutting". en. In: *International Journal for Numerical Methods in Engineering* 126.14 (July 2025), e70087. ISSN: 0029-5981, 1097-0207. DOI: 10.1002/nme.70087
- Jaewook Lee, Sebastian Hube, and Stefanie Elgeti. "Neural networks vs. splines: advances in numerical extruder design". en. In: *Engineering with Computers* 40.2 (Apr. 2024), pp. 989–1004. ISSN: 0177-0667, 1435-5663. DOI: 10.1007/s00366-023-01839-2

1 Introduction

In computational engineering, a central and persistent challenge lies in replacing legacy "black-box" models with simulations that are truly predictive. For decades, many complex industrial processes –from fluid dynamics to structural mechanics – have been successfully approximated using empirical models, calibrated from expensive, real-world data. While effective within their calibrated boundaries, these models are fundamentally inflexible; they fail when new materials, new geometries, or new operating conditions are introduced. Therefore, the natural way forward is the development of predictive simulations based on physics that are not reliant on scenario-specific empirical calibration.

This dissertation is a direct engagement with that challenge, set in the demanding world of high-performance metal cutting. This process is dominated by extreme thermo-mechanical loads and complex tribological phenomena – such as friction and wear – that are concentrated at the tool-chip interface. The behavior in this microscopic region, where the tool, chip, and any present lubricants interact, is the determining factor for critical outcomes like tool life, final product quality, and overall process efficiency. A complete understanding of its behavior is limited by a persistent gap in knowledge at this microscopic contact zone. The critical phenomena occurring in this region are hidden from direct observation, which is why existing simulation tools typically substitute these complex, coupled physics with simplified, empirical "black-box" models. This work documents the series of investigations to build a new framework of simulation tools to resolve the complex physics within this inaccessible region. It details the development of a novel, high-fidelity framework to model the complex, coupled physics of the process, and explores the methodologies required to achieve the computational speed necessary for practical engineering use.

1.1 Background

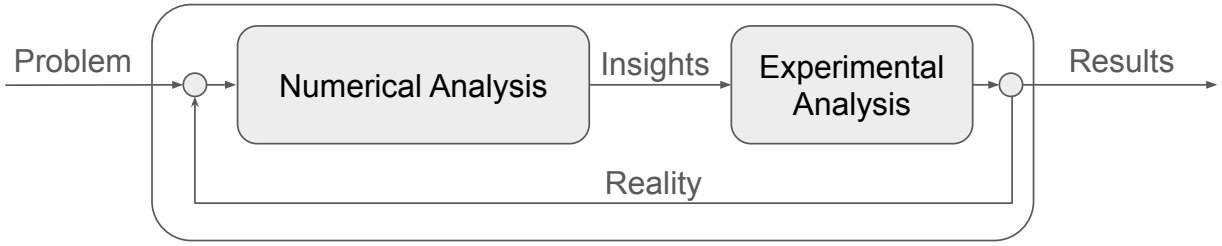
In any machining process there are three elements that interact: workpiece, tool, and chip. The workpiece describes the base metal, from which material will be subtracted in order to achieve a final form/product. Its counter body is the tool, with which one can subtract the designated volume. During machining, the material volume or mass that is removed is referred to as chips. Every type of machining operation – such as milling, turning, or drilling – produces these chips. The formation of chips is not merely a byproduct; it serves as a direct indicator of process stability and overall operational efficiency [25, 22]. Consequently, understanding and controlling chip formation is one of the central objectives in manufacturing research. To study this phenomenon, researchers often employ orthogonal cutting, which represents a generalized, idealized two-dimensional abstraction of the machining process. This approach provides a controlled environment that allows individual process variables to be isolated and analyzed more effectively.

In orthogonal cutting, the transformation of the workpiece into a chip occurs across two distinct regions. The process begins in the Primary Shear Zone (PSZ), where the material undergoes intense plastic deformation in a very short time. This is where the majority of the mechanical work is converted into heat, fundamentally defining the initial chip geometry and flow [25, 22]. However, while the PSZ sets the initial stage, it is the subsequent interaction in the Secondary Shear Zone (SSZ) that governs the phenomena most critical to tool life and surface quality. In the SSZ, the newly formed chip slides against the tool face, creating an interface dominated by extreme friction. It is this complex tribological phenomenon that lubrication aims to control. Therefore, a high-fidelity, physics-based model of the SSZ is crucial to better understand the friction mechanisms and the effectiveness of the lubricants.

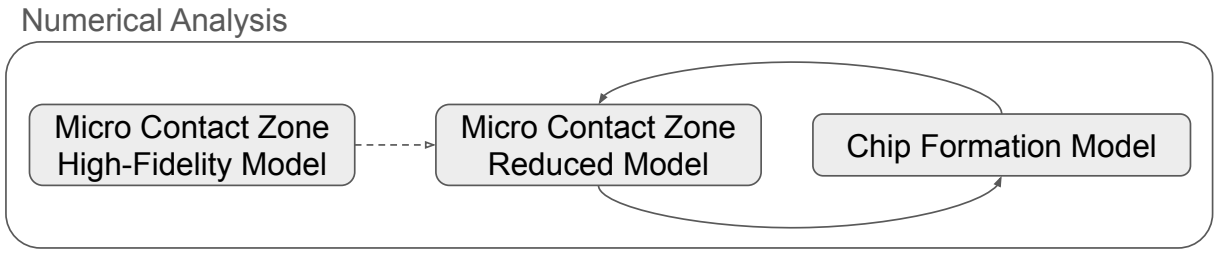
The essence of lubrication is that it reduces thermo-mechanical loads at the contact zone [10, 12], which most importantly improves product quality and tool life span. Regarding lubrication strategies, one conventional approach is the mass application of lubricant during production [26]. However, the benefits of this method are offset by its high production costs and negative environmental impact. To mitigate these drawbacks, research has increasingly focused on Minimum Quantity Lubrication (MQL) – an approach that seeks to achieve maximum lubrication effectiveness with minimal lubricant usage. How MQL can be achieved naturally depends on the production process; consequently, there is a need to better understand effects of the lubrication and its mechanisms. To this end, experimental investigations by Saelzer et al. [20] have provided valuable information on the effects of lubrication. Their work demonstrated that while lubrication effectively reduces forces at lower cutting speeds, its benefits diminish significantly at industrially-relevant high speeds. This suggests that thermal barriers may prevent the lubricant from reaching the critical tool-chip interface, highlighting a critical gap in our understanding of the lubrication mechanisms under these extreme conditions.

The main difficulty in the experimental analysis of such tribological interactions is observability. Accurately measuring complex dynamics of tool, chip, and lubricants becomes physically intractable – for example, observations are only achievable through pausing the interaction, which alters the true state that we aim to measure. To this end, numerical analysis – a complementary counterpart of experimental analysis – becomes a powerful alternative. Instead of designing elaborate experiments to observe the effects of specific changes, we can adopt a modeling approach using numerical methods to incorporate and

analyze the factors of interest. These numerical models can provide us with insight that is otherwise not available, making the aforementioned tribological interactions ideal candidates for numerical modeling, as they are difficult to assess experimentally. This is the starting point for this work. However, current simulation frameworks are not ideal to model the underlying physics of the lubricated contact zone. Commercial software packages [4, 21, 1], which are widely used for meso-scale chip formation analysis, typically treat complex tribological interactions in the secondary shear zone as a "black-box". These simulations rely on simplified empirical models, such as a state-dependent friction coefficient, which must be calibrated from expensive and sparse experimental data. While useful for reproducing known scenarios, these models lack the predictive power and generalizability required for true process optimization. This fundamental limitation creates a demand for a more flexible and robust numerical framework capable of modeling the coupled fluid-structure-contact interaction from the ground up.



(a) Description of experimental and numerical analysis feedback loop.



(b) Context of numerical analysis framework in this work.

Figure 1.1: The engineering analysis workflow. The overall process (a) involves an iterative feedback loop, where numerical and experimental modules provide inputs to each other until convergence. This dissertation focuses on developing the numerical framework (b) that serves as the high-fidelity computational counterpart in the loop.

This work is driven by a long-term vision to establish a tightly coupled experiment-simulation feedback loop for tool and process optimization (Figure 1.1a). To establish the computational foundations for this goal, this dissertation develops a framework combining micro-scale multiphysics and reduced modeling to balance fidelity with efficiency.

The core of this strategy is the development of a high-fidelity Fluid-Structure-Contact Interaction (FSCI) framework designed to resolve the microscopic contact physics. This framework also serves as the foundation for efficient reduced models. This work explores two approaches: a physics-based scalar model for immediate multi-scale coupling, and a data-driven full-field methodology. The latter leverages deep learning architectures developed in a parallel project [11], whose applicability to vector fields was successfully validated in a dedicated Master's thesis [19].

The following sections will detail the objectives, methodologies, and the specific contributions of the articles that investigate the aforementioned goals.

1.2 Research Questions and Objectives

Based on the challenges and research gaps identified in the previous section – namely, the limitations of experimental observation at high cutting speeds and the lack of a predictive, physics-based model for the secondary shear zone – this dissertation is driven by two central questions: How can a high-fidelity, physics-based numerical framework be developed to model the FSCI in the lubricated secondary shear zone, and how can the complex data from this model be translated into computationally efficient reduced models for practical engineering use?

To answer these questions, the following specific objectives were established:

O.1 High-Fidelity FSCI Modeling Framework

Develop and implement a foundational, high-fidelity, spline-based numerical framework capable of modeling fully coupled FSCI in the secondary shear zone, including a realistic thermo-elasto-plastic material model.

O.2 Tool-Chip Interface Tribology Investigation

Apply the developed framework to investigate the tribological phenomena at the tool–chip interface. This objective involves a systematic analysis of the influence of key process parameters (e.g., pressure, relative velocity, temperature) and an evaluation of the central "mechanics-dominant" friction hypothesis.

O.3 Efficient Reduced Model Development

Investigate and propose computationally efficient pathways for practical application by developing two distinct model reduction strategies: a) A scalar surrogate model based on Gaussian Process Regression (GPR), designed to provide a friction coefficient for FE^2 coupling with very low computational effort; b) A full-field surrogate methodology based on DNN, aimed at developing a novel approach for compressing and reconstructing complex vector field data.

1.3 A Framework for Modeling Microscopic Contact Zones

The research objectives requires a model that can resolve microscopic surface topographies while simultaneously simulating the strongly coupled physics of solids (the chip and the tool) and an incompressible fluid (the lubricant). We aim to address this FSCI problem, with spline-based approaches for both geometric representation and numerical formulation. Splines are chosen for two primary reasons: first, they provide a geometrically exact representation of the measured microscopic topographies; second, their inherent high-order continuity (smoothness) is essential for accurately modeling the complex interface physics, such as contact mechanics and fluid boundary layers, which rely on well-defined surface normals.

With respect to the numerical methods, the focus is placed on boundary-conforming techniques. This choice ensures that the fluid mesh aligns precisely with the solid boundary's spline-based geometry. For a strongly coupled FSCI problem, this conforming interface is a significant advantage, as it allows for the direct and accurate exchange of physical quantities defined in space – such as displacements and tractions – without the geometric approximation errors or complex projection algorithms associated with non-conforming or immersed methods.

This section provides a high-level overview of the framework's architecture, which constitutes the core methodological contribution of this dissertation. We will introduce B-Splines, present the model equations for both solid and fluid mechanics, as well as treatments for interface interactions. At last, two types of reduced models will be described.

1.3.1 An Introduction to B-Splines

A B-Spline is a function constructed from a piecewise polynomial basis. To understand the general multi-dimensional mapping, it is essential to first define the fundamental one-dimensional (univariate) building blocks.

Univariate B-Spline. A 1D B-Spline, which can be visualized as a curve, is defined by a polynomial degree p and a knot vector $\Xi = \{\xi_1, \ldots, \xi_{k+p+1}\}$ (where k is the number of control points). The knot vector is a non-decreasing sequence of coordinates in the parameter space that divides the domain into intervals. It serves two purposes: it defines the parametric range of the curve and controls the continuity of the basis functions. Specifically, at a knot value with multiplicity m (i.e., a knot repeated m times), the basis function is C^{p-m} continuous.

Given a knot vector and degree, the univariate B-Spline basis functions $N_{i,p}(\xi)$ are defined recursively using the Cox-de Boor recursion formula [18]. A 1D curve is then constructed as a linear combination of these basis functions and a set of control points.

Tensor-Product B-Spline. This work utilizes tensor-product B-Splines to represent the solid domain. It generalizes the 1D concept to define a mapping from a D-dimensional parameter space $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_D)$ to an M-dimensional physical space. This mapping is constructed by combining two key components:

1. **Tensor-Product Basis Functions $\mathcal{N}_{\mathbf{i}}(\boldsymbol{\xi})$**: These are the scalar-valued piecewise polynomials defined over the multi-dimensional parameter space. They are formed by the product of D independent univariate basis functions:

$$\mathcal{N}_{\mathbf{i}}(\boldsymbol{\xi}) = \prod_{d=1}^{D} N_{i_d,p_d}(\xi_d), \tag{1.1}$$

   where each direction d has its associated degree p_d and knot vector Ξ_d.

2. **Control Points $\mathbf{C}_{\mathbf{i}}$**: These coefficients form a D-dimensional regular grid within the target space (e.g., $\mathbb{R}^M$). The physical meaning of these coefficients depends on the context. While in geometric design they represent spatial coordinates to define shapes, in the context of Isogeometric Analysis (IGA), they serve as generalized degrees of freedom. In this work, they are used to define both the geometry and the physical solution fields.

The final mapping $\mathcal{S}(\boldsymbol{\xi})$ is computed as a sum of weighted coefficients, analogous to the 1D case, but summed over the multi-index $\mathbf{i} = (i_1, \ldots, i_D)$:

$$\mathcal{S}(\boldsymbol{\xi}) = \sum_{\mathbf{i}} \mathcal{N}_{\mathbf{i}}(\boldsymbol{\xi}) \mathbf{C}_{\mathbf{i}}. \tag{1.2}$$

This work focuses on analysis in a two-dimensional physical space ($M = 2$). The geometry is defined in two topological contexts:

- The solid domain is represented as bivariate B-Spline surfaces (where $D = 2$),
- The contact and fluid-structure interfaces are modeled as univariate B-Spline curves (where $D = 1$).

Crucially, the latter are directly extracted from the domain boundaries, maintaining exact geometric conformity.

1.3.2 Solid Mechanics

The solid domains that represent the chip and tool are constructed directly from experimental surface topography measurements. As these measurements consist of height values mapped to planar coordinates, they effectively constitute point cloud data. Consequently, this raw input is converted into B-Spline geometries using surface fitting techniques [18], ensuring that the model accurately reflects the surfaces measured experimentally. To generate the full computational domain from this boundary representation, the fitted B-Spline geometry is extruded to form the volumetric solid domain required for the analysis.

Following standard mechanical assumptions for orthogonal cutting, the deformation is assumed to be uniform along the cutting width and constrained from lateral expansion. This allows the problem to be idealized as a state of plane strain, reducing the physical space to two dimensions.

Adopting a total Lagrangian framework, the displacement field $\mathbf{u}$ is expressed with respect to reference (undeformed) configuration Ω^s_{ref}. The equation of motion governing the solid domain is given by:

$$\begin{aligned} \rho^s \frac{d^2\mathbf{u}}{dt^2} &= \nabla_{ref} \cdot \mathbf{P} + \mathbf{b}^s && \text{in} \quad \Omega^s_{ref}, \\ \mathbf{u}(t=0) &= \mathbf{u}_0 && \text{in} \quad \Omega^s_{ref}, \\ \mathbf{u} &= \mathbf{g}^s && \text{on} \quad \Gamma^s_D, \\ \mathbf{P}\mathbf{n}^s_{ref} &= \mathbf{h}^s && \text{on} \quad \Gamma^s_N, \end{aligned}$$

where:

- ρ^s is the density,
- $\mathbf{b}^s$ is the body force density,
- $\mathbf{P}$ is the first Piola-Kirchhoff stress,
- ∇_{ref} is the gradient with respect to the reference configuration,
- Γ^s_D and Γ^s_N are the Dirichlet and Neumann parts of the boundary, respectively,
- $\mathbf{g}^s$ and $\mathbf{h}^s$ are the prescribed boundary displacement and traction,
- $\mathbf{n}^s_{ref}$ is the outward unit normal vector.

The plasticity formulation utilizes the J_2 flow theory at finite strains presented by Simo and Hughes [23]. Thermal effects are treated as adiabatic. This simplification is justified by the high cutting speeds, where the timescale of plastic deformation is significantly shorter than that of thermal conduction, effectively confining the generated heat to the local deformation zone throughout the microsecond-scale duration of the simulation. To capture the resulting thermal softening, strain hardening, and strain-rate sensitivity, the yield stress σ_y is defined by the Johnson-Cook hardening law [7]:

$$\sigma_y = \left[A + B\left(\bar{\epsilon}^{pl}\right)^n\right]\left[1 + C\ \ln\left(\frac{\dot{\bar{\epsilon}}^{pl}}{\dot{\epsilon}_0}\right)\right]\left(1 - \hat{\theta}^m\right), \tag{1.3}$$

with

$$\hat{\theta} \equiv \begin{cases} 0 & \text{for} \quad \theta \leq \theta_{transition}\ , \\ {}^{(\theta-\theta_{transition})}/_{(\theta_{melt}-\theta_{transition})} & \text{for} \quad \theta_{transition} \leq \theta < \theta_{melt}\ , \\ 1 & \text{for} \quad \theta_{melt} \leq \theta\ , \end{cases} \tag{1.4}$$

where:

- σ_y is the yield stress,
- $\bar{\epsilon}^{pl}$ is the equivalent plastic strain,
- $\dot{\bar{\epsilon}}^{pl}$ is the equivalent plastic strain rate,
- $\hat{\theta}$ is non-dimensional temperature,
- $A, B, n, m, \theta_{melt}$, and $\theta_{transition}$ are material parameters.

The solid-solid interaction between the tool and the chip is modeled using a mortar knot-to-surface formulation [24, 5]. A critical challenge within spline-based contact lies

in the Nearest Point Projection (NPP), particularly when dealing with high-order splines. This step involves finding the closest point on the opposing surface, which is formulated as a nonlinear optimization problem that minimizes the squared distance between the query point and its projection point. To ensure robust convergence, a fallback sequence of solution algorithms is employed: Newton's method as the primary method, followed by sequential quadratic programming [8], and, if necessary, the Levenberg-Marquardt algorithm [15]. The use of high-order splines with a minimum of C^2-continuity is critical in this context; it guarantees that both the normal vector (required for the contact formulation) and the second derivatives (required for Newton's method) are continuously defined across the entire contact surface.

The governing equations for solid mechanics, including contact constraints, are discretized in space using IGA, which employs the same tensor-product B-Spline basis functions defined previously (Section 1.3.1). For time integration, a *generalized-α* scheme [3] is used.

1.3.3 Fluid Mechanics

The fluid domain represents the lubricant. The primary focus of this work is on the mechanically dominant phenomena driving the contact interaction. To clearly observe these mechanical influences – specifically the transmission of the fluid forces to the solid – the lubricant is modeled as an incompressible Newtonian fluid. Using Eulerian viewpoint, the Navier-Stokes equation is expressed using fluid velocity $\mathbf{v}$ and pressure p:

$$\rho^f \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) - \nabla \cdot \boldsymbol{\sigma} = \rho^f \mathbf{b}^f \qquad \text{in } \Omega^f, \tag{1.5}$$

$$\nabla \cdot \mathbf{v} = 0 \qquad \text{in } \Omega^f, \tag{1.6}$$

$$\mathbf{v}(t=0) = \mathbf{v}_0 \qquad \text{in } \Omega^f, \tag{1.7}$$

$$\mathbf{v} = \mathbf{g}^f \qquad \text{on } \Gamma_D^f, \tag{1.8}$$

$$\boldsymbol{\sigma}\mathbf{n}^f = \mathbf{h}^f \qquad \text{on } \Gamma_N^f, \tag{1.9}$$

$$\alpha^R \mathbf{v} + \boldsymbol{\sigma}\mathbf{n}^f = \mathbf{h}^f + \alpha^R \mathbf{g}^f \qquad \text{on } \Gamma_R^f, \tag{1.10}$$

with stress tensor:

$$\boldsymbol{\sigma} = -p\mathbf{I} + \eta\big(\nabla \mathbf{v} + (\nabla \mathbf{v})^T\big) \tag{1.11}$$

where:

- ρ^f is the fluid density,
- $\mathbf{b}^f$ is the body force density,
- p is the pressure,
- η is the dynamic viscosity,
- $\mathbf{I}$ is the identity tensor,
- $\mathbf{v}_0$ is the initial velocity,
- $\mathbf{g}^f$ and $\mathbf{h}^f$ are prescribed boundary velocities and tractions,

- $\mathbf{n}^f$ is the outward unit normal vector,
- Γ_D^f, Γ_N^f, and Γ_R^f are the Dirichlet, Neumann, and Robin parts of the fluid boundary Ω^f,
- α^R is the Robin boundary condition coefficient.

Since the lubricant is located between the moving tool and the deforming chip, the fluid domain itself changes shape over time. To account for this moving domain within the Eulerian framework, the governing equations are modified using an Arbitrary Lagrangian-Eulerian (ALE) formulation. Equation (1.5) is adjusted to incorporate the mesh velocity $\mathbf{v}_\#$ through an additional reference coordinate χ:

$$\rho^f \left(\frac{\partial \mathbf{v}}{\partial t}|\chi + [(\mathbf{v} - \mathbf{v}_\#) \cdot \nabla]\mathbf{v} \right) - \nabla \cdot \boldsymbol{\sigma} = \rho^f \mathbf{b}^f \quad \text{in} \quad \Omega^f(t), \tag{1.12}$$

where $\frac{\partial \mathbf{v}}{\partial t}|\chi$ denotes the time derivative taken at a point moving with the mesh velocity $\mathbf{v}_\#$. The continuity equation (1.6) remains unchanged. In this coupled framework, the mesh velocity arises from the motion prescribed at the fluid boundaries, which is obtained from the solid's deformation. The Elastic Mesh Update Method (EMUM) is employed to smoothly propagate these boundary displacements into the interior of the domain, thereby determining the mesh velocity field needed for the ALE formulation.

The governing equations for fluid mechanics are discretized in space using FEM. The time integration is performed with an implicit Euler scheme. The moving boundaries use the B-Spline representation, which allows the boundary displacement – defined on the control points and obtained from the solid mechanics solver – to be directly incorporated into the fluid mesh motion.

1.3.4 Fluid-Structure Interaction: Partitioned Approach

Computing the coupled behavior – where the fluid forces act on the solid and the resulting solid deformation alters the fluid domain boundaries – requires specific techniques. This work adopts a partitioned approach, in which dedicated solvers for each physical domain exchange information iteratively across the shared fluid–solid interface within each time step until convergence is achieved. Ensuring convergence in these strongly coupled iterations involves satisfying two complementary conditions: temporal coupling, which maintains physical consistency at the interface over time, and spatial coupling, which ensures the accurate transfer of data between different discretizations/domains.

Temporal coupling ensures physical consistency at the interface by requiring both kinematic compatibility (matching velocity and displacement) and dynamic equilibrium (balanced forces) at the end of each time step. This framework utilizes a Robin-Neumann scheme [13], which is crucial for avoiding the "incompressibility dilemma" [9] that arises from the fully enclosed fluid pockets characteristic of these micro-topographies. In a standard partitioned approach (i.e., Dirichlet-Neumann), imposing boundary displacements on such pockets violates the incompressibility constraint by attempting to compress the fluid, leading to solver failure. The Robin condition resolves this by allowing a temporary artificial flux through the interface during the coupling iterations, which vanishes once the solution converges.

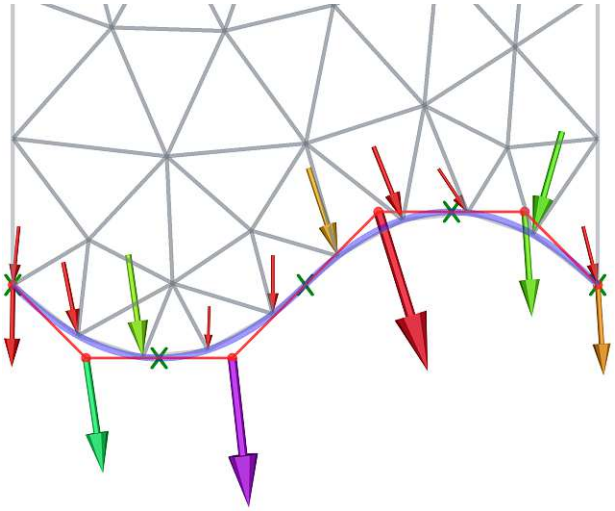


Figure 1.2: Description of the spatial coupling. Fluid tractions are projected onto the conforming spline's control points using the finite interpolation method, ensuring consistent load transfer.

Spatial coupling addresses the accurate transfer of data – such as displacements, velocities, and tractions – between the solid and fluid domains across their shared interface during the coupling iterations. Using the finite interpolation method [2, 6], the framework employs a conforming B-Spline representation at the interface, allowing the required quantities to be projected onto specific points through the evaluation of their corresponding basis functions.

1.3.5 Reduced Model

While the high-fidelity FSCI model provides detailed physical insights, its computational cost makes it impractical for direct integration into multi-scale frameworks or iterative design loops. To bridge this gap between high-fidelity and computational feasibility, this dissertation explores two distinct reduced modeling strategies, aimed at developing computationally efficient surrogates for the outputs of the high-fidelity model.

Scalar Surrogate Model

The first strategy focuses on creating a surrogate for the scalar friction coefficient μ, which is the primary output needed for existing meso-scale chip formation simulations. This is achieved using Gaussian Process Regression (GPR). GPR is a non-parametric Bayesian method that learns a mapping from input parameters (pressure, relative velocity, temperature) to the output friction coefficient based on training data generated by the high-fidelity model. Its purpose is to provide near-instantaneous predictions of μ for efficient multi-query scenarios.

Vector-Field Surrogate Model

The second strategy explores a more advanced approach capable of capturing the full complexity of the vector and tensor fields (e.g., stress, velocity) provided by the high-

fidelity model. This investigation leverages methodologies originally developed in a parallel project focused on geometry parameterization [11]. That work demonstrated the potential of using DNN, specifically an auto-decoder architecture based on DeepSDF [16], to learn a compressed, low-dimensional latent space representation of continuous implicit functions. Recognizing the applicability of this data-driven technique to any continuous functions defined over a spatial domain, e.g., field data, its potential for creating full-field reduced models was initially validated in a Master's thesis [19].

1.4 Overview of Scientific Contributions

This dissertation is comprised of four peer-reviewed publications that, taken together, document the development, application, and extension of a novel numerical framework for modeling the SSZ in orthogonal cutting. Each subsequent subsection presents one of these publications, which are organized to logically follow the research progression from experimental problem identification to data-driven reduced models.

P.1 Problem Identification and Methodological Requirements

J Saelzer, Y Alammari, A Zabel, D Biermann, J Lee, and S Elgeti. "Characterisation and modelling of friction depending on the tool topography and the intermediate medium". en. In: *Procedia CIRP* 102 (2021), pp. 435–440. ISSN: 22128271. DOI: 10.1016/j.procir.2021.09.074

Objective: The primary goal of this initial experimental study was to identify and quantify the physical phenomena governing lubricated contact in the secondary shear zone, thereby defining the essential requirements for the numerical framework (Objective O.1). The aim was to identify and understand the dynamics between lubricant, tool topography, and cutting conditions to establish the necessary physics for the subsequent modeling effort.

Methods: To isolate the effects of lubrication, two main experimental approaches were used. First, a novel orthogonal cutting test with discontinuous wetting was designed; slots were machined into the workpiece and filled with lubricant, allowing the tool to be momentarily wetted during the cut, simulating the presence of a lubricant film. Second, systematic friction characterization experiments were conducted using a specialized setup under conditions mimicking machining, explicitly varying the tool surface topography (polished vs. wet compressed air blasted) and the lubrication state (dry vs. wet) across a range of relative speeds. These experiments provided the core data needed to formulate the requirements for the numerical framework.

Findings & Contribution: The experimental results provided the crucial observation and requirements for the numerical framework. The friction characterization tests quantified the key influencing factors, confirming the friction coefficient's dependence on relative speed, tool topography, and lubrication state. The discontinuous wetting tests further confirmed that a controlled lubricant film can drastically reduce forces and improve surface quality. These findings established that any high-fidelity simulation must be able to resolve these specific coupled components. The primary contribution of this work was thus to use these findings to define the concrete requirements for the numerical framework and propose the spline-based FSCI methods as the solution. As a first step, the paper also presented a proof-of-concept for the spline-based geometric modeling of the measured surface topographies.

P.2 Development and Verification of the Numerical Building Blocks

J Lee, S Elgeti, J Saelzer, and A Zabel. "Development and Validation of a Micro-Tribology Model for the Chip Formation Zone". en. In: *Procedia CIRP* 117 (2023), pp. 317–322. ISSN: 22128271. DOI: 10.1016/j.procir.2023.03.054

Objective: Following the requirements established in Publication P.1, this work focused on the foundational development of the proposed spline-based FSCI framework (addressing Objective O.1). The goal was to implement and individually verify the core numerical "building blocks" necessary to model the solid, fluid, and contact components, ensuring their suitability before full integration.

Methods: This work details the initial implementation of the key methods described in Section 1.3. The solid domains (chip and workpiece) were modeled using IGA, the chip was modeled with a simplified hyper-elastic material as a stand-in for plasticity. The fluid domain was discretized using FEM with spline-based boundaries. To couple these domains, a partitioned Robin-Neumann scheme was employed to robustly handle the incompressibility constraints of the enclosed fluid domain. For contact, a Gaussian-point-to-segment penalty method was implemented, with the contact search algorithm solving the NPP problem. A phantom-domain approach was used for the moving mesh, which can incorporate large deformations as well as separation of the computational domain. Each of these components was then tested in standalone scenarios to validate its function.

Findings & Contribution: The test cases demonstrated that the chosen methods were functional and suitable for their intended purposes within the FSCI context. The spline-based contact algorithm correctly captured deformation based on topography, and the FSI coupling could handle enclosed fluid domains. The primary contribution of this work was providing the verified numerical components required for the full system. The results also clearly defined the next steps: integrating these components into a fully coupled solver and, most importantly, implementing a realistic thermo-elasto-plastic material model, e.g., Johnson-Cook, to replace the hyper-elastic placeholder.

P.3 The Fully Coupled Framework and Scalar Model Reduction

Jaewook Lee, Jannis Saelzer, Jacques Zwar, Florian Zwicke, Felipe Gonzalez, Thomas Spenke, Norbert Hosters, Gero Polus, Andreas Zabel, and Stefanie Elgeti. "Spline-Based Framework for Microscopic Contact Zone Modeling in Lubricated Orthogonal Cutting". en. In: *International Journal for Numerical Methods in Engineering* 126.14 (July 2025), e70087. ISSN: 0029-5981, 1097-0207. DOI: 10.1002/nme.70087

Objective: This work represents the core of the dissertation, aiming to integrate the verified components into a single, fully functional high-fidelity FSCI framework (completing Objective O.1). A second goal was to apply this framework to test the initial "mechanics-dominant" friction hypothesis (addressing Objective O.2). Finally, it tackles the inevitable high computational cost by developing a practical scalar reduced model (addressing Objective O.3-a).

Methods: The final framework was assembled by coupling the spline-based solvers. Key methodological advances over Publication P.2 included implementing the Johnson-Cook

thermo-elasto-plastic model with adiabatic thermal effects for the chip and upgrading the contact algorithm to a more robust mortar knot-to-surface formulation. For the moving mesh, EMUM was employed; this was a more direct approach than the phantom-domain method explored in Publication P.2, as the model focused specifically on simulating the predefined, enclosed fluid pockets. The FSI coupling employed the Robin-Neumann scheme and EMUM as detailed in Section 1.3. To create the reduced model, numerous high-fidelity simulations were performed, and this data was used to train a GPR surrogate model.

Findings & Contribution: This work yielded three major contributions. First, it delivered a complete, functional, high-fidelity FSCI simulation framework. Second, its application led to a critical insight: as detailed in the paper, when the micro-model's behavior was compared with experimentally observed tendencies (e.g., [14]), fundamental discrepancies were noted. The model predicted that the friction coefficient decreased with increasing contact pressure and increased with relative velocity, which is the opposite of what is observed in experiments. This led to the conclusion that the initial "mechanics-dominant" assumption (based on surface interlocking) is not the major factor, strongly indicating that other, more complex physical mechanisms are at play in the secondary shear zone. Third, the GPR surrogate model was highly successful, accurately predicting the model's output with a speed-up of over 10^7 at runtime, proving a viable pathway for coupling high-fidelity physics to meso-scale simulations.

P.4 A Novel Methodology for Full-Field Model Reduction

Jaewook Lee, Sebastian Hube, and Stefanie Elgeti. "Neural networks vs. splines: advances in numerical extruder design". en. In: *Engineering with Computers* 40.2 (Apr. 2024), pp. 989–1004. ISSN: 0177-0667, 1435-5663. DOI: 10.1007/s00366-023-01839-2

Objective: This work was conducted as a parallel project to investigate advanced data-driven methods for handling large spatial data by learning continuous implicit functions. The specific objective of this paper was to develop a novel geometry parameterization technique for shape optimization that could overcome the limitations of traditional, manually-defined methods, such as Free-Form Deformation (FFD). For the dissertation, this study serves as the foundational investigation for Objective O.3-b. It establishes a method for learning a low-dimensional latent-space representation of a continuous function. Since a high-fidelity, full-field simulation result (e.g., a stress or velocity field) is itself a continuous function defined over a spatial domain, this methodology is directly transferable to the challenge of full-field reduced modeling.

Methods: The core method involved using a DNN, specifically an autodecoder architecture based on DeepSDF [17], to create a "geometric filter". This network was trained on a large dataset of 2659 different 3D shapes (mixing elements). Instead of manually defining parameters, the network learned to compress and represent these complex geometries in a feature-rich, low-dimensional latent space (e.g., 4- , 8-, or 16-dimensional). The network's ability to represent shapes was then tested in a full shape-optimization loop for an extruder mixing element.

Findings & Contribution: The study successfully demonstrated that this DNN autodecoder architecture could learn a compact and continuous latent-space representation for industrially relevant geometries and be applied to shape optimization. The primary

contribution to this dissertation is therefore methodological: it provides a proof-of-concept that this architecture excels at learning continuous functions defined over a spatial domain. This capability is directly generalizable to other continuous functions, such as the stress and velocity fields generated by the high-fidelity FSCI model. To concretely bridge this methodology to the core problem of this thesis, its applicability was successfully validated in a dedicated Master's thesis conducted within the scope of this project [19]. This supervised work confirmed that the same autodecoder architecture could learn a compact latent-space representation for complex, transient fluid velocity fields – a problem analogous to the vector fields in the FSCI model – thus confirming this as a viable pathway for creating the full-field surrogate models proposed in Objective O.3-b

2 Publications

2.1 Characterisation and modelling of friction depending on the tool topography and the intermediate medium

J Saelzer, Y Alammari, A Zabel, D Biermann, J Lee, and S Elgeti. "Characterisation and modelling of friction depending on the tool topography and the intermediate medium". en. In: *Procedia CIRP* 102 (2021), pp. 435–440. ISSN: 22128271. DOI: 10.1016/j.procir.2021.09.074

Statement of Contribution: This work was a collaboration with the experimental team at TU Dortmund. My role was to define and conceptualize the numerical simulation framework, with specific contributions including:

- analyzing the experimental results from the cutting and friction characterization tests (Sections 2–3) to establish the core requirements for the new simulation model,
- authoring Section 5 ("Microscopic tribology model"), which defined the numerical part of the project,
- defining the necessary components of the numerical framework and developing the prototype for the spline-based modeling of the surface topography,
- contributing to the overall review and final writing of the manuscript.

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 102 (2021) 435–440

www.elsevier.com/locate/procedia

18th CIRP Conference on Modeling of Machining Operations

Characterisation and modelling of friction depending on the tool topography and the intermediate medium

Saelzer J*[a], Alammari Y[a], Zabel A[a], Biermann D[a], Lee J[b], Elgeti S[b]

[a]TU Dortmund University, Institute of Machining Technology, Dortmund, Germany
[b]TU Wien, Institute of Lightweight Design and Structural Biomechanics, Vienna, Austria
* Corresponding author. Tel.: +49-231-755-90163; fax: +49-231-755-5141. *E-mail address:* jannis.saelzer@tu-dortmund.de

Abstract

The friction in the contact zones during chip formation has a significant effect on tool wear and surface integrity. Especially the interaction of cooling lubricants and different surface preparations is rather unexplored. In the context of the present publication, the friction between tool and workpiece is therefore investigated in a special test under conditions close to machining for specific prepared tool surfaces when using a cooling lubricant. The results of the investigation are used to develop friction models depending on relative speed, surface preparation and intermediate medium, which will be implemented in new finite element formulations for chip formation simulation.

Keywords: Friction; Modelling; Cutting fluids; Surface topography; Chip formation

1. Introduction

Although the use of cooling lubricants in metal-cutting often leads to a great increase in productivity and allows the machining of difficult-to-machine materials, it is increasingly considered as critical in the form of flood cooling lubrication. The main disadvantages are the high costs of operating the cooling lubricant systems and the considerable potential damages to health and environment [1]. In order to counter these problems and at the same time to be able to use the productivity and quality increase from cutting fluids, methods must be developed to use them in a controlled way. One possibility for this is the minimum quantity lubrication (MQL), which serves in particular to reduce the coefficient of friction [2]. Because only a small amount of cutting fluid is used in MQL, the supply to the effective point is of central importance [3]. With the help of cutting tests, the macroscopic force-reducing effect of cutting fluids could be proven, whereby the chemical composition of the fluid has an influence on the mode of action [4].

High-quality chip formation simulations have become an indispensable tool for process and tool design, with friction modelling being an essential factor. In this context, the design of cutting fluid concepts is also an important task, for which the fluid-structure interaction behaviour must be modelled [5]. The focus can be on the fluid flow and the cooling effect as well as the mechanical effect on chip formation due to collision of a fluid jet with the chip [6]. In other works, the effect of cutting fluids in relation to the tribological system was investigated and applied for models in chip formation simulation [7]. For both dry friction and wet friction, various variables influencing the coefficient of friction have also been investigated. These include the relative speed, the contact pressure, the temperature, and in addition the substrate and the coating of the tool as well [8].

Within the scope of the present work, the interaction of a cutting fluid with different surface topographies of the rake face is to be investigated. For this purpose, cutting tests are first carried out in which the effect of a cutting fluid on the mechanical tool loads and chip formation is analysed. Subsequently, experiments to characterise friction by varying

2212-8271
10.1016/j.procir.2021.09.074

the lubrication concept as well as the tool surface and the relative speed will be carried conducted. The results are used to map the previously experimentally determined influence of the cutting fluid in a friction model for chip formation simulations. Finally, the concept of a numerical tribology simulation is presented, which is to be used in the future to map the mode of action of the cutting fluid to the micro level in order to derive new data for friction modelling in chip formation simulations.

Nomenclature

MQL	Minimum quantity lubrication
WCAB	Wet compressed air blasted
A	Initial flow stress in MPa
B	Pre-exponential factor in MPa
C	Strain rate factor
F_c	Cutting force in N
F_p	Passive force in N
Rz	Average roughness in µm
T	Temperature in °C
T_0	Reference temperature in °C
T_m	Melting temperature in °C
a	Infeed in mm
b	Width of cut in mm
b_t	Width of tool in mm
b_{wp}	Width of workpiece in mm
h	Uncut chip thickness in mm
h_c	Chip thickness in mm
$h_{c,m}$	Mean chip thickness in mm
h_t	Height of tool in mm
l	Length of cut in mm
l_c	Contact length of chip and tool in mm
l_e	Length of element in mm
l_{wp}	Length of workpiece in mm
m	Thermal-softening exponent
n	Work hardening exponent
r_e	Cutting edge radius in µm
v_c	Cutting speed in m/min
v_r	Relative speed in m/min
α	Clearance angle in °
γ	Rake angle in °
ε	Strain
$\dot{\varepsilon}$	Strain rate in s^{-1}
$\dot{\varepsilon}_0$	Reference strain rate in s^{-1}
µ	Friction coefficient

2. Cutting experiments

2.1. Experimental setup

As part of the first series of tests, machining experiments were carried out in orthogonal cutting. The special machine for chip formation analysis from *Heinz Berger Maschinenfabrik GmbH* is used for this purpose. Figure 1 shows the test set-up used. The workpiece sample is clamped on a table that is equipped with a linear direct drive and allows relative speeds of up to $v_r = 180$ m/min. The tool is mounted on the traverse where it is clamped in a piezoelectric dynamometer from *Kistler*, type 9263, which is used to record the mechanical tool loads.

A cuboid specimen of normalised AISI 1045 with a width of $b_{wp} = 2$ mm and a length of $l_{wp} = 140$ mm was used. Self-prepared cemented carbide bodies of the grade K40 UF act as the tools. These were designed to be as suitable as possible for investigating friction phenomena. For this purpose, the rake angle was chosen to $\gamma = 0°$ so that the passive force is essentially determined by the friction in the secondary shear zone. Moreover, a relatively large clearance angle of $\alpha = 10°$ was also applied and the cutting edge was provided with a small cutting edge radius of $r_e = 10$ µm in order to minimise the proportion of the force resulting from ploughing processes.

During the tests, two different variants of surface topography (see figure 1) on the rake face were applied in order to investigate their influence on the contact situation during machining with and without an intermediate medium. On the one hand, a polished surface, with a very smooth topography and an average roughness of Rz = 0.29 µm in chip flow direction and Rz = 0.26 µm in orthogonal direction, came into operation. On the other hand, wet compressed air blasted (WCAB) rake faces were used. This surface has a dimple structure and an isotropic average roughness of Rz = 0.96 µm. The ester-based minimal quantity lubricating oil *Vascomill MMS HD*1 from *Blaser* served as intermediate medium.

2.1. Results

The aim of the cutting tests was to investigate the influence of the wetting of the rake face on the friction behaviour in the secondary shear zone and thus on the passive force. For this purpose, a series of tests was first carried out with different cutting speeds between $v_c = 60$ m/min and $v_c = 180$ m/min, in which the intermediate medium was initially applied to the tool in different ways. Regardless of the type of oil application, no difference in the mechanical loads could be identified. It can therefore be assumed, that the intermediate medium was mechanically displaced from the contact zone by the chip flow the temperatures that occurred in the process removed it thermically. Under specific boundary conditions, it was possible to detect an influence of the tool wetting on the mechanical tool loads. For this purpose, the cutting speed was reduced to $v_c = 5$ m/min and a discontinuous supply of the intermediate medium was realised during the cut.

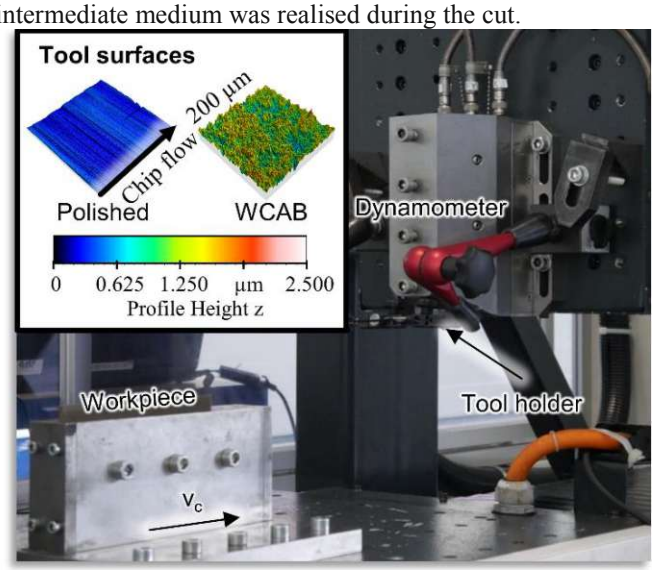
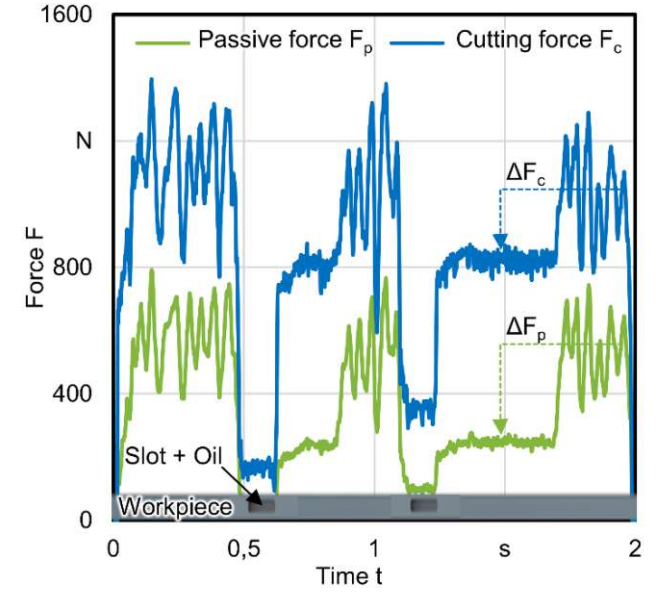


Fig. 1. Experimental setup of the cutting experiments and different types of tool surface topographies.

This was achieved by means of two oil-filled slots in the workpiece, which represent a local and temporary interruption of the cut, during which the tool passes through the oil and is wetted. The principle is illustrated for the example of a polished rake face in figure 2 using the recorded cutting and passive forces of a single test as well as a sketch of the workpiece with the oil-filled slots. This test was repeated three times for both surface preparations for statistical validation.

At the beginning of the test, dry cutting is present. The forces are at a high level and show large variations around a stationary value. As soon as the slot reaches the tool, the forces suddenly drop, as the cross-section of the undeformed chip is significantly reduced. After the end of the slot, the forces increase again. In contrast to the section before the slot, however, the forces are at a significantly lower level and show smaller variations. After a certain cutting distance, the forces increase quickly until they reach the level of the dry cut again. The observations suggest that immediately behind the slot there is a load-bearing lubricating film over the entire contact surface between the tool and the chip/workpiece. As long as this lubricating film is present, the frictional stress between the contact partners is reduced, whereby the cutting force and the passive force drop considerably compared to the dry state. It is remarkable that the effect of the lubricating film on the forces dissipates abruptly. There is almost no transition area towards the force level corresponding to the dry cut. In addition, both force components rise again to the previous level at the same time. It can be concluded that the lubricating film effect depends on reaching a critical film thickness and that this must be present over the entire contact length, meaning that no localised presence of a lubricating film occurs.

For the WCAB surface, the mean decrease of the cutting force caused by the lubrication compared to dry cutting is $\Delta F_c = 323$ N, which corresponds to a percentage decrease of -27 %.

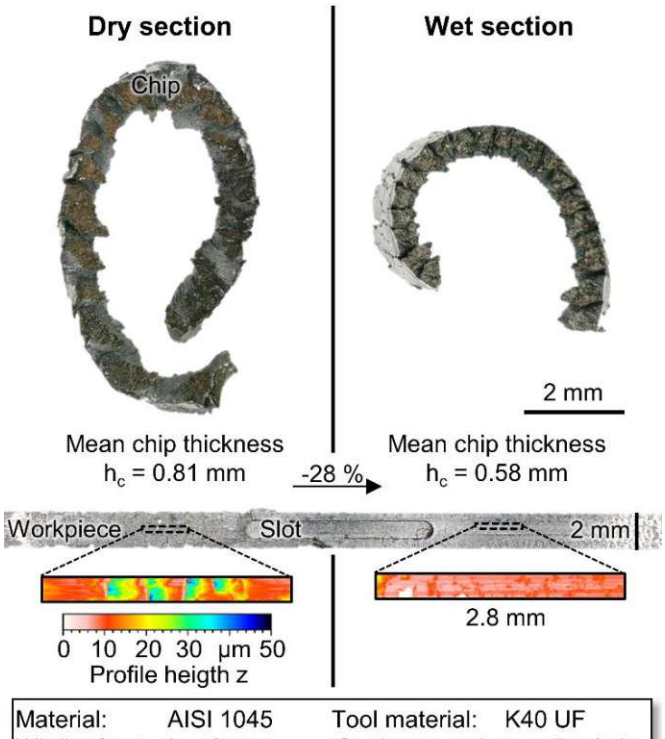


Material:	AISI 1045	Tool material:	K40 UF
Witdh of cut:	b = 2 mm	Cutting speed:	v_c = 5 m/min
Rake angle:	γ = 0°	Uncut chip th.:	h = 0.15 mm
Rake face:	Polished	Cutting fluid:	Vascomill HD

Fig. 2. Principle of cutting experiments with discontinuous tool wetting.

Regarding the passive force, this difference is $\Delta F_p = 375$ N, which is a decrease of -58 %. In the case of the polished surface, the cutting force difference between dry and wet is $\Delta F_c = 291$ N and thus -25 %. The passive force difference is $\Delta F_p = 312$ N, which corresponds to a percentage difference of -50 %. In both cases, the mechanical tool loads are significantly reduced, whereby the force level is lowered more for the passive force. This is due to the fact that under the boundary conditions used, the passive force results to a large extent from friction. For the WCAB surface, a greater reduction in force than for the polished surface could be detected, especially for the passive force. This is presumably due to the fact that the dimple-structured surface has micro-lubrication pockets in which the intermediate medium can stabilise and thus form a more load-bearing lubricating film than in the case of the very smooth polished surface.

For further analysis of the contact behaviour, the chips and the machined workpiece surface were examined. Exemplary results are shown in figure 3 for the polished rake face. The comparison of the chip produced during dry machining with the chip produced during the presence of the lubricant film shows that there are significant differences both in terms of the qualitative chip morphology and the mean chip thickness. In the case of dry machining, a thick, discontinuous chip occurs. This is consistent with the observation that the forces in this area are subject to greater variations. Machining with the lubricating film effect lead to a chip that shows a more homogeneous geometry in comparison. The average chip thickness in the dry case is $h_c = 0.81$ mm and it decreases by -28 % to $h_c = 0.58$ mm due to the lubricating film.



Material:	AISI 1045	Tool material:	K40 UF
Witdh of cut:	b = 2 mm	Cutting speed:	v_c = 5 m/min
Rake angle:	γ = 0°	Uncut chip th.:	h = 0.15 mm
Rake face:	Polished	Cutting fluid:	Vascomill HD

Fig. 3. Chip morphology and workpiece surface for dry and wet cutting.

Based on high-speed chip formation recordings, it could be qualitatively shown that the change in chip compression can be explained by an increase in the chip flow speed in the presence of the lubricant film. The decreasing chip thickness is probably a major cause for the lower forces during the lubricated cut.

The influence of the lubricating film can also be identified on the machined workpiece surface. Overall, the surface quality is poor due to the very low cutting speed. Dry machining produces a very rough surface with a more or less regular chipping, which indicates stick-slip contact between tool and workpiece. The mean average roughness depth is Rz = 28.9 µm. In comparison, the surface created during the existing lubricant film is significantly smoother and more homogeneous. This is also reflected in the mean average roughness depth, which in this case is Rz = 7.6 µm. The immediate contact during dry machining results in a mixture of abrasive and adhesive friction, which leads to a stick-slip effect and the chipping on the workpiece. As soon as a lubricating film is present, the chip and the tool slide against each other with significantly less resistance, which gives the workpiece surface its homogeneous topography.

Based on the experiments carried out, it was possible to show an enormous potential in the use of an intermediate medium to reduce forces and increase the workpiece quality. This potential is probably rarely used in practical applications due to insufficient supply. Since a corresponding effect could not be generated experimentally due to the lack of supply at practically relevant cutting values, an estimation of the potential at higher cutting speeds can only be achieved by means of a process simulation. For this purpose, the friction behaviour in the dry and lubricated state is to be characterised and made accessible for chip formation simulations.

3. Friction characterisation

A tribological test rig based on the example of Puls et al. [9] and already described in Saelzer et al. [10] was used to investigate the friction behaviour in form of the friction coefficient. The conditions prevailing in this test rig with regard to temperature, contact pressure and relative speed approximate well the conditions in the chip formation zone. In the series of tests carried out, the topographies of the carbide surfaces were varied according to the tools from the cutting tests. In addition, the wetting state was varied between dry and fully wet. This resulted in four test series, within each of which the relative speed was varied on eight levels. All tests were carried out twice to statistically validate the results and the measured values were averaged over both tests. A plot of the final results is shown in figure 4.

The results of the dry friction characterisation were already discussed in [10]. Accordingly, the measured values here only serve as a comparison for the wet state. It can be seen that the wetting of the tool surface has a great influence on the coefficient of friction. In the range of low relative speeds up to 40 m/min, there is a decrease in the coefficient of friction of up to -70 % compared to dry friction. As the relative speed increases, the curves converge, but there is still a significant difference. This behaviour can be explained generally by the fact that the lubricating film results in a partial separation of the surfaces and the friction mechanisms occurring in dry conditions are prevented or weakened. The partial separation of the surfaces also nearly prevents the effect of the specific topography on the friction coefficient. While the influence of the relative speed in dry friction arises from the variable friction mechanisms [10], which occur with reduced intensity in the presence of the lubricating film.

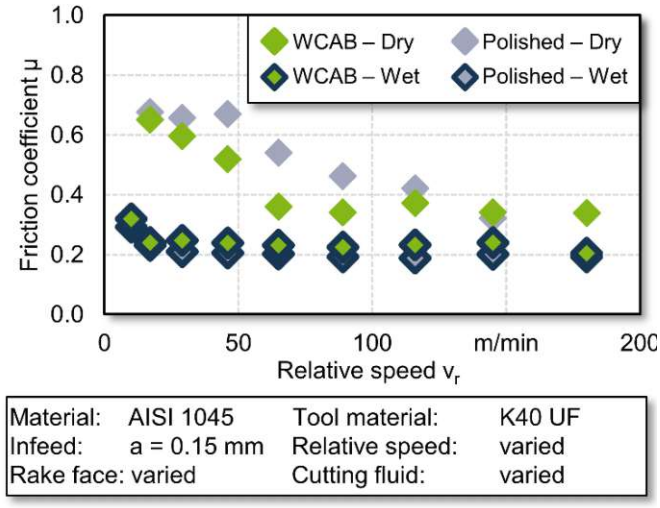


Fig. 4. Results of the friction characterisation.

4. Chip formation simulation

4.1. Simulation setup

The findings from the friction characterisation are used in the following to develop finite element chip formation simulations with which it is possible to map the different lubrication states during an orthogonal cut. This will be illustrated using the polished surface. All simulations were carried out in DEFORM 2D V12.0. The simulation setup is shown in Figure 5. A section of the tool with a height of h_t = 2 mm and a width of b_t = 1 mm was modelled.

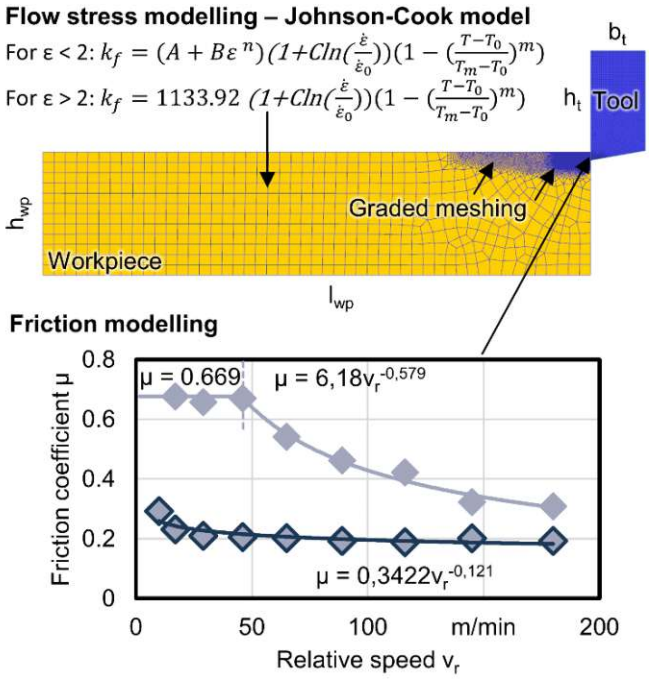


Fig. 5. Setup of the finite element chip formation simulation.

The height of the workpiece section was selected as h_{wp} = 2.25 mm and the length as l_{wp} = 10 mm. An adaptive meshing was provided in both bodies to enable an efficient calculation and the detailed resolution of the effective zones. The length of the elements in the workpiece in the area of the primary shear zone and secondary shear zone were chosen to be l_e = 0.007 mm, since the largest gradients of the state variables are present here. For the area of the chip forming and the preliminary deformation zone, the length of the elements was l_e = 0.02 mm. In the rest of the workpiece it was set to l_e = 0.2 mm. In the case of the tool, a gradation with l_e = 0.13 mm in the area of the cutting edge and l_e = 0.033 in the rest of the tool was provided. For the flow stress modelling, a sectionally defined model according to Johnson and Cook was used, which corresponds to the standard model up to a strain of ε = 2 and assumes the Ludwik term to be constant above a strain of ε = 2, since a flow stress that continues to increase with strain does not represent the real material behaviour [11]. The model equations are shown in Figure 5 and the parameters values in table 1:

Table 1. Parameters for Johnson-Cook flow stress model.

Parameter	Value
Initial flow stress	A = 288 MPa
Pre-exponential factor	B = 695 MPa
Work hardening exponent	n = 0.2835
Strain rate factor	C = 0.034
Thermal-softening exponent	m = 1.3558
Reference strain rate	$\dot{\varepsilon}_0 = 0.004\ s^{-1}$
Reference temperature	T_0 = 20 °C
Melting temperature	T_m = 1500 °C

For the close-to-reality-consideration of the friction behaviour in the simulation, the friction model according to Coulomb with speed-dependent friction coefficients is mapped based on the measurement results from chapter 3. In the case of dry friction, a model defined in different sections is used, assuming the friction coefficient to be constant up to a relative speed of v_r = 46 m/min and afterwards calculating with the power function shown in Figure 5. For the wet contact state, the behaviour of the friction coefficient is described as a function of the relative speed over the entire speed interval by a power function.

4.2. Simulation results

First, simulations were carried out at a cutting speed of v_c = 5 m/min and an uncut chip thickness of h = 0.15 mm. The results are shown in figure 6, where in addition to the cutting and passive forces, the mean chip thicknesses and the contact lengths of the tool and the chip were also evaluated. Compared to the experimentally determined results, the forces are approx. 40 % higher, which can be attributed to the fact that the flow stress model is not designed for such low deformation speeds as they occur at v_c = 5 m/min. However, the relative difference of the cutting and passive force between the dry and the wet state is very similar to the experiment with -27 % and -59 %. The average chip thickness also decreases by -27 % and therefore to a similar extent as in the experiment. This correlates with the decrease in cutting forces due to the reduced size of the primary shear zone. An additional information from the simulation is the reduction in the contact length between the chip and the rake face, which is -34 % in the wet case compared to the dry. Presumably this is similar in real machining and contributes significantly to the decrease in passive forces. In another simulation, a cutting speed of v_c = 120 m/min was investigated. Here, a percentage difference between dry and wet conditions for the cutting force of -10 % and for the passive force of -47 % was shown. Although the wet condition for this high cutting speed could not be validated, it shows a great potential for the use of a cutting fluid for reduction of friction.

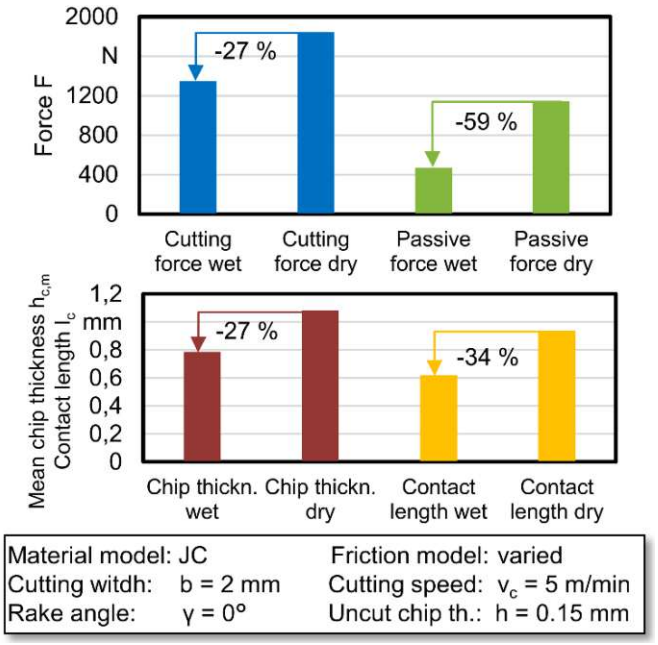


Fig. 6. Results of the finite element chip formation simulation.

5. Microscopic tribology model

In order to extend the work shown above, an additional microscopic tribology model is developed using innovative numerical simulation methods. To incorporate lubricant flow phenomena, fluid-structure-contact-coupling needs to be established. Using a partitioned approach, the fluid-structure-contact-interaction (FSCI) can be solved by iteratively coupling two independent solvers for each field. Here, NURBS enhanced finite element method (NEFEM) [12, 13] and isogeometric analysis (IGA) [14] are utilised for fluid and structure respectively. Adopting the coupling scheme presented by Hosters et al. [15], the interface between fluid and structure is represented using the identical spline description. As a result, necessary quantities can be directly transferred between fluid and structure. In this context, appropriate solutions for contact problems will be further investigated. The measured surface topography from figure 1 is used as a basis to generate a spline geometry for both NEFEM and IGA. The measurement is given as a rasterised height map, i.e., the height of the surface is measured at uniformly distributed locations throughout the surface. This can be interpreted as a point cloud that can be directly used as an input to perform a spline fitting. For such an evenly spaced point set, the global interpolation algorithm given by Piegl and Tiller [16] can be applied to obtain a B-Spline that captures the roughness of the surface. For example,

for 2D analysis, a spline surface that represents a 2D structure can be acquired by first performing a B-spline curve fitting. While keeping the knot vector of the first dimension the same, a new knot vector for the second dimension and additional control points can be defined to transform a spline curve into a spline surface. An illustration of this process is shown in figure 7.

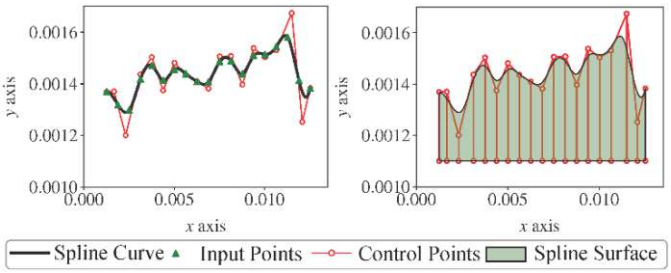


Fig. 7. Spline curve fitting and its extension to a spline surface.

Figure 8 depicts the general idea for the computational domain around the contact zone. The tool is assumed to be a rigid body, while the workpiece behaves according to the Johnson-Cook-Model (figure 5). The lubricant is modelled by the Carreau-Model. With this setup, the interaction between the tool, the fluid, and the workpiece under vertical pressure and horizontal translation is computed. During this process, the change in coolant distribution is captured using level-set method [17].

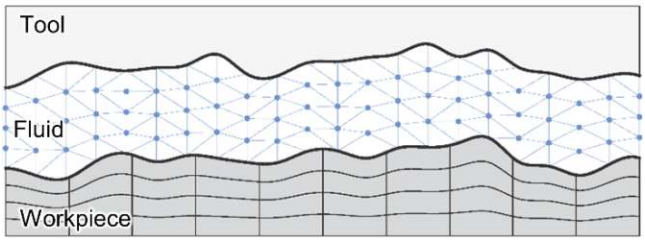


Fig. 8. A computational domain of the contact zone.

6. Conclusion and outlook

Within the scope of the present work, the influence of a rake face lubrication in interaction with different surface topographies was investigated. First, orthogonal cutting tests with discontinuous wetting of the tool were carried out. Through this wetting, a load-bearing lubricating film was temporary formed over the contact surface of tool and chip/workpiece. While the lubricating film was present, a significant reduction in chip thickness and mechanical tool loads could be observed. In order to visualise the different wetting conditions in a chip formation simulation, friction characterisation tests were subsequently conducted. Using the example of a polished tool surface, the results were transferred to a finite element chip formation simulation based on empirical models. Within the simulations, a similar difference between the dry and the wet condition was found with regard to chip thickness and mechanical tool loads. Finally, the first steps in the development of a microscopic tribological simulation were presented, which will be used in the future to model fluid-structure interaction in the contact zone through a coupling with the chip formation simulation. On the experimental side, a continuous supply of fluid will be used to investigate the effect of the lubrication at higher cutting speeds.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 439919057.

References

[1] Weinert K, Inasaki I, Sutherland J W, Wakabayashi. Dry Machining and Minimum Quantity Lubrication. In: CIRP Annals 2004;53 (2):511–537.

[2] Cabanettes F, Rolland J, Dumont F, Rech J, Dimkovski Z. Influence of Minimum Quantity Lubrication on Friction Characterizing Tool–Aluminum Alloy Contact. In: J. Tribol 2016;138 (2): 21107.

[3] Chetan, Ghosh S, Rao P V. Specific cutting energy modeling for turning nickel-based Nimonic 90 alloy under MQL condition. In: International Journal of Mechanical Sciences 2018;146-147: 25–38.

[4] Min S, Inasaki I, Fujimura S, Wada T, Suda S, Wakabayashi T. A Study on Tribology in Minimal Quantity Lubrication Cutting. In: CIRP Annals 2005;54 (1):105–108.

[5] Melkote S N, Grzesik W, Outeiro J, Rech J, Schulze V, Attia H et al. Advances in material and friction data for modelling of metal machining. In: CIRP Annals 2017;66 (2):731–754.

[6] Oezkaya E, Iovkov I, Biermann D. Fluid structure interaction (FSI) modelling of deep hole twist drilling with internal cutting fluid supply. In: CIRP Annals 2019;68 (1):81–84.

[7] Claudin C, Mondelin A, Rech J, Fromentin G. Effects of a straight oil on friction at the tool–workmaterial interface in machining. In: Int J Mach Tools Manuf 2010;50 (8):S. 681–688.

[8] Peng B, Bergs T, Schraknepper D, Smigielski T, Klocke F. Development and validation of a new friction model for cutting processes. In: Int J Adv Manuf Technol 2020;107 (11-12): 4357–4369.

[9] Puls H, Klocke F, Lung, D. Experimental investigation on friction under metal cutting conditions. In: Wear 2014;310:63–71.

[10] Saelzer J, Zabel A, Biermann D. Experimental Analysis of the Friction Behaviour in Cutting. In: Production at the leading edge of technology - Proceedings of the 9th Congress of the German Academic Association for Production Technology (WGP), Hamburg, 2019: p. 297-305.

[11] Hockett J E, Sherby O D. Large strain deformation of polycrystalline metals at low homologous temperatures. J. Mech. Phys. Solids 1975;23:87-98.

[12] Sevilla R, Fernández-Méndez S, Huerta A. NURBS ‐ enhanced finite element method (NEFEM). International Journal for Numerical Methods in Engineering 2008;76 (1):56-83.

[13] Stavrev A, Knechtges P, Elgeti S, Huerta A. Space ‐ time NURBS ‐ enhanced finite elements for free ‐ surface flows in 2D. International Journal for Numerical Methods in Fluids 2016;81/7: 426-450

[14] Hughes T J R, Cottrell J A, Bazilevs Y. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. Computer Methods in Applied Mechanical Engineering 2005;194:4135-4195.

[15] Hosters N, Helmig J, Stavrev A, Behr M, Elgeti, S. Fluid–structure interaction with NURBS-based coupling. Computer Methods in Applied Mechanics and Engineering 2018;332:520-539.

[16] Piegl L, W Tiller. The NURBS Book. 2nd ed. Berlin/Heidelberg: Springer; 1997.

[17] Sethian J A. Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science. Vol. 3. Cambridge university press; 1999.

2.2 Development and Validation of a Micro-Tribology Model for the Chip Formation Zone

J Lee, S Elgeti, J Saelzer, and A Zabel. "Development and Validation of a Micro-Tribology Model for the Chip Formation Zone". en. In: *Procedia CIRP* 117 (2023), pp. 317–322. ISSN: 22128271. DOI: 10.1016/j.procir.2023.03.054

Statement of Contribution: As first author, I was the primary contributor to this work. This research was conducted in collaboration with the experimental team at TU Dortmund, who provided the experimental context. My specific contributions were:

- conceptualizing and implementing the core numerical building blocks of the FSCI framework; this involved developing the IGA solid solver with its spline-based penalty contact algorithm (including NPP), while integrating an in-house FEM fluid solver,
- verifying the function of these individual components through the numerical test cases presented,
- serving as the lead author of the manuscript.

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 117 (2023) 317–322

www.elsevier.com/locate/procedia

19th CIRP Conference on Modeling of Machining Operations

Development and Validation of a Micro-Tribology Model for the Chip Formation Zone

Lee J*[a], Elgeti S[a], Saelzer J[b], Zabel A[b]

[a]TU Wien, Institute of Lightweight Design and Structural Biomechanics, Vienna, Austria
[b]TU Dortmund University, Institute of Machining Technology, Dortmund, Germany

* Corresponding author. Tel.: +43-1-58801-31735; fax: +43-1-58801-31799. *E-mail address:* jaewook.lee@tuwien.ac.at

Abstract

The modelling of friction is essential for the accuracy of chip formation simulations. Current models are not suitable to represent the influence of lubricants in interaction with different tool surfaces in its complexity. The present work aims on the development of a spline-based fluid-structure-contact interaction framework to model the friction of tool, chip and intermediate medium under various conditions. Using this framework, microscopic tribology simulations of the secondary shear zone can be performed to implement a friction coefficient distribution depending on relative speed, temperature and contact normal pressure into a chip formation simulation, which will be validated by experimental data.

Keywords: Type your keywords here, separated by semicolons ;

1. Introduction

Particularly in the light of rising prices for energy and resources, industry is faced with the demand for savings while at the same time maintaining productivity and quality. The positive influence of cutting fluids on the machining process has been proven in many studies. If the fluid can penetrate into the contact zone and therefore provide a lubrication, the shear angle is increased, which leads to a decreasing chip thickness, hence less mechanical tool loads. This usually results in less thermal tool loads and an extended tool life [1]. In addition, cutting fluids can improve the cutting process with respect to part quality. For example, both the cooling and the lubrication effects can reduce the surface roughness of the produced components [2]. In the field of machining, there is particularly great potential in the reduction of lubricant quantities, which are fed into the process in energy-intensive and cost-intensive circulation systems, especially in the case of difficult-to-cut materials [2]. In order to be able to meet these conflicting demands of keeping the positive influence of the cutting fluid while reducing the amount of fluid used in the processes, it is necessary that the cutting fluids are used in a more controlled manner, especially concerning the lubrication function of the media. This in turn requires a better understanding of the working mechanisms of cutting fluids and, based on this, the qualification of tools and processes for more effective use. Improved utilization of the lubrication potential of cutting fluids is possible through the specific functionalization of tool surfaces in combination with a supply system close to the working point [3]. In this context, already known influences on the friction in the chip formation zone should be taken into account. For example, the macro geometry of the tool has a significant influence on the friction in the secondary shear zone. In case of negative rake angles, the frictional load increases with increasing rake angle, while for positive rake angle one observes the opposite trend [4]. In addition, the micro geometry of the cutting edge has an impact on the friction in the chip

2212-8271
10.1016/j.procir.2023.03.054

formation zone. The larger the cutting edge rounding and the more it is oriented towards the flank face, the larger the friction is at the cutting edge [5]. Other factors relevant to tool design that play a significant role in chip formation friction are the surface topography and the surface roughness of the contact surfaces. These affect not only the dry friction, but also especially the interaction with a cutting fluid [6]. These detailed tribological aspects of cutting fluid application are of limited experimental accessibility; accordingly, the further development of powerful chip formation and tribology models is of central importance [7]. The standard for friction modeling in chip formation simulations, until today, is the use of relatively simple models (e.g., the Coulomb model with constant friction coefficients for the corresponding material pairing). In recent years, however, models have been developed that take into account important factors influencing friction, such as the relative velocity of the friction partners, the contact normal stress and the temperature [8]. In addition, approaches have been developed that allow the contact in the secondary shear zone to be modeled differentially into a sticking and a sliding zone [9] and distinguish from global friction along the entire contact area or averaged friction in the secondary shear zone as an expression of the influence of cutting edge micro geometry on friction [10]. The effect of appropriate models on the outcome of chip formation simulations has been extensively studied, also for different simulation environments [11]. Corresponding modeling approaches, which are often empirical in nature, exist for various materials and allow to take into account their sometimes very different friction properties in chip formation simulations [12].

While models of the macroscopic mechanical effect of cutting fluids on chip formation are already relatively mature [13], there is a large gap in the modelability of the cutting fluid effect with respect to friction reduction. Approaches such as that according to Banerjee and Sharma, which depicts the friction coefficient using minimum quantity lubrication in an empirical model as a function of cutting speed and feed rate [14], show a poor relationship to the tribological mechanisms that actually occur and are accordingly only applicable to changed boundary conditions to a very limited extent. Cabanettes et al. developed an analytical model of a contact lubricated by minimum quantity lubrication on a tribometer under conditions similar to machining, which is used to calculate the proportion of dry and lubricated contact areas at the microscopic level as a function of the deformation state of the workpiece surface and the quantity of oil [15]. Although this model represents a promising approach and considers the tribological interaction at the micro level, the prediction quality is not sufficient and the usability in the context of chip formation simulations is significantly limited by the exclusive specification of the dry and lubricated portions.

In tribology science, there are a number of more advanced friction models, some of which also take lubrication effects into account [16]. For example, Zheng et al. have developed a molecular dynamics simulation which can represent the contact of rough technical surfaces in different lubrication states [17]. Although this corresponds in the basic constellation to the processes in the secondary shear zone, the extreme boundary conditions that occur in the chip formation zone are not taken into account and the models are too computationally intensive for coupling with numerical chip formation simulations. This is the starting point of the present work, which aims at developing a micro-tribological model of the contact between tool and workpiece/chip, using innovative numerical methods to predict the effect of a cutting fluid on the friction coefficient. This model will calculate the coefficient of friction based on a combination of the material's flow stress model and a model for the fluid behavior. It can be used for varying relative speeds, contact normal pressures, temperatures, tool surfaces and local lubrication conditions. On this basis, it will be integrated into a numerical chip formation simulation in the future.

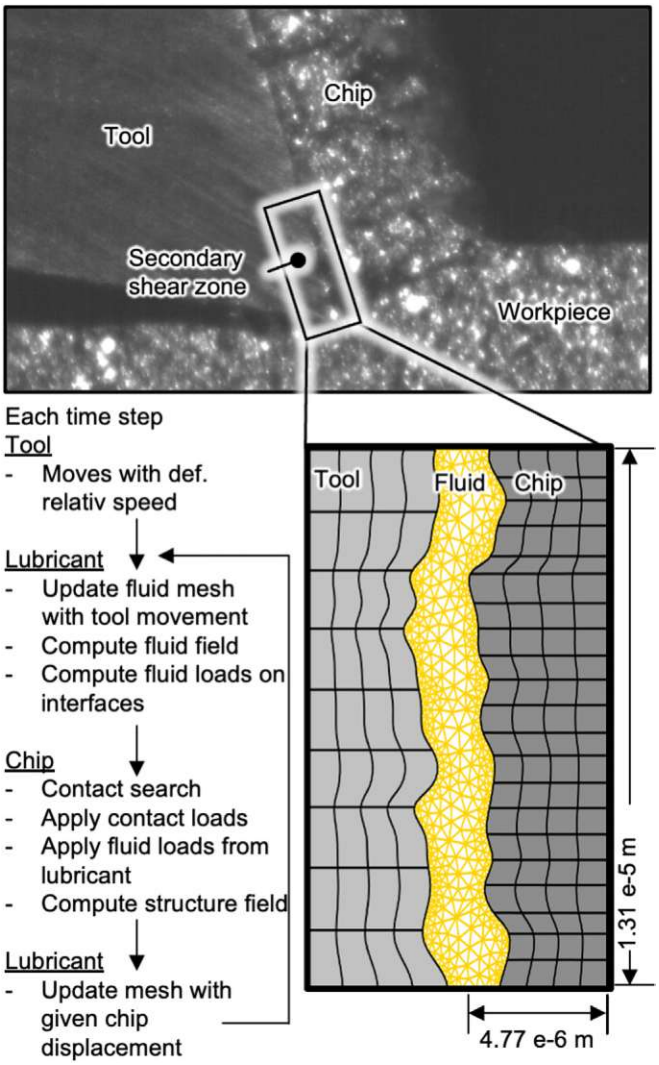


Fig. 1. Computational domain and solution process at each time step. The process is repeated until the solutions of both fluid and structure fields have converged.

In the following sections, a spline-based framework for the micro-tribological modelling is introduced. Next, primary test cases using the current state of the framework are presented. Section 4 serves as an extended outlook, where the calibration of the model based on experimental results is discussed.

Nomenclature

FSCI	Fluid-Structure-Contact Interaction
FEM	Finite Element Method
IGA	Isogeometric Analysis
NURBS	Non-Uniform Rational Basis-Spline

NEFEM NURBS-Enhanced FEM

$\boldsymbol{t}_c$	contact traction vector
ϵ	Penalty coefficient
$\boldsymbol{x}_q$	Gaussian quadrature point in current configuration
$\boldsymbol{x}_p$	Nearest projection point on contacting body
$\boldsymbol{n}_p$	Unit normal vector of at $\boldsymbol{x}_p$
ξ	Parametric coordinate of a spline
$S(\cdot)$	Spline mapping from parametric to physical space
$\boldsymbol{x}_s$	Physical coordinates mapped with spline from ξ
J	Objective function for nearest projection point search
Γ	Boundary of the fluid domain

2. Spline-Based Fluid-Structure-Contact Interaction Framework

From numerical analysis perspective, the contact zone during the lubricated cutting poses two major interactions: structure-structure and fluid-structure interactions, which corresponds to tool-chip and tool-lubricant-chip interactions, respectively. This naturally requires methods that can incorporate Fluid-Structure-Contact Interaction (FSCI). This work aims to develop a spline-based FSCI framework for a high-fidelity micro-tribological model, to consider aforementioned aspects of the interaction. Within the framework, tool, chip and lubricant are considered as rigid body, plastic solids (e.g., Johnson-Cook), and incompressible fluid, respectively.

In previous work [18], the geometrical modelling that is based on actual measurement of the surface topography was introduced. The resulting geometry is a spline that serves a base for further numerical analysis thereof. The fluid solver employs the FEM with a boundary description using splines, while the structural solver utilizes Isogeometric Analysis (IGA). This leads to conforming boundaries for both solvers in the partitions approach, as they can utilize identical splines for the interface [19]. As fluid loads are applied at certain points of the spline, they can be projected to control points of the structure by multiplying the load with corresponding basis functions. To further improve the boundary conformity, NURBS-Enhanced FEM (NEFEM) can also be used for the fluid solver, where boundary elements directly use splines to describe boundary sub-elements. As a results, fluid loads can directly be computed on control points. For detailed information, the reader is referred to [19].

The direct interaction of tool and chip, without lubricants, can be considered as a contact problem. The contact constraints are formulated as a gaussian-point-to-segment contact using the penalty method [20], where the contact traction $\boldsymbol{t}_c$ depends on a penalty coefficient ϵ, the gaussian quadrature point in current configuration $\boldsymbol{x}_q$, the nearest projection point $\boldsymbol{x}_p$ on a foreign structure from $\boldsymbol{x}_q$, and the normal vector $\boldsymbol{n}_p$:

$$\boldsymbol{t}_c = \epsilon\left(\left(\boldsymbol{x}_p - \boldsymbol{x}_q\right)\cdot \boldsymbol{n}_p\right)\boldsymbol{n}_p. \tag{1}$$

In the context of the micro-tribological model, the foreign structure is a tool, and quadrature points of interest are defined on the workpiece. There are two things to note regarding the use of splines: the normal vector and the nearest projection point. Approximating the normal vector can be avoided; due to the high continuity of spline basis functions, normal vectors are clearly defined everywhere except for the corners of a patch. However, finding the nearest projection point needs to be formulated as an optimization problem, since a closed form for the search does not exist. The implementation for this framework defines an objective function as a squared distance between query and mapped point. Then, it searches for a point where the derivative of the objective function is zero using newton methods. This formulation is applicable to any combination of a spline's parametric and physical dimension. Note that the nearest projection point can be acquired, if the search is performed on the boundary only. Using spline mapping S, and the parametric coordinate of spline ξ, the objective function J and its derivatives can be formulated as follows:

$$J(\xi) = \sum_{i=1}^{d}\left(S_i(\xi) - \boldsymbol{x}_{q_i}\right)^2 \tag{2}$$

$$\frac{\partial J}{\partial \xi_j} = 2\cdot\sum_{i=1}^{d}\left(S_i(\xi) - \boldsymbol{x}_{q_i}\right)\frac{\partial S_i}{\partial \xi_j} \tag{3}$$

$$\frac{\partial^2 J}{\partial \xi_j \partial \xi_k} = 2\cdot\sum_{i=1}^{d}\left(\frac{\partial S_i}{\partial \xi_k}\frac{\partial S_i}{\partial \xi_j} + \left(S_i(\xi) - \boldsymbol{x}_{q_i}\right)\frac{\partial^2 S_i}{\partial \xi_j \partial \xi_k}\right). \tag{4}$$

In practice, having a good initial guess is crucial for the search convergence. This can be achieved by, for example, using k-d trees [21] built with sampled spline points that are equally distributed in parametric space. As the application for this work requires numerous queries on a same non-deforming geometry, e.g., the tool, at every time step, the overhead for building a k-d tree is negligible compared to the solution process itself.

The surface roughness, which is considered in the model, can introduce the transitions between the FSI and the contact domain. Consequently, it can separate fluid domains, which may be also fully enclosed. The separating fluid domains can be efficiently tracked using, for example, PD-MUM [22], whereas using Robin-Neumann coupling scheme [23] satisfies incompressibility constraints of the fluids within the partitioned approach.

Finally, with all the aforementioned methods as building blocks, the solution process at each time step is described in Fig. 1.

3. Test Cases

This section will discuss test cases that were conducted using current implementations of the framework. The computational domains and materials for lubricant and tool are identical to the aforementioned geometrical modelling. Due to the absence of a plastic material model within the framework, the chip is assumed to be a hyper-elastic solid. However, it provides a possibility to test the methods with large and dynamic deformations. Without the metal-like material model, each test case focuses on an individual component of the framework, to examine its suitability as a building block at this

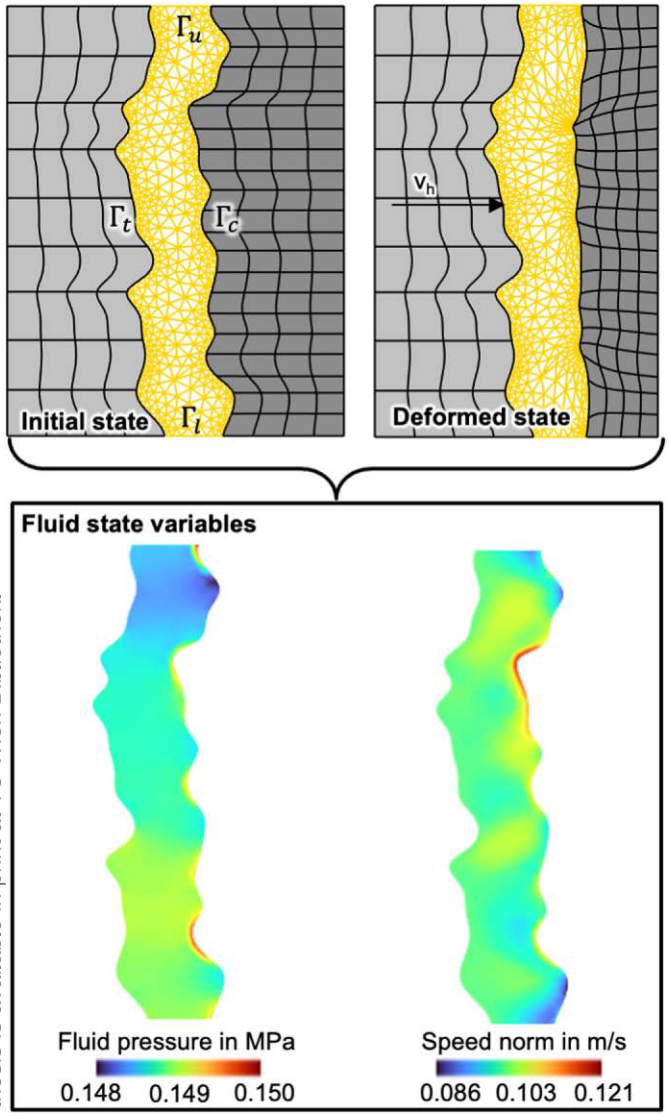


Fig. 2. FSI test case - enclosed domain with horizontal tool movement.

stage of the development. Note that the roughness is scaled by a factor of five to better observe the influence thereof.

First, the FSI implementations regarding an enclosed domain is tested. Starting from the computational domain shown in Fig. 2, a horizontal motion is applied at the tool interface Γ_t, and the speed of the fluid is restricted in vertical direction at the upper Γ_u and lower boundaries Γ_l. The last, Γ_c indicates the interface boundary where fluid-structure interaction takes place. The final state shows that the fluid load caused by the horizontal motion caused a chip deformation, while fulfilling the incompressibility requirements. As mentioned in the previous section, this component will be mainly responsible for accounting the influences of the lubricant that is expected to be in an enclosed domain. Furthermore, the horizontal motion will be replaced by the horizontal pressure, with which the results can be used as initial conditions for vertical motions that represents various stages, e.g., sticking and sliding, in the secondary shear zone.

On the same geometry and the same motion, a horizontal contact computation is tested. The results show that the chip is deformed based on the geometry of the tool, meaning that the roughness embedded in the geometry has an influence. Due to the nature of the penalty contact formulation, there exists a small overlap at the interface. The contact module will play a

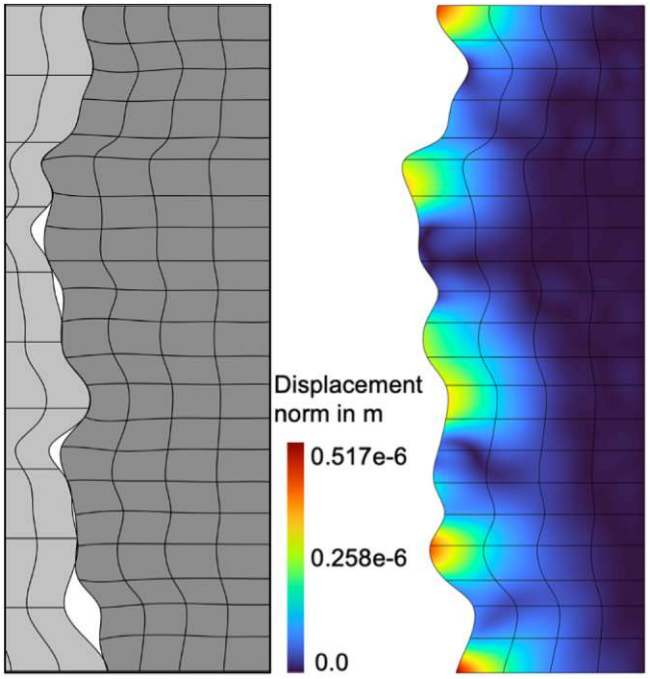


Fig. 3. Contact test case - chip deformation under horizontal tool movement.

crucial role within the framework as the impact of surface roughness, temperature, and relative speed on the chip behavior is significant in contact dynamics. Note that the current state of the contact formulation does not include oxide films and abrasive wearing.

Based on the previous contact deformation, a mesh update method can be tested. At each solution step, a fluid mesh can be updated based on the current state of the tool and chip. Utilizing the phantom-domain approach [22], first, a set of elements are defined on a virtual region where the fluid domain may be located. By distinguishing activated and deactivated zones, it is possible to avoid remeshing the whole fluid domain. In this context, the activated zones for lubricants correspond to

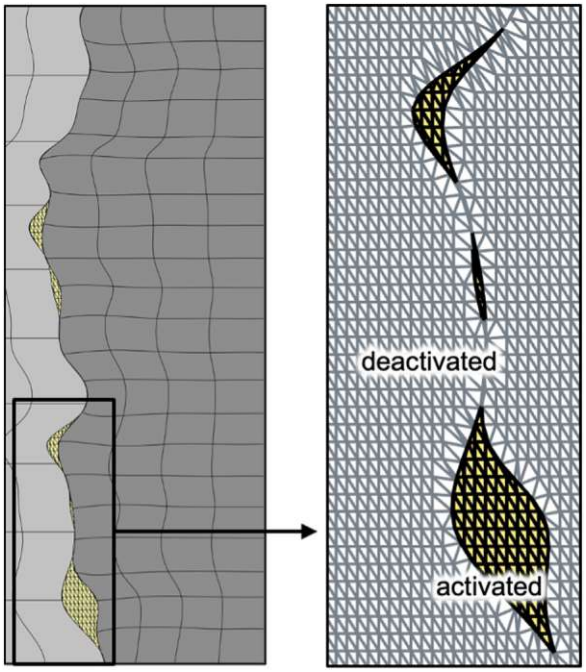


Fig. 4. Mesh update method test case. Yellow elements with black edges describe activated elements, on which current fluid domain is defined. The elements with grey edges describe deactivated elements, which are excluded during computation.

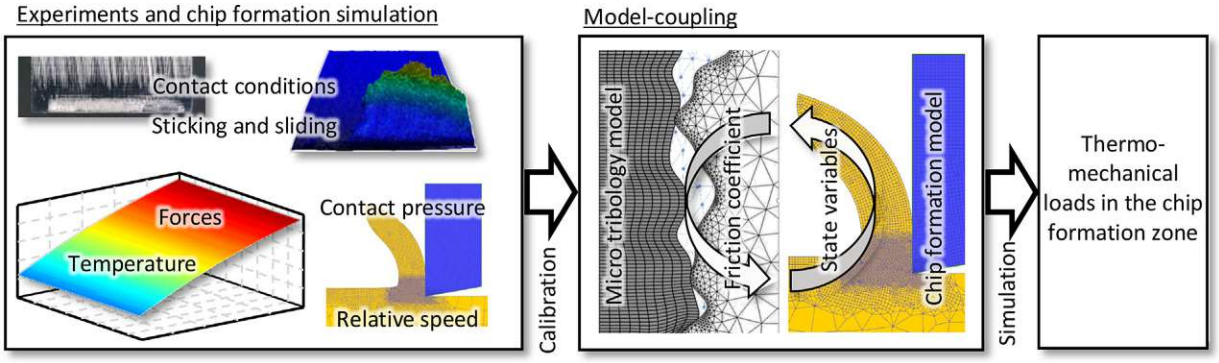


Fig. 5. Calibration and validation of micro-tribology model based on experimental results. Afterward, the model will be coupled with chip formation model.

the region between tool and chip. They are tracked by updating their boundaries based on the deforming structural domain, where the boundaries can be found using exactly the same formulation as the contact search.

The result of each test case indicates that the chosen methods for each component are capable of serving their purpose as a part of the framework. Aiming towards the micro-tribological analysis of the lubricated contact zone, the current work will be extended to model real world scenarios; the presented elements will be fully coupled and perform the steps described in Fig. 1 for vertical tool motions, i.e., tangential motions; the implementation of an appropriate thermo-elasto-plastic material model, e.g., the Johnson-Cook model, to better represent realistic chip behaviors; adhesion mechanisms; as well as extension to 3D simulations.

4. Experimental Calibration

Comprehensive experimental calibration is necessary to develop the numerical framework into a reliable model of the secondary shear zone. For this purpose, different levels of temperature, contact normal stress and relative speed from experimental measurement data as well as chip formation simulations [17, 24] already experimentally validated within the project are first imposed on the micromodel and the resulting cavities for the introduction of cooling lubricant are considered. This results in a division into sticking and sliding zones predicted by the model, which can be compared to studies still pending on the distribution of sticking and sliding fractions in the secondary shear zone. An initial validation is possible using passive forces from machining investigations in orthogonal cutting [17, 25]. After calibrating the friction coefficient depending on strain, strain-rate and temperature in the workpiece material, which is the output of the fluid-structure-contact interaction model, as well as reducing the high-fidelity model in order to save computation time, a coupling with a numerical chip formation simulation is planned. The research hypothesis is that with the newly developed model it is possible to represent the friction state in a locally graded manner, depending on the distribution of temperatures, contact normal stresses and relative speeds as well as the fluid distribution in the contact zone and thus to achieve an improved prediction of the thermomechanical load spectrum.

5. Conclusion and Outlook

In this work, the development of a spline-based framework for micro-tribology modelling and experimental validation methods for cutting applications is presented. The framework consists of numerical methods from various fields, e.g., FSI, contact, and moving mesh, to model complex phenomena that take place at the lubricated contact zone. Related to the state of research, this model is mainly applicable to intermittent cutting processes like milling, since in continuous cutting processes, the possibility of lubrication is very limited. Nevertheless, as a future perspective, the model can also be applied to interrupted turning or drilling as well as minimum quantity lubrication. Exemplary cases were computed to test the current state of the development, which showed that applied methods were suitable for their intended purposes. This in turn opens door to the next step, where all the components, with appropriate extensions, will be fully coupled to analyze lubricated contact zone. Finally, experimental methods to validate and calibrate the numerical models, as well as a brief outlook of following steps of the research were introduced.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 439919057. Furthermore, the authors appreciate Jacques Zwar and Florian Zwicke for their support in formulating nearest point search.

References

[1] Gajrani KK, Suvin PS, Kailas SV, Rajurkar KP, Sankar MR (2021) Machining of hard materials using textured tool with minimum quantity nano-green cutting fluid. *CIRP Journal of Manufacturing Science and Technology* (35):410–421.

[2] Pimenov DY, Mia M, Gupta MK, Machado AR, Tomaz ÍV, Sarikaya M, Wojciechowski S, Mikolajczyk T, Kapłonek W (2021) Improvement of machinability of Ti and its alloys using cooling-lubrication techniques: a review and future prospect. *Journal of Materials Research and Technology* 11:719–53.

[3] Chetan, Ghosh S, Venkateswara Rao P (2015) Application of sustainable techniques in metal cutting for enhanced machinability: a review. *Journal of Cleaner Production* 100:17–34.

[4] Fang N (2005) Tool-chip friction in machining with a large negative rake angle tool. *Wear* 258 (5-6):890–897.

[5] Denkena B, Biermann D (2014) Cutting edge geometries. *CIRP Annals* 63 (2):631–653.

[6] Hou Q, Yang X, Li D, Cheng J, Wang S, Xiao J, Li W (2022) Tribological performance of hydrophobic and micro/nano triangle textured rake face of cutting tools. *Applied Surface Science* (571):151250.
[7] Melkote SN, Grzesik W, Outeiro J, Rech J, Schulze V, Attia H, Arrazola P-J, M'Saoubi R, Saldana C (2017) Advances in material and friction data for modelling of metal machining. *CIRP Annals* 66(2):731–54.
[8] Zanger F, Bollig P, Schulze V (2017) Simulative Investigations on Different Friction Coefficient Models. *Procedia CIRP* 58:140–5.
[9] Atkins T (2015) Prediction of sticking and sliding lengths on the rake faces of tools using cutting forces. *International Journal of Mechanical Sciences* 91:33–45.
[10] Molinari A, Cheriguene R, Miguelez H (2011) Numerical and analytical modeling of orthogonal cutting: The link between local variables and global contact characteristics. *International Journal of Mechanical Sciences* 53(3):183–206.
[11] Malakizadi A, Hosseinkhani K, Mariano E, Ng E, Del Prete A, Nyborg L (2017) Influence of friction models on FE simulation results of orthogonal cutting process. *Int J Adv Manuf Technol* 88(9-12):3217–32.
[12] Courbon C, Sterle L, Cici M, Pusavec F (2020) Tribological Effect of Lubricated Liquid Carbon Dioxide on TiAl6V4 and AISI1045 under Extreme Contact Conditions. *Procedia Manufacturing* 47:511–6.
[13] Klocke F, Döbbeler B, Peng B, Lakner T (2017) FE-simulation of the Cutting Process under Consideration of Cutting Fluid. *Procedia CIRP* 58:341–6.
[14] Banerjee N, Sharma A (2014) Identification of a friction model for minimum quantity lubrication machining. *Journal of Cleaner Production* 83:437–43.
[15] Cabanettes F, Rolland J, Dumont F, Rech J, Dimkovski Z (2016) Influence of Minimum Quantity Lubrication on Friction Characterizing Tool–Aluminum Alloy Contact. *Journal of Tribology* 138(2).
[16] Vakis AI, Yastrebov VA, Scheibert J, Nicola L, Dini D, Minfray C, Almqvist A, Paggi M, Lee S, Limbert G, Molinari JF, Anciaux G, Aghababaei R, Echeverri Restrepo S, Papangelo A, Cammarata A, Nicolini P, Putignano C, Carbone G, Stupkiewicz S, Lengiewicz J, Costagliola G, Bosia F, Guarino R, Pugno NM, Müser MH, Ciavarella M (2018) Modeling and simulation in tribology across scales: An overview. *Tribology International* 125:169–99.
[17] Zheng X, Zhu H, Kiet Tieu A, Kosasih B (2013) A molecular dynamics simulation of 3D rough lubricated contact. *Tribology International* 67:217–21.
[18] Saelzer J, Alammari Y, Zabel A, Biermann D, Lee J, Elgeti S (2021) Characterisation and modelling of friction depending on the tool topography and the intermediate medium. *Procedia CIRP* 102:435–40.
[19] Hosters, N.; Helmig, J.; Stavrev, A.; Behr, M.; Elgeti, S.: Fluid–structure interaction with NURBS-based coupling. (2018) Computer Methods in Applied Mechanics and Engineering 332, S. 520-539.
[20] Wriggers, P.: Contact Boundary Value Problem and Weak Form. In: Computational Contact Mechanics. (2006) Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-32609-0_6
[21] Bentley, J. (1975). Multidimensional Binary Search Trees Used for Associative Searching. *Commun. ACM, 18(9), 509–517.*
[22] Hilger, D., Hosters, N., Key, F., Elgeti, S., & Behr, M. (2021). A Novel Approach to Fluid-Structure Interaction Simulations Involving Large Translation and Contact. *Isogeometric Analysis and Applications 2018*, 39–56. doi:10.1007/978-3-030-49836-8_3
[23] Make, M., Spenke, T., Hosters, N., Behr, M. (2022). Spline-based space-time finite element approach for fluid-structure interaction problems with a focus on fully enclosed domains. *Computers & Mathematics with Applications, 114, 210-224.*
[24] Afrasiabi M, Saelzer J, Berger S, Iovkov I, Klippel H, Röthlin M, Zabel A, Biermann D, Wegener K (2021) A Numerical-Experimental Study on Orthogonal Cutting of AISI 1045 Steel and Ti6Al4V Alloy: SPH and FEM Modeling with Newly Identified Friction Coefficients. *Metals* 11(11):1683.
[25] Saelzer J, Berger S, Iovkov I, Zabel A, Biermann D (2021) Modelling of the friction in the chip formation zone depending on the rake face topography. *Wear* 477:20380

2.3 Spline-Based Framework for Microscopic Contact Zone Modeling in Lubricated Orthogonal Cutting

Jaewook Lee, Jannis Saelzer, Jacques Zwar, Florian Zwicke, Felipe Gonzalez, Thomas Spenke, Norbert Hosters, Gero Polus, Andreas Zabel, and Stefanie Elgeti. "Spline-Based Framework for Microscopic Contact Zone Modeling in Lubricated Orthogonal Cutting". en. In: *International Journal for Numerical Methods in Engineering* 126.14 (July 2025), e70087. ISSN: 0029-5981, 1097-0207. DOI: 10.1002/nme.70087

Statement of Contribution: As first author, I was the primary architect of the full framework, coordinating with a team of co-authors on specific modules and with the TU Dortmund team for experimental context. My primary contributions were:

- integrating all individual components into the final, fully-coupled FSCI simulation tool,
- implementing the complete Johnson-Cook thermo-elasto-plastic model within the IGA solver,
- upgrading the contact algorithm to the more robust mortar knot-to-surface formulation and implementing the Robin-Neumann coupling scheme,
- developing, training, and validating the GPR surrogate model based on high-fidelity simulation data,
- performing all high-fidelity simulations and analyzing the physical findings that revealed the fundamental discrepancies with experimental trends,
- serving as the lead author of the manuscript.

International Journal for Numerical Methods in Engineering

WILEY

RESEARCH ARTICLE

Spline-Based Framework for Microscopic Contact Zone Modeling in Lubricated Orthogonal Cutting

Jaewook Lee[1] | Jannis Saelzer[2] | Jacques Zwar[1] | Florian Zwicke[1] | Felipe Gonzalez[3] | Thomas Spenke[3] | Norbert Hosters[3] | Gero Polus[2] | Andreas Zabel[2] | Stefanie Elgeti[1,3]

[1]Institute of Lightweight Design and Structural Biomechanics, TU Wien, Vienna, Austria | [2]Institute of Machining Technology, TU Dortmund, Dortmund, Germany | [3]Chair for Computational Analysis of Technical Systems, RWTH Aachen, Aachen, Germany

Correspondence: Jaewook Lee (jaewook.lee@tuwien.ac.at)

Received: 28 September 2024 | **Revised:** 4 June 2025 | **Accepted:** 2 July 2025

Funding: This work was supported by the Deutsche Forschungsgemeinschaft (Grant No. 439919057 and 523163077).

Keywords: fluid-structure-contact interaction | microscopic contact | orthogonal cutting | spline

ABSTRACT

Accurately predicting the coefficient of friction between tool, chip, and workpiece during machining is essential for a reliable and cost-effective process. In this context, current numerical methods are often based on homogenized approaches with friction models that use constant friction coefficients; thus, failing to capture local effects. In addition, often neither the effect of lubricants nor the local tool and chip topographies is accounted for. Towards improving the state of the art in both respects, in this paper, we present a micro-scale friction model that can be coupled with a meso-scale chip formation model. The micro model determines a local friction coefficient based on local temperature, contact pressure, cutting speed, and lubricant wetting. It also incorporates the experimentally determined tool and chip topography. A key assumption of the model is that the main contribution to the friction coefficient is the interlocking of local roughness peaks. Our numerical implementation uses a combination of isogeometric analysis (IGA) for the chip and finite elements with spline-based boundaries for the fluid. This approach ensures a smooth, conformal interface between the cooling fluid and the chip, allowing for direct spatial coupling. Temporally, a Robin–Neumann coupling is used, which is critical for handling fully enclosed fluid pockets. The direct contact between the tool and the chip is modeled using a mortar knot-to-surface approach. To ensure computational affordability in this multi-query FE^2 scenario, a surrogate model for the micro-scale model is created using Gaussian process regression.

1 | Introduction

In machining, the friction between the tool and chip, as well as the tool and workpiece, significantly influences the manufacturing process. On the one hand, the resulting tool wear has an impact on process reliability and production costs [1]. On the other hand, the chip-surface integrity has a decisive influence on the application properties of the manufactured components [2]. The tribological loads have a significant influence on both of these effects during the manufacturing process. Accordingly, the tribological boundary conditions of the machining process must be designed to minimize loads in the critical contact areas. To this end, an effective solution is the application of cooling lubricants. When used properly, these can significantly reduce the thermo-mechanical load collectives on the tool and chip [3, 4]. However, the use of cooling lubricants also has disadvantages, such as high production costs, as well as environmental and health hazards [5, 6]. Accordingly, production engineering

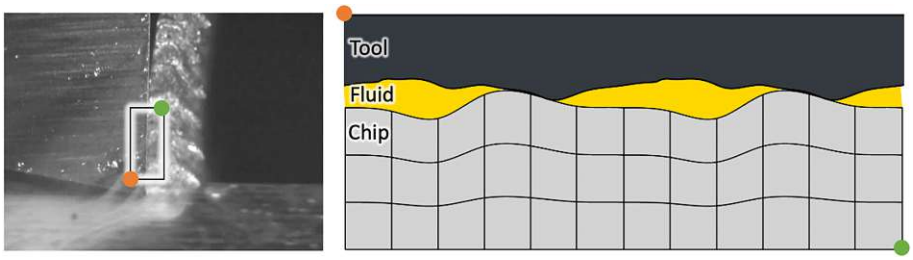


FIGURE 1 | Depiction of the secondary shear-zone in orthogonal cutting and the corresponding domain in the computational model. Orange and green dots are references for orientation.

research focuses on developing methods to minimize the use of cooling lubricants while maximizing their effectiveness. Achieving this requires comprehensive studies of machining processes, such as analyzing the impact of various cooling lubricants or tool surface topographies [7].

Numerical simulations have become an essential tool to gain a deeper understanding of the effects of cooling lubricants under specific processing conditions, offering insights into areas inaccessible to experimental observation. Depending on the scenario, different discretization methods are employed. In general, mesh-based approaches outperform particle-based methods regarding numerical efficiency and accuracy. In Reference [8], the impact of nanoparticle cooling is analyzed by predicting an accumulated heat transfer coefficient using a Lagrangian Discrete Phase Model (DPM), which is then integrated into a finite element-based chip formation simulation. By contrast, [9] investigates cryogenic Minimum Quantity Lubrication (MQL) during high-speed milling within a finite element context. Purely dry contact with varying cooling geometries is the focus of [10]. A finite difference method for analyzing the thermal behavior of tools during orthogonal cutting is presented in Reference [11], while [12] explores local contact conditions. Mesh-free methods, on the other hand, excel in handling topological changes. The focus of [13] is a Smoothed Particle Hydrodynamics (SPH) simulation of orthogonal metal cutting, emphasizing dynamic refinement for chip formation simulation and analysis. Similarly, SPH is employed in both [14] and [15]. In Reference [16], the Finite Pointset Method (FPM) is utilized to simulate orthogonal metal cutting, incorporating fluid-structure interaction (FSI) to account for cooling effects.

As established, numerical chip formation models are not only valuable tools for developing machining processes but are also frequently studied and refined in the literature. However, state-of-the-art models still face significant limitations, particularly in capturing the full complexity of the chip formation process. Accurately modeling phenomena such as the lubricating effects of cooling lubricants or microscopic adjustments to tool topography remains challenging. The primary issue lies in the friction models commonly used, as described in References [17, 18]. These friction models are often based on either the Coulomb model or Zorev's hybrid model [19], neither of which adequately account for the complex factors affecting friction in the chip formation zone, especially when constant model coefficients are employed [20, 21]. While empirical models, such as those presented in Reference [22], attempt to incorporate the effects of cooling lubricants or specific tool surface characteristics, they are difficult to calibrate and only valid under specific boundary conditions.

Therefore, the objective of this paper is to present a micro-scale friction model that can be integrated into a multi-scale framework for simulating chip formation, integrating both meso-scale and micro-scale simulations. At the meso-scale, any simulation model that can predict chip formation with an adjustable friction model can be used. The friction coefficients are then no longer constant, but are determined via a Python interface that links to the micro-scale model, which is the primary focus of this paper (cf. Figure 1). Within an FE^2 framework, the micro-scale model is evaluated at each Gauss point. The meso-scale model provides localized values for contact pressure, velocity, and temperature, while the micro-scale model determines corresponding friction coefficients.

Based on the assumption that the main contribution to the friction coefficient is the interlocking of local roughness peaks, the micro-scale model performs a high-resolution fluid-structure-contact interaction simulation involving the tool, modeled as a rigid body, the lubricating fluid, treated as an incompressible Newtonian fluid, and the chip, represented using the Johnson–Cook plasticity model. Fluid-structure-contact interaction (FSCI) forms the foundational basis for modeling tribological phenomena, particularly where lubrication and interface deformation are involved. The handling of such coupled problems remains an active area of research, especially in the finite element context, due to its inherent complexity and relevance in scientific and engineering applications [23]. Within monolithic frameworks, Burman et al. proposed Nitsche-based coupling schemes using a fully Eulerian approach [24–27] also employed Nitsche-type formulations, discretizing fluid domains via the CutFEM method, while [28] developed a unified-field fictitious domain method. In contrast, partitioned strategies have also been explored: [29] introduced a fully Eulerian variational framework for multiphase FSCI, and [30] employed a boundary-conforming space-time formulation [31] addressed FSCI using a fictitious domain approach with penalization. Notably, porous medium formulations have been adopted to model mechanically consistent transitions in fluid-structure and structure-structure interactions [24–26, 32], with Ager et al. [26] specifically addressing rough surface contact. It is known that the friction in the chip formation zone results from a superposition of different mechanisms of action. The effects in the sticking zone between the chip and rake face are determined by material adhesion and mechanical interlocking [33], although relatively little is known about the extent to which each of the mechanisms has an effect. The

study by Gregorio et al. suggests that for most metals (with the exception of titanium, which has an high affinity for reaction), a significant proportion of the frictional load can be attributed to mechanical interlocking [34], which is why this is envisaged as the defining mechanism for static friction in this study. Key to this approach is the local roughness topography, which has been determined experimentally using confocal microscopy. For the numerical implementation, we employ a combination of isogeometric analysis (IGA) for the chip and finite elements with spline-based boundaries for the fluid. This approach ensures a smooth, conformal interface between the cooling fluid and the chip, allowing for a direct spatial coupling. Temporally, a Robin–Neumann coupling is used, which is crucial for handling completely enclosed fluid pockets. The direct contact between the tool and chip is modeled using a mortar knot-to-surface approach. To guarantee computational affordability in this multi-query scenario of FE^2, a surrogate model for the micro-scale model is created using Gaussian Process regression.

In the remainder of the paper, we first present the numerical framework, detailing the physical models, the high-fidelity numerical methods, and the surrogate model. Furthermore, the behavior of the framework is demonstrated via numerical experiments.

2 | Numerical Framework: Spline-Based Fluid-Structure-Contact Interaction

In this section, we present a spline-based fluid-structure-contact interaction model that is generally applicable, but for the purpose of this paper, has been applied as a micro-model for lubricated contact in orthogonal cutting. The section is organized according to the individual components of the framework: Structure, fluid, and the interaction thereof. For each problem, we present theoretical backgrounds and solution methods.

2.1 | Structural Problem

In the framework, the behavior of a structural body is expressed in terms of a displacement field. Based on a total Lagrangian viewpoint, the displacement field $\boldsymbol{u}$ is formulated with respect to a reference (undeformed) configuration Ω^s_{ref} and is governed by the equation of motion

$$\rho^s \frac{d^2\boldsymbol{u}}{dt^2} = \nabla_{ref} \cdot \boldsymbol{P} + \boldsymbol{b}^s \qquad \text{in} \quad \Omega^s_{ref} \tag{1}$$

where ρ^s, $\boldsymbol{b}^s$, $\boldsymbol{P}$ denote structural density, body force, and first Piola–Kirchhoff stress, respectively; ∇_{ref} implies the gradient with respect to the reference configuration. The stress here is a function of the deformation gradient, i.e., $\boldsymbol{P}(\boldsymbol{F})$, which represents a constitutive response. Additionally, the structural problem is subjected to initial conditions as well as Dirichlet and Neumann boundary conditions:

$$\boldsymbol{u}(t=0) = \boldsymbol{u}_0 \qquad \text{in} \quad \Omega^s_{ref} \tag{2}$$

$$\boldsymbol{u} = \boldsymbol{g}^s \qquad \text{on} \quad \Gamma^s_D \tag{3}$$

$$\boldsymbol{P}\boldsymbol{n}^s_{ref} = \boldsymbol{h}^s \qquad \text{on} \quad \Gamma^s_N \tag{4}$$

where Γ^s_D and Γ^s_N denote Dirichlet and Neumann part of the boundary; $\boldsymbol{g}^s$ and $\boldsymbol{h}^s$ are prescribed boundary displacement and traction, respectively; $\boldsymbol{n}^s_{ref}$ is the outward pointing unit normal vector in the reference configuration.

The structural problem is solved numerically using the MFEM library [35] with NURBS-based isogeometric analysis (IGA) [36] in space and a *generalized-α* time integration scheme [37].

There are two structures to consider in the micro model: The tool and the chip. While the tool is modeled as a rigid body, the chip is represented by a nonlinear plasticity model with Johnson–Cook hardening, which reads as follows:

$$\sigma_y = \left[A + B\left(\bar{\epsilon}^{pl}\right)^n\right]\left[1 + C ln\left(\frac{\dot{\bar{\epsilon}}^{pl}}{\dot{\epsilon}_0}\right)\right]\left(1 - \hat{\theta}^m\right) \tag{5}$$

with

$$\hat{\theta} \equiv \begin{cases} 0 & \text{for} \quad \theta \le \theta_{transition}, \\ (\theta - \theta_{transition})/(\theta_{melt} - \theta_{transition}) & \text{for} \quad \theta_{transition} \le \theta < \theta_{melt}, \\ 1 & \text{for} \quad \theta_{melt} \le \theta \end{cases} \tag{6}$$

where σ_y is the yield stress, $\bar{\epsilon}^{pl}$ is the equivalent plastic strain, $\dot{\bar{\epsilon}}^{pl}$ is the equivalent plastic strain rate, $\hat{\theta}$ is non-dimensional temperature. $A, B, n, m, \theta_{melt}$, and $\theta_{transition}$ denote material parameters. For plasticity, we follow J_2 flow theory at finite strains formulation presented in Reference [38]. Finally, thermal effects are considered adiabatic.

2.2 | Fluid Problem

The cooling lubricant in the micro model is represented using the incompressible Navier–Stokes equations, which describe the dynamics of a viscous fluid in terms of velocity $\boldsymbol{v}$ and pressure p:

$$\rho^f\left(\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v} \cdot \nabla \boldsymbol{v}\right) - \nabla \cdot \boldsymbol{\sigma} = \rho^f \boldsymbol{b}^f \qquad \text{in} \quad \Omega^f \tag{7}$$

$$\nabla \cdot \boldsymbol{v} = 0 \qquad \text{in} \quad \Omega^f \tag{8}$$

where ρ^f is the fluid density and $\boldsymbol{b}^f$ the volume forces. We consider the cooling lubricant to be Newtonian, leading to a stress tensor $\boldsymbol{\sigma}$ defined as:

$$\boldsymbol{\sigma} = -p\boldsymbol{I} + \eta\left(\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T\right) \tag{9}$$

where η is the dynamic viscosity. The fluid problem is also subjected to initial conditions as well as Dirichlet and Neumann, and Robin boundary conditions:

$$\boldsymbol{v}(t=0) = \boldsymbol{v}_0 \qquad \text{in} \quad \Omega^f \tag{10}$$

$$\boldsymbol{v} = \boldsymbol{g}^f \qquad \text{on} \quad \Gamma^f_D \tag{11}$$

$$\boldsymbol{\sigma}\boldsymbol{n}^f = \boldsymbol{h}^f \qquad \text{on} \quad \Gamma^f_N \tag{12}$$

$$\alpha^R \boldsymbol{v} + \boldsymbol{\sigma}\boldsymbol{n}^f = \boldsymbol{h}^f + \alpha^R \boldsymbol{g}^f \qquad \text{on} \quad \Gamma^f_R \tag{13}$$

where Γ^f_D, Γ^f_N, and Γ^f_R denote Dirichlet, Neumann, Robin part of the fluid boundary with prescribed velocity $\boldsymbol{g}^f$ and traction $\boldsymbol{h}^f$; α^R is the scalar weighting factor for Robin boundary condition.

The fluid problem is solved using stabilized finite elements in space [39] and an implicit Euler time integration scheme. Moreover, an arbitrary Lagrangian–Eulerian formulation is employed to accommodate a moving fluid domain, where the mesh is adapted using a linear elastic mesh-update method (EMUM) [40]. Moving boundaries are represented using NURBS, which enables the definition of displacements directly through the control points. Further information about the fluid solver can be found in References [41, 42].

2.3 | Interface Interactions

2.3.1 | Partitioned FSI Approach

As the fluid and structural domains share an interface, the solution fields become naturally dependent on each other and require coupling. In this work, we adopt a partitioned approach: Two individual solvers—one for each domain—are coupled through a coupling module. Where necessary, properties at the interface are exchanged iteratively to fulfill temporal and spatial coupling conditions.

To fulfill temporal coupling conditions at each time step, the solvers are strongly coupled through fixed-point iterations [43]. Starting with initial guesses on both sides, the following procedure is repeated until convergence: (1) given structural displacement at the interface, compute fluid load (kinematic continuity); (2) given the fluid load at the interface, compute structural displacement (dynamic continuity). This is often referred to as Dirichlet–Neumann coupling [44]. Note that a Robin–Neumann coupling is used to circumvent the so-called incompressibility dilemma of classical Dirichlet–Neumann schemes for fully-enclosed incompressible fluids [45, 46]. Accordingly, in step (1), a Robin boundary condition is imposed on the flow problem, formed from the current deformation state and the previous interface tractions, as proposed by [46].

The spatial coupling regards the transfer of the related properties that become boundary conditions for the other domain. This is especially relevant when different discretization techniques are employed, requiring a conservative projection method. With the finite interpolation method [44, 47], arbitrary data associated at specific points of the interface can be projected to the destination domain by multiplying their corresponding basis functions. Within this work, we apply the spline-based formulation presented in Reference [44].

2.3.2 | Contact Enforcement

The interactions between structural bodies are considered a contact problem. Since two bodies cannot physically overlap and the location is unknown a priori, this becomes a nonlinear optimization problem to solve at the interface level [48]. Throughout this work, we adopt the mortar knot-to-surface formulation [49, 50]. Enforcing the impenetrability condition starts by defining a normal gap g_n between a point $\boldsymbol{x}_q$ on contact boundary Γ_C^s and its nearest projection point $\boldsymbol{x}_p$ on the opposing contact boundary Γ_{CO}^s:

$$g_n = (\boldsymbol{x}_p - \boldsymbol{x}_q) \cdot \boldsymbol{n}_p \tag{14}$$

where $\boldsymbol{n}_p$ denotes the outward pointing normal at the projection point. Following the mortar knot-to-surface formulation [49, 50], we then compute the contact pressure p_{na} using weighted average of the normal gaps g_{na}, associated to the individual control points:

$$g_{na} = \frac{\int_{\Gamma_C^s} N \; g_n \; \mathrm{d}\Gamma}{\int_{\Gamma_C^s} N \; \mathrm{d}\Gamma} \tag{15}$$

$$p_{na} = \epsilon g_{na} \tag{16}$$

where N denotes the basis function for a given control point and ϵ is a penalty factor. It is important to note that for splines of higher degrees than one, finding the nearest projection point becomes an optimization problem. Within this work, we minimize a squared distance function $\boldsymbol{J}_d(\xi) = (\boldsymbol{x}_q - \boldsymbol{x}_p(\xi))^2$ to determine the parametric position ξ of the closest projection point and employ the Newton method, Levenberg–Marquardt algorithm [51], and sequential quadratic programming [52] to ensure convergence.

3 | Micro-Model Towards Friction Modeling

The micro model is intended to provide insight into the complex tribological interactions in the secondary shear zone (see Figure 1). In particular, we aim to develop a robust and extensible setup that can incorporate all factors influencing friction coefficients and behavior referenced in characterization experiments [22, 53]: Tool topography, lubricants, normal pressure, and relative tangential velocity. As a result, we can use the model to characterize friction on any topography and boundary conditions of interest, e.g., to improve the simulation of chip formation used for tool development.

Using the proposed spline-based fluid-structure-contact interaction framework detailed in Section 2, we have developed a model that represents the contact zone at the microscopic level and can incorporate the above factors. In the remainder of this paper, we investigate its capability to characterize friction.

3.1 | High-Fidelity Model

The starting point of the numerical model is the geometric modeling of the domain, incorporating the microscopic tool topography into the analysis. Based on actual measurements of the surface topography of both the tool and chip, we create a spline representation using a spline fitting technique [54]. The fitted surface is then extruded to form a volumetric structural domain suitable for analysis using IGA.

The two structural domains, Ω_T and Ω_W, correspond to the tool and chip, respectively. The tool is treated as a rigid body, while the chip is modeled as a Johnson–Cook material [55]. Within this work, we adopt the material parameters presented in Reference [53]. The fluid within the domain Ω_F is incompressible. At the chip boundary, Γ_W, the chip can interact with both the fluid and the tool. It is important to note that this model permits only a single type of interaction at each physical location of the boundary at any given time. However, due to the high continuity of the B-spline basis function, it is possible for loading from both sources to manifest in the same location.

A partitioned approach is used to couple the fluid and structural domains. For spatial coupling, we utilize a spline-based finite interpolation method [44], which leverages identical spline representations at the interface. The Robin–Neumann temporal coupling scheme mitigates the incompressibility dilemma [46], arising from the fluid domain being surrounded by the structural domain. Domain deformations caused by the moving fluid interface are accounted for using EMUM [40].

To account for pressure and relative velocity, we apply normal pressure to the vertical degrees of freedom (y) and prescribe horizontal (x) displacement at Γ_B. On $\Gamma_{periodic}$, a periodic boundary condition is implemented—for both fluid and structural domains—to reflect that the model represents a portion of a larger statistically repeating domain. Lastly, we incorporate a temperature introduced as an initial condition to characterize the material state.

With these conditions established, the computation begins by moving the chip at a specified relative velocity while applying pressure on the bottom of the chip domain and initial fluid loads. The structural solver produces a displacement field that satisfies the contact condition, which is then relayed to the fluid solver. Based on the interface displacement, the fluid solver deforms the domain and computes the resulting loads at the interface, which are passed back to the structural solver. This process iterates at each time step until convergence is achieved for both solvers.

3.2 | Reduced Model

The objective of the high-fidelity model is to predict the friction behavior based on the microscopic topography, material state, and boundary conditions. In particular, the coefficients of friction under certain boundary conditions—such as pressure, temperature, and relative velocity—can be used in chip formation simulations to provide valuable insight into cutting operations. Within our micro-model, the friction coefficient μ is calculated as the ratio of the vertical T_x and horizontal T_y components of the integrated applied traction vector $\boldsymbol{T}$ along the interface:

$$\mu = \frac{\int_{\Gamma_W} T_x \, \mathrm{d}\Gamma}{\int_{\Gamma_W} T_y \, \mathrm{d}\Gamma} \tag{17}$$

For dynamic simulations, a time-averaged value μ within given time steps is used. However, calculating these coefficients locally for every element of the meso simulation using a high-fidelity model is computationally too demanding. Therefore, a reduced model is needed to efficiently provide this information with significantly less computational effort.

In this work, we use Gaussian process regression [56] to develop a reduced model for the characterized friction coefficient. This model approximates the friction coefficient as a function of pressure, relative velocity, and temperature. We employ the implementation provided in *scikit-learn* [57] package with radial basis function kernel. For the training set, 900 data points are prepared, and hyperparameters of the kernel are also optimized during the training.

4 | Numerical Examples

In this section, we examine the model's behavior with respect to application-related boundary conditions. The main goal of the micro-model is to provide a flexible friction modeling framework that serves as the basis for further chip formation simulation and tool design. Given pressure, relative velocity, and initial temperature, the model computes the chip's deformation and applied forces, which can then be used to determine friction coefficients (see Equation (17)).

We utilize the geometry created based on surface topography measurement (see Figure 2). Then, we define the range of required boundary conditions based on the process conditions reported in previous experiments, see [22, 53]. The prescribed pressure and relative velocity at Γ_B range from 300 MPa to 1,300 MPa and 22.5 m/min to 200 m/min, respectively. The initial temperature of the chip ranges from 25°C to 900°C. From these ranges, we uniformly sample values for characterization computations.

The penalty factor for computing contact pressure (see Equation (16)) is 2×10^{18}. The geometry close to the contact region is uniformly refined until the friction coefficient computation converges with less than 0.7%. As discussed at the end of Section 3.1, periodic boundary conditions are applied so only a single interval represents our structural domain in the model problem. We employed Newton's method with a line search technique—specifically, the Quadratic Variation of the Norm of the Residual Line Search [58]—to solve the nonlinear system of equations, and used a direct solver for the associated linear systems. For visualization purposes, Figures 3 and 4 show a section that is two intervals long to help illustrate the structural deformation at a given time step. The structural material parameters used in the computations are shown in Table 1 and fluid material parameters are shown in Table 2.

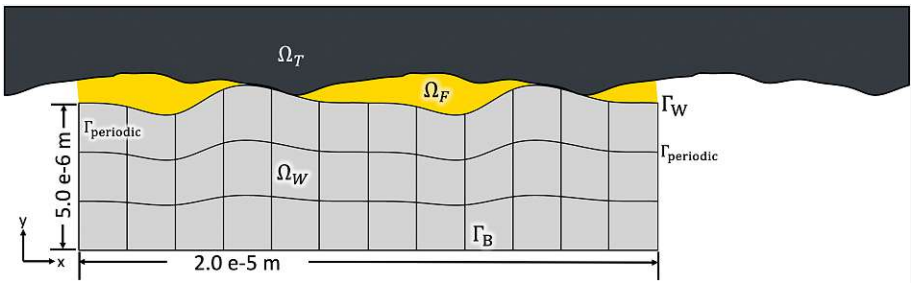
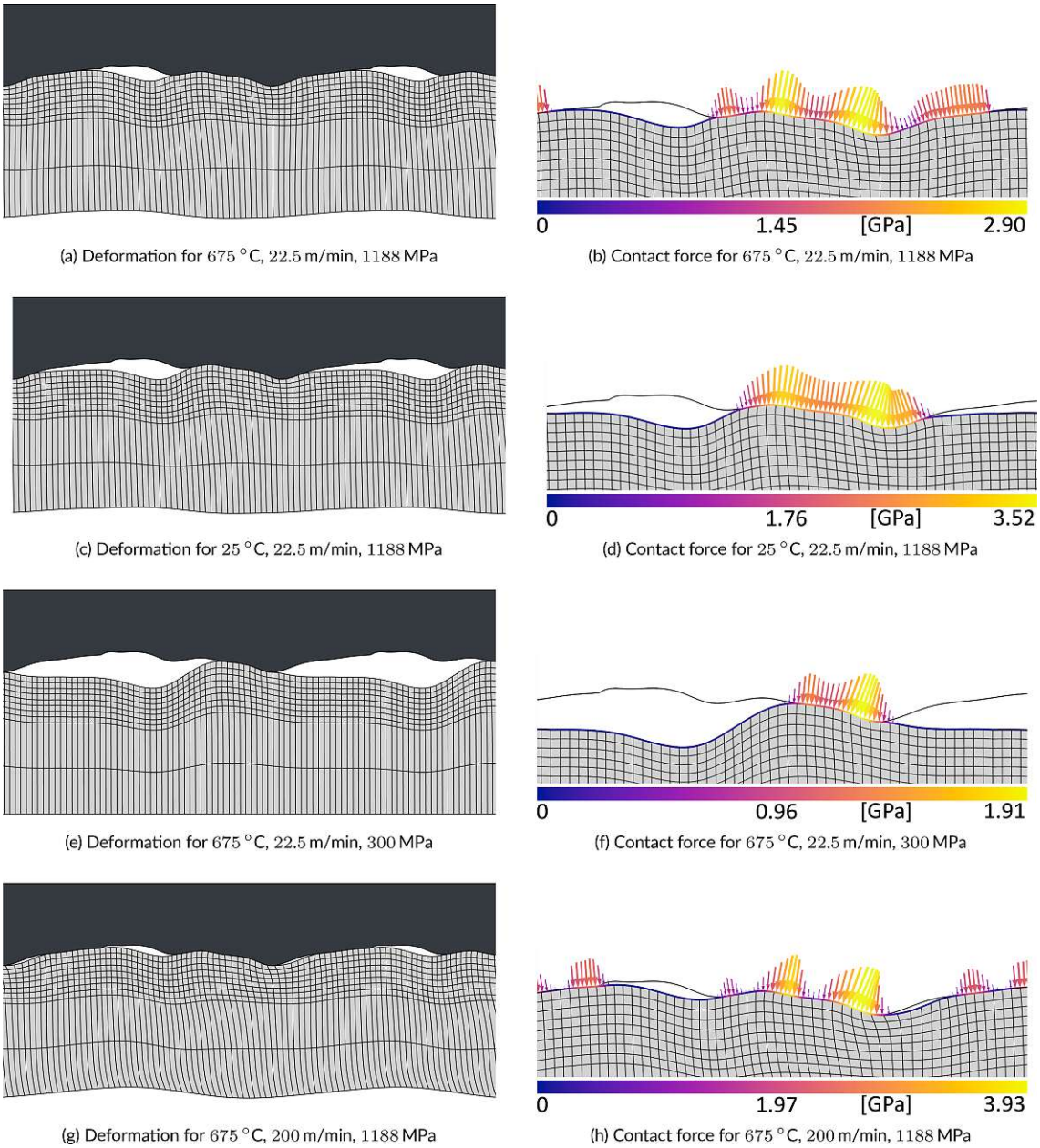


FIGURE 2 | Computational domain of the high-fidelity model.



(a) Deformation for 675 °C, 22.5 m/min, 1188 MPa

(b) Contact force for 675 °C, 22.5 m/min, 1188 MPa

(c) Deformation for 25 °C, 22.5 m/min, 1188 MPa

(d) Contact force for 25 °C, 22.5 m/min, 1188 MPa

(e) Deformation for 675 °C, 22.5 m/min, 300 MPa

(f) Contact force for 675 °C, 22.5 m/min, 300 MPa

(g) Deformation for 675 °C, 200 m/min, 1188 MPa

(h) Contact force for 675 °C, 200 m/min, 1188 MPa

FIGURE 3 | Results of the high-fidelity model computation with dry contact for various boundary conditions after a complete time step of 2.5×10^{-7} s, showing both the deformation state of the chip as well as the associated contact force. The left column shows two intervals of the deformed domain, spanning a total width of 2×10^{-5} m, while the right column presents a magnified view of approximately one such interval, focusing on the geometric peaks of the tool. The local normal contact force on the tool-chip interface achieves a maximum value of 3.93 GPa.

The resulting deformed states of the chips under various boundary conditions for dry contact are shown in Figure 3. Due to the prescribed pressure, the chip deforms and closes the void between itself and the tool. This also causes the bottom of the chip to deform, given a sufficient amount of pressure proportional to the size of the void. With increasing temperature (Figure 3a,c), the deformation becomes more significant. A similar trend can be observed with increasing boundary pressure (see Figure 3a,e).

Most notably, the force distribution changes considerably with increased relative tangential velocity (Figure 3b,h). At higher velocities, the contact zone between chip and tool becomes less concentrated, resulting in an uneven distribution of the contact forces with local peaks. These high-pressure zones are located at geometric spikes of the chip, resulting in steep normal vectors at the regions of contact with high magnitudes of the horizontal forces. Accordingly, these regions contribute more importantly to the overall resistance of the chip movement, resulting in higher friction coefficients.

Unlike dry contact, lubricated contact, as shown in Figure 4, under the same boundary conditions, exhibits more evenly distributed forces on the chip at the boundary Γ_W. Force peaks are only observed at the initial minimal gap that separates the tool and the chip, which consequently leads to a local deformation. The incompressibility of the fluid results in a more uniform force

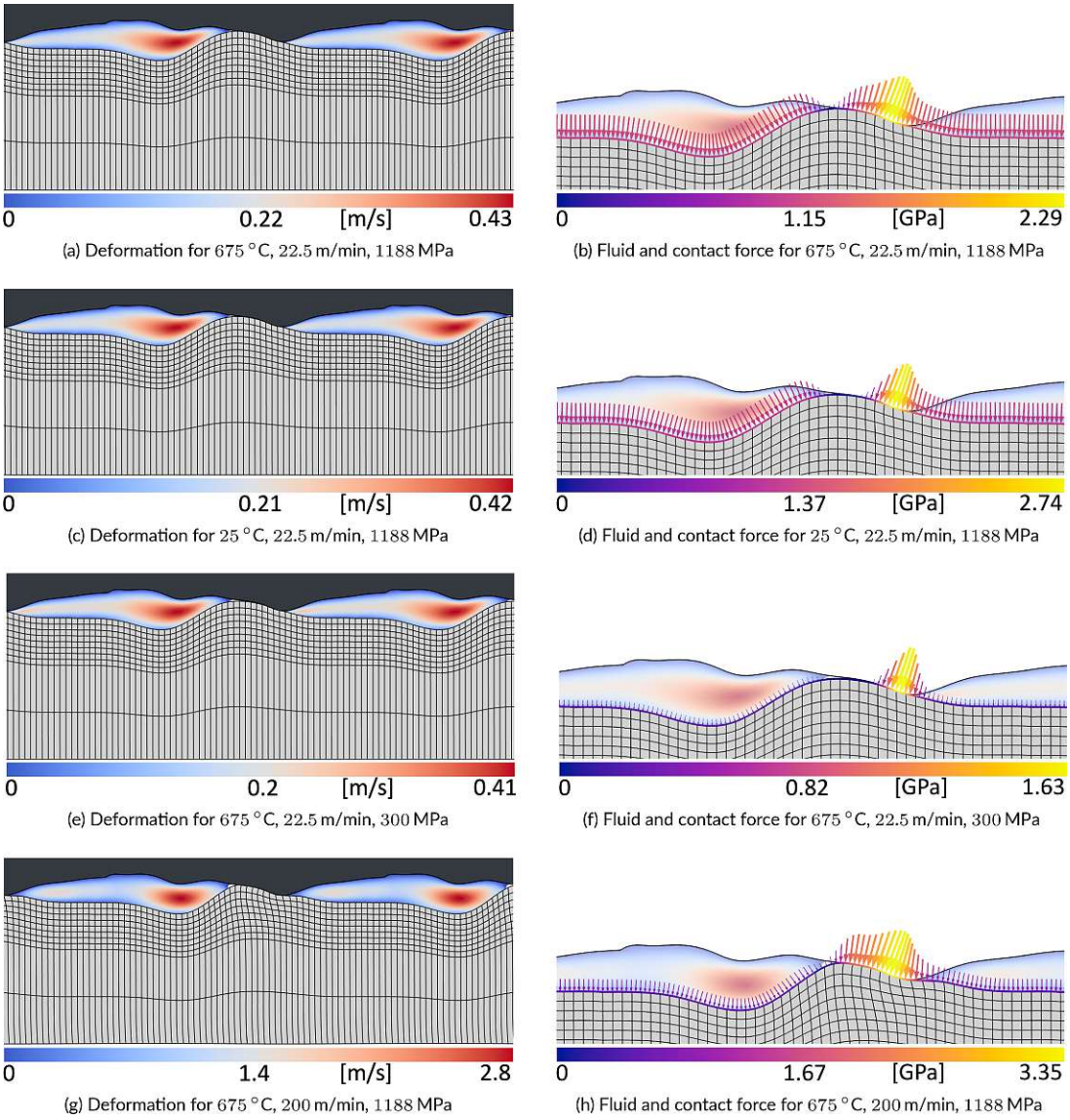


FIGURE 4 | Results of the high-fidelity model computation with lubricated contact for various boundary conditions after 2.5×10^{-7} s. The fluid is encapsulated in the gaps between the chip and tool; the colors depict the velocity magnitude of the fluid. Viewpoints are consistent with those in Figure 3. The local normal contact force on the tool achieves a maximum value of 3.35 GPa.

TABLE 1 | Material parameters for structure.

Structure parameters	Value
Density	7,800 kg m^{-3}
Young's Modulus	205 GPa
Poisson's Ratio	0.29
A	288 MPa
B	695 MPa
C	0.034
n	0.2835
m	1.3558
$\dot{\epsilon}_0$	0.004 S^{-1}
$\theta_{transition}$	20
θ_{melt}	1,500
Specific Heat Capacity	450 J kg^{-1} K^{-1}
Heat Fraction	0.9

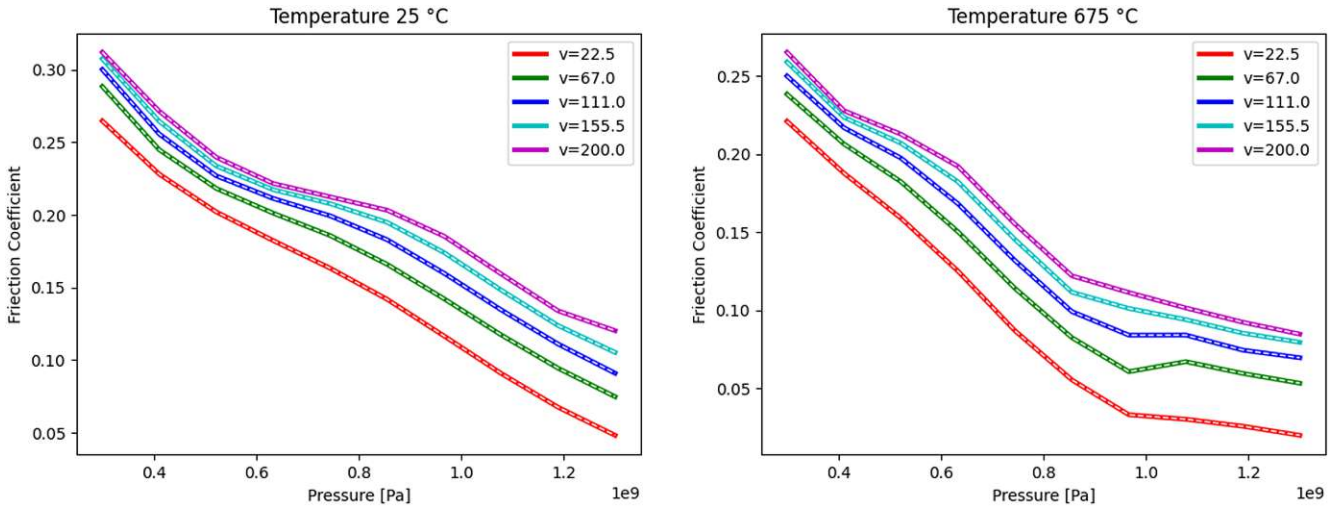
TABLE 2 | Material parameters for incompressible fluid.

Fluid parameters	Value
Density	800 kg m^{-3}
Dynamic viscosity	0.032 kg m^{-1} s^{-1}

distribution in all areas that are touched by the fluid. Additionally, this incompressibility prevents further deformation of the chip, which can occur in dry contact due to the presence of voids instead of fluid. By increasing relative velocity, an increase in the force magnitude can be seen at the contact boundary. In some regions, the fluid force points in the same direction as the chip motion. Unless the chip has a flat surface, such an area can always be observed. This is an additional effect of the lubrication fluid, which now causes thrust and therefore reduces the coefficient of friction.

(a) Velocity magnitude (b) Pressure magnitude

FIGURE 5 | Fluid field under boundary conditions at 675°C, 22.5 m min^{-1} and 300 MPa after 2.5×10^{-7} s. The full domain is shown in Figure 4e.



(a) Friction coefficient at 25 °C (b) Friction coefficient at 675 °C

FIGURE 6 | Friction coefficient (dry) from high-fidelity model and reconstruction with reduced model (white dashed line) at varying boundary conditions.

A detailed view of the internal flow field is shown in Figure 5. The magnitude of the velocity is within the order of the relative velocity between chip and tool, which is consistent with the mesh velocity boundary condition during FSI coupling. Furthermore, the velocity vector is aligned with the direction of chip motion. The pressure field shows a homogeneous profile. While the left side, where the chip is fed forward, causing a "squeezing" effect, shows a slightly higher pressure, the relative difference between the maximum and minimum value is 27%.

As briefly mentioned in Section 3.2, the friction coefficient in this work is computed as a ratio between the integrated force in the horizontal and vertical directions (see Equation (17)). Figure 6 shows the result of friction characterization without lubrication. We observe a general decrease in the friction coefficient with increasing pressure and increasing relative velocity. On the other hand, an increase in temperature results in a decrease in the friction coefficient.

Friction coefficients from the lubricated model, shown in Figure 7, share similar tendencies with the dry model. The friction coefficient decreases with increasing pressure, increasing relative velocity, and increasing temperature.

The corresponding reduced models are trained using the 900 sampled results for dry and lubricated contact. The trained models are tested with 200 randomly sampled datasets, which resulted in 1% (dry) and 1.8% (lubricated) mean relative error. The reconstructed data from the reduced models are shown as dashed lines in Figures 6 and 7, which agree well with the high-fidelity data. Training data reconstruction resulted in 2.3×10^{-5}% (dry) and 9.8×10^{-4}% (lubricated) mean relative error.

The trained reduced model's predictions, along with the 95% confidence intervals, are illustrated in Figures 8 and 9. They show the predicted mean values as a solid line, while the shaded regions represent the confidence intervals, indicating the uncertainty of the model's predictions. Since the training data is uniformly sampled, the uncertainties remain low within the data range. A slight increase in uncertainty is observed towards both ends of the interval, particularly for the dry model. As expected, the uncertainty grows more significantly outside the data range.

The computational cost of the high-fidelity model is measured and compared with the reduced model by evaluating their execution times on a single core of an AMD Ryzen 7 3700X processor. The average execution times for the high-fidelity dry and lubricated models are 1.3 h and 6.4 h, respectively. In contrast, the reduced model consistently completes in under 0.00015 s, offering speed-ups of 3.12×10^7 and 1.54×10^8, respectively. It is important to note that evaluating the reduced model requires a preliminary offline phase, which involves preparing the dataset and training the model. Running high-fidelity simulations for all sampling points—both dry and lubricated—required approximately 7,000 h of wall-clock time. With a high-fidelity dataset in place, training the reduced model took only 20 s. Potentially, active learning could reduce the cost of high-fidelity simulations by iteratively adding sampling points in high-uncertainty regions of the input space and retraining the model.

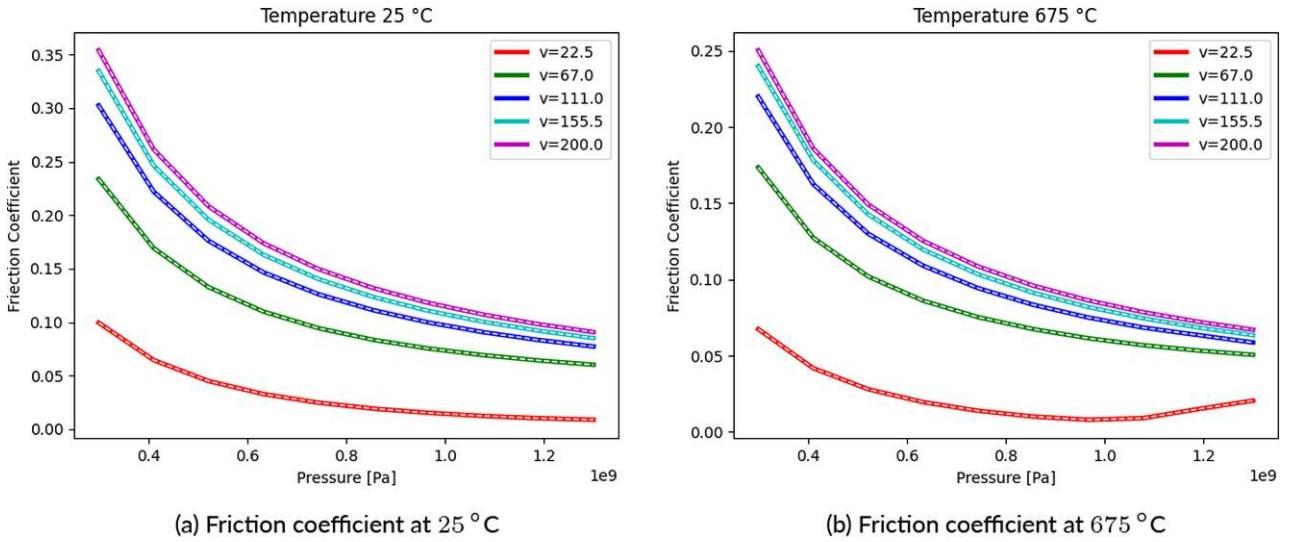


FIGURE 7 | Friction coefficient (lubricated) from high-fidelity model and reconstruction with reduced model (white dashed line) at varying boundary conditions.

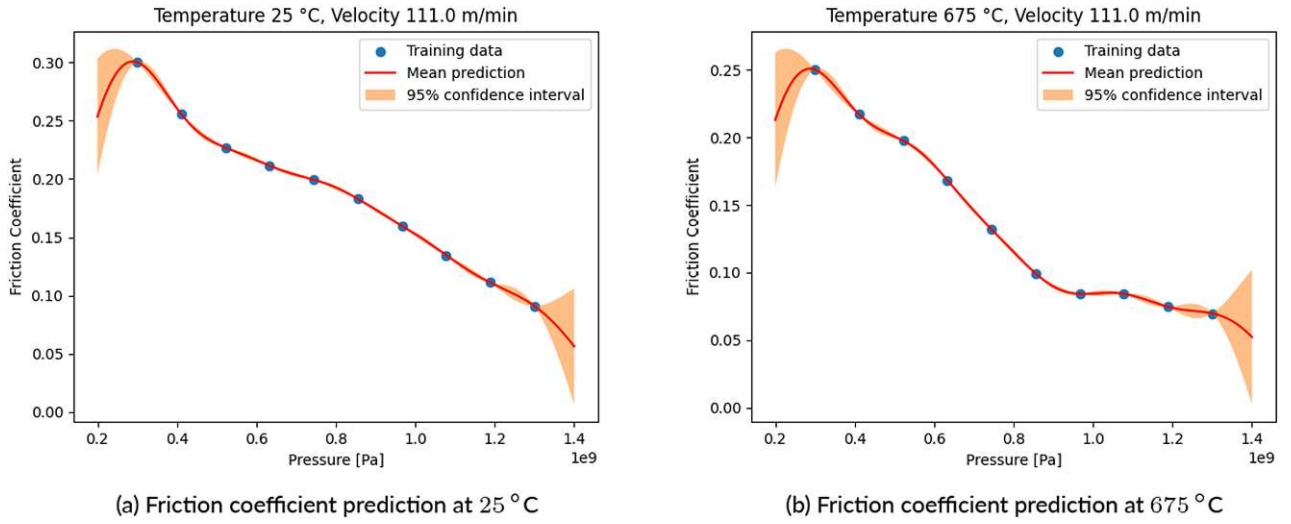


FIGURE 8 | Friction coefficient predictions (dry) and their 95% confidence interval from reduced model at 111 m min^{-1}. The sample range includes regions that are not present in the training data.

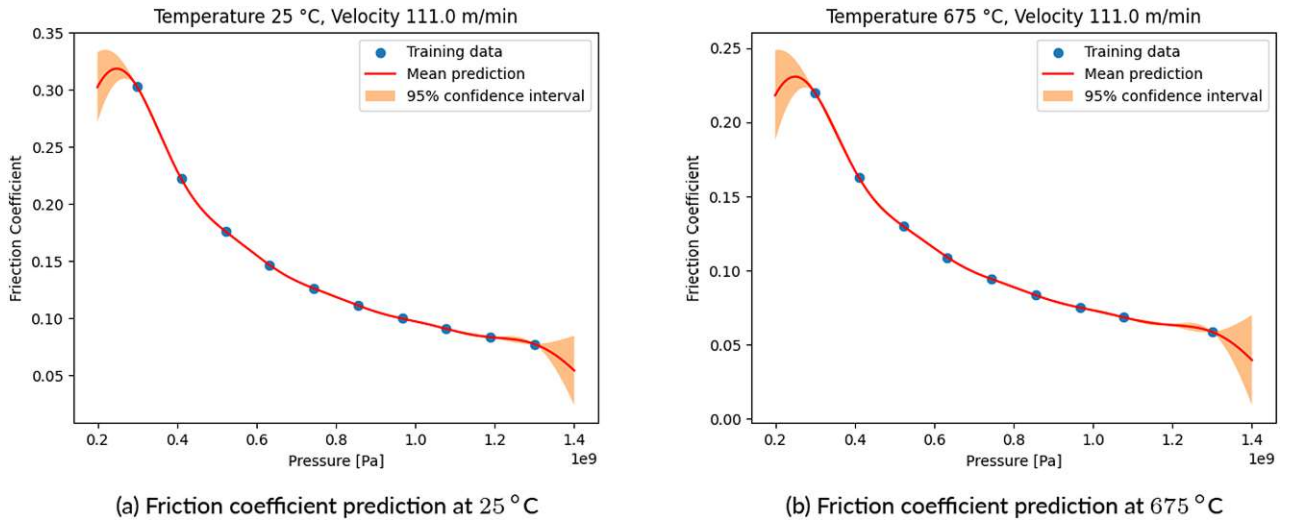


FIGURE 9 | Friction coefficient predictions (lubricated) and their 95% confidence interval from reduced model at 111 m min^{-1}. The sample range includes regions that are not present in the training data.

5 | Discussion and Outlook

In this work, we have investigated the applicability of a high-fidelity fluid-structure-contact interaction model and a corresponding reduced model for friction modeling in chip formation simulations. The numerical experiments show that, in contrast to existing models, our micro-model can (1) provide a localized friction coefficient based on boundary conditions for temperature, contact pressure, and relative velocity; (2) incorporate the effects of lubricating fluids; and (3) account for an experimentally determined characteristic tool and chip topography. Numerically, a key feature of the model is the use of spline-based methods, which accurately represent the measured topographies, provide continuous and uniquely defined normal vectors for the contact formulation, and also exhibit higher accuracy per degree of freedom compared to standard finite elements. Another important aspect of the presented method is the temporal Robin–Neumann coupling during the partitioned FSI approach, which is the only way to handle the fully enclosed fluid pockets, which are inevitable in this modeling approach. Overall, a fully functional fluid-structure-contact interaction framework has been developed that provides a number of features required by the micro-model. Further, the framework is fully modular, facilitating process-specific extensions if necessary. Furthermore, a reduced model based on Gaussian process regression was presented. It provides accurate results at a significantly lower computational cost, allowing multi-scale coupled simulation in a reasonable time. The reduced model is entirely non-intrusive and can be continuously extended as novel data points become available.

Using both the high-fidelity and the reduced model, we investigate the resulting friction coefficients for varying temperature, relative velocity, and contact pressure over a wide range of experimentally deducted process-relevant values. For the dry model, it is observed that an increase in temperature leads to a decrease in the coefficient of friction. This is due to the temperature dependence of the plasticity model, where a higher temperature results in higher deformations and thus a larger contact area between the tool and chip. The effect of pressure is similar to that of temperature, as an increase in pressure results in a bigger contact area. Combined with the more uniform force distribution, the horizontal components on opposite sides of the local peaks counteract one another, leading to a reduction in the overall friction coefficient. The relative velocity, however, has the opposite effect. As the velocity increases, the distribution of contact zones becomes sparser, leading to more irregular and locally concentrated forces at the chip interface, particularly around the spikes of the microscopic surface geometry, where steeper slopes generate greater ratios between the horizontal and vertical force components. This ultimately results in higher coefficients of friction.

It is important to note that to ensure higher accuracy and lower uncertainty in the reduced model, training data should include samples that cover the entire expected range of the boundary conditions, i.e., input state variables that originate from the meso model, during the chip formation simulation process. In cases where runtime query is out of the bounds of the training data, i.e., if the model needs to extrapolate, it can be trained iteratively by adding samples in areas where uncertainties are high, continuing this process until the desired accuracy is achieved.

The lubricated contact model generally shows a coefficient of friction behavior similar to the dry model. An important difference, however, is that the deformation pattern of the chip is now less localized. This is due to the fact that the forces applied at the interface are more homogeneous, and the contact area is limited to areas without lubricant. In fact, it is these areas that contribute most to the coefficient of friction.

If one compares the described behavior of the micro model with experimentally observed tendencies, e.g., in Reference [59], one notices discrepancies with respect to the influence of velocity and contact pressure: In the real system, the coefficient of friction increases with normal contact pressure until it levels off at some point. Furthermore, as the relative velocity increases, the coefficient of friction decreases. One can conclude that neither effects is dominated by the mechanical interaction between the rough surfaces, as was our initial assumption.

In future work, the presented modular framework may therefore be extended to include additional tribological mechanisms. One example is an adhesion model, which can help to capture interatomic forces. The influence of thermal effects can be handled more accurately by considering a temperature field for each domain rather than a single adiabatic temperature field for workpieces. This can also allow for heat transfer between domains, both fluid-structure and structure-structure interactions, incorporating cooling/heating effects into the model. In addition, the fluid could be modeled as a shear-thinning fluid to more accurately capture its behavior. Boundary conditions, e.g., pressure-based resistance, for the fluid can be further explored to accommodate domains with open ends. The magnitude may be validated through experimental studies. Finally, since the secondary shear zone has already undergone plastic deformation, information about the accumulated plastic strain can be used as an initial condition in the model. On the numerical side, the spatial coupling could benefit from the adoption of the NURBS-enhanced finite-element method [60], which could further improve the accuracy of the fluid simulation. This could render the implementation of more advanced mesh refinement methods possible, resulting in a higher accuracy of the high-fidelity model.

Subsequently, the micro-model will be integrated into meso-scale chip formation simulations. Our goal is to adopt this multi-scale framework to enhance the tool design process, enabling predictions of tool, chip, and product behavior in relation to the previously mentioned influencing factors. The framework's applicability will be assessed through experimental validation.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) — Projektnummer 439919057, 523163077.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. P. Mativenga, J. Schoop, I. Jawahir, et al., "Engi . . . mance and Customized Applications: A Review," *CIRP Journal of Manufacturing Science and Technology* 52 (2024): 212–228, https://doi.org/10.1016/j.cirpj.2024.06.001.

2. A. Malakizadi, R. Bertolini, F. Ducobu, Z. Kilic, M. C. Magnanini, and A. Shokrani, "Recent Advances in Modelling and Simulation of Surface Integrity in Machining-A Review," *Procedia CIRP* 115 (2022): 232–240. 10th CIRP Global Web Conference-Material Aspects of Manufacturing Processes, https://doi.org/10.1016/j.procir.2022.10.079.

3. F. Pape, G. Poll, L. Ellersiek, B. Denkena, and H. Liu, "Tribological Effects of Metalworking Fluids in Cutting Processes," *Lubricants* 11, no. 5 (2023): 224, https://doi.org/10.3390/lubricants11050224.

4. R. Teti, D. D'Addona, and T. Segreto, "Microbial-Based Cutting Fluids as Bio-Integration Manufacturing Solution for Green and Sustainable Machining," *CIRP Journal of Manufacturing Science and Technology* 32 (2021): 16–25, https://doi.org/10.1016/j.cirpj.2020.09.016.

5. M. Chen, R. Peng, L. Zhao, R. Chen, and Y. Luo, "Effects of Minimum Quantity Lubrication Strategy With Internal Cooling Tool on Machining Performance in Turning of Nickel-Based Superalloy GH4169," *International Journal of Advanced Manufacturing Technology* 118, no. 11–12 (2022): 3673–3689, https://doi.org/10.1007/s00170-021-08194-w.

6. T. He, N. Liu, H. Xia, et al., "Progress and Trend of Minimum Quantity Lubrication (Mql): A Comprehensive Review," *Journal of Cleaner Production* 386 (2023): 135809, https://doi.org/10.1016/j.jclepro.2022.135809.

7. I. Jawahir, H. Attia, D. Biermann, et al., "Cryogenic Manufacturing Processes," *CIRP Annals* 65, no. 2 (2016): 713–736, https://doi.org/10.1016/j.cirp.2016.06.007.

8. I. Nouzil, A. Eltaggaz, I. Deiab, and S. Pervaiz, "Numerical Cfd-Fem Model for Machining Titanium Ti-6al-4v With Nano Minimum Quantity Lubrication: A Step Towards Digital Twin," *Journal of Materials Processing Technology* 312 (2023): 117867, https://doi.org/10.1016/j.jmatprotec.2023.117867.

9. H.-P. Zhang, Q.-Y. Zhang, Y. Ren, T. Shay, and G.-L. Liu, "Simulation and Experiments on Cutting Forces and Cutting Temperature in High Speed Milling of 300m Steel Under Cmql and Dry Conditions," *International Journal of Precision Engineering and Manufacturing* 19 (2018): 1245–1251.

10. S. Saketi, J. Östby, and M. Olsson, "Influence of Tool Surface Topography on the Material Transfer Tendency and Tool Wear in the Turning of 316l Stainless Steel," *Wear* 368-369 (2016): 239–252, https://doi.org/10.1016/j.wear.2016.09.023.

11. W. Grzesik, M. Bartoszuk, and P. Nieslony, "Finite Difference Analysis of the Thermal Behaviour of Coated Tools in Orthogonal Cutting of Steels," *International Journal of Machine Tools & Manufacture* 44, no. 14 (2004): 1451–1462, https://doi.org/10.1016/j.ijmachtools.2004.05.008.

12. B. Denkena, H. Liu, F. Pape, et al., "Simulation of Local Contact Conditions in the Secondary Shear Zone in Dry and Wet Metal Cutting," *Procedia CIRP* 117 (2023): 293–298. 19th CIRP Conference on Modeling of Machining Operations, https://doi.org/10.1016/j.procir.2023.03.050.

13. M. Afrasiabi, M. Roethlin, H. Klippel, and K. Wegener, "Meshfree Simulation of Metal Cutting: An Updated Lagrangian Approach With Dynamic Refinement," *International Journal of Mechanical Sciences* 160 (2019): 451–466, https://doi.org/10.1016/j.ijmecsci.2019.06.045.

14. J. Limido, C. Espinosa, M. Salauen, and J. L. Lacome, "Sph Method Applied to High Speed Cutting Modelling," *International Journal of Mechanical Sciences* 49, no. 7 (2007): 898–908, https://doi.org/10.1016/j.ijmecsci.2006.11.005.

15. M. Dehghani, A. Shafiei, and M. M. Abootorabi, "Analyzing Orthogonal Cutting Process Using SPH Method by Kinematic Cutting Tool," *Comptes Rendus. Mécanique* 348, no. 2 (2020): 149–174, https://doi.org/10.5802/crmeca.6.

16. E. Uhlmann, E. Barth, T. Seifarth, M. Höchel, J. Kuhnert, and A. Eisenträger, "Simulation of Metal Cutting With Cutting Fluid Using the Finite-Pointset-Method," *Procedia CIRP* 101 (2021): 98–101, https://doi.org/10.1016/j.procir.2021.02.013.

17. P. Arrazola, T. Özel, D. Umbrello, M. Davies, and I. Jawahir, "Recent Advances in Modelling of Metal Machining Processes," *CIRP Annals* 62, no. 2 (2013): 695–718, https://doi.org/10.1016/j.cirp.2013.05.006.

18. W. Grzesik and J. Rech, "Influence of Machining Conditions on Friction in Metal Cutting Process-A Review," *Mechanik* 4 (2019): 242–248, https://doi.org/10.17814/mechanik.2019.4.33.

19. N. Zorev, "Inter-Relationship Between Shear Processes Occurring Along Tool Face and Shear Plane in Metal Cutting," in *International Research in Production Engineering* (ASME, 1963).

20. B. Peng, T. Bergs, D. Schraknepper, T. Smigielski, and F. Klocke, "Development and Validation of a New Friction Model for Cutting Processes," *International Journal of Advanced Manufacturing Technology* 107, no. 11–12 (2020): 4357–4369, https://doi.org/10.1007/s00170-019-04709-8.

21. C. Bonnet, F. Valiorgue, J. Rech, et al., "Identification of a Friction Model–Application to the Context of Dry Cutting of an AISI 316L Austenitic Stainless Steel With a TiN Coated Carbide Tool," *International Journal of Machine Tools and Manufacture* 48, no. 11 (2008): 1211–1223, https://doi.org/10.1016/j.ijmachtools.2008.03.011.

22. J. Saelzer, S. Berger, I. Iovkov, A. Zabel, and D. Biermann, "Modelling of the Friction in the Chip Formation Zone Depending on the Rake Face Topography," *Wear* 477 (2021): 203802, https://doi.org/10.1016/j.wear.2021.203802.

23. A. Vakis, V. Yastrebov, J. Scheibert, et al., "Modeling and Simulation in Tribology Across Scales: An Overview," *Tribology International* 125 (2018): 169–199, https://doi.org/10.1016/j.triboint.2018.02.005.

24. E. Burman, M. A. Fernández, and S. Frei, "A Nitsche-Based Formulation for Fluid-Structure Interactions With Contact. *ESAIM*," *Mathematical Modelling and Numerical Analysis* 54, no. 2 (2020): 531–564, https://doi.org/10.1051/m2an/2019072.

25. E. Burman, M. A. Fernández, S. Frei, and F. M. Gerosa, "A Mechanically Consistent Model for Fluid-Structure Interactions With Contact Including Seepage," *Computer Methods in Applied Mechanics and Engineering* 392 (2022): 114637, https://doi.org/10.1016/j.cma.2022.114637.

26. C. Ager, B. Schott, A. Vuong, A. Popp, and W. A. Wall, "A Consistent Approach for Fluid-Structure-Contact Interaction Based on a Porous Flow Model for Rough Surface Contact," *International Journal for Numerical Methods in Engineering* 119, no. 13 (2019): 1345–1378, https://doi.org/10.1002/nme.6094.

27. C. Ager, A. Seitz, and W. A. Wall, "A Consistent and Versatile Computational Approach for General Fluid-Structure-Contact Interaction Problems," *International Journal for Numerical Methods in Engineering* 122, no. 19 (2020): 5279–5312, https://doi.org/10.1002/nme.6556.

28. C. Wang, P. Sun, Y. Zhang, J. Xu, Y. Chen, and J. Han, "A Unified-Field Monolithic Fictitious Domain-Finite Element Method for Fluid-Structure-Contact Interactions and Applications to Deterministic Lateral Displacement Problems," *Journal of Computational Physics* 510 (2024): 113083, https://doi.org/10.1016/j.jcp.2024.113083.

29. X. Mao, B. Rath, and R. Jaiman, "A 3d Phase-Field Based Eulerian Variational Framework for Multiphase Fluid-Structure Interaction With Contact Dynamics," *Computer Methods in Applied Mechanics and Engineering* 429 (2024): 117172, https://doi.org/10.1016/j.cma.2024.117172.

30. D. Hilger, N. Hosters, F. Key, S. Elgeti, and M. Behr, "A Novel Approach to Fluid-Structure Interaction Simulations Involving Large Translation and Contact," in *Isogeometric Analysis and Applications 2018* (Springer International Publishing, 2020), 39–56, https://doi.org/10.1007/978-3-030-49836-83.

31. O. Yakhlef and C. M. Murea, "Numerical Simulation of Dynamic Fluid-Structure Interaction With Elastic Structure-Rigid Obstacle Contact," *Fluids* 6, no. 2 (2021): 51, https://doi.org/10.3390/fluids6020051.

32. F. M. Gerosa and A. L. Marsden, "A Mechanically Consistent Unified Formulation for Fluid-Porous-Structure-Contact Interaction," *Computer Methods in Applied Mechanics and Engineering* 425 (2024): 116942, https://doi.org/10.1016/j.cma.2024.116942.

33. H. K. Tönshoff, *Werkzeuge für die moderne fertigung* (Expert Verlag, 1993).

34. A. Gregorio, T. Santos, R. Rossi, A. Jesus, J. Outeiro, and P. Rosa, "Tribology of Metal Cutting: Newly Formed Underside of Chip," *Procedia CIRP* 82 (2019): 136–141, https://doi.org/10.1016/j.procir.2019.04.034.

35. R. Anderson, J. Andrej, A. Barker, et al., "MFEM: A Modular Finite Element Methods Library," *Computers & Mathematics with Applications* 81 (2021): 42–74, https://doi.org/10.1016/j.camwa.2020.06.009.

36. J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs, *Isogeometric Analysis: Toward Integration of CAD and FEA* (Wiley. OCLC, 2009), ocn335682757.

37. J. Chung and G. M. Hulbert, "A Time Integration Algorithm for Structural Dynamics With Improved Numerical Dissipation: The Generalized-Method," *Journal of Applied Mechanics* 60, no. 2 (1993): 371–375, https://doi.org/10.1115/1.2900803.

38. J. Simo and T. Hughes, *Computational Inelasticity* (Springer, 2006).

39. J. Donea and A. Huerta, *Finite Element Methods for Flow Problems*, 1st ed. (Wiley, 2003), https://doi.org/10.1002/0470013826.

40. A. Johnson and T. Tezduyar, "Mesh Update Strategies in Parallel Finite Element Computations of Flow Problems With Moving Boundaries and Interfaces," *Computer Methods in Applied Mechanics and Engineering* 119, no. 1–2 (1994): 73–94, https://doi.org/10.1016/0045-7825(94)00077-8.

41. T. Spenke, N. Hosters, and M. Behr, "A Multi-Vector Interface Quasi-Newton Method With Linear Complexity for Partitioned Fluid-Structure Interaction," *Computer Methods in Applied Mechanics and Engineering* 361 (2020): 112810. arXiv:2001.07947 [cs], https://doi.org/10.1016/j.cma.2019.112810.

42. T. Spenke, M. Make, and N. Hosters, "A Robin-Neumann Scheme With Quasi-Newton Acceleration for Partitioned Fluid-Structure Interaction," *International Journal for Numerical Methods in Engineering* 124, no. 4 (2023): 979–997, https://doi.org/10.1002/nme.7151.

43. U. Küttler and W. A. Wall, "Fixed-Point Fluid-Structure Interaction Solvers With Dynamic Relaxation," *Computational Mechanics* 43, no. 1 (2008): 61–72, https://doi.org/10.1007/s00466-008-0255-5.

44. N. Hosters, J. Helmig, A. Stavrev, M. Behr, and S. Elgeti, "Fluid-Structure Interaction With NURBS-Based Coupling," *Computer Methods in Applied Mechanics and Engineering* 332 (2018): 520–539, https://doi.org/10.1016/j.cma.2018.01.003.

45. U. Küttler, C. Förster, and W. A. Wall, "A Solution for the Incompressibility Dilemma in Partitioned Fluid-Structure Interaction With Pure Dirichlet Fluid Domains," *Computational Mechanics* 38, no. 4–5 (2006): 417–429, https://doi.org/10.1007/s00466-006-0066-5.

46. M. Make, T. Spenke, N. Hosters, and M. Behr, "Spline-Based Space-Time Finite Element Approach for Fluid-Structure Interaction Problems With a Focus on Fully Enclosed Domains," *Computers & Mathematics with Applications* 114 (2022): 210–224, https://doi.org/10.1016/j.camwa.2022.03.035.

47. A. Beckert, "Coupling Fluid (CFD) and Structural (FE) Models Using Finite Interpolation Elements," *Aerospace Science and Technology* 4, no. 1 (2000): 13–22, https://doi.org/10.1016/S1270-9638(00)00111-5.

48. P. Wriggers, *Computational Contact Mechanics* (Springer, 2006), https://books.google.at/books?id=oBMBy4eMYsQC.

49. F. F. Temizer, P. Wriggers, and T. Hughes, "Contact Treatment in Isogeometric Analysis With NURBS," *Computer Methods in Applied Mechanics and Engineering* 200, no. 9–12 (2011): 1100–1112, https://doi.org/10.1016/j.cma.2010.11.020.

50. L. De Lorenzis, F. F. Temizer, P. Wriggers, and G. Zavarise, "A Large Deformation Frictional Contact Formulation Using NURBS-Based Isogeometric Analysis: APPLICATION OF NURBS-BASED ISOGEOMETRIC ANALYSIS," *International Journal for Numerical Methods in Engineering* 87, no. 13 (2011): 1278–1300, https://doi.org/10.1002/nme.3159.

51. J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer. OCLC, 2006), ocm68629100.

52. D. Kraft, *A Software Package for Sequential Quadratic Programming* (DFVLR, 1988).

53. J. Saelzer, Y. Alammari, A. Zabel, D. Biermann, J. Lee, and S. Elgeti, "Characterisation and Modelling of Friction Depending on the Tool Topography and the Intermediate Medium," *Procedia CIRP* 102 (2021): 435–440, https://doi.org/10.1016/j.procir.2021.09.074.

54. L. A. Piegl and W. Tiller, *The NURBS Book*, 2nd ed. (Springer, 1997).

55. G. Johnson and W. Cook, "A Constitutive Model and Data for Metals Subjected to Large Strains, High Strain Rates and High Temperatures," in *Proceedings of the Seventh International Symposium on Ballistics, The Hague, The Netherlands*, vol. 21 (American Defense Preparedness Association, 1983), 541–548.

56. C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, 2005), https://doi.org/10.7551/mitpress/3206.001.0001.

57. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830.

58. F. T. Tracy, "Testing Line Search Techniques for Finite Element Discretizations for Unsaturated Flow," in *Computational Science-ICCS 2009*, ed. G. Allen, J. Nabrzyski, E. Seidel, G. D. van Albada, J. Dongarra, and P. M. A. Sloot (Springer Berlin Heidelberg, 2009), 473–480.

59. A. Malakizadi, K. Hosseinkhani, E. Mariano, E. Ng, A. Del Prete, and L. Nyborg, "Influence of Friction Models on Fe Simulation Results of Orthogonal Cutting Process," *International Journal of Advanced Manufacturing Technology* 88, no. 9–12 (2016): 3217–3232, https://doi.org/10.1007/s00170-016-9023-4.

60. R. Sevilla, S. Fernández-Méndez, and A. Huerta, "Nurbs-Enhanced Finite Element Method (Nefem): A Seamless Bridge Between Cad and Fem," *Archives of Computational Methods in Engineering* 18, no. 4 (2011): 441–484, https://doi.org/10.1007/s11831-011-9066-5.

2.4 Neural networks vs. splines: advances in numerical extruder design

Jaewook Lee, Sebastian Hube, and Stefanie Elgeti. "Neural networks vs. splines: advances in numerical extruder design". en. In: *Engineering with Computers* 40.2 (Apr. 2024), pp. 989–1004. ISSN: 0177-0667, 1435-5663. DOI: 10.1007/s00366-023-01839-2

Statement of Contribution: As the first author, I was the primary contributor to this research, which established a key methodology for this dissertation. My specific contributions were:

- conceptualizing the DNN-based geometry parameterization method, which involved generating the 3D shape training dataset and integrating the DeepSDF autodecoder architecture,
- performing all numerical experiments for the shape optimization runs and the subsequent latent space analysis,
- serving as the lead author of the manuscript, with co-authors contributing to the flow solver framework and supervision.

Engineering with Computers (2024) 40:989–1004
https://doi.org/10.1007/s00366-023-01839-2

ORIGINAL ARTICLE

Neural networks vs. splines: advances in numerical extruder design

Jaewook Lee[1] · **Sebastian Hube**[2] · **Stefanie Elgeti**[1,2]

Received: 26 January 2023 / Accepted: 23 April 2023 / Published online: 18 May 2023

Abstract
In this paper, we present a novel approach to geometry parameterization that we apply to the design of mixing elements for single-screw extruders. The approach uses neural networks of a specific architecture to automatically learn an appropriate parameterization. This stands in contrast to the so far common user-defined parameterizations. Geometry parameterization is crucial in enabling efficient shape optimization as it allows for optimizing complex shapes using only a few design variables. Recent approaches often utilize computer-aided design (CAD) data in conjunction with spline-based methods where the spline's control points serve as design variables. Consequently, these approaches rely on the design variables specified by the human designer. This approach results in a significant amount of manual tuning to define a suitable parameterization. In addition, despite this effort, many times the optimization space is often limited to shapes in close proximity to the initial shape. In particular, topological changes are usually not feasible. In this work, we propose a method that circumvents this dilemma by providing low-dimensional, yet flexible shape parametrization using a neural network, which is independent of any computational mesh or analysis methods. Using the neural network for the geometry parameterization extends state-of-the-art methods in that the resulting design space is not restricted to user-prescribed modifications of certain basis shapes. Instead, within the same optimization space, we can interpolate between and explore seemingly unrelated designs. To show the performance of this new approach, we integrate the developed shape parameterization into our numerical design framework for dynamic mixing elements in plastics' extrusion. Finally, we challenge the novel method in a competitive setting against current free-form deformation-based approaches and demonstrate the method's performance even at this early stage.

Keywords Shape optimization · Single-screw extruder · Neural networks · Mixing · Filter · Geometry parameterization

1 Introduction

Modern numerical design is boosted by high-performance computers and the advent of neural networks. While neural networks are well-established in fields such as image recognition, their power to further polymer processing is yet to be fully discovered. This work attempts to contribute towards this goal. We combine deep neural networks with established shape-optimization methods to enhance mixing in single-screw extruders via a novel numerical design.

In many polymer processing steps, screw-based machines play a crucial role. Screws are, e.g., used as plasticators to prepare polymer melts for injection molding or in extruders in profile extrusion. For simplicity, we will, in the remainder, summarize all such screw-based machines as *extruders*. Single-screw extruders (SSEs) are especially widespread among the many variants of extruders for their economic advantages and simple operation. Economics also drives current attempts to further increase the throughput. This increase is achieved using fast-rotating extruders. However, the current SSE's poor mixing ability has limited the advances and, therefore, improving the mixing ability is a topic of research [1–6].

Special focus is put on improved mixing elements that alleviate this limitation. Approaches to improve mixing

✉ Jaewook Lee
jaewook.lee@tuwien.ac.at

Sebastian Hube
hube@cats.rwth-aachen.de

Stefanie Elgeti
stefanie.elgeti@tuwien.ac.at

[1] Institute of Lightweight Design and Structural Biomechanics (E317), TU Wien, Gumpendorfer Str. 7, 1060 Vienna, Austria

[2] Chair for Computational Analysis of Technical Systems (CATS), RWTH Aachen University, Schinkelstr. 2, 52062 Aachen, Germany

elements have been proposed based on analytical derivations, experimental, and simulation-based works. In the following, we review recent developments in these three areas. Subsequently, we outline relevant developments in the field of neural networks and, finally, motivate the use of neural nets in the numerical design of mixing elements.

Due to the high pressures and temperatures, analyzing the flow inside extruders is a difficult task. Early studies thus focus on analytical models and geometrically simpler screw sections, e.g., the metering section [7]. Experiments complement these theoretical derivations and allow extending the analysis to more complex screw sections. As reported by Gale, typical configurations rely on photomicrographs of the solidified melt [2] that allow either investigating cross sections of the flow channel or the extrudate. One example of such flow channel photomicrographs is Kim and Kwon's pioneering work on barrier screws via cold-screw extrusion [8]. Apart from investigating solidified melt streams, attempts to analyze the melt flow during the actual operation of extruders are occasionally reported, e.g., by Wong et al. [9]. Despite the great success of such experiments, a standard limitation is their focus on a single operating condition. In contrast, numerical analysis allows studying different designs and operating points at significantly reduced costs and, therefore, proliferates. In the following, we give an overview of such numerical analyses.

One early example is Kim and Kwon's quasi-three-dimensional finite-element (FE) simulation of the striation formation, studying the influence of the barrier flight [10]. Another example is the work by Domingues et al., who obtain global mixing indices for dispersive and distributive mixing in both liquid–liquid and solid–liquid systems [11]. Utilizing a two-dimensional simplification, their simulation domain extends from the hopper to the metering section, and their framework even allows for design optimization.

While these early works typically neglect mixing sections, studying the influence of mixers has recently become a vital research topic. Celik et al. use three-dimensional flow simulation coupled with a particle-tracking approach to determine the degree of mixing based on a deformation-based index [1]. Another example is Marschik et al.'s study comparing different Block-Head mixing screws in distributive and dispersive mixing [6]. A comparable study—focused on the mixing capabilities of different pineapple mixers—is reported by Roland et al. [3]. Both works rely on three-dimensional non-Newtonian flow simulations. Besides such works towards the numerical assessment of *given* screw designs, numerical *design* is also reported, however, partially in other fields of polymer processing. For example, Elgeti et al. aim for balanced dies and reduced die swell by applying shape optimization [12, 13]. Design by optimization is also reported by Gaspar-Cunha and Covas, who alter the length of the feed and compression zones, the internal screw diameters of the feed and metering zone, the screw pitch, and the flight clearance [14]. Potente and Többen report another recent study devoted to mixing elements that develops empirical models for shearing sections' pressure-throughput and power consumption for numerical design [15]. Finally, a first approach combining the shape-optimization methods inspired by [12] with a mixing-quantifying objective function to design mixing sections is reported in [16].

However, the shape optimizations above share one commonality: they essentially only modify predefined geometry features. This is accepted in many cases like die or mold design, where the final product's shape is close to the initial one (i.e., the shape variation is small). However, topologically flexible shape parameterizations offer far greater optimization gains for mixing element design, because the optimal geometry might differ significantly from the initial shape. The achievable improvements motivate research on geometry parametrization.

Established shape-parameterization approaches include radial basis functions (RBF) [17], surface parameterizations using Bezier surfaces [18], and surface splines [19]. All these methods may be understood as *filters* that parameterize a geometry by a few variables at the price of a lack of local control. The use of surface splines in shape optimizations can also be found in [12, 13]. A similar concept to surface splines is free-form deformation (FFD) [20] that encapsulates the body-to-deform in a volumetric spline, which allows tailoring the spline further towards an efficient optimization. An alternative approach that does, however, not parameterize the geometry as a filter is given using the computational grid's mesh nodes as shape parameters [21]. Fortunately, with the advent of neural networks, novel means of shape parameterizations offering outstanding flexibility emerged. Finalizing the introduction, we will summarize the most relevant works in this field.

Many neural networks are essentially classifiers. These neural networks are non-linear algorithms that are optimized, (i.e., trained), to determine—possibly counterintuitive—similarities and dissimilarities to discriminate between objects. One typical use case is image recognition using red–green–blue (RGB) pixel data. Neural networks can, however, be trained to classify features far beyond RGB-pixel values. One example is style transfer or texture synthesis [22]: instead of aiming at reproducing *pixel* data, output images are generated in combination with *perceptual* data. This allows image transformations, where one image's style is transferred to the motive of another. An extension of these ideas to three-dimensional shapes is first reported by Friedrich et al. [23]. Comparing different shape representations, the authors find that style transfer is applicable to shapes as well.

Our work is especially inspired by Liu et al. [24], who utilize a so-called *Variational Shape Learner*, that learns a voxel representation of three-dimensional shapes. *Learning* here refers to creating a so-called *latent space*, a low-dimensional, feature-rich embedding space to represent and morph between various shapes. Even beyond simple shape interpolation, it is shown that—using the latent representation—geometry features can be transferred from one to another shape. Successful learning of voxel-based shapes can also be found in [25, 26]. In terms of shape representations, pointcloud-based approaches [27–29], which utilize coordinates of three-dimensional point sets, as well as polygonal mesh-based approaches with either template meshes [30, 31] or multiple mesh planes [32] are widely adopted.

While previously mentioned representations show that learning an embedding space of three-dimensional shapes is possible, each work lacks at least one of the following properties: water-tight surfaces, flexible output resolution, and smooth and continuous surface details. Recent works satisfy the aforementioned properties by learning shapes represented by continuous implicit functions, such as signed-distance functions (SDFs) [33] and binary occupancies [34, 35], from which the shapes are extracted as isosurfaces. This work investigates the shape-parameterization capabilities of the DeepSDF auto-decoder [33].

We exploit the feature richness of this latent space as an aid to reduce the optimization space's dimension for the given mixing-element shape optimization. The important novelty compared to recent spline-based filters is that the neural network finds—possibly counterintuitive—ways to commonly parameterize a set of significantly different shapes irrespective of user-defined design features. This abstraction from the human designer yields low-dimensional yet far more flexible shape parameterizations, which sets the motivation for the work presented here.

This paper is structured as follows: We start in Sect. 2 by summarizing numerical shape optimization and splines, which leads to the concept of geometric filters. Based on that, we explain in Sect. 3 how neural networks can be utilized to create suitable geometry parameterizations for shape optimization. In Sect. 4, we review the utilized software components, summarize the proposed framework's building blocks, and detail the specific differences to spline-based shape optimization setups. The results obtained from the new approach are presented in Sect. 5, including comparisons to current spline-based designs. Finally, we discuss the results and outline further developments in Sect. 6.

2 Geometric filters as a component of shape-optimization frameworks

The following section discusses shape parameterizations as one building block of numerical shape-optimization frameworks. Therefore, we first introduce the general shape optimization problem. After that, we recall spline-based shape parameterizations. Based on this general introduction of shape-optimization frameworks, we will continue by discussing the specific changes needed to adapt neural nets in Sect. 3.

2.1 Building blocks of numerical shape-optimization frameworks

The general optimization problem is formulated as the minimization of a cost function J that relates the design variables $\boldsymbol{\sigma}$ to some output—here, the degree of mixing ability obtained with a specific mixing element, (i.e., a particular design). In shape optimization, this minimization problem is typically solved subject to two sets of constraints: (1) inequality and equality conditions, as well as bound constraints on the design variables and (2) partial differential equations (PDEs) that need to be fulfilled by each design to qualify as a feasible solution. This results in the following formulation:

$$J : \mathbb{R}^{n_\sigma} \mapsto \mathbb{R}, \tag{1a}$$

$$\underset{\boldsymbol{\sigma}\in\Sigma\subset\mathbb{R}^n}{\arg\min} J(\boldsymbol{\sigma}), \tag{1b}$$

$$\text{s.t.} \mathbf{F}(\boldsymbol{\sigma}) = \mathbf{0} \qquad \text{in}\Omega(\boldsymbol{\sigma}), \tag{1c}$$

$$\sigma_i \geq \sigma_{\min,i}, \quad i = 1, \ldots n_\sigma, \tag{1d}$$

$$\sigma_i \leq \sigma_{\max,i}, \quad i = 1, \ldots n_\sigma. \tag{1e}$$

Here, (1d) and (1e) describe bound constraints on the optimization variables $\boldsymbol{\sigma}$, whereas (1c) denotes the set of governing PDEs. One approach to numerically solve such a *PDE-constraint* design problem is to alternately compute (1) shape updates and (2) the cost function value. For the studied use case of mixing-element design, this results in the computational steps depicted in Fig. 1.

First, we update the shape (i.e., the simulation domain covering the mixing element). We use this modified computational domain to compute the flow field from which we afterwards infer the objective (i.e., the cost function). The design loop is closed by feeding back the cost function value to the optimization algorithm that now computes an updated shape. This loop continues until any termination criterion,

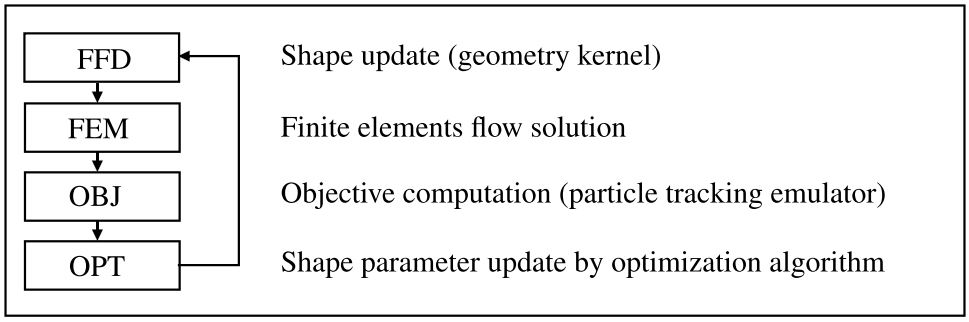


Fig. 1 Building blocks of a shape-optimization framework. The shape is updated by a geometry kernel, such as FFD. Subsequently, the flow field is computed using this updated shape and given as input to the objective calculator. Based on the current design variables and the design's objective value, the optimization algorithm computes optimized shape parameters and restarts the design loop until at least one termination criterion for the design loop is met

such as a minimal objective decrease, a maximum number of iterations, or another condition, is met.

2.2 Spline-based shape parameterizations

In classical shape-optimization frameworks, the actual shape parameterization, or geometry filtering, is often achieved using splines. The following paragraph, therefore, first provides a summary of splines illustrating how one achieves the filtering. For a detailed description of B-splines, we refer the reader to the book of Piegl and Tiller [19]. After that, we detail on *boundary splines* and FFD as two particular use cases of spline parameterizations.

Splines belong to the group of parametric shape representations. Therefore, each coordinate in the parametric space is connected to one point in physical space. This mapping is best understood using a simple *B-spline* surface that is written as

$$\mathbf{S}(\xi,\eta) = \sum_{j=1}^{m}\sum_{i=1}^{n} N_{i,r}N_{j,p}(\xi,\eta)\mathbf{B}_{i,j}, \tag{2}$$

where ξ and η denote the parametric coordinates (two for the surface), $N_{i,r}$ denote the interpolation or *basis functions* of order r in the first parametric direction, $N_{j,p}$ denote the basis functions of order p in the second parametric direction, and finally, $\mathbf{B}$ denotes the support or *control points*. Figure 2 illustrates the concept and visualizes how single control points affect the geometry.

The control grid (i.e., the polygon spanned by the control point) aligns with the ξ and η directions, and any parametric coordinate (within the spline's parametric bounds) maps to one point of the blue shape. Consequently, the spline mapping allows controlling an arbitrary number of parametric points by a constant, typically low, number of control points. Being able to control a high number of points with few control points will be the basic idea of filtering using splines.

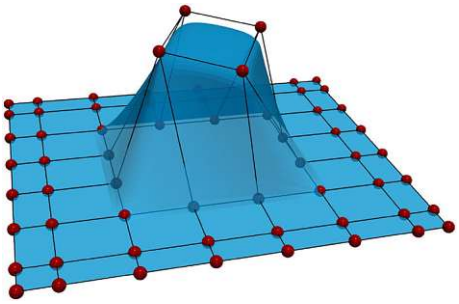


Fig. 2 B-spline representation (blue) obtained from control points (red) for a bi-quadratic B-spline. The upper four control points are rotated, illustrating a possible deformation (colour figure online)

One can obtain geometry parameterizations from splines in multiple ways. As shown in Fig. 2, one way uses the B-splines as a boundary representation. Such spline-based boundary representations are common in CAD. Using these CAD representations, their control points (i.e., the red points in Fig. 2) can be directly used as design variables in shape optimization. However, this use of the CAD's geometry parameterization limits the design process, because a given spline may not be able to represent shapes substantially different from the initial design. Consequently, if modifications of the spline's parameterization, such as inserting additional control point lines, are to be avoided, this limitation restricts the use of the CAD spline to use cases that deal with small shape updates such as *die* or *mold design* [12].

An alternative to using boundary B-splines is FFD [20]. In FFD, first, an—often volumetric—spline is constructed around the body to be deformed. Second, this volumetric spline is deformed, and finally, the resulting deformation field is imposed on the enclosed body. Figure 3 visualizes this process.

The advantage of FFD is that the spline is constructed irrespective of the enclosed shape, which gives complete

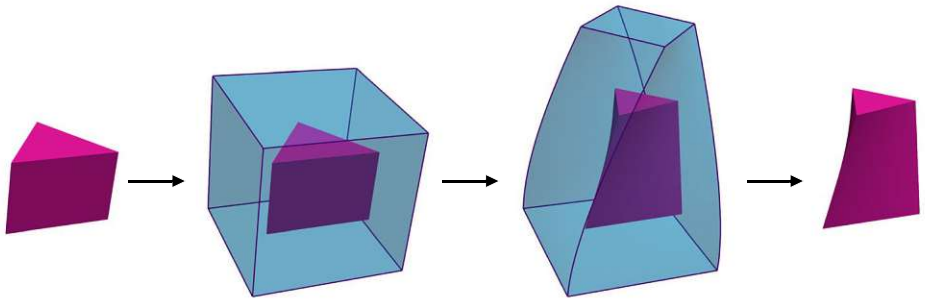


Fig. 3 Free-form deformation using a volumetric spline (light blue) applied to a mixing element (pink). The control points are omitted in this figure. The embedded shape deforms correspondingly to the embedding, simple, volumetric spline (colour figure online)

freedom in choosing degree and resolution. This freedom allows tailoring the spline to the designer's needs (rather than using a given parameterization optimized for CAD usage). Therefore, FFD is widely applied, with just one example being the recent works by Lassila and Rozza combining FFD and reduced order modeling [36]. A combination of both methods, boundary B-splines and FFD, will be compared against the novel shape parameterization based on neural networks that use FFD as a generic interface to modify any given CAD spline, which in turn is used to update the boundary of the simulation domain [16].

3 Shape parametrization using neural networks

As explained in Sect. 2, the prime objective of this work is to investigate how neural networks can be used to encode different shapes in a single set of a few continuous variables. To train the network, thereby determining such a condensed representation, it has to be provided with suitable data. *Suitable* here means that the input data (i.e., shapes) are provided in such a way that the network can learn from these data. In addition—using the same data format—we need to be able to produce high-quality computational meshes from the neural network's output.

In the following, we first introduce deep generative models and then describe a shape representation meeting these two requirements. Finally, we discuss the training data generation and utilization of neural networks as shape generators.

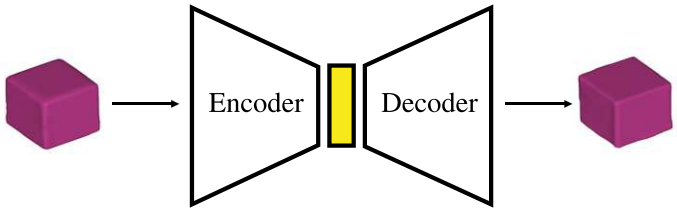
3.1 Deep generative models

With the advent of *generative models*, an alternative approach to shape parameterization emerged. In this subsection, we review two of the most common approaches of generative models, explain their basic concepts and use, and detail how they can be employed for geometric filtering.

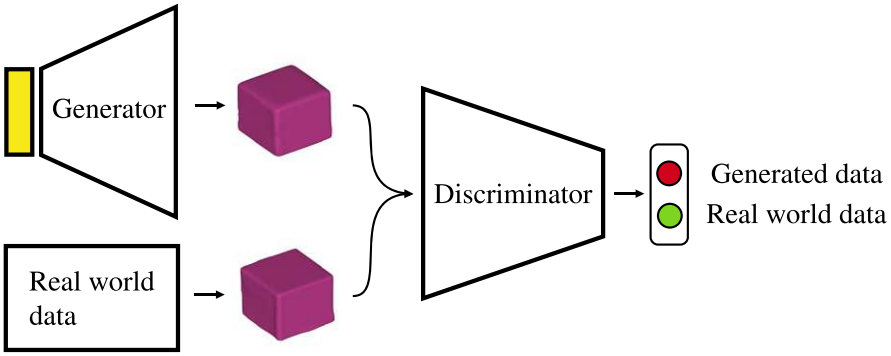
Generative models are an application of neural networks and, thus, in essence, classification algorithms. *Classification* here means the ability to determine whether a certain object is in some measure *close* to a specified input. Conversely to just classifying input, such models can also be used to generate an output that resembles an input. *Resemble*, however, needs to be explained. In most applications, the user is not interested in reproducing a given input exactly. Instead, the output should only be *like* the input (i.e., the output should feature a slight variation). Generative models attempt to achieve this goal via statistical modeling. An excellent guide to generative models is found in [37], with special focus on the *Variational Autoencoder* (VAE).

The VAE, like the traditional autoencoder, consists of an *encoder* and a *decoder* and aims to reproduce any given data while passing the input through a bottleneck. However, its probabilistic formulation using the so-called "reparametrization trick" provides an exceptional advantage over the traditional autoencoder in practice [38]. The roles of the encoder and the decoder can be interpreted as two separate processes. The encoder learns relations in the given data and encodes them in the so-called *latent variables*, $\mathbf{z}$. Given these latent variables, the decoder, in turn, learns to produce data that are *likely* to match the input. Once trained, the user can omit the encoder and directly generate new data from sampling the latent space. For details, we refer to [37, 38], and for applications, we refer to [24] and [39].

The difference between the spline-based approach and generative models is the choice of latent variables. When the human designer creates a spline parameterization that allows modifying geometry in the desired way, the optimization variables are the control points, which are *intuitively* placed in $\mathbb{R}^3$ by the designer. Generative models, in contrast, *learn* a latent space and explicitly assume that the single latent variables do not have an intuitive interpretation. As a result, data are compressed from a high-dimensional intuitive design space, in our case, $\chi \subset \mathbb{R}^{3\times n}$, onto a hardly interpretable, feature-dense, low-dimensional latent space Z. In short, generative models use the computational power of neural networks to find a dense classification space that one can sample to produce new data. For the VAE, this process is depicted in Fig. 4a.



(a) Autoencoder providing an input-to-output mapping while passing data through a bottleneck, i.e., the low-dimensional latent representation.



(b) Generative adversarial model learning latent space by inferring representations that enable generating output indistinguishable from the input.

Fig. 4 **a** Autoencoder providing an input-to-output mapping while passing data through a bottleneck, i.e., the low-dimensional latent representation. **b** Generative adversarial model learning latent space by inferring representations that enable generating output indistinguishable from the input. Two main concepts of deep generative networks: variational autoencoders and generative adversarial networks

A competing concept to VAEs are *Generative Adversarial Networks (GANs)*. Their basic structure is shown in Fig. 4b. GANs, first introduced by Goodfellow et al. [40], follow a different concept and train two adversarial nets, the *generator* and the *discriminator*. In GANs, the generator is trained to create data that mimics real-world data, while the discriminator tries to determine whether or not a dataset was artificially created. In a minimax fashion, the generator's learning goal is to maximize the probability of the discriminator making a wrong decision.

GANs have proven to be an excellent tool for shape modeling. Wu et al., for example, apply a GAN for 3D shape generation and demonstrate their superior performance compared to three-dimensional VAEs. They even use a GAN to reconstruct three-dimensional models from two-dimensional images based on the a VAE output that is used to infer a latent representation for these images [41]. As in [24], Wu et al. also demonstrate the ability to apply shape interpolation and shape arithmetic to the learned latent representation. More recently, Ramasinghe et al. [28] utilize a GAN to model high-resolution three-dimensional shapes using point clouds.

3.2 Implicit shape representation

The neural network learns a mapping between the low-dimensional latent space and a three-dimensional body. To construct such a mapping, we first need to define how to represent our shapes (i.e., define what data the neural network actually has to learn). Before presenting the approach chosen in this work, we review standard methods and their limitation.

Three ways of shape representation are common in machine learning: (1) voxels, (2) point clouds, and (3) meshes [33]. The problem with meshes is that the mesh topology also prescribes the possible shape topologies. Point clouds, in contrast, can represent arbitrary topologies, but prescribe a given resolution. Finally, voxels can represent arbitrary topologies and vary in resolution, but, unfortunately, the memory consumption scales cubically with the resolution. Because of these drawbacks, the network utilized in this work learns SDFs following a network configuration originally proposed by Park et al. [33].

SDFs provide the distance to the closest point on the to-be-encoded surface for every point in space. Furthermore, encoded in the sign, information on whether the point lies inside or outside the surface is available. Using such continuous SDF data, a shape is then extracted—at an arbitrary resolution suitable for meshing—as its zero-valued isosurface.

3.3 Training set generation

As mentioned in Sect. 3.1, training a neural network requires a set of source shapes. However, to the authors' knowledge, no shape library exists for mixing elements in single screw extruders. Thus, we explain an approach to building custom training sets.

To generate a suitable training set, we first select categories of basis shapes that should be considered—pin and pineapple mixers in our case. From this choice, we arbitrarily infer a total of four basis shapes (i.e., triangle, square, hexagon, and cylinder—cf. Figure 5). At the same time, we define a set of deformations, which should be considered within the design space. Examples of applied deformations are given in Fig. 5.

We start by creating basis shapes represented as triangular meshes. For each basis shape, we apply the aforementioned deformations and their combinations in varying magnitudes using FFD to gather a rich set of shapes. To obtain SDF-training data from these shapes, we follow the approach by Park et al. [33]: first normalize each shape to fit into a unit sphere, and then sample 500,000 pairs of spatial coordinates and their corresponding SDF values using the trimesh library [42] from each shape. In total, 2659 training shapes are generated, which constitute the accessible deformations within the design space.

3.4 Shape generator

As explained, the shape generator's task is to provide a mixing element given a set of optimization variables. The shape generator—in this work—is thus built around the neural network, which is presented in the following.

The utilized neural network is based on DeepSDF autodecoder [33]: a feed-forward network with ten fully connected layers, with each of the eight hidden (i.e., internal) layers having 256 neurons and ReLU activation functions. In contrast to autoencoders, the autodecoder only trains the decoder using a simultaneous optimization of the network parameters and the latent code during training. We investigate 4, 8, and 16 as latent dimensions, l. The input layer consists of these l neurons concatenated with a three-dimensional query location. The output layer has only one neuron with a tanh activation function. For details on the chosen SDF network, we, again, refer to [33]. To train the network, we use the ADAM optimization algorithm [43]. To utilize improved learning rates, we follow a progressive approach with the initial rates $\varepsilon_0 = 5e-4$ for $\boldsymbol{\theta}$, and $\varepsilon_0 = 1e-3$ for z, and a decay as

$$\varepsilon = \varepsilon_0 \cdot \left(0.5^{e\%500}\right), \tag{3}$$

where e denotes the current training iteration (i.e., *epochs*)—and % denotes integer division. The network's training can be seen as the parametrization of the shapes.

To extract isosurfaces (i.e., to generate new mixing elements) from the trained network's SDF output, we sample a discrete SDF field and apply a marching cube algorithm [44] in the implementation of [45]. Finally, we apply automated meshing using TetGen [46] to obtain a simulation domain as depicted in Fig. 6, including the new mixing element.

4 The developed shape-optimization framework

In general, our framework consists of three building blocks: (1) *shape generator*, (2) *flow solver*, and (3) *optimizer*, which will be described in the following.

Starting with an initial set of optimization variables, $\boldsymbol{\sigma}_0$, the shape generator creates a new mixing element $\Omega(\boldsymbol{\sigma}_0)$. The flow solver then computes the flow field around this

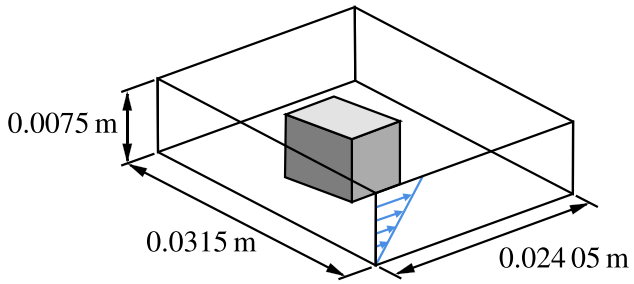


Fig. 6 Simulation domain with single mixing element resembling the flow around a single mixing element in the unwound screw channel. Flow conditions are shown in blue using a barrel rotation setup. For a detailed description of the objective function and governing equations, we refer the reader to [16]

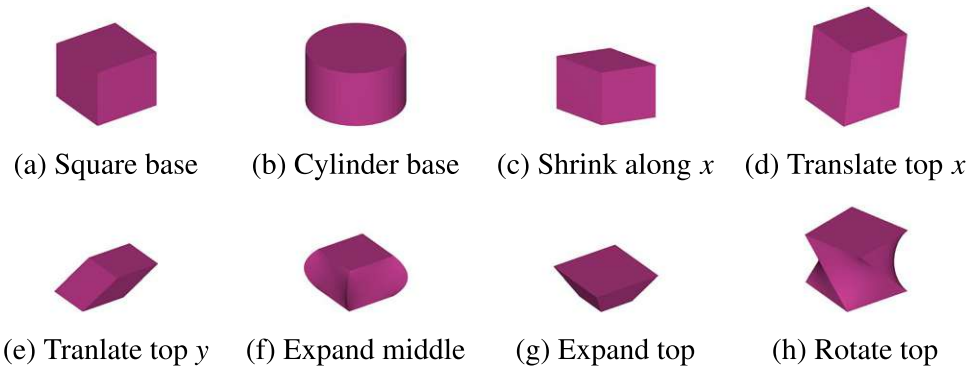


Fig. 5 **a** Square base, **b** cylinder base, **c** shrink along x, **d** translate top x, **e** translate top y, **f** expand middle, **g** expand top, and **h** rotate top. Examples of basic shapes and applied deformations. In total, a triangle, a square, a cylinder, and a hexagon are used as basis shapes

mixing element, which the optimizer evaluates to determine the flow's degree of mixing. Based on the obtained mixing value and by comparison to previous iterations, an optimization algorithm determines a new set of optimization variables. This sequence is iteratively re-run until either a maximum number of iterations is reached or any other termination criterion—typically a good objective value or insignificant objective decrease—is met.

4.1 Flow solver and simulation model

The flow solver and simulation model is identical to the one introduced in [16] and therefore only summarized in the following. The flow field induced by the various mixing elements is obtained from solving the steady, incompressible non-isothermal Navier–Stokes equations using a Carreau model and WLF temperature correction. The governing equations are discretized with linear stabilized finite elements and solved using a Newton linearization and a GMRES iterative solver. Subsequently, we solve a set of advection equations using the identical configuration to mimic particle tracking, which we use as an input to our objective function. All methods are implemented in an in-house flow solver.

We make two simplifications to our simulation model (i.e., the single-screw-extruder flow channel): first, we simulate the flow around only a single mixing element instead of simulating the entire mixing section. Second, we assume barrel rotation in an unwound flow channel section. Both assumptions yield significantly reduced computational costs while allowing a qualitative mixing improvement. To assess mixing, we mimic particle tracking by solving a series of advection equations yielding an inflow–outflow mapping for particles advected by the melt flow. We process this advection information by subdividing a portion of the inflow domain into smaller rectangular subdomains. In each of these rectangles, we select a set of particles, such that the particle set's bounding box coincides with the rectangular subdomain. Then, we follow each particle as they are conveyed through the domain, store each particle's position at the outflow domain, and finally construct a convex hull at the outflow around the same sets of points. Averaging the convex hull's length increments between inflow and outflow yields a simple yet robust objective function inspired by interfacial area measurements. Using this objective function, we found that such a simulation model provides a good balance between accuracy and computational efficiency [16]. Figure 6 depicts the chosen simulation domain.

4.2 Optimizer

We utilize the open-source optimization library Dakota [47] to drive the design process. Two different algorithms are selected and described in the following. The first algorithm is the Dividing RECTangle (DIRECT) algorithm, first introduced in [48]. DIRECT belongs to the category of *branch-and-bound* methods and uses n-dimensional trisection to iteratively partition the design space. To find minima, it follows the approach of Lipschitzian optimization, which identifies the design space partition that should be further sampled by evaluating a lower bound to the objective value in each partition. The partition with the lowest lower bound is chosen and further sampled. DIRECT modifies that concept and computes multiple lower bounds that weight the current sampling value (i.e., the objective value in the partition center). This promotes to further sample partitions with good objective values against the partition size, which permits to effectively sample large areas of unexplored design space. Thereby, DIRECT identifies *multiple* partitions that are *possibly optimal* and allows for global convergence.

The second algorithm utilized in this work is the single-objective genetic algorithm (SOGA) introduced (as its multi-objective variant) in the JEGA package [49]. As it belongs to the class of *genetic* algorithms, it solves optimization problems by recreating biological evolution. Therefore, each optimization run consists of numerous samples referred to as the *population*. Members of the population are paired and recombined in such ways that the *fitness* (i.e., the objective value) is successively improved. Regarding its application in this work, it is especially noteworthy that the recreation of evolution includes a *mutation* step, which modifies or re-initializes design variables randomly. The added randomness allows the algorithm to escape locally convex regions of the design space. Such evolutionary optimization approaches generally converge slower yielding higher computational costs. However, they are often able to find better results than non-evolutionary algorithms. For both DIRECT and SOGA, we rely on the default convergence criterion and a maximum of 1000 iterations as a termination criterion. Additionally, for SOGA, we choose a population size of ten times the number of design dimensions. The complete computational framework is depicted in Fig. 7.

5 Numerical results

This section presents the results obtained using shape parameterizations from neural networks.

Thereby, Sect. 5.1 focuses on the results of the offline phase, i.e., the training of the shape-representing neural network. In particular, we will discuss the differences in the constructed latent space based on its dimension using the widely used data reduction technique *t-Distributed Stochastic Neighbor Embedding (t-SNE)* to visualize the learned, n-dimensional shape parameterization. In Sect. 5.2, we then

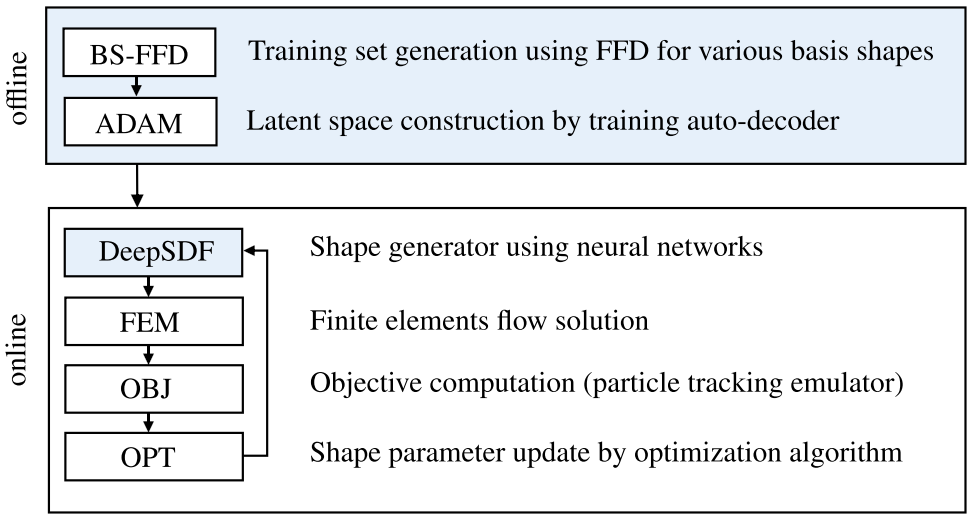


Fig. 7 Pipeline with building blocks of the proposed computational framework. The process is split into two parts: a one-time computationally intensive training part and the actual optimization, including the quick filter evaluation. To create a training set, FFD is applied to a set of basis shapes. Subsequently, we train the network using the ADAM optimizer, which concludes the *offline* phase. During optimization (i.e., the *online* phase), first, a new shape is created from the neural net. Then, a new computational mesh is created around this shape, and based on FEM simulations, the new design's mixing is assessed. Depending on the objective value, the optimization loop is re-initiated using altered latent variables. Building blocks that are modified compared to the general, geometry-kernel-based approach (cf. Figure 1) are highlighted in blue

present the mixing shapes that could be obtained using our shape-optimization approach.

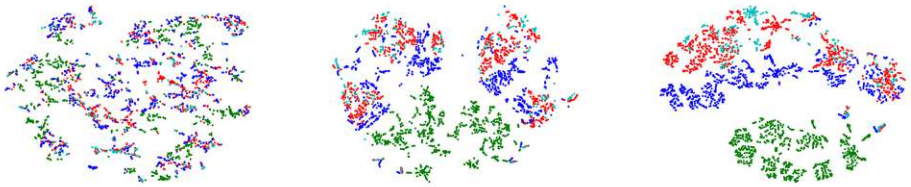
5.1 Latent space dimension

One of the most important choices is the target dimension of the embedding space l. In all established filtering mechanisms like radial basis functions, free-form deformation, CAD-based approaches, and even mesh-based methods, the practitioner has to balance improved flexibility against the computational demand. Despite a potentially more compact and dense embedding with neural networks, this is still of relevance and manifests itself in the dimension of the chosen latent space. Previous works utilized only a very small number of optimization variables. Elgeti et al. vary between only one and two parameters [12]. Other works by the authors, however, showed that also for six design variables good results are obtained [16]. To obtain a competitively small number of optimization variables, we investigate embedding spaces of dimension 4, 8, and 16, respectively, and compare against a free-form-deformation approach using nine variables.

Even though the latent space, as discussed in Sect. 3.1, in general, obtained lacks an intuitive interpretation, we are still interested in evaluating the quality of the learned embedding space. We do so in three different ways which we present in the following: (1) we show a data reduction technique that allows us to visually investigate the latent space; (2) we apply an interpolation between the latent representation of two training shapes and compare with the expected result; (3) we apply shape arithmetics, i.e., we isolate a specific modification of a basis shape and impose it onto another basis shape to inspect whether or not features are also recognized by the latent space.

(1) For the visualization of the high-dimensional latent space, a dimension reduction technique is required. An intuitive choice might be principal component analysis (PCA), but PCA tries to primarily preserve global structures and thus data points which are far apart in the high-dimensional data will also be drawn far apart in the 2D plot. Conversely, the correlation between similar points is often lost. This loss of correlation in similar data is problematic, since we aim to investigate whether—from a human's perspective—similar shapes are represented by similar latent code. The problem of loss in local correlation is, however, alleviated by t-SNE [50]. Using t-SNE, we plot each training shape's obtained latent code and—due to the preservation of local similarities—similar latent code will form clusters in the scatter plot. These clusters can then be sampled to verify that the latent code clusters resemble similar shapes. t-SNE plots for all three latent dimensions—4, 8, and 16—are shown in Fig. 8.

Figure 8 shows how an increased latent dimension leads to increased classification performance of the neural net. Specifically, the four chosen basis shapes are clustered with their respective modifications more and more densely as the latent dimension increases. This improved classification performance indicates that the neural net was able to learn the



(a) Training set encoded in 4D latent code (b) Training set encoded in 8D latent code (c) Training set encoded in 16D latent code

Fig. 8 **a** Training set encoded in 4D latent code. **b** Training set encoded in 8D latent code. **c** Training set encoded in 16D latent code. t-SNE plots obtained using different latent space dimensions. Increased latent dimension resembles in increased classification performance of the neural net. Each color corresponds to one base training shape: green corresponds a triangular base, dark blue is the cube, red is the hexahedron, and light blue is a tessellated version of the cylinder (colour figure online)

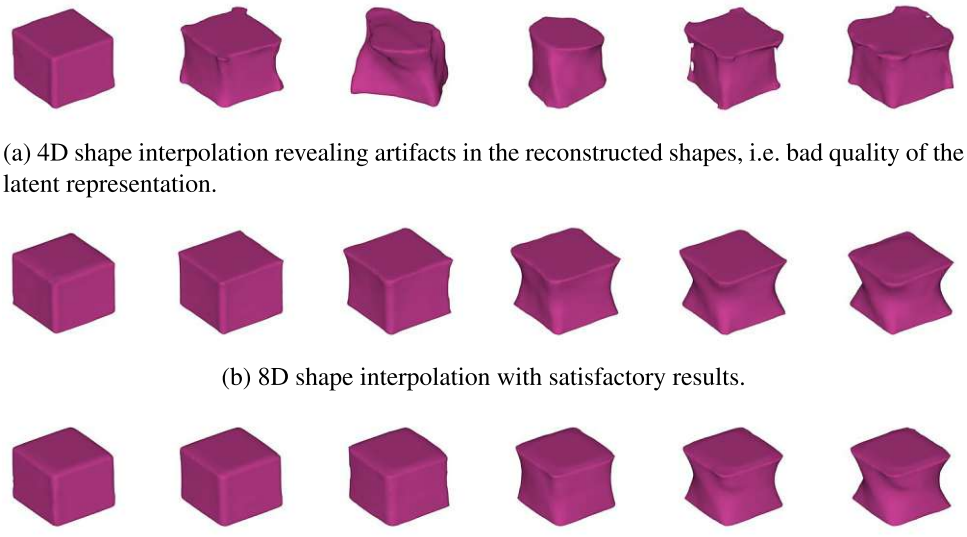
similarities between similar shapes properly for the case of 8 and 16 dimensions.

(2) In addition to comparing clusters of similar shapes in physical and latent space, we also investigate how well the latent space is suited to represent shapes that have not been included in the training set. We do so by *interpolation* between two shapes. Figure 9 shows the obtained results for all three latent spaces.

Consistent to the observed lack in classification ability of the four-dimensional latent space, Fig. 9a shows that interpolation between shapes yields unsatisfactory results. In particular, shape defects are observed. This might be a result of the fact that the twisted cube is not at all well represented in the latent space as seen in the rightmost figure. However, both the 8- and the 16-dimensional latent space show a visually smooth transition between the regular and the twisted cube shape.

(3) The above two analyses investigated the overall classification ability of the neural net and the suitability to represent intermediate shapes. A final test is given by applying *shape arithmetic*. Using arithmetic operations applied to the latent code, we extract an exemplary feature—here a stretching along the center plane—by taking the component-wise difference of a stretched and a regular cube. This difference represents center-plane expansion and can then be applied to any other basis shape—here the undeformed hexahedron. Figure 10 shows the resulting shapes. Again, the four-dimensional latent space performs significantly worse, since the basis shapes are not represented in detail. Contrary to the interpolation case, the 16-dimensional latent space now shows better results than the 8-dimensional case.

All three investigations, t-SNE plots, interpolation, and arithmetic, indicate that the four-dimensional latent space fails in producing a suitable latent representation. It should



(a) 4D shape interpolation revealing artifacts in the reconstructed shapes, i.e. bad quality of the latent representation.

(b) 8D shape interpolation with satisfactory results.

(c) 16D shape interpolation, which brings only slight improvement in shape representation compared to eight-dimensional latent representation.

Fig. 9 **a** 4D shape interpolation revealing artifacts in the reconstructed shapes, i.e., bad quality of the latent representation. **b** 8D shape interpolation with satisfactory results. **c** 16D shape interpolation, which brings only slight improvement in shape representation compared to eight-dimensional latent representation. Shape interpolation using different latent dimensions. An interpolated shape is obtained using $z_{\text{interp}} = z_a + \frac{z_b - z_a}{N+1} n$ with z_a and z_b denoting the latent code between shapes a—here the undeformed cube—and b—here the twisted cube. With $N = 20$, the shown examples represent $n \in [1, 3, 7, 13, 17, 20]$

(a) 4D shape arithmetics with significant representation errors, especially for the hexahedron and the final shape.

(b) 8D shape arithmetic with improved representation compared to 4D latent code but still yielding slightly imprecise results.

(c) 16D shape arithmetic showing perfect resemblance of all training shape and also a clean resulting shape.

Fig. 10 **a** 4D shape arithmetics with significant representation errors, especially for the hexahedron and the final shape. **b** 8D shape arithmetic with improved representation compared to 4D latent code but still yielding slightly imprecise results. **c** 16D shape arithmetic showing perfect resemblance of all training shape and also a clean resulting shape. Shape arithmetics for different latent dimensions. A linear thickening in the center plane is imposed on a hexagonal base body by evaluation of the latent code as $z_{E4_{\text{thick}}} - z_{E4} + z_{E6}$, where $z_{E4_{\text{thick}}}$, z_{E4}, and z_{E6} denote the latent codes of the thickened cube, the regular cube, and the regular hexahedron, respectively

Table 1 Different optimization algorithms and latent space dimensions compared by best objective value and contrasted to a nine-dimensional FFD 1b

	4	8	16	(FFD)
SOGA	−0.0726	−0.0710	−0.0750	–
DIRECT	−0.0645	−0.0738	−0.0769	−0.0422

Smaller values correspond to better results using the aforementioned objective formulation

be noted though that in view of the doubled number of optimization variables, the attainable gains in using 16 latent variables compared to 8 appear unattractively small.

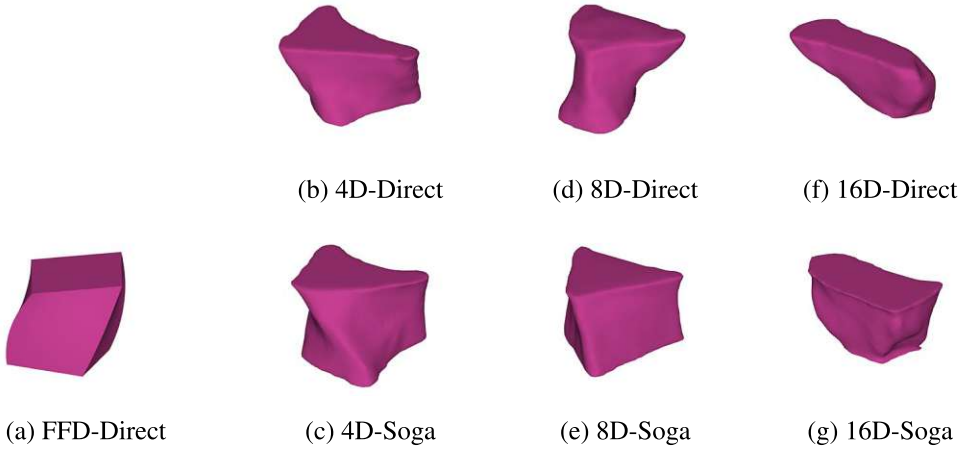
5.2 Optimization results

To study the effects of the novel shape-parameterization technique, we compare configurations that vary in latent space dimensions and optimization algorithms, as shown in Table 1. l furthermore, we require all generated shapes to have the exact same volume as the undeformed rhombic mixing element utilized in the spline-based optimization (cf. Sect. 4.1). We choose such scaling to avoid convergence towards merely enlarged shapes that yield good objective values but do not deliver helpful insights. Table 1 lists the obtained results, and Table 2 gives insights into the corresponding computational effort.

The obtained best shapes are shown in Fig. 11.

Table 2 Different optimization algorithms and latent space dimensions compared by the final iteration count, the obtained objective value, and the number of total iterations; contrasted to a nine-dimensional (FFD)

	4		8		16		(FFD)	
# Iteration(s)	Optimal	Total	Optimal	Total	Optimal	Total	Optimal	Total
SOGA	768	1000	752	1000	534	1000	–	–
DIRECT	96	113	129	143	138	149	16	67



(b) 4D-Direct (d) 8D-Direct (f) 16D-Direct

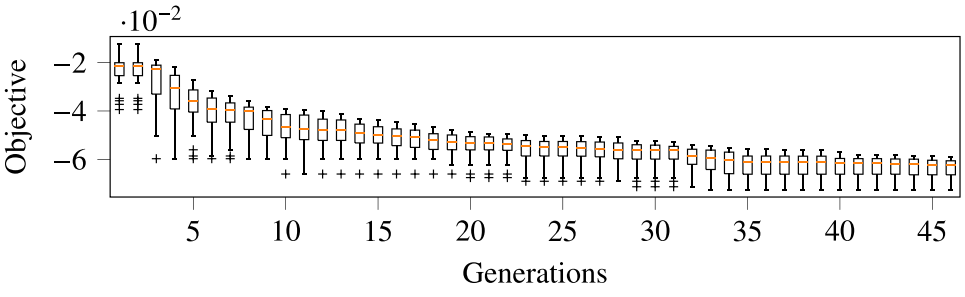
(a) FFD-Direct (c) 4D-Soga (e) 8D-Soga (g) 16D-Soga

Fig. 11 **a** FFD-Direct, **b** 4D-Direct, **c** 4D-Soga, **d** 8D-Direct, **e** 8D-Soga, **f** 16D-Direct, and **g** 16D-Soga. Optimization results obtained for all different latent codes and optimization algorithms compared to an existing (FFD)-based shape optimization

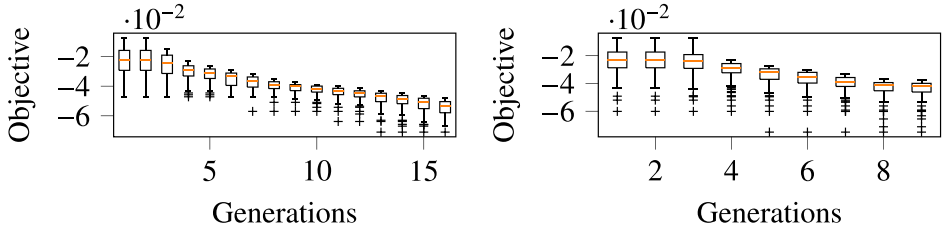
Comparing the optimized geometries shows interesting results from a plastics processing point of view. On the one hand, the triangular shape and a mixing element that widens towards the top appear advantageous. One should note, however, that these deformations do not correspond to a general optimum for plastics engineering but are merely the best possible deformations within the range permitted by the training set. Choosing an even more diverse training set is expected to yield even further improved shapes.

More relevant for this study (with a focus on neural nets as shape parameterizations) is the comparison of convergence, the achieved mixing, and the difference and similarities in the results. Table 1 shows that for the chosen shape optimization problem, the DIRECT algorithm has no disadvantages compared to SOGA and converges reliably. Simultaneously, the shape parameterization's dimensionality appears to influence the optimization because the four- and eight-dimensional neural networks lead to optimized triangles. In contrast, the 16-dimensional case renders the top-expanded quadrilateral optimal. Common to all results is a skewed and slightly twisted geometry.

The objective value per generation of the SOGA algorithm is shown in Fig. 12. For all dimensions, each generation tends to evolve towards better objective values. We investigate if improvements in objective values can be related to certain trends in terms of geometric features. Within the latent space, features are expressed as multi-dimensional vectors, as opposed to single-dimensional values; thus, latent code arithmetics [24, 41], for example, are often used to explore learned features (cf. Figure 10). Furthermore, the



(a) 4D latent code with 40 populations per generation



(b) 8D latent code with 80 populations per generation (c) 16D latent code with 160 populations per generation

Fig. 12 **a** 4D latent code with 40 populations per generation. **b** 8D latent code with 80 populations per generation. **c** 16D latent code with 160 populations per generation. Comparison of objective value per population among different latent dimensions from SOGA algorithm

latent code alone does not provide information about the associated geometric features. To determine the corresponding features, we begin by searching for the nearest neighbor in the training set based on Euclidean distance. From the nearest neighbor, we can then quantify applied features, i.e., deformations. The mean values of the approximated features are shown in Fig. 13. It shows that diverse features, even with less significance, were visited at the beginning of the iterations. Across all generations, relatively high magnitudes of "rotate", "rotate top", and "shrink along y", as well as a low magnitude of "expand top", consistently appear in all the results The four-dimensional latent space also includes consistent application of "shrink top". Interestingly, all of the consistently applied features are reflected in Fig. 11.

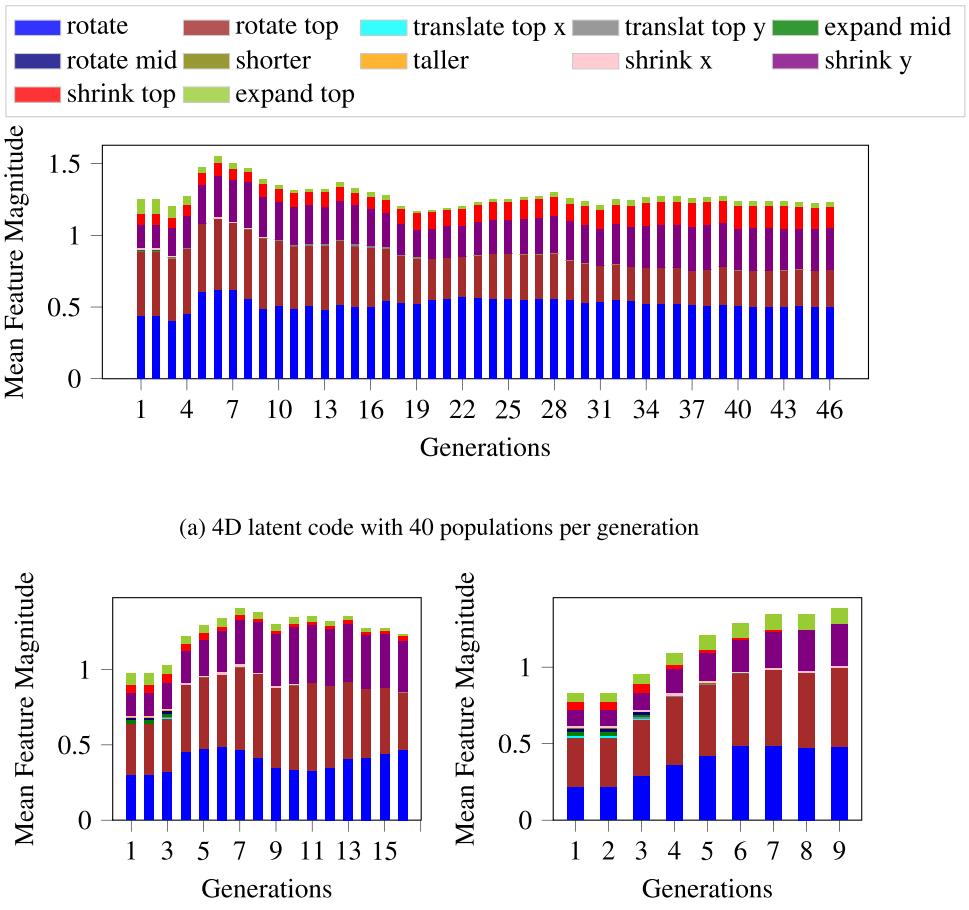
The behavior of objective values for DIRECT is not separately investigated, as the objective values and geometric features can highly fluctuate during partitioned sampling of the latent space. Consequently, DIRECT's iteration history does not yield significant insight regarding the deformation trend.

A noticeable difference between the spline-based and neural-net-based shape optimization is that the neural-net-based shape parameterization encodes several shapes, of which multiple may mix the melt equally well. Because of this, from the practitioner's point of view, it does make sense to not only look at the best result but rather compare numerous equally optimal designs and derive design rules from that comparison. Figure 14 shows such a comparison and reveals one advantage of evolutionary algorithms.

While the DIRECT algorithm converges locally and, therefore, the ten best designs are geometrically similar, the generative nature of SOGA allows the practitioner to identify possibly equally well-working designs (cf. Figures 14f and g) amongst which the most economical option may be chosen. Such a choice allows one to account for further restrictions regarding screw cleaning, manufacturability, and others.

6 Discussion and outlook

In this work, we studied the applicability of generative models as shape parameterizations. We chose numerical shape optimization of dynamic mixing elements as a use



(a) 4D latent code with 40 populations per generation

(b) 8D latent code with 80 populations per generation

(c) 16D latent code with 160 populations per generation

Fig. 13 **a** 4D latent code with 40 populations per generation. **b** 8D latent code with 80 populations per generation. **c** 16D latent code with 160 populations per generation. Comparison of mean value of approximated applied deformation per population among different latent dimensions from SOGA algorithm

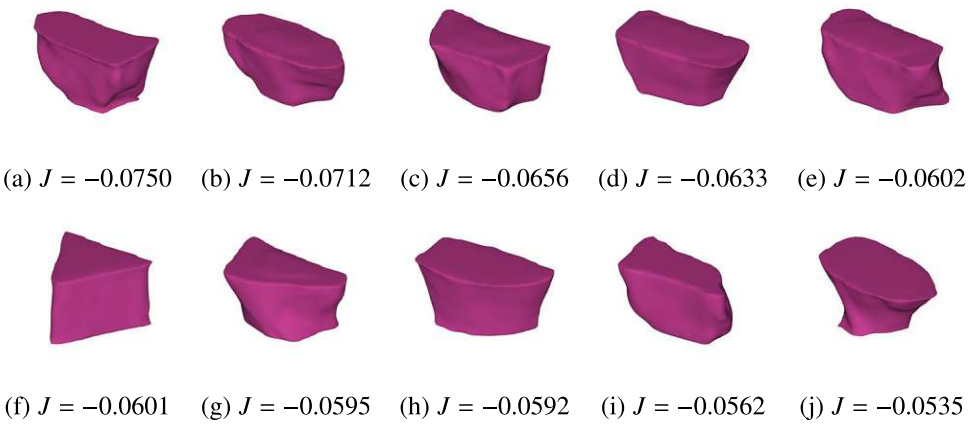


Fig. 14 **a** $J = -0.0750$; **b** $J = -0.0712$; **c** $J = -0.0656$; **d** $J = -0.0633$; **e** $J = -0.0602$; **f** $J = -0.0601$; **g** $J = -0.0595$; **h** $J = -0.0592$; **i** $J = -0.0562$; **j** $J = -0.0535$. Ten best shapes obtained from 16D SOGA optimization. Except for the sixth-best shape (h), all shapes feature an expanded top, similar orientation, and appear widened in y direction (i.e., perpendicular to the main flow direction)

case. The developed shape parameterization's fundamental principle is to exploit neural nets' ability to construct a dimension reduction onto a feature-dense, low-dimensional latent space. It should be noted that other interpolation methods may be able to construct similar interpolation spaces. However, this work demonstrates neural networks' exceptional generalization power yields excellent shape parametrization, which particularly allows interpolation between shapes of varying topologies.

First, the nature of this low-dimensional space is studied by t-SNE-plots. These plots give visual evidence that the generative models create *smooth* shape parameterizations that enable one to use classical, heuristic optimization algorithms. Comparing genetic to such heuristic algorithms, Table 2 reveals that the SOGA algorithm required significantly more iterations (i.e., simulations). Additionally, Table 1 shows that in the studied examples, this additional computational effort is not reflected proportionally by improved mixing. One may expect that the SOGA algorithm's random nature may be better suited to explore the hardly interpretable latent space. However, the results suggest a smoothness of the learned parameterization that renders deterministic methods like DIRECT equally well suited for optimization in the latent space.

In addition, to the general applicability of generative models, we study the influence of different latent dimensions. While the actual optimization results appear pleasing, Figs. 9 and 10 suggest that very compressed (i.e., four-dimensional) latent spaces may not be used for optimization purposes. Analogously, no direct preference between the 8- and 16-dimensional results can be drawn from the optimization results. Similarly, Fig. 10 indicates that higher dimensional latent spaces yield more precise shape encoding, which seems generally preferable. Since the overall number of iterations until convergence of the optimization problem is comparable, the 16-dimensional parameterization might be chosen over the 8-dimensional variant.

As intended, a fundamental improvement over established low-dimensional shape parameterizations is that the new approach covers a much broader design area in a single optimization. Since its fundamental concept is to encode diverse shapes, optimizations lead to numerous, nearly equally optimal shapes. Consequently, this novel approach extends on the existing methods in that it allows the practitioner to *derive* design features that enhance mixing most and for a wide range of basis shapes. Therefore, rather than creating complex shape parameterizations, the crucial step towards optimal design reduces to the creative definition of a training set.

A significant challenge in using neural-net-based shape parameterization is proper control of the output shapes' size. This work implements a volume constraint to avoid simple size maximization of the mixing elements. However, a reformulated objective, such as penalizing pressure loss, may circumvent such adverse designs. Alternatively, a scale factor may be added as an additional optimization variable. Both size control and efficient training set generation may be topics of further studies.

Given the presented results, utilizing the feature-rich latent representations and their immense generalization power has a significant potential to improve established industrial designs.

Finally, the presented framework can be extended by adopting neural-network-based simulators, as recent works [51–55] have shown promising results. Using neural-network-based simulators, the evaluations of objective functions during the online phase become mere forward pass(es) of neural nets that replace costly numerical simulations. Combining the presented parametrization and aforementioned simulators, one could also benefit from the differentiability of neural networks to acquire gradients, which naturally opens the door to other gradient-based optimization algorithms. It is worth noting that if the objective function is formulated as part of the network, automatic differentiation can be leveraged to its full potential, allowing for the

efficient computation of derivative values with respect to design parameters. Both neural-network-based simulators and development of suitable objective functions will be investigated in the future work.

Acknowledgements The German Research Foundation (DFG) fundings under the DFG grant "Automated design and optimization of dynamic mixing and shear elements for single-screw extruders" and priority program 2231 "Efficient cooling, lubrication and transportation—coupled mechanical and fluid-dynamical simulation methods for efficient production processes (FLUSIMPRO)"—under Project No. 439919057 are gratefully acknowledged. Implementation was done on the HPC cluster provided by IT Center at RWTH Aachen. Simulations were performed with computing resources granted by RWTH Aachen University under Project Nos. jara0185 and thes0735.

Funding Open access funding provided by TU Wien (TUW).

Data availability Not applicable.

Declarations

Conflict of interest The authors have no conflict of interest related to this manuscript.

References

1. Celik O, Erb T, Bonten C (2017) Mischgüte in Einschneckenextrudern vorhersagen. Kunststoffe 71:175–177
2. Gale M (2009) Mixing in single screw extrusion. Handbook series. Smithers Rapra Publishing, Shrewsbury
3. Roland W, Marschik C, Miethlinger J, Chung C (2019) Mixing study on different pineapple mixer designs—simulation results 1. In: SPE-ANTEC Technical Papers
4. Sun X, Spalding MA, Womer TW, Uzelac N (2017) Design optimization of maddock mixers for single-screw extrusion using numerical simulation. In: Proceedings of the 75th Annual Technical Conference of the Society of Plastic Engineers (ANTEC)
5. Campbell GA, Spalding MA (2021) Analyzing and troubleshooting single-screw extruders. In: Analyzing and troubleshooting single-screw extruders (second edition), Second edition edn., pp 1–19. Hanser, Munich. https://doi.org/10.3139/9781569907856.fm
6. Marschik C, Osswald T, Roland W, Albrecht H, Skrabala O, Miethlinger J (2018) Numerical analysis of mixing in block-head mixing screws. Polym Eng Sci. https://doi.org/10.1002/pen.24968
7. Böhme G (1981) Strömungsmechanik Nicht-newtonscher Fluide, 1st ed. 1981 edn. Leitfäden der angewandten Mathematik und Mechanik - Teubner Studienbücher. Vieweg+Teubner Verlag, Wiesbaden
8. Kim SJ, Kwon TH (1996) Enhancement of mixing performance of single-screw extrusion processes via chaotic flows: Part i. basic concepts and experimental study. Adv Polym Technol 15(1):41–54. https://doi.org/10.1002/(SICI)1098-2329(199621)15:1<41::AID-ADV4>3.0.CO;2-K
9. Wong AC-Y, Lam Y, Wong ACM (2009) Quantification of dynamic mixing performance of single screws of different configurations by visualization and image analysis. Adv Polym Technol 28(1):1–15. https://doi.org/10.1002/adv.20142
10. Kim SJ, Kwon TH (1996) Enhancement of mixing performance of single-screw extrusion processes via chaotic flows: part ii. numerical study. Adv Polym Technol 15(1):55–69. https://doi.org/10.1002/(SICI)1098-2329(199621)15:1<55::AID-ADV5>3.0.CO;2-J
11. Domingues N, Gaspar-Cunha A, Covas J (2012) A quantitative approach to assess the mixing ability of single-screw extruders for polymer extrusion. J Polym Eng. https://doi.org/10.1515/polyeng-2012-0501
12. Elgeti S, Probst M, Windeck C, Behr M, Michaeli W, Hopmann C (2012) Numerical shape optimization as an approach to extrusion die design. Finite Elem Anal Des 61:35–43. https://doi.org/10.1016/j.finel.2012.06.008
13. Siegbert R, Behr M, Elgeti S (2016) Die swell as an objective in the design of polymer extrusion dies. AIP Conf Proc 1769(1):140003. https://doi.org/10.1063/1.4963540
14. Gaspar-Cunha A, Covas J (2001) The design of extrusion screws: an optimization approach. Int Polym Process. https://doi.org/10.3139/217.1652
15. Potente H, Többen WH (2002) Improved design of shearing sections with new calculation models based on 3d finite-element simulations. Macromol Mater Eng 287(11):808–814. https://doi.org/10.1002/mame.200290010
16. Hube S, Behr M, Elgeti S, Schön M, Sasse J, Hopmann C (2022) Numerical design of distributive mixing elements. Finite Elem Anal Des 204:103733. https://doi.org/10.1016/j.finel.2022.103733
17. Botsch M, Kobbelt L (2004) An intuitive framework for real-time freeform modeling. ACM Trans Graph 23(3):630–634. https://doi.org/10.1145/1015706.1015772
18. Farin G (2002) 5 - the bernstein form of a bézier curve. In: Farin G (ed) Curves and Surfaces for CAGD (Fifth Edition), Fifth edition edn. The Morgan Kaufmann Series in Computer Graphics, pp 57–79. Morgan Kaufmann, San Francisco. https://doi.org/10.1016/B978-155860737-8/50005-3
19. Piegl L, Tiller W (1996) The NURBS book. Monographs in visual communication. Springer, Berlin
20. Sederberg TW, Parry SR (1986) Free-form deformation of solid geometric models. SIGGRAPH Comput Graph 20(4):151–160. https://doi.org/10.1145/15886.15903
21. Hojjat M, Stavropoulou E, Bletzinger K-U (2014) The vertex morphing method for node-based shape optimization. Comput Methods Appl Mech Eng 268:494–513. https://doi.org/10.1016/j.cma.2013.10.015
22. Johnson J, Alahi A, Li F (2016) Perceptual losses for real-time style transfer and super-resolution. CoRR arXiv:1603.08155
23. Friedrich T, Aulig N, Menzel S (2018) On the potential and challenges of neural style transfer for three-dimensional shape data. In: EngOpt 2018. Springer
24. Liu S, Giles CL, Ororbia A (2018) Learning a hierarchical latent-variable model of 3d shapes. In: 2018 International Conference on 3D Vision, 3DV 2018, Verona, Italy, September 5–8, 2018, pp 542–551. IEEE Computer Society, your homne. https://doi.org/10.1109/3DV.2018.00068

25. Wu J, Zhang C, Xue T, Freeman WT, Tenenbaum JB (2016) Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16, pp 82–90. Curran Associates Inc., Red Hook, NY, USA
26. Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, Xiao J (2015) 3d shapenets: a deep representation for volumetric shapes. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1912–1920. https://doi.org/10.1109/CVPR.2015.7298801
27. Charles RQ, Su H, Kaichun M, Guibas LJ (2017) Pointnet: deep learning on point sets for 3d classification and segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 77–85. https://doi.org/10.1109/CVPR.2017.16
28. Ramasinghe S, Khan S, Barnes N, Gould S (2020) Spectral-gans for high-resolution 3d point-cloud generation. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 8169–8176. https://doi.org/10.1109/IROS45743.2020.9341265
29. Yang Y, Feng C, Shen Y, Tian D (2018) Foldingnet: point cloud auto-encoder via deep grid deformation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
30. Bagautdinov T, Wu C, Saragih J, Fua P, Sheikh Y (2018) Modeling facial geometry using compositional vaes. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 3877–3886
31. Tan Q, Gao L, Lai Y-K, Xia S (2018) Variational autoencoders for deforming 3d mesh models, pp 5841–5850. https://doi.org/10.1109/CVPR.2018.00612
32. Groueix T, Fisher M, Kim VG, Russell BC, Aubry M (2018) A papier-mache approach to learning 3d surface generation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 216–224. https://doi.org/10.1109/CVPR.2018.00030
33. Park JJ, Florence P, Straub J, Newcombe R, Lovegrove S (2019) Deepsdf: learning continuous signed distance functions for shape representation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
34. Chen Z, Zhang H (2019) Learning implicit fields for generative shape modeling. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
35. Mescheder L, Oechsle M, Niemeyer M, Nowozin S, Geiger A (2019) Occupancy networks: learning 3d reconstruction in function space. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
36. Lassila T, Rozza G (2010) Parametric free-form shape design with pde models and reduced basis method. Comput Methods Appl Mech Eng 199(23):1583–1592. https://doi.org/10.1016/j.cma.2010.01.007
37. Doersch C (2016) Tutorial on Variational Autoencoders. arXiv preprint. arXiv:1606.05908
38. Kingma DP, Welling M (2013) Auto-Encoding Variational Bayes. arXiv preprint. arXiv:1312.6114
39. Tan Q, Gao L, Lai Y, Xia S (2018) Variational autoencoders for deforming 3d mesh models. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 5841–5850. https://doi.org/10.1109/CVPR.2018.00612
40. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ (eds) Advances in neural information processing systems, vol 27. Curran Associates Inc, New York, pp 2672–2680
41. Wu J, Zhang C, Xue T, Freeman B, Tenenbaum J (2016) Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In: Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R (eds) Advances in neural information processing systems, vol 29. Curran Associates Inc, New York, pp 82–90
42. Dawson-Haggerty et al.: trimesh. https://trimsh.org/
43. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint. arXiv:1412.6980
44. Lorensen WE, Cline HE (1987) Marching cubes: a high resolution 3d surface construction algorithm. SIGGRAPH Comput Graph 21(4):163–169. https://doi.org/10.1145/37402.37422
45. Lewiner T, Lopes H, Vieira AW, Tavares G (2003) Efficient implementation of marching cubes' cases with topological guarantees. J Graph Tools 8:2003
46. Si H (2015) Tetgen, a delaunay-based quality tetrahedral mesh generator. ACM Trans Math Softw. https://doi.org/10.1145/2629697
47. Adams BM, Eldred MS, Geraci G, Hooper RW, JD, J, Maupin KA, Monschke JA, Rushdi AA, Stephens JA, Swiler LP, Wildey TM (July 2014; updated May 2019) Dakota, A multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: version 6.10 user's manual
48. Jones D, Perttunen C, Stuckman B (1993) Lipschitzian optimisation without the lipschitz constant. J Optim Theory Appl 79:157–181. https://doi.org/10.1007/BF00941892
49. Eddy J, Lewis K (2001) Effective generation of pareto sets using genetic programming. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. Volume 2B: 27th Design Automation Conference, pp 783–791. https://doi.org/10.1115/DETC2001/DAC-21094
50. Maaten LJP, Hinton GE (2008) Visualizing high-dimensional data using t-sne. J Mach Learn Res 9(11):2579–2605
51. Li A, Zhang YJ (2023) Isogeometric analysis-based physics-informed graph neural network for studying traffic jam in neurons. Comput Methods Appl Mech Eng 403:115757. https://doi.org/10.1016/j.cma.2022.115757
52. Mallik W, Farvolden N, Jelovica J, Jaiman RK (2022) Deep convolutional neural network for shape optimization using level-set approach. arXiv preprint. arXiv:2201.06210
53. Remelli E, Lukoianov A, Richter S, Guillard B, Bagautdinov T, Baque P, Fua P (2020) Meshsdf: differentiable iso-surface extraction. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H (eds) Advances in Neural Information Processing Systems, vol 33. Curran Associates Inc, New York, pp 22468–22478 https://proceedings.neurips.cc/paper_files/paper/2020/file/fe40fb944ee700392ed51bfe84dd4e3d-Paper.pdf
54. Shukla K, Oommen V, Peyvan A, Penwarden M, Bravo L, Ghoshal A, Kirby RM, Karniadakis GE (2023) Deep neural operators can serve as accurate surrogates for shape optimization: a case study for airfoils. arXiv preprint. arXiv:2302.00807
55. Woldseth RV, Aage N, Bærentzen JA, Sigmund O (2022) On the use of artificial neural networks in topology optimisation. Struct Multidiscip Optim 65(10):294. https://doi.org/10.1007/s00158-022-03347-1

3 Outlook

The central contribution of this dissertation is the conception, development, and validation of a novel, modular, and fully coupled FSCI framework. This framework is not an endpoint; it serves as an extensible basis designed to accommodate additional physical complexity.

The initial application of this framework has already proven its value by revealing a fundamental contradiction between purely mechanical friction models and experimental reality, demonstrating that other physics are involved. This finding, combined with the successful development of two distinct reduced modeling methods, provides clear next steps for future research. This outlook is threefold: completing the physics, applying the current models, and advancing the data-driven surrogates.

1. Completing the Physics: Extending the High-Fidelity Model
The most critical finding of this work was that a purely mechanics-based model, while capturing surface interlocking, is insufficient to describe the friction in the secondary shear zone. The framework was built to be modular specifically to address this. The most urgent extension of the high-fidelity FSCI model is the incorporation of missing physical phenomena.

Adhesion Model: Investigating interatomic forces via an adhesion model should be the first step to see if it corrects the contradictory friction trends.

Full Thermal Coupling: The current adiabatic assumption should be upgraded to a fully-coupled, non-adiabatic thermal model. This would allow for the simulation of heat transfer between the chip, tool, and fluid, which is essential for modeling the lubricant's cooling effect and potential "thermal barrier" phenomena identified in Publication P.1.

Shear-Thinning Fluid Model: The fluid model can be improved from a simple Newtonian fluid to a non-Newtonian, shear-thinning model, e.g., Carreau model, to more accurately capture lubricant behavior.

2. Application: Integrating the Scalar Reduced Model (GPR)
The scalar surrogate model developed in Publication P.3 achieved speed-ups of 10^7, while preserving accuracy. This model is ready for application. The next logical step is to integrate this GPR model into a meso-scale chip formation simulation (the FE^2 loop). This would enable large-scale simulations that have a physics-based (rather than empirical) friction law at their core. This coupled multi-scale framework could then be used to run optimization studies on tool topography and MQL parameters.

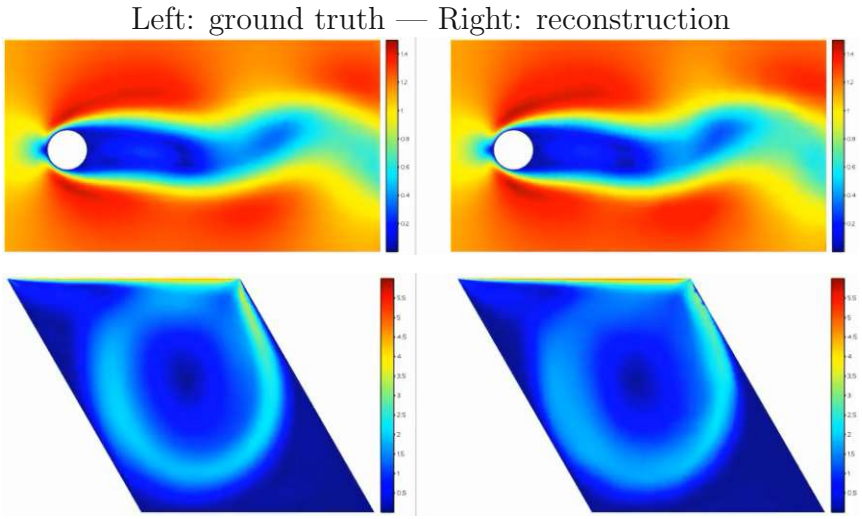
3. Developing the Full-Field Surrogate

The long-term vision of this research is to move beyond a single scalar friction coefficient and create a surrogate model that captures the entire full-field physics of the contact zone. The methodological foundation for this was established in Publication P.4, which demonstrated that an autodecoder could effectively learn a compact latent-space representation of continuous functions defined over a spatial domain.

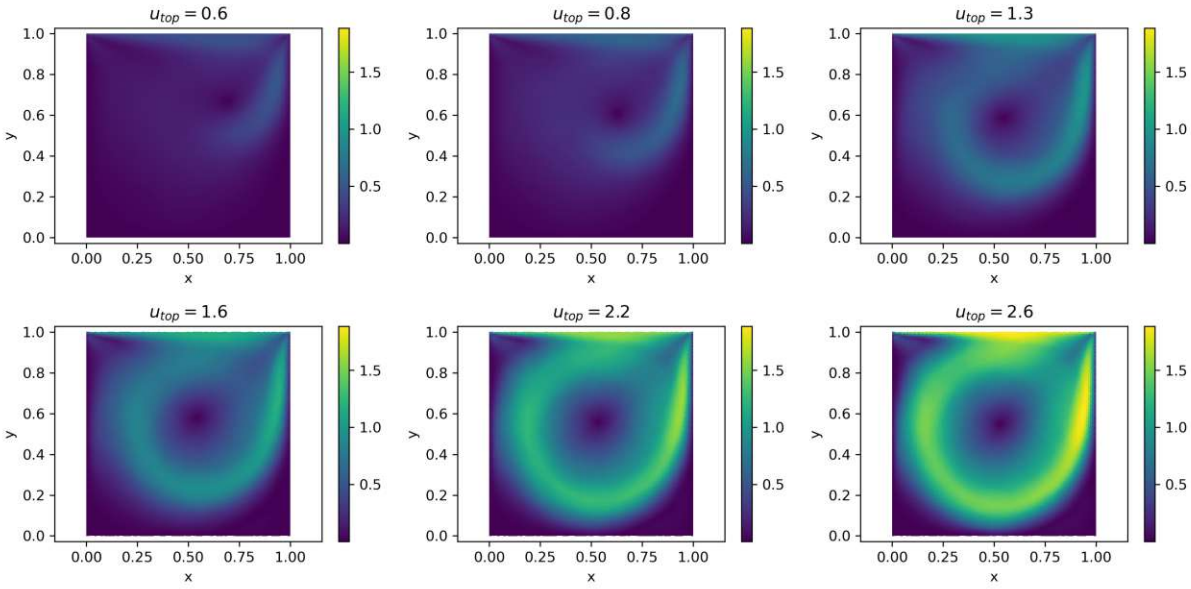
As this dissertation's key bridge to future work, this methodology was validated on the specific type of problem central to this thesis – complex, non-linear vector fields. This validation was performed in a supervised Master's thesis [19]. That study applied the autodecoder architecture to the dynamic velocity fields from Computational Fluid Dynamics (CFD) simulations, such as the Karman vortex street and lid-driven cavity.

The results of that study, shown in Figures 3.1a and 3.1b, demonstrate the suitability of the architecture. The network successfully learned to reconstruct the full velocity field from the latent space and to accurately capture flow structures even under unseen boundary conditions.

This successful validation confirms that the autodecoder is a viable path to create a full-field reduced model for the FSCI problem. The next step is therefore to apply this method to the newly extended high-fidelity model (as proposed above) to create a true "digital twin" of the contact zone, enabling real-time, physics-based optimization of tool design.



(a) Reconstruction of training data. Top: Karman vortex street; Bottom: Lid-driven cavity.



(b) Latent space interpolation. Adapted from [19].

Figure 3.1: Validation of the deep learning-based full-field surrogate model. (a) Comparison of ground truth (left) and network reconstruction (right) for the Karman vortex street and Lid-driven cavity test cases, demonstrating accurate capture of flow features. (b) Latent space interpolation between two different boundary conditions (lid velocities), illustrating the evolution of the flow field as the lid velocity increases, which demonstrates the smoothness of the learned manifold.

Bibliography

[1] ANSYS, Inc. *ANSYS Mechanical User's Guide*. Release 2024 R1. ANSYS, Inc. Canonsburg, PA, USA, 2024 (cit. on p. 3).

[2] Armin Beckert. "Coupling fluid (CFD) and structural (FE) models using finite interpolation elements". en. In: *Aerospace Science and Technology* 4.1 (Jan. 2000), pp. 13–22. ISSN: 12709638. DOI: 10.1016/S1270-9638(00)00111-5 (cit. on p. 11).

[3] J. Chung and G. M. Hulbert. "A Time Integration Algorithm for Structural Dynamics With Improved Numerical Dissipation: The Generalized-αMethod". en. In: *Journal of Applied Mechanics* 60.2 (June 1993), pp. 371–375. ISSN: 0021-8936, 1528-9036. DOI: 10.1115/1.2900803 (cit. on p. 9).

[4] Dassault Systèmes. *Abaqus/Standard User's Manual*. Version 2024. Dassault Systèmes. Providence, RI, USA, 2024 (cit. on p. 3).

[5] L. De Lorenzis, İ. Temizer, P. Wriggers, and G. Zavarise. "A large deformation frictional contact formulation using NURBS-based isogeometric analysis: APPLICATION OF NURBS-BASED ISOGEOMETRIC ANALYSIS". en. In: *International Journal for Numerical Methods in Engineering* 87.13 (Sept. 2011), pp. 1278–1300. ISSN: 00295981. DOI: 10.1002/nme.3159 (cit. on p. 8).

[6] Norbert Hosters, Jan Helmig, Atanas Stavrev, Marek Behr, and Stefanie Elgeti. "Fluid–structure interaction with NURBS-based coupling". en. In: *Computer Methods in Applied Mechanics and Engineering* 332 (Apr. 2018), pp. 520–539. ISSN: 00457825. DOI: 10.1016/j.cma.2018.01.003 (cit. on p. 11).

[7] G.R. Johnson and W.H. Cook. "A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures". In: *Seventh International Symposium on Ballistics, the Hague, the Netherlands* 21 (Jan. 1983), pp. 541–548 (cit. on p. 8).

[8] D. Kraft. *A Software Package for Sequential Quadratic Programming*. Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt Köln: Forschungsbericht. Wiss. Berichtswesen d. DFVLR, 1988 (cit. on p. 9).

[9] Ulrich Küttler, Christiane Förster, and Wolfgang A. Wall. "A Solution for the Incompressibility Dilemma in Partitioned Fluid–Structure Interaction with Pure Dirichlet Fluid Domains". en. In: *Computational Mechanics* 38.4-5 (Sept. 2006), pp. 417–429. ISSN: 0178-7675, 1432-0924. DOI: 10.1007/s00466-006-0066-5 (cit. on p. 10).

[10] J Lee, S Elgeti, J Saelzer, and A Zabel. "Development and Validation of a Micro-Tribology Model for the Chip Formation Zone". en. In: *Procedia CIRP* 117 (2023), pp. 317–322. ISSN: 22128271. DOI: 10.1016/j.procir.2023.03.054 (cit. on pp. XVII, 2, 14, 25).

[11] Jaewook Lee, Sebastian Hube, and Stefanie Elgeti. "Neural networks vs. splines: advances in numerical extruder design". en. In: *Engineering with Computers* 40.2 (Apr. 2024), pp. 989–1004. ISSN: 0177-0667, 1435-5663. DOI: 10.1007/s00366-023-01839-2 (cit. on pp. XVII, 3, 12, 15, 45).

[12] Jaewook Lee, Jannis Saelzer, Jacques Zwar, Florian Zwicke, Felipe Gonzalez, Thomas Spenke, Norbert Hosters, Gero Polus, Andreas Zabel, and Stefanie Elgeti. "Spline-Based Framework for Microscopic Contact Zone Modeling in Lubricated Orthogonal Cutting". en. In: *International Journal for Numerical Methods in Engineering* 126.14 (July 2025), e70087. ISSN: 0029-5981, 1097-0207. DOI: 10.1002/nme.70087 (cit. on pp. XVII, 2, 14, 32).

[13] Michel Make, Thomas Spenke, Norbert Hosters, and Marek Behr. "Spline-based space-time finite element approach for fluid-structure interaction problems with a focus on fully enclosed domains". en. In: *Computers & Mathematics with Applications* 114 (May 2022), pp. 210–224. ISSN: 08981221. DOI: 10.1016/j.camwa.2022.03.035 (cit. on p. 10).

[14] Amir Malakizadi, Keyvan Hosseinkhani, Emilia Mariano, E. Ng, Antonio Del Prete, and Lars Nyborg. "Influence of friction models on FE simulation results of orthogonal cutting process". In: *The International Journal of Advanced Manufacturing Technology* 88.9–12 (June 2016), pp. 3217–3232. ISSN: 1433-3015. DOI: 10.1007/s00170-016-9023-4 (cit. on p. 15).

[15] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. en. 2nd ed. Springer series in operations research. OCLC: ocm68629100. New York: Springer, 2006. ISBN: 978-0-387-30303-1 (cit. on p. 9).

[16] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019 (cit. on p. 12).

[17] Kyusic Park and Matthew S. Allen. "A Gaussian process regression reduced order model for geometrically nonlinear structures". In: *MECHANICAL SYSTEMS AND SIGNAL PROCESSING* 184 (Feb. 2023). ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2022.109720 (cit. on p. 15).

[18] Les A. Piegl and Wayne Tiller. *The NURBS book*. en. 2nd ed. Monographs in visual communications. Springer, 1997. ISBN: 978-3-540-61545-3 (cit. on pp. 6, 7).

[19] Mark Riegler. "Data-driven reduced models for numerical simulations". MA thesis. Wien: Technische Universität Wien, 2023. DOI: 10.34726/hss.2023.112843 (cit. on pp. 3, 12, 16, 64, 65).

[20] J Saelzer, Y Alammari, A Zabel, D Biermann, J Lee, and S Elgeti. "Characterisation and modelling of friction depending on the tool topography and the intermediate medium". en. In: *Procedia CIRP* 102 (2021), pp. 435–440. ISSN: 22128271. DOI: 10.1016/j.procir.2021.09.074 (cit. on pp. XVII, 2, 13, 18).

[21] Scientific Forming Technologies Corporation (SFTC). *DEFORM-3D User's Manual*. Version 13.0. Scientific Forming Technologies Corporation (SFTC). Columbus, OH, USA, 2023 (cit. on p. 3).

[22] Milton C Shaw. *Metal Cutting Principles*. en. 2nd ed. Oxford Series on Advanced Manufacturing. New York, NY: Oxford University Press, June 2004 (cit. on p. 2).

[23] J.C. Simo and T.J.R. Hughes. *Computational Inelasticity*. Interdisciplinary Applied Mathematics. Springer New York, 2006. ISBN: 9780387227634 (cit. on p. 8).

[24] İ. Temizer, P. Wriggers, and T.J.R. Hughes. "Contact treatment in isogeometric analysis with NURBS". en. In: *Computer Methods in Applied Mechanics and Engineering* 200.9-12 (Feb. 2011), pp. 1100–1112. ISSN: 00457825. DOI: 10.1016/j.cma.2010.11.020 (cit. on p. 8).

[25] E M Trent and P K Wright. *Metal Cutting*. en. 4th ed. Oxford, England: Butterworth-Heinemann, Jan. 2000 (cit. on p. 2).

[26] K. Weinert, I. Inasaki, J.W. Sutherland, and T. Wakabayashi. "Dry Machining and Minimum Quantity Lubrication". In: *CIRP Annals* 53.2 (2004), pp. 511–537. ISSN: 0007-8506. DOI: 10.1016/s0007-8506(07)60027-4 (cit. on p. 2).

[27] Andreas Zabel, Jannis Saelzer, Gerhard Poll, Florian Pape, Haichao Liu, Berend Denkena, Lars Ellersiek, Stefanie Elgeti, and Jaewook Lee. "Effizientes Schmieren der Spanbildungszone/Controlled use of cooling lubricants for tribological optimization of machining processes – Efficient lubrication of the chip formation zone". In: *wt Werkstattstechnik online* 112.01–02 (2022), pp. 44–49. ISSN: 1436-4980. DOI: 10.37544/1436-4980-2022-01-02-48 (cit. on p. XVII).