
A COMPLEMENT TO NEURAL NETWORKS FOR ANISOTROPIC INELASTICITY AT FINITE STRAINS

A PREPRINT

Hagen Holthusen
Institute of Applied Mechanics
University of Erlangen-Nuremberg
Erlangen, 91058, Germany
hagen.holthusen@fau.de

Ellen Kuhl
Department of Mechanical Engineering
Stanford University
Stanford, CA 94305, United States
ekuhl@stanford.edu

October 7, 2025

ABSTRACT

We propose a complement to constitutive modeling that augments neural networks with material principles to capture anisotropy and inelasticity at finite strains. The key element is a dual potential that governs dissipation, consistently incorporates anisotropy, and—unlike conventional convex formulations—satisfies the dissipation inequality without requiring convexity.

Our neural network architecture employs invariant-based input representations in terms of mixed elastic, inelastic and structural tensors. It adapts Input Convex Neural Networks, and introduces Input Monotonic Neural Networks to broaden the admissible potential class. To bypass exponential-map time integration in the finite strain regime and stabilize the training of inelastic materials, we employ recurrent Liquid Neural Networks.

The approach is evaluated at both material point and structural scales. We benchmark against recurrent models without physical constraints and validate predictions of deformation and reaction forces for unseen boundary value problems. In all cases, the method delivers accurate and stable performance beyond the training regime. The neural network and finite element implementations are available as open-source and are accessible to the public via Zenodo.org.

Keywords Neural Networks · Inelasticity · Anisotropy · Finite Strains · Automated Model Discovery · Finite Element Simulation

1 Introduction

Artificial intelligence (AI) has increasingly permeated diverse areas of computational mechanics. Its main advantage lies in its intrinsic ability to handle complex and high-dimensional data while remaining continuously adaptable. Applications range from inverse design [1, 2], topology optimization [3], multiscale modeling [4, 5, 6], and multiphysics problems [7, 8], to model order reduction [9, 10], the identification of material behavior [11, 12, 13], and real-time digital twins [14]. In all these contexts, the development of improved models requires more data to enhance predictive accuracy and deepen our understanding of the underlying mechanics. At the same time, however, larger datasets generally increase the complexity of the models themselves. Overcoming this apparent contradiction highlights the particular promise of AI in mechanics.

Beyond its methodological benefits, the current momentum in AI research—both in terms of algorithmic development and in its application across disciplines—represents an additional advantage [15, 16]. One may question whether AI always constitutes the most efficient solution, or whether in certain cases its application may be disproportionate to the complexity of the problem. Nevertheless, disregarding the accumulated expertise and its adaptation to computational mechanics would be a missed opportunity.

Compared to fields such as the social sciences, AI in mechanics benefits from the existence of uncertainty-free physical laws. Recent years have therefore seen considerable efforts to incorporate these laws into AI models. Although early

applications of neural networks to material modeling date back more than 30 years [17], a major boost was provided by the introduction of physics-informed neural networks (PINNs) [18], where governing equations are weakly enforced through additional loss function terms. An alternative paradigm designs neural network architectures that satisfy physical principles *a priori*, thereby guaranteeing that predictions adhere to the underlying laws. Prominent approaches include the unsupervised EUCLID framework [11], Thermodynamics-based Artificial Neural Networks (TANNs) [19, 20], Constitutive Artificial Neural Networks (CANNs) [12, 21], and Physics-Augmented Neural Networks (PANNs) [13, 22]. While all of these methods enforce physical consistency, they differ in the balance between interpretability and expressivity: CANNs favor sparse networks that allow tracing each computational path, whereas PANNs rely on denser architectures with potentially greater representational power. Although originally developed for elastic materials, recent research has focused on extending these frameworks to (anisotropic) inelastic behavior—representing a natural progression towards handling ever-increasing model complexity in order to explain increasingly rich datasets.

1.1 State-of-the-art in neural networks for inelastic and anisotropic material modeling

Inelastic materials. The EUCLID framework has been extended in recent years to cover a broad spectrum of inelastic phenomena. Early developments demonstrated its applicability to the discovery of plasticity [23], with particular emphasis on non-associative, pressure-sensitive plasticity [24], as well as to the modeling of linear viscoelastic material behavior at small strains [25]. In a more general context, [26] embedded EUCLID in the theoretical setting of Generalized Standard Materials (GSM) [27, 28], a widely recognized framework for capturing inelasticity. GSM is based on the assumption of an inelastic potential complementing the Helmholtz free energy, thereby ensuring thermodynamic consistency.

The idea of postulating a potential that governs the evolution of inelastic deformations has inspired a variety of neural network architectures aiming to replicate this principle; see, for example, [29, 30, 31, 32, 33, 34]. While GSM typically relies on potentials expressed in terms of inelastic strain rates, one may alternatively consider potentials formulated in terms of stresses. This stress-based perspective is often employed when dealing with yield functions, such as the von Mises criterion, and naturally connects to convex analysis. Neural networks exploiting such formulations have recently been applied to a wide range of inelastic processes, including finite viscoelasticity [35, 36], elasto-plasticity [37, 38], fracture mechanics [39], and growth and remodeling phenomena [40]. Notably, stress-based potentials and inelastic strain rate-based potentials share an intrinsic connection: if both are convex, they can be transformed into one another by means of a Legendre transformation, a relation that also holds at finite strains [36].

In the spirit of [37], elasto-plasticity at small strains was recently extended to intrinsically incorporate anisotropy. Two strategies were proposed in [41]: the first augments the classical Hill-48 yield criterion with an Input Convex Neural Network (ICNN) [42], thereby preserving convexity and guaranteeing stability, while the second relies entirely on a neural network in the principal stress space, where anisotropy is embedded through additional permutation-invariant networks.

In the training of neural networks for inelastic material modeling, the availability of high-quality data is a central challenge. While most physics-informed networks are formulated at the material point level, the primary experimental data typically available consists of macroscopic force–displacement curves. Consequently, finite element model updating strategies are frequently employed to generate synthetic data for training, also in the inelastic regime [43].

Beyond GSM-inspired formulations, several alternative paradigms have emerged for tackling inelastic material behavior with neural networks. A first direction focuses on dynamical system representations: [44] introduced a neural ordinary differential equation framework that models inelastic stress response via internal state variables, effectively replacing classical evolution equations by a data-driven continuous-time formulation. Building on this idea, [45] extended the neural ODE framework by incorporating attention mechanisms, thereby enhancing its capability to capture complex history-dependent inelastic processes with improved generalization across diverse loading paths. Similarly, recurrent architectures have been explored, as in [46], where a physics-informed recurrent neural network was developed for multiaxial plasticity, ensuring thermodynamic consistency by enforcing monotonic plastic work. Building upon this time-series perspective, long short-term memory (LSTM) networks were proposed in [47] to capture strain path dependence in crystal plasticity, which offers the possibility of orders of magnitude speed-ups compared to traditional simulations while preserving accuracy in path-dependent effects such as the Bauschinger phenomenon.

A second direction emphasizes operator learning. The work of [48] proposed the History-Aware Neural Operator (HANO), combining Fourier neural operators with hierarchical self-attention to achieve discretization-invariant modeling of path-dependent inelasticity, including anisotropic damage. This approach alleviates the limitations of recurrent models by avoiding hidden state inconsistencies and by generalizing across varying load histories. In parallel, graph-based models have been employed to capture spatially distributed responses: [49] demonstrated that graph neural

networks can accurately predict stress, strain, and displacement fields in materials and structures, highlighting their potential for bridging microstructure-informed modeling with structural-scale predictions.

Anisotropic materials. The modeling of anisotropy with neural networks has gained increasing attention due to the ubiquity of direction-dependent material behavior in both engineering and biological applications [50]. A central idea is to encode preferred material directions through structural tensors, thereby ensuring frame invariance and compatibility with material symmetry groups. Several approaches have been developed along these lines. For instance, [51] introduced polyconvex neural ordinary differential equations for tissue mechanics, ensuring thermodynamic admissibility and convexity while capturing the nonlinear anisotropic response of skin. Complementary to this, probabilistic machine learning strategies have been proposed by [52], who employed Gaussian process regression and physics-informed sampling techniques to build anisotropic hyperelastic models that respect material symmetries and thermodynamic consistency. More recently, [53] presented a framework of Physics-Augmented Neural Networks (PANNs) that leverage generalized structural tensors of higher order. This formulation allows for the simultaneous calibration of network parameters and structural tensors, thereby enabling the detection and accurate representation of complex anisotropy arising from heterogeneous microstructures.

In addition to tensor-based formulations, automated model discovery has been explored for highly anisotropic engineering materials. A notable example is the work of [54], who combined biaxial testing with constitutive neural networks to uncover the unique anisotropic mechanical signature of warp-knitted fabrics. Their study emphasized the sensitivity of constitutive laws to microstructural directions and demonstrated that data-driven discovery can outperform classical orthotropic models.

Anisotropy is particularly crucial in biomechanics, where soft tissues inherently exhibit direction-dependent behavior. In [55], operator-learning strategies were applied to model anisotropic biological tissues, while physics-informed neural networks have been employed to capture the anisotropic hyperelastic response of the human myocardium [56, 57] as well as the constitutive behavior of pulmonary arteries [58]. These studies underline the importance of integrating physics-based constraints with data-driven methods to achieve both predictive accuracy and generalizability in biologically inspired anisotropic material models.

Overall, these works demonstrate that neural networks provide a versatile toolkit for modeling anisotropy, ranging from physics-augmented formulations with structural tensors, to automated discovery of textile mechanics, and advanced PINN-based approaches for cardiovascular tissues. Together, they illustrate the broad applicability of AI-enhanced constitutive modeling in addressing anisotropy across scales and domains.

This literature review is by no means exhaustive, as the integration of machine learning into computational material modeling is currently being studied with great intensity. A particularly comprehensive overview is provided in the recent reviews by [59, 60], which systematically discuss data-driven constitutive laws for solids.

1.2 Research gap and aim of the study

Research gap. In recent years, substantial progress has been made in the development of physics-embedded neural networks for modeling inelastic and anisotropic materials. Nevertheless, most approaches remain confined to the small-strain regime and typically impose convexity of the underlying inelastic potential. The systematic incorporation of anisotropy, both in the Helmholtz free energy and in the corresponding inelastic potential, has received less consistent attention. Furthermore, physics-embedded neural network architectures for inelasticity continue to encounter issues of stability and robustness during training. As a result, a comprehensive framework for anisotropic inelasticity in the finite-strain regime that is robust throughout training is still missing.

Aim of this study. The present contribution aims to provide a complement to the existing theoretical foundations by establishing a consistent strategy for incorporating anisotropy into the architecture of neural networks for inelastic materials at finite strains. This complement is formulated independently of a specific network architecture, thereby broadening its applicability. To further account for the increasing complexity of materials—such as architected microstructures—we propose an extension to non-convex yet physically admissible formulations of the inelastic potential. For stabilizing the training procedure, we complement recent advances in time-integration schemes for inelastic materials by extending them to the finite-strain regime. Together, these complementary developments are expected to substantially advance the use of neural networks in computational material modeling.

1.3 Outline

This contribution is structured as follows: In Section 2.2, we introduce the underlying thermodynamic requirement in a simplified setting. We demonstrate how this condition can be satisfied in one dimension and explain why convexity of

the dual potential is a sufficient, but not a necessary, constraint. Section 3 reviews several neural network architectures. We briefly recap Input Convex Neural Networks and introduce Input Monotonic Neural Networks, whose combination generalizes the findings from the illustrative example to a broader class of network architectures. In addition, we present Liquid Neural Networks as a specialized class of recurrent architectures. The continuum mechanical foundations of finite strain anisotropic inelasticity are developed in Section 4. Sections 4.1 and 4.2 then specify how to represent the Helmholtz free energy and the dual potential in terms of invariants. Building on these foundations, Section 5 integrates the theoretical principles with the neural architectures introduced earlier. Here, we also describe the prediction of inelastic variable updates using Liquid Neural Networks. The effectiveness of the proposed framework is demonstrated in Section 6 through benchmark problems at both the material point and structural level. We investigate isotropic and anisotropic responses and compare the results against recurrent neural networks that lack physics-based constraints. Section 7 provides a critical assessment of the approach, outlining its limitations and highlighting open research questions related to anisotropic inelasticity modeled with neural networks. Finally, Section 8 summarizes the key findings and conclusions of this work.

2 Preliminary Problem Description

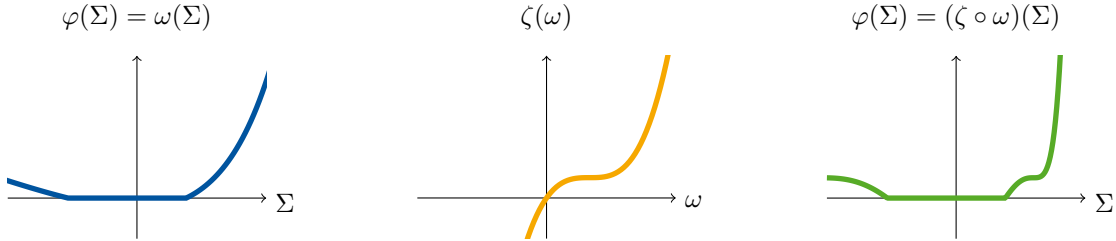


Figure 1: Schematic illustration of possible constructions for a function $\varphi(\Sigma)$ satisfying Inequality (3). The function ω is *convex, non-negative, and zero-valued at the origin*, whereas ζ is *monotonically increasing and zero-valued with respect to ω* . Their composition $(\zeta \circ \omega)$ ensures that the sign of the subgradient $\partial_{\Sigma}\varphi$ coincides with the sign of Σ in the negative and positive regimes, respectively.

In this section, we introduce the fundamental mathematical problem underlying finite-strain inelasticity. This preliminary discussion will facilitate a deeper understanding of the challenges involved and motivate the design of the neural network architectures employed throughout this manuscript.

2.1 Notation

Let a , $\mathbf{a}$, and $\mathbf{A}$ denote tensors of order zero (scalars), one (vectors), and two (second-order tensors), respectively. To distinguish tensor notation from standard linear algebra notation, we write $\mathbf{a}$ and $\mathbf{A}$ for vectors and matrices in the usual sense. The set of unit vector is denoted by $\mathbb{S}^{n-1} := \{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v}\| = 1\}$. Further, the set of real $n \times n$ matrices is denoted by $\mathbb{M}^{n \times n}$, and the group of invertible matrices is given by $\text{GL}(n, \mathbb{R}) = \{\mathbf{A} \in \mathbb{M}^{n \times n} : \det \mathbf{A} \neq 0\}$, with the subgroup of orientation-preserving transformations $\text{GL}^+(n, \mathbb{R}) = \{\mathbf{A} \in \text{GL}(n, \mathbb{R}) : \det \mathbf{A} > 0\}$. The set of symmetric matrices is denoted by $\text{Sym}^{n \times n} = \{\mathbf{A} \in \mathbb{M}^{n \times n} : \mathbf{A}^T = \mathbf{A}\}$, with the subset of symmetric positive definite matrices $\text{Sym}_+^{n \times n} = \{\mathbf{A} \in \text{Sym}^{n \times n} : \mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \ \forall \mathbf{x} \neq \mathbf{0}\}$. Any two tensors $\mathbf{A}, \mathbf{B} \in \mathbb{M}^{n \times n}$ related by $\mathbf{A} = \mathbf{P} \mathbf{B} \mathbf{P}^{-1}$ for $\mathbf{P} \in \text{GL}(n, \mathbb{R})$ are called *similar*, i.e., they share the same eigenvalues. The cofactor of a tensor $\mathbf{A} \in \text{GL}(n, \mathbb{R})$ is defined as $\text{cof } \mathbf{A} = \det(\mathbf{A}) \mathbf{A}^{-T}$, and $\text{cof } \mathbf{A} = \det(\mathbf{A}) \mathbf{A}^{-1}$ if $\mathbf{A} \in \text{Sym}_+^{n \times n}$. The set of orthogonal tensors is given by $\text{O}(n) = \{\mathbf{Q} \in \mathbb{M}^{n \times n} : \mathbf{Q}^T \mathbf{Q} = \mathbf{I}, \det \mathbf{Q} = \pm 1\}$, while the subgroup of proper orthogonal tensors (rotations) is denoted by $\text{SO}(n) = \{\mathbf{Q} \in \text{O}(n) : \det \mathbf{Q} = 1\}$. The second-order identity tensor is expressed as $\mathbf{I} = \delta_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ with $\{\mathbf{e}_i\}$ denoting the Cartesian basis. For consistency, the symbol $\mathbf{1}$ will be used whenever the identity is required in a more general context. In addition, the deviatoric operator dev is defined as $\text{dev } \mathbf{A} = \mathbf{A} - \frac{1}{3} \text{tr}(\mathbf{A}) \mathbf{I}$. The inner product of two objects is denoted by $\langle \cdot, \cdot \rangle$, and a superposed dot $(\dot{\cdot})$ indicates the total time derivative. In tensor notation, $:$ denotes a double contraction, while $\cdot$ refers to a single contraction. Finally, $\partial_A f$ denotes the derivative of f with respect to an argument A (understood as the gradient in the smooth case, or as the subgradient in the nonsmooth case). Moreover, the operator $\mathcal{N}$ indicates that a function is represented by a neural network.

Throughout this manuscript, we do not introduce new symbols for functions whose arguments have been modified; instead, we use the same symbol whenever the context makes the meaning unambiguous.

2.2 One-dimensional illustrative example

We begin by defining two scalar-valued functions

$$\psi : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (\varepsilon, \varepsilon_i) \mapsto \psi(\varepsilon, \varepsilon_i), \quad \varphi : \mathbb{R} \rightarrow \mathbb{R}, \quad \Sigma \mapsto \varphi(\Sigma),$$

where ε denotes the total strain and ε_i is an internal (inelastic) variable that accounts for history dependency. We further introduce the following constitutive relations

$$\sigma = \frac{\partial \psi}{\partial \varepsilon}, \quad \Sigma = -\frac{\partial \psi}{\partial \varepsilon_i}, \quad (1)$$

where Σ is referred to as the *thermodynamically consistent driving force* conjugate to $\dot{\varepsilon}_i$ and σ denotes the mechanical stress state. The essential thermodynamic requirement is

$$\Sigma \dot{\varepsilon}_i \geq 0. \quad (2)$$

This inequality expresses the non-negativity of the internal dissipation and must be satisfied at all times.

To close the system, we postulate that the evolution of the internal variable ε_i can be derived from the scalar potential φ , i.e.,

$$\Sigma \dot{\varepsilon}_i = \Sigma \partial_\Sigma \varphi \geq 0. \quad (3)$$

where $\partial_\Sigma \varphi$ is generally nonsmooth, i.e., the subgradient of φ . The central question is then how to construct φ such that the dissipation inequality (3) is always satisfied.

Convex functions. A standard choice in material modeling is to assume that φ is *convex*, *non-negative*, and *zero-valued at the origin* [61]. Such a function ensures that the dissipation inequality is satisfied automatically, due to the properties of subgradients [62], namely

$$\omega(0) \geq \omega(\Sigma) - \langle \partial_\Sigma \omega, \Sigma \rangle, \quad (4)$$

where ω denotes a convex function with $\omega(0) = 0$ and $\omega(\Sigma) \geq 0$. Inequality (3) is then satisfied *a priori*. A schematic illustration of such a potential is given in Figure 1.

Monotonic functions. While convexity is *sufficient*¹, it is not *necessary* to guarantee positive dissipation. Indeed, Equation (3) only requires that the signs of Σ and $\partial_\Sigma \varphi$ coincide. This can be achieved if φ is *monotonically decreasing* for $\Sigma < 0$ and *monotonically increasing* for $\Sigma > 0$ ². To formalize this, let $\zeta : \mathbb{R} \rightarrow \mathbb{R}$ be a monotonically increasing, zero-valued function. We can then define

$$\varphi(\Sigma) = (\zeta \circ \omega)(\Sigma).$$

By the chain rule, the dissipation reads

$$\Sigma \partial_\Sigma \varphi = \Sigma \left(\frac{\partial \zeta}{\partial \omega} \partial_\Sigma \omega \right) = \underbrace{\frac{\partial \zeta}{\partial \omega}}_{\geq 0} \underbrace{(\partial_\Sigma \omega \Sigma)}_{\geq 0} \geq 0, \quad (5)$$

which is guaranteed by the monotonicity of ζ with respect to ω and the convexity property (4). This generalized construction allows greater flexibility in designing constitutive potentials while maintaining thermodynamic consistency. Figure 1 illustrates the different possibilities: the blue (left) graph represents the convex (more constraining) option, while the green graph is monotonically increasing in the positive regime and monotonically decreasing within the negative regime (more flexible).

3 Neural Network Architectures

In the previous Section 2, we demonstrated how to construct a function $\varphi(\Sigma)$ that satisfies the reduced dissipation inequality *a priori*. In the setting of finite strain inelasticity, these considerations can be extended in a straightforward manner to three-dimensional problems. Since our ultimate goal is to approximate both the Helmholtz free energy ψ and the dual potential φ , this section introduces the neural network architectures employed for this purpose. Moreover, inelastic materials additionally require the solution of evolution equations for history-dependent internal variables. To this end, we further make use of a specialized recurrent architecture known as *Liquid Neural Networks*.

¹In combination with zero-valued at the origin and non-negativity.

²Note that convex, zero-valued, and non-negative functions are a special case of this class.

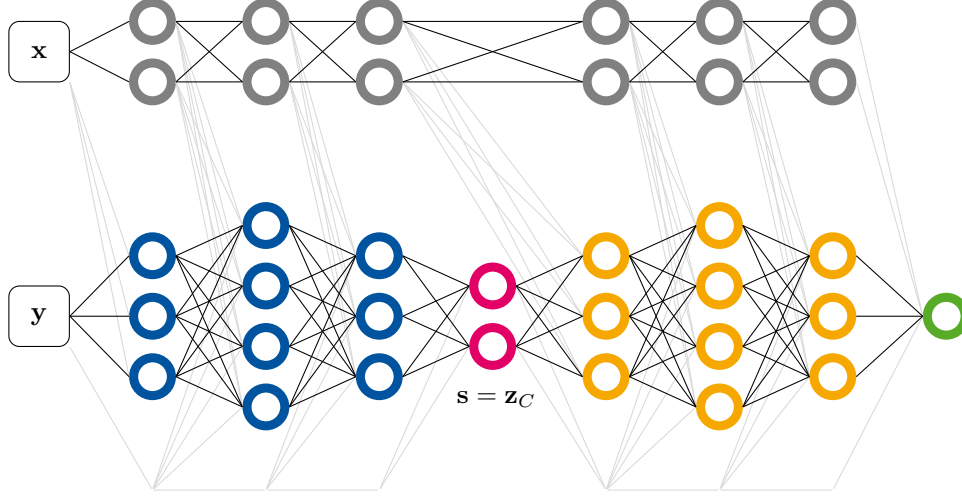


Figure 2: Schematic illustration of the composition network $\mathcal{N}_o$. The blue part represents the Input Convex Neural Network $\mathcal{N}_c$, which is convex in $\mathbf{y}$ but not in $\mathbf{x}$. Each layer takes as input the output of the previous layer and of the parallel network in gray carrying $\mathbf{x}$ as well as $\mathbf{y}$ itself. Its final output $\mathbf{z}_C$ highlighted in red serves as the input $\mathbf{s}$ to the Input Monotonic Neural Network $\mathcal{N}_m$ shown in orange, whose architecture mirrors that of $\mathcal{N}_c$. To guarantee a non-negative $\mathbf{z}_C$, the activation between $\mathbf{z}_{C-1}$ and $\mathbf{z}_C$ is chosen as ReLU.

Input Convex Neural Networks. We begin by introducing (partially) Input Convex Neural Networks (ICNNs) $(\mathbf{x}, \mathbf{y}) \mapsto \mathcal{N}_c(\mathbf{x}, \mathbf{y})$, originally proposed in [42]. These networks are designed such that the output is convex with respect to $\mathbf{y}$ while remaining unconstrained in $\mathbf{x}$. The operator $\mathcal{N}_c$ denotes a neural network with hidden layers $l = 0, \dots, C-1$ and an output layer C . The key idea is to employ two parallel computational branches: one for the convex variables and one for the non-convex variables. The non-convex branch may feed into the convex branch in a way that preserves convexity with respect to $\mathbf{y}$. The architecture proposed in [42] reads

$$\begin{aligned} \mathbf{u}_{l+1} &= k_l^c (\mathbf{W}_l' \mathbf{u}_l + \mathbf{b}_l'), \quad \mathbf{u}_0 = \mathbf{x}, \\ \mathbf{z}_{l+1} &= g_l^c \left(\mathbf{W}_l^z (\mathbf{z}_l \odot [\mathbf{W}_l^{zu} \mathbf{u}_l + \mathbf{b}_l^{zu}]_+) + \mathbf{W}_l^y (\mathbf{y} \odot [\mathbf{W}_l^{yu} \mathbf{u}_l + \mathbf{b}_l^{yu}]) + \mathbf{W}_l^u \mathbf{u}_l + \mathbf{b}_l^z \right), \\ \mathbf{z}_1 &= g_0^c \left(\mathbf{W}_0^y (\mathbf{y} \odot [\mathbf{W}_0^{yu} \mathbf{x} + \mathbf{b}_0^{yu}]) + \mathbf{W}_0^u \mathbf{x} + \mathbf{b}_0^z \right), \end{aligned} \quad (6)$$

where $\odot$ denotes the Hadamard product and $[\bullet]_+$ the non-negative part, typically implemented using the ReLU activation function. To ensure that each component of the final output layer $\mathbf{z}_C$ is convex in $\mathbf{y}$, the weights $\mathbf{W}^z$ must satisfy $\mathbf{W}^z \in \mathbb{R}_{\geq 0}$, while the activation functions g_l^c are required to be *convex* and *non-decreasing*. The remaining weights $\mathbf{W}$ and biases $\mathbf{b}$ are unconstrained, and the activations k_l^c are not restricted.

In the context of finite strain inelasticity, we employ ICNNs to approximate both the Helmholtz free energy and the dual potential. Since these functions are zero-valued at the origin, an adjustment to the architecture is necessary. Moreover, if the input $\mathbf{y}$ is already passed through a convex function f , i.e. $(\mathbf{x}, f(\mathbf{y})) \mapsto \mathcal{N}_c(\mathbf{x}, \mathbf{y})$, then also $\mathbf{W}^y$ must satisfy $\mathbf{W}^y \in \mathbb{R}_{\geq 0}$. To incorporate these aspects, we slightly modify the architecture (6), yielding

$$\begin{aligned} \mathbf{u}_{l+1} &= \tilde{k}_l^c (\mathbf{W}_l' \mathbf{u}_l + \mathbf{b}_l'), \quad \mathbf{u}_0 = \mathbf{x}, \\ \mathbf{z}_{l+1} &= \tilde{g}_l^c \left(\mathbf{W}_l^z (\mathbf{z}_l \odot [\mathbf{W}_l^{zu} \mathbf{u}_l + \mathbf{b}_l^{zu}]_+) + \mathbf{W}_l^y (\mathbf{y} \odot [\mathbf{W}_l^{yu} \mathbf{u}_l + \mathbf{b}_l^{yu}]_+) + \mathbf{b}_l^c \right), \\ \mathbf{z}_1 &= \tilde{g}_0^c \left(\mathbf{W}_0^y (\mathbf{y} \odot [\mathbf{W}_0^{yu} \mathbf{u}_0 + \mathbf{b}_0^{yu}]_+) + \mathbf{b}_0^c \right), \end{aligned} \quad (7)$$

with $\mathbf{b}_l^c := \mathbf{W}_l^u \mathbf{u}_l + \mathbf{b}_l^z$, and where $\mathbf{W}^z, \mathbf{W}^y \in \mathbb{R}_{\geq 0}$. The modified activations

$$\tilde{k}_l^c := k_l^c((\bullet)_l) - k_l^c(\mathbf{b}_l'), \quad \tilde{g}_l^c := g_l^c((\bullet)_l) - g_l^c(\mathbf{b}_l^c), \quad (8)$$

ensure that $\mathbf{u}_L = \mathbf{0}$ if $\mathbf{x} = \mathbf{0}$ and $\mathbf{z}_C = \mathbf{0}$ if $\mathbf{y} = \mathbf{0}$. This guarantees zero-valuedness of the scalar functions to be approximated. Note that we made the architectural choice that $\mathbf{z}_C = \mathbf{0}$ if $\mathbf{y} = \mathbf{0}$ holds independently of the value of $\mathbf{x}$. Lastly, it is worth noting that although the constraint $\mathbf{W}^y \in \mathbb{R}_{\geq 0}$ appears to limit the expressive power of ICNNs, architectures of the form (6) can still be represented by augmenting the input as $(\mathbf{x}, \mathbf{y}, -\mathbf{y}) \mapsto \mathcal{N}_c(\mathbf{x}, \mathbf{y})$.

Input Monotonic Neural Networks. Next, we propose a new architecture, termed *Input Monotonic Neural Networks* (IMNNs), in analogy to ICNNs. These networks are designed to produce outputs that are monotonically increasing with respect to a specific subset of inputs $\mathbf{s}$, while remaining unconstrained in other inputs $\mathbf{v}$. Formally, IMNNs implement mappings $(\mathbf{v}, \mathbf{s}) \mapsto \mathcal{N}_m(\mathbf{v}, \mathbf{s})$. The architecture again employs two parallel branches:

$$\begin{aligned} \mathbf{r}_{l+1} &= \tilde{k}_l^m (\mathbf{W}_l^* \mathbf{r}_l + \mathbf{b}_l^*), \quad \mathbf{r}_0 = \mathbf{v}, \\ \mathbf{p}_{l+1} &= \tilde{g}_l^m \left(\mathbf{W}_l^p (\mathbf{p}_l \odot [\mathbf{W}_l^{pr} \mathbf{r}_l + \mathbf{b}_l^{pr}]_+) + \mathbf{W}_l^s (\mathbf{s} \odot [\mathbf{W}_l^{sr} \mathbf{r}_l + \mathbf{b}_l^{sr}]_+) + \mathbf{b}_l^m \right), \\ \mathbf{p}_1 &= \tilde{g}_0^m \left(\mathbf{W}_0^w (\mathbf{s} \odot [\mathbf{W}_0^{sr} \mathbf{r}_0 + \mathbf{b}_0^{sr}]_+) + \mathbf{b}_0^m \right), \end{aligned} \quad (9)$$

where $\mathbf{b}_l^m := \mathbf{W}_l^r \mathbf{r}_l + \mathbf{b}_l^p$. If $\tilde{g}_l^m$ are monotonically increasing activation functions and $\mathbf{W}^p, \mathbf{W}^s \in \mathbb{R}_{\geq 0}$, then each output component of $\mathbf{p}_M$ is guaranteed to be monotonically increasing in each component of $\mathbf{s}$; see Appendix A.1. As before, shifted activations analogous to Equation (8) are used to enforce zero-valuedness. A monotonically decreasing output can be obtained by simply replacing $\mathbf{s}$ with $-\mathbf{s}$.

Noteworthy, a three-layer architecture with min-max pooling, which enforces monotonicity, was first proposed in [63] and later extended in [64] to partially input specific monotonic networks. More recently, [65] combined three types of layers—calibrators, linear embeddings, and lattices—within a deep lattice framework to enforce monotonicity.

Composition of ICNNs and IMNNs. Section 2 and Figure 1 demonstrated that the dual potential φ can be expressed as a composition of convex and monotonic functions. Accordingly, we construct a neural network architecture $\mathcal{N}_o$ by composing ICNNs and IMNNs:

$$\mathcal{N}_o(\mathbf{x}, \mathbf{y}) = \mathcal{N}_m(\mathbf{u}_{C-1}, \mathbf{z}_C), \quad \text{with } (\mathbf{u}_{C-1}, \mathbf{z}_C) = \mathcal{N}_c(\mathbf{x}, \mathbf{y}). \quad (10)$$

Here, $\mathbf{u}_{C-1}$ is chosen instead of $\mathbf{u}_C$, since $\mathbf{u}_C$ is typically not computed in ICNNs as it does not influence the output $\mathbf{z}_C$. Figure 2 schematically illustrates the composition.

Liquid Neural Networks. Unlike purely elastic materials, inelasticity requires not only an additional dual potential but also the solution of evolution equations for the internal variables. Explicit time discretization schemes are computationally efficient but often unstable, while implicit schemes are stable but computationally costly and thus impractical for neural network training, which typically requires many epochs. An alternative, explored in [33] for feed-forward networks and long-short-term memory (LSTM) networks, is to approximate the update of the internal variables using auxiliary networks [66].

In this work, we employ *Liquid Neural Networks* (LiNNs) [67], a class of recurrent networks that update their hidden state $\mathbf{h}$ using dynamics inspired by differential equations of continuous systems. This formulation naturally reflects the temporal evolution typical for inelastic internal variables. In continuous time, the update reads

$$\frac{d\mathbf{h}(t)}{dt} = -\boldsymbol{\alpha}(\mathbf{h}(t), \mathbf{q}(t)) \odot \mathbf{h}(t) + \mathbf{f}(\mathbf{h}(t), \mathbf{q}(t)), \quad (11)$$

with $\alpha_i \geq 0$ and $\mathbf{q}$ denoting an external input. Since $\boldsymbol{\alpha}$ and $\mathbf{f}$ may vary with time, these networks are called *liquid*.

For numerical implementation, Equation (11) can be discretized within $t \in [t_n, t_{n+1}]$, for example by a (semi-)explicit Euler scheme

$$\mathbf{h}_{n+1} = (\mathbf{1} - \Delta t \boldsymbol{\alpha}(\mathbf{h}_n, \mathbf{q}_{n+1})) \odot \mathbf{h}_n + \Delta t \mathbf{f}(\mathbf{h}_n, \mathbf{q}_{n+1}), \quad (12)$$

or by a (semi-)explicit exponential time integration scheme [68]

$$\mathbf{h}_{n+1} = \exp(-\Delta t \boldsymbol{\alpha}(\mathbf{h}_n, \mathbf{q}_{n+1})) \odot \mathbf{h}_n + \Delta t \xi(-\Delta t \boldsymbol{\alpha}(\mathbf{h}_n, \mathbf{q}_{n+1})) \odot \mathbf{f}(\mathbf{h}_n, \mathbf{q}_{n+1}), \quad \xi(x) = \frac{\exp(x) - 1}{x} \quad (13)$$

where $\Delta t := t_{n+1} - t_n$.

In the neural network formulation, both $\boldsymbol{\alpha}$ and $\mathbf{f}$ are learned by independent feed-forward networks, i.e. $(\mathbf{h}_n, \mathbf{q}_{n+1}) \mapsto \mathcal{N}_\alpha(\mathbf{h}_n, \mathbf{q}_{n+1})$ and $(\mathbf{h}_n, \mathbf{q}_{n+1}) \mapsto \mathcal{N}_f(\mathbf{h}_n, \mathbf{q}_{n+1})$. There are no constraints on weights, biases, or activation functions, except that we again apply shifted activations (analogous to Equation (8)) to ensure that the state remains unchanged if both the previous state and external input are zero. Further, the last activation function of $\mathcal{N}_\alpha$ should ensure the non-negativity of $\boldsymbol{\alpha}$. Although both discretization schemes are viable, we observed that the (semi-)explicit Euler method yields superior stability for the problems considered in this work. Accordingly, we restrict our investigation to this scheme.

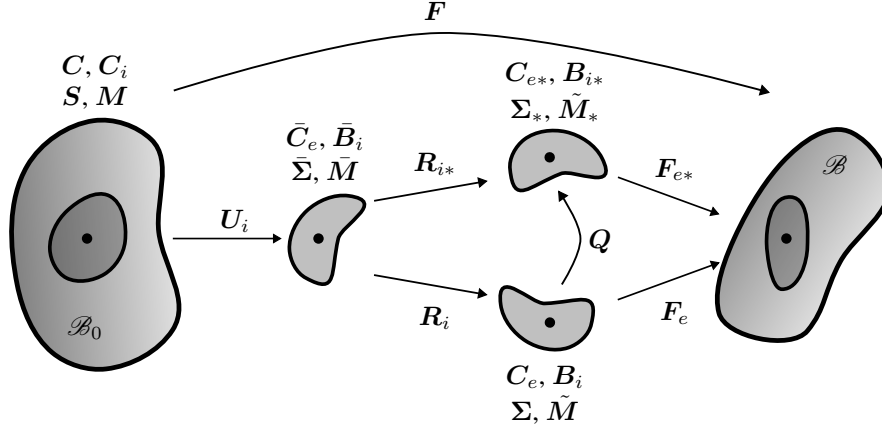


Figure 3: Motion from the reference configuration $\mathcal{B}_0$ to the current configuration $\mathcal{B}$. An infinitesimal material element is described by the deformation gradient $\mathbf{F}$, which admits a multiplicative decomposition into elastic $\mathbf{F}_e$ and inelastic $\mathbf{F}_i$ parts. Due to the non-uniqueness, one may equivalently write $\mathbf{F} = \mathbf{F}_{e*} \mathbf{F}_{i*}$ with $\mathbf{F}_{e*} = \mathbf{F}_e \mathbf{Q}^T$ and $\mathbf{F}_{i*} = \mathbf{Q} \mathbf{F}_i$ with $\mathbf{Q} \in \text{SO}(3)$. From the polar decompositions of $\mathbf{F}_i$ and $\mathbf{F}_{i*}$, it follows that $\mathbf{R}_i^* = \mathbf{Q} \mathbf{R}_i$. Noteworthy, the shown quantities referred to one intermediate configuration share their eigenvalues with the corresponding quantities in any other intermediate configuration.

4 Fundamentals of Constitutive Relations

After introducing the neural network architectures in the previous section, which will later serve as functional representations of φ and ψ , we now turn to the foundations of constitutive modeling for inelastic materials at finite strains. Our objective is to formulate a framework for anisotropic materials that captures the essential kinematics and satisfies the dissipation inequality.

Noteworthy, the term *anisotropy* is not unambiguous in inelasticity. Unlike in elasticity, where anisotropy does not necessarily result in coaxiality of the stress–strain response, inelasticity exhibits additional sources of anisotropic behavior. For example, kinematic hardening or the coupling between elastic and inelastic strains can both induce effective anisotropy. In this context, one may further distinguish between *initial* anisotropy (present in the undeformed material, e.g., due to texture or fibers) and *induced* anisotropy (evolving during inelastic deformation). Here, we use anisotropy exclusively to denote the presence of preferred directions—either in the energy (e.g., due to embedded fibers) or in the dual potential (e.g., Hill’s plasticity). The latter case is closely related to what is commonly referred to as *distortional hardening*.

In this contribution, we limit ourselves to the class of initial anisotropy that can be characterized by several line or outward normal vectors $\mathbf{n}^i$ perpendicular to a plane. For brevity, we derive the equations for one family of vectors $\mathbf{n}$ (transversal isotropy), however, the extension to more such classes of anisotropy is straightforward.

To this end, we will first introduce the kinematics of the multiplicative decomposition and the material principles that the Helmholtz free energy must satisfy. We will then derive the evolution equation for isotropic materials based on a dual potential and extend this concept to anisotropic materials. The scalar functions, namely the energy and the dual potential, will be discovered by the neural networks introduced in the previous section as the manuscript progresses.

Kinematics. Let $\chi : \mathcal{B}_0 \subset \mathbb{R}^3 \rightarrow \mathcal{B}$ denote the motion that maps the reference configuration $\mathcal{B}_0$ onto the current configuration $\mathcal{B}$ in Euclidean space. The deformation is characterized by the deformation gradient

$$\mathbf{F} = \text{Grad } \chi, \quad J := \det \mathbf{F}.$$

To ensure an admissible, orientation-preserving motion one requires $J > 0$ in $\mathcal{B}_0$, such that $\mathbf{F} \in \text{GL}^+(3, \mathbb{R})$. In the finite strain regime, inelastic effects are commonly described by postulating a multiplicative decomposition³ of the

³As shown in [69], the multiplicative decomposition is superior to the additive decomposition in case of finite elasto-plasticity regardless of the elastic material law.

deformation gradient [70, 71, 72, 73, 74], irrespective of the specific inelastic mechanism,

$$\mathbf{F} = \mathbf{F}_e \mathbf{F}_i = \mathbf{F}_{e*} \mathbf{F}_{i*}. \quad (14)$$

Here, $\mathbf{F}_e$ and $\mathbf{F}_i$ denote the elastic and inelastic parts, respectively. Constitutive admissibility requires $J_e := \det \mathbf{F}_e > 0$ and $J_i := \det \mathbf{F}_i > 0$. The decomposition is not unique, since any rotation $\mathbf{Q} \in \text{SO}(3)$ yields an equivalent representation with $\mathbf{F}_{e*} = \mathbf{F}_e \mathbf{Q}^T$ and $\mathbf{F}_{i*} = \mathbf{Q} \mathbf{F}_i$ (see, e.g., [75]). Since $\mathbf{F}_e, \mathbf{F}_i \in \text{GL}^+(3, \mathbb{R})$, both admit polar decompositions, $\mathbf{F}_e = \mathbf{R}_e \mathbf{U}_e$ and $\mathbf{F}_i = \mathbf{R}_i \mathbf{U}_i$, with $\mathbf{R}_e, \mathbf{R}_i \in \text{SO}(3)$ and $\mathbf{U}_e, \mathbf{U}_i \in \text{Sym}_+^{3 \times 3}$. It follows that the elastic right stretch tensor in the rotated configuration, $\mathbf{U}_{e*} = \mathbf{Q} \mathbf{U}_e \mathbf{Q}^T$, is orthogonally similar to $\mathbf{U}_e$, while the inelastic stretch remains unchanged, $\mathbf{U}_{i*} = \mathbf{U}_i$. Figure 3 shows the various mappings involved.

Material principles. For anisotropic materials, we introduce a symmetry class $\mathcal{S} \subset \text{O}(3)$ consisting of all orthogonal tensors $\mathbf{Q}$ relative to which the material response remains invariant. In this work, following the structural-tensor approach (cf. [76, 77]), we focus on a single preferred direction $\mathbf{n} \in \mathbb{S}^2$ in the reference configuration and set the structural tensor $\mathbf{M} = \mathbf{n} \otimes \mathbf{n}$. In analogy to [78], we introduce the group $\mathcal{G} = \{\mathbf{Q} \in \text{SO}(3) : \mathbf{Q}\mathbf{n} = \mathbf{n}\}$, i.e., the set of proper rotations leaving $\mathbf{n}$ invariant. As mentioned at the beginning, we set $\mathcal{S} = \mathcal{G}$, and thus, the material is invariant about $\mathbf{n}$. For isotropic materials, in contrast, the symmetry class is $\mathcal{S} = \text{SO}(3)$ (often $\text{O}(3)$ is admitted as well).

We derive the constitutive equations in the sense of hyperelasticity and postulate the existence of a Helmholtz free energy

$$\psi : \text{GL}^+(3, \mathbb{R}) \times \text{GL}^+(3, \mathbb{R}) \times \mathcal{M} \rightarrow \mathbb{R}, \quad (\mathbf{F}, \mathbf{F}_i, \mathbf{M}) \mapsto \psi(\mathbf{F}, \mathbf{F}_i, \mathbf{M}),$$

where $\mathcal{M} = \{\mathbf{M} = \mathbf{n} \otimes \mathbf{n} : \mathbf{n} \in \mathbb{S}^2\}$. For inelastic materials with multiplicative split $\mathbf{F} = \mathbf{F}_e \mathbf{F}_i$, the free energy is required to satisfy the following principles:

- Objectivity (frame indifference)

$$\psi(\mathbf{F}, \mathbf{F}_i, \mathbf{M}) = \psi(\mathbf{Q}\mathbf{F}, \mathbf{F}_i, \mathbf{M}) \quad \forall \mathbf{Q} \in \text{SO}(3).$$

- Material symmetry

$$\psi(\mathbf{F}, \mathbf{F}_i, \mathbf{M}) = \psi(\mathbf{F}\mathbf{Q}^T, \mathbf{F}_i\mathbf{Q}^T, \mathbf{Q}\mathbf{M}\mathbf{Q}^T) \quad \forall \mathbf{Q} \in \mathcal{G}.$$

- Indifference to the choice of intermediate configuration

$$\psi(\mathbf{F}, \mathbf{F}_i, \mathbf{M}) = \psi(\mathbf{F}, \mathbf{Q}\mathbf{F}_i, \mathbf{M}) \quad \forall \mathbf{Q} \in \text{SO}(3).$$

Consequently, ψ can be expressed as a scalar-valued isotropic function in terms of the right Cauchy–Green tensors $\mathbf{C} = \mathbf{F}^T \mathbf{F}$ and $\mathbf{C}_i = \mathbf{F}_i^T \mathbf{F}_i$ and the structural tensor $\mathbf{M}$

$$\psi(\mathbf{F}, \mathbf{F}_i, \mathbf{M}) = \psi(\mathbf{C}, \mathbf{C}_i, \mathbf{M}) = \psi(\mathbf{Q}\mathbf{C}\mathbf{Q}^T, \mathbf{Q}\mathbf{C}_i\mathbf{Q}^T, \mathbf{Q}\mathbf{M}\mathbf{Q}^T) \quad \forall \mathbf{Q} \in \text{SO}(3). \quad (15)$$

Further, we impose volumetric growth (coercivity) conditions

$$\psi \rightarrow \infty \quad \text{as} \quad J \searrow 0 \text{ or } J \rightarrow \infty, \quad (16)$$

$$\psi \rightarrow \infty \quad \text{as} \quad J_e \searrow 0 \text{ or } J_e \rightarrow \infty. \quad (17)$$

In addition, non-essential conditions are employed: the stress is usually normalized such that it vanishes for $\mathbf{F}\mathbf{F}_i^{-1} = \mathbf{I}$, and we set the reference level $\psi(\mathbf{I}, \mathbf{I}, \mathbf{M}) = 0$.

To ensure existence of minimizers, we design ψ to be polyconvex in $\mathbf{F}$ [79, 80], i.e., there exists a function $W = W(\mathbf{F}, \text{cof } \mathbf{F}, J; \mathbf{F}_i, \mathbf{M})$ that is convex in $(\mathbf{F}, \text{cof } \mathbf{F}, J)$ (with $\mathbf{F}_i$ and $\mathbf{M}$ acting as parameters) and satisfies

$$\psi(\mathbf{F}, \mathbf{F}_i, \mathbf{M}) = W(\mathbf{F}, \text{cof } \mathbf{F}, J; \mathbf{F}_i, \mathbf{M}).$$

For background on polyconvex hyperelastic formulations in terms of the deformation gradient for anisotropic solids we refer the reader to [78]; see also [81] for classical symmetry principles and [82] for a discussion of isomorphism concepts in inelastic anisotropy.

Thermodynamic consistency for isotropic materials. Considering Equation (15), we evaluate the Clausius–Planck inequality $-\dot{\psi} + \frac{1}{2} \mathbf{S} : \dot{\mathbf{C}} \geq 0$, where $\mathbf{S}$ denotes the second Piola–Kirchhoff stress tensor. Following the arguments of [83, 84, 85], this yields the state law $\mathbf{S} = 2 \partial_{\mathbf{C}} \psi$ as well as the reduced dissipation inequality

$$\mathcal{D} := \Sigma : \mathbf{D}_i \geq 0, \quad \Sigma := -2 \mathbf{F}_i \frac{\partial \psi}{\partial \mathbf{C}_i} \mathbf{F}_i^T, \quad (18)$$

where D_i denotes the symmetric part of $\dot{F}_i F_i^{-1}$ and is related to C_i by $\dot{C}_i = 2 F_i^T D_i F_i$. As discussed in [86], Σ shares the same eigenvalues as the Kirchhoff stress $F S F^T$ if ψ depends solely on the elastic part of the deformation⁴. Note that neither D_i nor Σ is invariant under the choice of the intermediate configuration; indeed, $D_{i*} = Q D_i Q^T$, $\Sigma_* = Q \Sigma Q^T$ for any $Q \in \text{SO}(3)$. However, the resulting dissipation is invariant with respect to the intermediate configuration, i.e., $\Sigma_* : D_{i*} = \Sigma : D_i$.

We postulate the existence of a dual potential [88]

$$\varphi : \text{Sym}^{3 \times 3} \longrightarrow \mathbb{R}, \quad \Sigma \mapsto \varphi(\Sigma).$$

Similar to the Helmholtz free energy, the dual potential must satisfy the principles of objectivity, material symmetry, and invariance with respect to the choice of intermediate configuration. The first two properties are automatically satisfied, while the latter implies that φ can be expressed as a scalar-valued isotropic function. We then postulate that the evolution equation follows from the dual potential

$$D_i \in \partial_{\Sigma} \varphi, \quad (19)$$

which in general is nonsmooth [89].

As shown, for instance, in [36], it suffices to formulate a potential ω that is convex, non-negative, and vanishes at the origin in terms of the invariants $\mathcal{S}_{\varphi}(\Sigma) := (S_1, S_2, S_3)$ with $S_2 := \sqrt{S'_2}$ and $S_3 := \sqrt[3]{S'_3}$, where⁵

$$S_1 := \text{tr } \Sigma, \quad S'_2 := \frac{1}{2} \text{tr}((\text{dev } \Sigma)^2), \quad S'_3 := \frac{1}{3} \text{tr}((\text{dev } \Sigma)^3), \quad (20)$$

and by setting $\varphi = \omega$; see Appendix A.3 for details⁶.

As outlined in Section 2 and Figure 1, this concept can be further generalized by introducing a monotonically increasing function ζ and defining

$$\varphi(\mathcal{S}_{\varphi}) = (\zeta \circ \omega)(\mathcal{S}_{\varphi}).$$

Since thermodynamic consistency already holds for $\varphi = \omega$, it is preserved under this composition provided that ζ is monotonically increasing; cf. Section 2 or see Appendix A.3.

Extension to anisotropic materials. We now extend the methodology developed for inelastic isotropic materials to the anisotropic case. To this end, we postulate a dual potential

$$\varphi : \text{Sym}^{3 \times 3} \times \mathbb{M}^{3 \times 3} \longrightarrow \mathbb{R}, \quad (\Sigma, \tilde{M}) \mapsto \varphi(\Sigma, \tilde{M}),$$

where $\tilde{M} = F_i \star M$ denotes an appropriate push-forward of M onto the intermediate configuration such that $Q F_i \star M = Q \tilde{M} Q^T \quad \forall Q \in \text{SO}(3)$, see [90]. The push-forward must be chosen such that φ remains invariant under the choice of the intermediate configuration.

Let us introduce a set of extended invariants

$$\mathcal{T}_{\varphi}(\Sigma, \tilde{M}) := (\mathcal{S}_{\varphi}(\Sigma), \mathcal{A}_{\varphi}(\Sigma, \tilde{M})),$$

where $\mathcal{A}_{\varphi}(\Sigma, \tilde{M}) := (A_1(\Sigma, \tilde{M}), \dots, A_m(\Sigma, \tilde{M}))$ accounts for anisotropy. Similarly to $\mathcal{S}_{\varphi}$, we assume that each anisotropic invariant A_k satisfies

$$\frac{\partial A_k}{\partial \Sigma} : \Sigma = A_k, \quad k = 1, \dots, m. \quad (21)$$

Thermodynamic consistency can then be established by the same arguments as in the isotropic case for

$$\varphi(\mathcal{T}_{\varphi}) = (\zeta \circ \omega)(\mathcal{T}_{\varphi}). \quad (22)$$

A suitable choice of invariants $\mathcal{A}_{\varphi}$ will be introduced in Section 4.2; see Appendix A.3 for the corresponding proof of consistency.

4.1 Helmholtz free energy

In the previous section, we have introduced several principles the Helmholtz free energy must respect. As a consequence, we express it in terms of invariants, which we now define. Inelastic materials are usually described relative to the

⁴Alternatively, one may derive the reduced dissipation inequality with respect to $\dot{C}_i$ and its conjugated force (see [87]). However, this approach loses the physical interpretation of Σ , which is considered a disadvantage here

⁵See Appendix A.2 for a note on the computation of roots.

⁶If φ is convex in Σ , the dual potential can be interpreted as the Legendre transform of the dissipation potential known from Generalized Standard Materials [27]; see, e.g., [36].

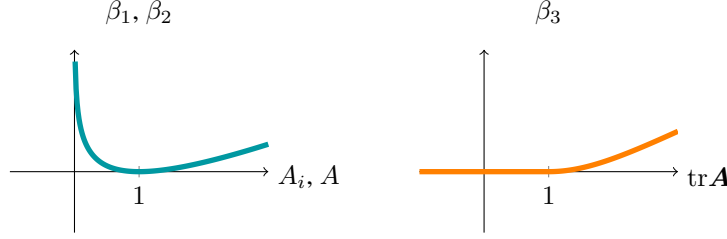


Figure 4: Illustration of the normalization functions for the isotropic $\mathcal{S}_\psi$ and anisotropic $\mathcal{A}_\psi$ sets for the Helmholtz free energy ψ .

intermediate configuration through elastic and inelastic invariants. To include anisotropy, the structural tensor is pushed to the intermediate configuration by an admissible mapping $\mathbf{F}_i \star \mathbf{M}$, as discussed previously. The structural vector $\tilde{\mathbf{n}}$ in the intermediate configuration is obtained via a covariant or contravariant mapping [90]

$$\tilde{\mathbf{n}} \in \left\{ \frac{\mathbf{F}_i \mathbf{n}}{\|\mathbf{F}_i \mathbf{n}\|}, \frac{\text{cof} \mathbf{F}_i \mathbf{n}}{\|\text{cof} \mathbf{F}_i \mathbf{n}\|} \right\} =: \{\tilde{\mathbf{n}}_1, \tilde{\mathbf{n}}_2\}, \quad (23)$$

which we normalize, since we do not attribute any physical meaning with the vector's length [91, 92]. Consequently, two structural second-order tensors follow

$$\tilde{\mathbf{M}} \in \{\tilde{\mathbf{n}}_1 \otimes \tilde{\mathbf{n}}_1, \tilde{\mathbf{n}}_2 \otimes \tilde{\mathbf{n}}_2\} =: \{\tilde{\mathbf{M}}_1, \tilde{\mathbf{M}}_2\}.$$

With the elastic right Cauchy–Green tensor $\mathbf{C}_e = \mathbf{F}_e^T \mathbf{F}_e$ and the inelastic left Cauchy–Green tensor $\mathbf{B}_i = \mathbf{F}_i^T \mathbf{F}_i$, the following *similarity* holds

$$\mathbf{C}_e^k \mathbf{B}_i^r \tilde{\mathbf{M}}_1^s \tilde{\mathbf{M}}_2^t = \mathbf{F}_i^{-T} (\mathbf{C} \mathbf{C}_i^{-1})^k \mathbf{C}_i^r \left(\frac{\mathbf{C}_i \mathbf{M}}{\mathbf{C}_i : \mathbf{M}} \right)^s \left(\frac{\mathbf{M} \mathbf{C}_i^{-1}}{\mathbf{C}_i^{-1} : \mathbf{M}} \right)^t \mathbf{F}_i^T, \quad k, r \in \mathbb{Z}, \quad s, t \in \mathbb{N}_0. \quad (24)$$

Thus every invariant relative to the intermediate configuration can be expressed in referential quantities while remaining independent of the choice of intermediate configuration. Since $\tilde{\mathbf{M}}_1$ and $\tilde{\mathbf{M}}_2$ are idempotent, we may restrict $s, t \in \{0, 1\}$ without loss.⁷

Using (24), we define the isotropic set $\mathcal{S}_\psi$ and the anisotropic set $\mathcal{A}_\psi$. In line with [93], the isotropic irreducible basis in the intermediate configuration is

$$\mathbf{C}_e, \text{cof} \mathbf{C}_e, J_e, \mathbf{B}_i, \text{cof} \mathbf{B}_i, J_i, \mathbf{C}_e \mathbf{B}_i, \text{cof} \mathbf{C}_e \text{cof} \mathbf{B}_i, \text{cof} \mathbf{C}_e \mathbf{B}_i, \mathbf{C}_e \text{cof} \mathbf{B}_i,$$

which maps to the reference configuration as

$$\mathbf{C} \mathbf{C}_i^{-1}, \frac{1}{J_i^2} \mathbf{C}_i \text{cof} \mathbf{C}, \frac{J}{J_i}, \mathbf{C}_i, \text{cof} \mathbf{C}_i, J, \mathbf{C}, \text{cof} \mathbf{C}, \frac{1}{J_i^2} \mathbf{C}_i \text{cof} \mathbf{C} \mathbf{C}_i, J_i^2 \mathbf{C} \mathbf{C}_i^{-2},$$

where J_i has been replaced by J to directly account for the volumetric growth condition (16). The trace of each second-order argument is convex in either $\mathbf{F}$ or $\text{cof} \mathbf{F}$, as they are of the form $f(\mathbf{F}, \mathbf{G}) = \text{tr}(\mathbf{F}^T \mathbf{F} \mathbf{G})$ (or analogously with $\text{cof} \mathbf{F}$) with $\mathbf{G} \succeq 0$; see Appendix A.4. In addition, the scalar invariants J/J_i and J are convex in J . Hence, the entire irreducible basis is polyconvex and defines $\mathcal{S}_\psi$. Notably, the seventh and eighth mixed invariants of elastic and inelastic arguments simplify to total deformation, which gives them a clear physical interpretation.

For the anisotropic set $\mathcal{A}_\psi$, we restrict ourselves to the mapping $\tilde{\mathbf{M}}_1$ for simplicity, though both can be included (e.g. letting the neural network select the more appropriate one). In line with [93], we introduce

$$\mathbf{C}_e \tilde{\mathbf{M}}_1, \text{cof} \mathbf{C}_e \tilde{\mathbf{M}}_1, \mathbf{B}_i \tilde{\mathbf{M}}_1, \text{cof} \mathbf{B}_i \tilde{\mathbf{M}}_1, \mathbf{C}_e (\mathbf{B}_i \# \tilde{\mathbf{M}}_1),$$

which map to

$$\frac{\mathbf{C} \mathbf{M}}{\mathbf{C}_i : \mathbf{M}}, \frac{1}{J_i^2} \mathbf{C}_i \text{cof} \mathbf{C} \frac{\mathbf{C}_i \mathbf{M}}{\mathbf{C}_i : \mathbf{M}}, \frac{\mathbf{C}_i^2 \mathbf{M}}{\mathbf{C}_i : \mathbf{M}}, \frac{J_i^2 \mathbf{M}}{\mathbf{C}_i : \mathbf{M}}, \mathbf{C} \mathbf{C}_i^{-1} \left(\mathbf{C}_i \# \frac{\mathbf{M} \mathbf{C}_i}{\mathbf{C}_i : \mathbf{M}} \right),$$

with $\#$ the outer product of second-order tensors⁸; see Appendix A.4. This unusual choice replaces the classical invariant $\mathbf{C}_e \mathbf{B}_i \tilde{\mathbf{M}}_1$, which is generally not polyconvex.

⁷Mixed-variant mappings are not discussed in this contribution; see Appendix A.4.

⁸For materials with two distinct in-plane preferred directions, the outer product of the corresponding structural tensors can be used to compute the structural tensor being normal to the plane they span; cf. [92].

To normalize the free energy and stresses, we introduce the convex functions

$$\beta_1(\mathbf{A}) := \text{tr} \mathbf{A} - 3 - \ln \det \mathbf{A}, \quad \beta_2(A) := A - 1 - \ln A, \quad \beta_3(\mathbf{A}) := \ln \cosh(\max(\text{tr} \mathbf{A} - 1, 0)).$$

For β_1 , one has $\beta_1 = \sum_{i=1}^3 (A_i - 1 - \ln A_i)$ if $\mathbf{A} \in \text{Sym}_+^{3 \times 3}$ with the eigenvalues A_i . All β_j are nonnegative, convex, and vanish at the normalized state ($\mathbf{A} = \mathbf{I}$ for β_1 , $A = 1$ for β_2 , and $\text{tr} \mathbf{A} = 1$ for β_3). Further, β_3 is non-decreasing. Figure 4 illustrates the normalization functions β_1 , β_2 , and β_3 .

We then define the isotropic invariant set

$$\begin{aligned} S_1 &= \beta_1(\mathbf{C}\mathbf{C}_i^{-1}), & S_2 &= \beta_1\left(\frac{1}{J_i^2} \mathbf{C}_i \text{cof} \mathbf{C}\right), & S_3 &= \beta_2\left(\frac{J}{J_i}\right), & S_4 &= \beta_1(\mathbf{C}_i), \\ S_5 &= \beta_1(\text{cof} \mathbf{C}_i), & S_6 &= \beta_2(J), & S_7 &= \beta_1(\mathbf{C}), & S_8 &= \beta_1(\text{cof} \mathbf{C}), \\ S_9 &= \beta_3\left(\frac{1}{J_i^2} \mathbf{C}_i \text{cof} \mathbf{C} \mathbf{C}_i\right), & S_{10} &= \beta_3(J_i^2 \mathbf{C}\mathbf{C}_i^{-2}). \end{aligned} \quad (25)$$

Each S_i is polyconvex⁹.

Analogously, the anisotropic set is

$$\begin{aligned} A_1 &= \beta_1\left(\frac{\mathbf{C}\mathbf{M}}{\mathbf{C}_i:\mathbf{M}}\right), & A_2 &= \beta_3\left(\frac{1}{J_i^2} \mathbf{C}_i \text{cof} \mathbf{C} \frac{\mathbf{C}_i\mathbf{M}}{\mathbf{C}_i:\mathbf{M}}\right), & A_3 &= \beta_3\left(\frac{\mathbf{C}_i^2\mathbf{M}}{\mathbf{C}_i:\mathbf{M}}\right), \\ A_4 &= \beta_3\left(\frac{J_i^2\mathbf{M}}{\mathbf{C}_i:\mathbf{M}}\right), & A_5 &= \beta_3\left(\mathbf{C}\mathbf{C}_i^{-1}\left(\mathbf{C}_i \# \frac{\mathbf{M}\mathbf{C}_i}{\mathbf{C}_i:\mathbf{M}}\right)\right). \end{aligned} \quad (26)$$

Again, each A_i is polyconvex.

Finally, the Helmholtz free energy is expressed as

$$\psi = \psi(\mathcal{T}_\psi), \quad \mathcal{T}_\psi(\mathbf{C}, \mathbf{C}_i, \mathbf{M}) := (S_\psi(\mathbf{C}, \mathbf{C}_i), \mathcal{A}_\psi(\mathbf{C}, \mathbf{C}_i, \mathbf{M})). \quad (27)$$

We follow the energy normalization condition, but slightly relax the stress normalization condition. That is, we allow for individual normalization conditions by choosing the β_i functions. However, if mixed invariants involving both elastic and inelastic parts are not employed, our choices for β_i are consistent with the classical assumption of an elastic, stress-free state.

4.2 Dual potential

We introduce the set $\mathcal{T}_\varphi$ (22) as a basis for describing the inelastic evolution of solids. Similarly to the Helmholtz free energy, the set is formulated in terms of invariants. In contrast to ψ , which must be polyconvex in $(\mathbf{F}, \text{cof} \mathbf{F}, J)$ (rather than merely in its invariants), thermodynamic consistency of the dual potential does not require convexity of the chosen invariants. This reflects the different mathematical roles of the two potentials; see [94, 95] for theoretical discussions of non-convex potentials and [96] for experimental evidence of non-convex yield surfaces in prestressed concrete.

The isotropic invariants S_φ are given in Equation (20). For the anisotropic contribution, we permit both push-forwards of the structural vector in (23). Following [93], this leads to

$$\begin{aligned} A_{11} &= \text{dev} \boldsymbol{\Sigma} : \tilde{\mathbf{M}}_1, & A_{12} &= \text{dev} \boldsymbol{\Sigma} : \tilde{\mathbf{M}}_2, & A_{21} &= \sqrt{\frac{1}{2} (\text{dev} \boldsymbol{\Sigma})^2 : \tilde{\mathbf{M}}_1}, \\ A_{22} &= \sqrt{\frac{1}{2} (\text{dev} \boldsymbol{\Sigma})^2 : \tilde{\mathbf{M}}_2}, & A_3 &= \boldsymbol{\Sigma} : \text{sym}(\tilde{\mathbf{n}}_1 \otimes \tilde{\mathbf{n}}_2), & A_4 &= \sqrt{\frac{1}{2} \boldsymbol{\Sigma}^2 : \text{sym}(\tilde{\mathbf{n}}_1 \otimes \tilde{\mathbf{n}}_2)}, \end{aligned} \quad (28)$$

and we set $\mathcal{A}_\varphi := (A_{11}, A_{12}, A_{21}, A_{22}, A_3, A_4)$. All these invariants are positively 1-homogeneous in $\boldsymbol{\Sigma}$.

As discussed earlier, the Mandel stress $\boldsymbol{\Sigma}$ shares its eigenvalues with the Kirchhoff stress under suitable conditions, while the Cauchy stress admits a clearer physical interpretation (being related by the determinant J). Using the positive 1-homogeneity of $\mathcal{T}_\varphi$ in $\boldsymbol{\Sigma}$, the evolution equation (19) can be written as

$$\mathbf{D}_i \in \partial_{\boldsymbol{\Sigma}} \varphi\left(\frac{\mathcal{T}_\varphi}{J}\right) = \frac{1}{J} \left[\partial_{\mathbf{z}} \varphi(\mathbf{z}) \right]_{\mathbf{z}=\mathcal{T}_\varphi/J} : \partial_{\boldsymbol{\Sigma}} \mathcal{T}_\varphi, \quad (29)$$

where the gradient of φ is evaluated at $\mathcal{T}_\varphi/J$. Thermodynamic consistency is preserved, because for each component T_k of $\mathcal{T}_\varphi$,

$$(\partial_{\boldsymbol{\Sigma}} T_k) : \boldsymbol{\Sigma} = T_k, \quad k = 1, \dots, 9$$

hence,

$$\frac{1}{J} \langle \partial_{\boldsymbol{\Sigma}} \mathcal{T}_\varphi, \boldsymbol{\Sigma} \rangle = \frac{\mathcal{T}_\varphi}{J}.$$

⁹For $f(J) = -\ln(aJ^b) = -\ln a - b \ln J$ with $J > 0$ and $a > 0$, one has $f''(J) = b/J^2$. Since $J^2 > 0$, the function is convex iff $b \geq 0$ (strictly convex for $b > 0$). The invariant $f(\mathbf{C}, \mathbf{G}) = \text{tr}(\mathbf{C}\mathbf{G})$ is polyconvex if $\mathbf{G} \succeq 0$, same holds when $\mathbf{C}$ is replaced by its cofactor; see Appendix A.4.

Co-rotated intermediate configuration. Direct computation of Σ , $\tilde{n}_1$, and $\tilde{n}_2$ would require the inelastic part F_i of the deformation gradient, which is non-unique. One may either establish a mapping from intermediate to referential invariants (analogous to Equation (24)) or adopt a co-rotated intermediate configuration [97] (similarly exploited in [98]) that maps all non-unique quantities to unique counterparts via (cf. Figure 3)

$$(\bar{\bullet}) := R_i^T \star (\bullet) = R_{i*}^T \star (\bullet)_*,$$

where R_i and R_{i*} denote the rotational parts associated with F_i and F_{i*} , respectively. This yields

$$\bar{\Sigma} = -2 U_i \frac{\partial \psi}{\partial C_i} U_i, \quad \bar{D}_i = \text{sym}(\dot{U}_i U_i^{-1}) = \frac{1}{2} U_i^{-1} \dot{C}_i U_i^{-1}, \quad \bar{n}_1 = \frac{U_i n}{\|U_i n\|}, \quad \bar{n}_2 = \frac{\text{cof} U_i n}{\|\text{cof} U_i n\|}, \quad (30)$$

An attractive feature of this representation is that the functional form of constitutive relations—such as the invariant set—remains unchanged under the transformation.

5 Constitutive Neural Network for Anisotropic Inelasticity

Building on the constitutive fundamentals in Section 4, the invariant sets $\mathcal{T}_\psi$ and $\mathcal{T}_\varphi$ introduced in Sections 4.1 and 4.2, and the network architectures of Section 3, we now combine neural networks with the constitutive framework of anisotropic inelasticity. Our aim is to discover the Helmholtz free energy and the dual potential in a thermodynamically consistent manner that respects the material principles. In fact, the approach to discovering the Helmholtz free energy is closely aligned with the Physics-Augmented Neural Networks (PANNs) framework [13, 99, 22]. Furthermore, we seek to replace the solution of the implicit time discretization scheme by a recurrent architecture, in analogy to [66, 33]. While the small strain regime is investigated in [33] and the finite strain regime is studied in [66], both assume an additive decomposition of the total strain and internal variables.

Helmholtz free energy. The Helmholtz free energy is represented by an Input Convex Neural Network (ICNN) that is polyconvex with respect to $(F, \text{cof} F, J)$. However, we pass $\mathcal{T}_\psi$ to the ICNN, which already provides a convex representation of $(F, \text{cof} F, J)$. This motivates the slight adaptation of the ICNN architecture in Equation (7) compared to the original (6). To enforce the volumetric growth conditions (16)–(17), we augment the network with a convex penalty contribution ψ_{gr} that prevents the discovery of spurious solutions without volumetric dependence. Employing the strictly convex function $\beta_2(A) = A - 1 - \ln A$ the energy takes the form

$$\psi = \mathcal{N}_c(\bullet, \mathcal{T}_\psi) + \psi_{\text{gr}}(J, \frac{J}{J_i}), \quad (31)$$

with

$$\psi_{\text{gr}}(J, \frac{J}{J_i}) = \lambda_{\text{gr}} (\beta_2(J) + \beta_2(\frac{J}{J_i})), \quad (32)$$

where λ_{gr} is a penalty parameter, fixed to 10^{-4} in all simulations. In Equation (31), $\bullet$ denotes the dependence of ψ on the non-convex quantities.

Dual potential. For the dual potential, we employ the composition of ICNNs and Input Monotonic Neural Networks proposed in (10):

$$\varphi = \mathcal{N}_o(\bullet, (\mathcal{T}_\varphi, -\mathcal{T}_\varphi)), \quad (33)$$

where $\bullet$ again denotes the dependence on non-convex quantities. Unlike the free energy, φ must only be convex with respect to the invariants themselves, not with respect to their arguments, i.e., the thermodynamic driving force Σ . To recover the structure of the original ICNN (6), we pass both $\mathcal{T}_\varphi$ and $-\mathcal{T}_\varphi$ through the convex branch of the network. To guarantee non-negativity of the ICNN output within the composition, the final activation function is chosen as the ReLU. Due to the invariants' definition, the zero-value constraint is automatically satisfied.

Time discretization. In the inelastic setting, the evolution problem must be discretized in time $t \in [t_n, t_{n+1}]$. For the finite strain regime, the exponential integrator scheme is well-suited [100], as it consistently preserves the underlying kinematics of the evolution of U_i [40]. The *explicit* integration step reads

$$C_{i_{n+1}} = U_{i_n} \exp(2 \Delta t D_{i_n}) U_{i_n}, \quad U_{i_{n+1}} = \sqrt{C_{i_{n+1}}}, \quad (34)$$

with $\Delta t := t_{n+1} - t_n$. The corresponding *implicit* scheme requires solving

$$\mathbf{r} := C_{i_n} - U_{i_{n+1}} \exp(-2 \Delta t D_{i_{n+1}}) U_{i_{n+1}} \stackrel{!}{=} \mathbf{0}, \quad (35)$$

where $D_{i_{n+1}}$ depends nonlinearly on $U_{i_{n+1}}$.

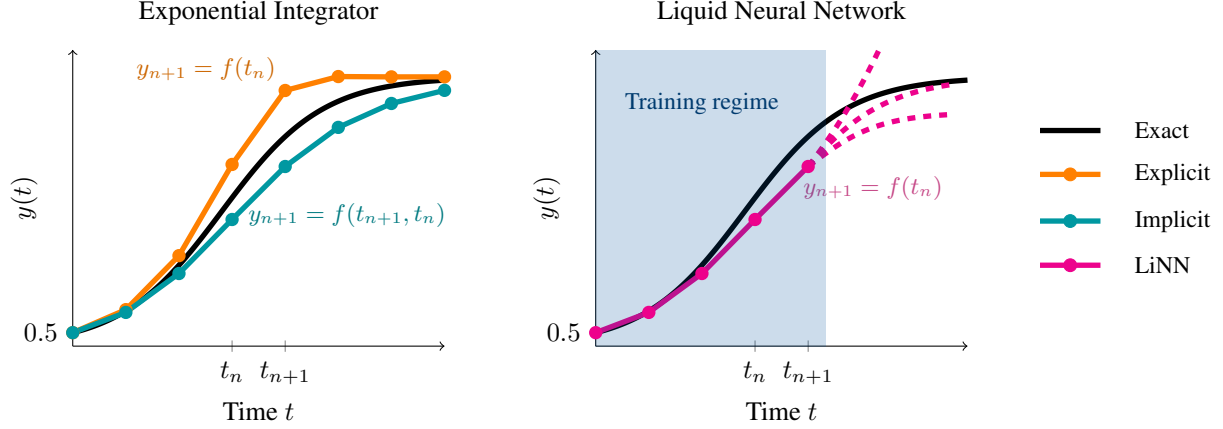


Figure 5: Schematic of time discretizations of the ordinary differential equation $\dot{y} = A(y)y$ with $A(y) = r(1 - y/k)$, where $r = 0.7$ and $k = 10$. The exponential integrator with time step $\Delta t = 1.5$ shows the explicit scheme $y_{n+1} = \exp(\Delta t A(y_n)) y_n$ and the implicit scheme $y_{n+1} = \exp(\Delta t A(y_{n+1})) y_n$, the latter must be solved numerically. The Liquid Neural Network (LiNN) approximates y_{n+1} by minimizing the loss $\mathcal{L} = (y_{n+1} - \exp(\Delta t A(y_{n+1})) y_n)^2$. Within the training domain, the LiNN attains accuracy comparable to the implicit scheme while remaining explicit; outside this domain, predictive reliability is not ensured.

Following [66, 33], we predict the internal variable by an auxiliary neural network during training. This facilitates stable training of the physically relevant networks for ψ and φ . We adopt an architecture based on LiNNs (12); Figure 5 illustrates the different time discretization schemes based on the explicit and implicit exponential integrators as well as the LiNN. The hidden states $\mathbf{h} \in \mathbb{R}^6$ are initialized as $\mathbf{h}_0 = \mathbf{0}$. We then learn $\alpha = \mathcal{N}_\alpha$ and $\mathbf{f} = \mathcal{N}_f$ via

$$(\mathbf{h}_n, (\mathbf{E}_{n+1}, \mathbf{D}_i^{\text{trial}})) \mapsto \mathcal{N}_\alpha(\mathbf{h}_n, (\mathbf{E}_{n+1}, \mathbf{D}_i^{\text{trial}})), \quad (\mathbf{h}_n, (\mathbf{E}_{n+1}, \mathbf{D}_i^{\text{trial}})) \mapsto \mathcal{N}_f(\mathbf{h}_n, (\mathbf{E}_{n+1}, \mathbf{D}_i^{\text{trial}})), \quad (36)$$

where $\mathbf{E}_{n+1}$ denotes the Green–Lagrange strain tensor and $\mathbf{D}_i^{\text{trial}}$ is the trial value of $\mathbf{D}_i$, i.e., $\mathbf{D}_i$ evaluated with the current $\mathbf{C}_{n+1}$ but freezing $\mathbf{U}_{i,n+1} = \mathbf{U}_{i,n}$. This choice follows the well-established trial step procedure in plasticity, ensuring that the input is close to the target output, which—to our experience—in practice stabilizes training. A systematic study of this effect is beyond the current scope.

Evaluating $\mathcal{N}_\alpha$ and $\mathcal{N}_f$ in Equation (36) allows us to update the hidden state according to Equation (12). Since these states do not generally coincide with the independent components of $\mathbf{U}_i \in \text{Sym}_+^{3 \times 3}$, positive definiteness must be enforced. For this purpose, we employ the Cholesky decomposition of $\mathbf{U}_i$ and map $\mathbf{h}$ to the lower-triangular factor of $\mathbf{U}_i$; see Appendix A.2.

Loss function. During training, the weights and biases $\theta := \{\mathbf{W}_\psi, \mathbf{b}_\psi, \mathbf{W}_\varphi, \mathbf{b}_\varphi, \mathbf{W}_h, \mathbf{b}_h\}$ of the three networks for ψ , φ , and the hidden states $\mathbf{h}$ are determined by solving the minimization problem

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\theta), \quad (37)$$

where the total loss is additively split into¹⁰

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{stress}}(\theta) + \lambda_{\text{evo}} \mathcal{L}_{\text{evo}}(\theta). \quad (38)$$

Here, $\mathcal{L}_{\text{stress}}$ measures the discrepancy between predicted and experimentally observed stresses, while $\mathcal{L}_{\text{evo}}$ quantifies the residual of the implicit evolution equation (35). The penalty parameter λ_{evo} , accounting for the relative magnitudes of the two losses, is heuristically set to 1000. Both contributions are evaluated via mean squared error. For the stress contribution we obtain

$$\mathcal{L}_{\text{stress}}(\theta) = \frac{1}{BTO} \sum_{b=1}^B \sum_{t=1}^T \sum_{o=1}^O \left(\underline{s}_{b,t,o}(\theta) - \hat{\underline{s}}_{b,t,o} \right)^2, \quad (39)$$

with the number of batches B , time steps T , independent stress components $O = 6$, the predicted normalized second Piola–Kirchhoff stress components $\underline{s}(\theta)$, and the normalized experimental stresses $\hat{\underline{s}}$. Normalization is performed by

¹⁰A regularization term can be added if desired.

dividing the raw experimental stress $\hat{\mathbf{S}}$ by the absolute maximum $\underline{S}$ across all batches and time steps, i.e., $\underline{\hat{\mathbf{S}}} = \hat{\mathbf{S}}/\underline{S}$. Thus, after training we have to multiply $\underline{\hat{\mathbf{S}}}$ by $\underline{S}$ to obtain $\mathbf{S}$. Analogously, the evolution loss is defined as

$$\mathcal{L}_{\text{evo}}(\boldsymbol{\theta}) = \frac{1}{BTO} \sum_{b=1}^B \sum_{t=1}^T \sum_{o=1}^O (r_{b,t,o}(\boldsymbol{\theta}))^2, \quad (40)$$

where $r(\boldsymbol{\theta})$ denotes the independent components of the residual of the implicit evolution equation (35).

6 Numerical results

In this section, we present a comprehensive evaluation of the proposed framework for discovering inelastic material behavior at finite strains. The analysis proceeds in three steps. First, we describe the generation of artificial training data using a classical constitutive model (Section 6.1). Second, we examine the predictive capabilities of the discovered networks at the material point level, both for isotropic and anisotropic materials, and compare them to recurrent neural network baselines that lack physics priors (Section 6.2). Finally, we assess whether the favorable material point performance translates to unseen structural boundary value problems, thereby testing the robustness and generalization capability of the framework (Section 6.3).

For clarity, the same neural architectures are employed across all studies. The network representing the Helmholtz free energy (31) is defined as

$$\mathcal{N}_c : \mathbb{R}^{15} \rightarrow \mathbb{R}, \quad \mathcal{T}_\psi \xrightarrow{\text{exp}} 16 \xrightarrow{\text{Softplus}} 16 \xrightarrow{\text{Softplus}} 16 \xrightarrow{\text{Softplus}} 16 \xrightarrow{\text{Softplus}} \psi,$$

while the network for the dual potential (33) is designed as (cf. Figure 2 for the color code)

$$\mathcal{N}_o : \mathbb{R}^{18} \rightarrow \mathbb{R}, \quad (\mathcal{T}_\varphi, -\mathcal{T}_\varphi) \xrightarrow{\text{exp}} 16 \xrightarrow{\text{ReLU}} 16 \xrightarrow{\text{ReLU}} 16 \xrightarrow{\text{ReLU}} 4 \xrightarrow{\text{tanh}} 16 \xrightarrow{\text{tanh}} 16 \xrightarrow{\text{Softplus}} 16 \xrightarrow{\text{Linear}} \varphi.$$

During training, the internal variables are predicted by an auxiliary Liquid Neural Network (LiNN), which serves to stabilize the evolution dynamics

$$\mathcal{N}_\alpha : \mathbb{R}^{18} \rightarrow \mathbb{R}^6, \quad (\mathbf{h}_n, (\mathbf{E}_{n+1}, \mathbf{D}_i^{\text{trial}})) \xrightarrow{\text{GELU}} 12 \xrightarrow{\text{GELU}} 12 \xrightarrow{\text{GELU}} 8 \xrightarrow{\text{GELU}} 8 \xrightarrow{\text{ReLU}} \boldsymbol{\alpha},$$

and

$$\mathcal{N}_f : \mathbb{R}^{18} \rightarrow \mathbb{R}^6, \quad (\mathbf{h}_n, (\mathbf{E}_{n+1}, \mathbf{D}_i^{\text{trial}})) \xrightarrow{\text{GELU}} 12 \xrightarrow{\text{GELU}} 12 \xrightarrow{\text{GELU}} 8 \xrightarrow{\text{GELU}} 8 \xrightarrow{\text{Linear}} \mathbf{f},$$

where GELU refers to the Gaussian Error Linear Unit activation function [101].

All networks are implemented in Flax on top of JAX [102]. Optimization uses the ADAM algorithm from Optax with a learning rate of 10^{-3} . To avoid instabilities due to excessively large updates, global-norm gradient clipping with a threshold of 10^{-3} is applied, similar to [36]. Finite element meshes are generated in Gmsh [103], and post-processing is performed in ParaView [104].

6.1 Generation of training data

Reference material model. To generate sufficiently rich and controlled data sets, we employ a classical constitutive model in which Helmholtz free energy and dual potential are additively split into isotropic and anisotropic parts

$$\psi = \psi_{\text{iso}} + \psi_{\text{ani}}, \quad \varphi = \varphi_{\text{iso}} + \varphi_{\text{ani}}. \quad (41)$$

For the isotropic study, anisotropic terms are omitted. The functional forms follow [105] for the free energy and [106] for the dual potential; detailed equations are given in Appendix A.5. Rheologically, the model corresponds to a three-element rheological system: an equilibrium spring in parallel with a non-equilibrium spring–dashpot branch. Noteworthy, anisotropy is introduced via a metric tensor rather than a structural tensor, and the dual potential uses a mixed-variant mapping without normalized vector lengths. Thus, the networks must approximate the constitutive response rather than reproducing it analytically.

Material parameters are listed in Table 1. For time integration of the evolution equations, we use the explicit scheme (34). Additionally, in the anisotropic case, the preferred direction is

$$\mathbf{n} = (1/\sqrt{2}, 1/\sqrt{2}, 0)^T.$$

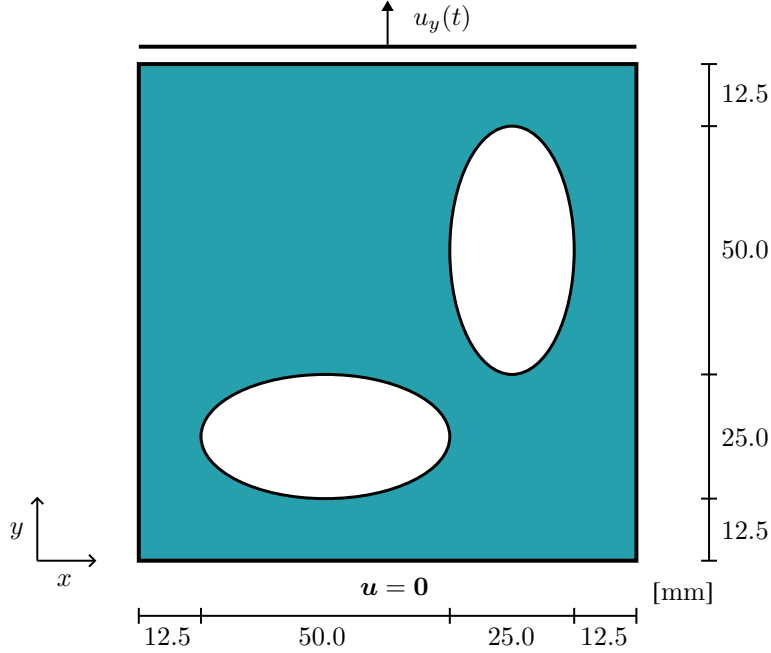


Figure 6: **Plate with two elliptic holes.** Geometry and boundary conditions of the BVP used to generate the training data set, adapted from [23]. The bottom edge is clamped, while on the top edge a vertical displacement $u_y(t)$ is prescribed. The displacements in x and z directions are free on the top, but the z -displacement is fixed at the front and rear surfaces. The plate has a thickness of 6 mm (three hexahedral elements across the thickness); in-plane, 710 elements are used.

Table 1: Material parameters of the reference model used to generate the artificial training data set. The model can be interpreted as an equilibrium branch (eq) in parallel with a non-equilibrium branch (neq).

Isotropic						Anisotropic				Relaxation Time	
eq			neq			eq		neq		β	τ
a	b	c	a	b	c	α	η	α	η		
MPa	MPa	MPa	MPa	MPa	MPa	—	MPa	—	MPa	—	s
80	100	100	40	50	50	2	10	2	10	2	12

Boundary value problem and loading. The simulation setup is illustrated in Figure 6. The plate is discretized with 2130 eight-node hexahedral elements and loaded over 120 time steps of size $\Delta t = 0.05$ s, resulting in a total duration of 6 s. The prescribed displacement $u_y(t)$ consists of three linear loading/unloading phases with two hold phases at constant displacement in between. Representative deformed configurations are shown in Figure 7. The isotropic and anisotropic specimens deform differently under identical loading, and during the hold phase the anisotropic case is more prone to relax due to anisotropic dissipation. At the end of the program, both specimens exhibit residual deformations, characteristic of inelasticity; see Figure 8.

Training data extraction and strategy. Stress–stretch data are extracted from one quadrature point per element. Here, we use only a subset of 1111 out of 2130 elements, yielding 133,320 data pairs (1111 elements with 120 time steps). In contrast to elastic problems, clustering the invariant space is not feasible [107], since the internal variables are not known a priori. Hence, the raw stress–stretch data are employed directly. An approach to cluster also in the case of inelasticity is desirable for real world application but beyond the scope of this contribution.

As reported in the literature [36], direct training on the full dataset proved unstable. We therefore adopt a two-stage procedure: first, pre-train for 500 epochs on a single element and the first 40 time steps; second, continue training with the best parameters from pre-training on the full dataset. This strategy significantly improves stability.

After pre-training, both isotropic and anisotropic networks are trained for 10,000 epochs. The resulting losses are shown in Figure 9. The isotropic model converges to a loss of order 10^{-6} , while the anisotropic one stabilizes around 10^{-5} . The auxiliary loss related to the evolution equation remains below the stress loss but, due to the penalty $\lambda_{\text{evo}} = 1000$, neither

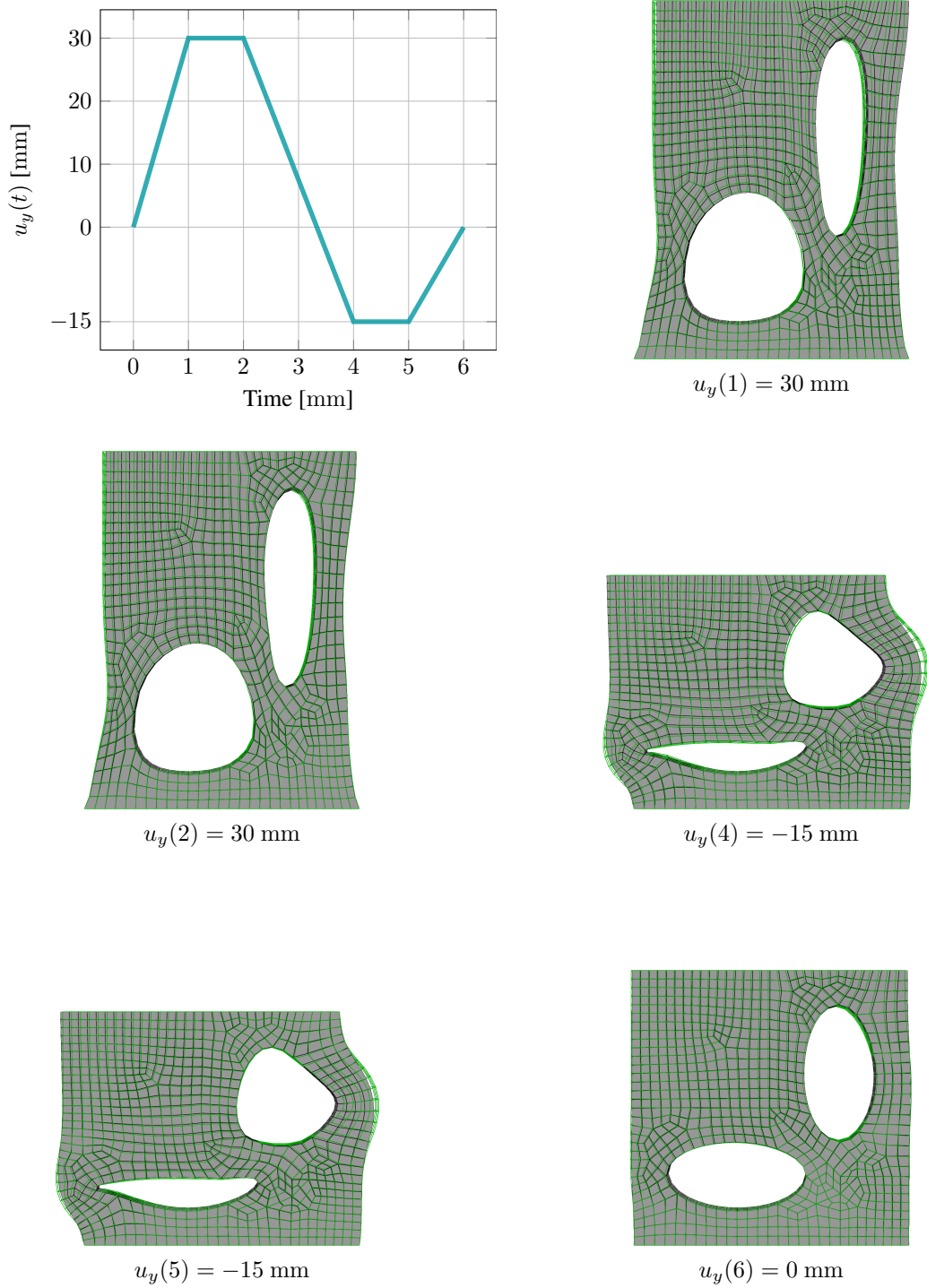


Figure 7: Prescribed loading program and representative deformed configurations for isotropic (green mesh) and anisotropic (gray body/black mesh) materials. The comparison highlights the influence of anisotropy under otherwise identical loading. At the end of the program, although the external displacement returns to its initial value, residual strains remain due to inelasticity and the unloaded configurations are not identical; see also Figure 8.

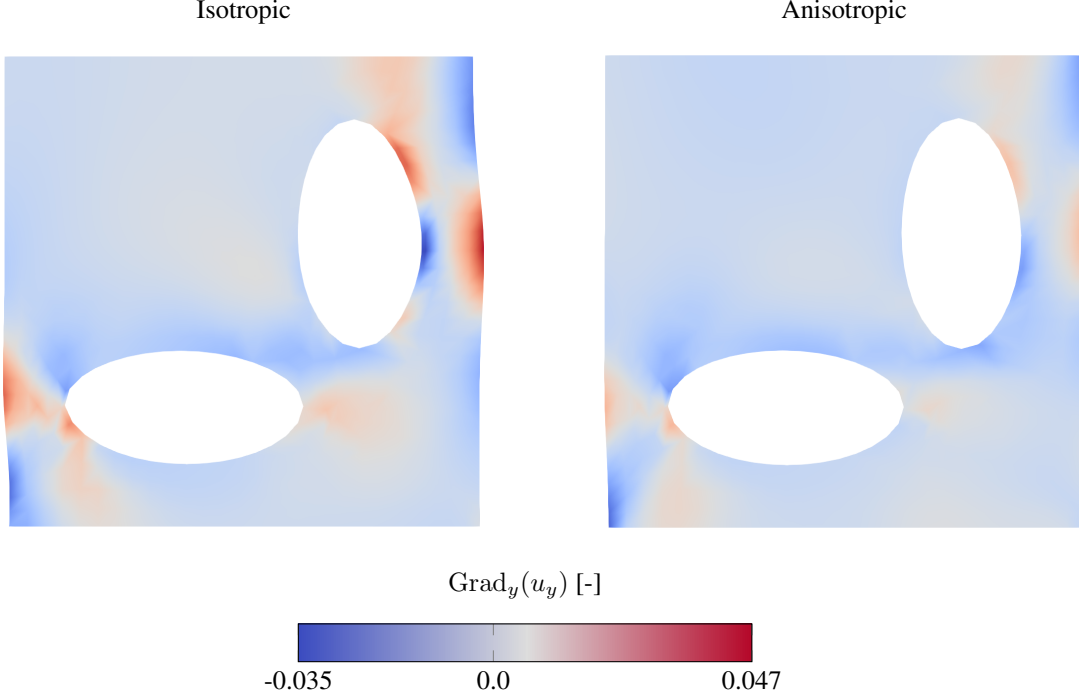


Figure 8: Gradient of the vertical displacement u_y along the y -direction at the end of the loading program with $u_y(6) = 0$ mm. For a purely elastic material, the gradient vanishes throughout the specimen. In contrast, due to inelastic effects, a residual deformation remains for both the isotropic and anisotropic materials.

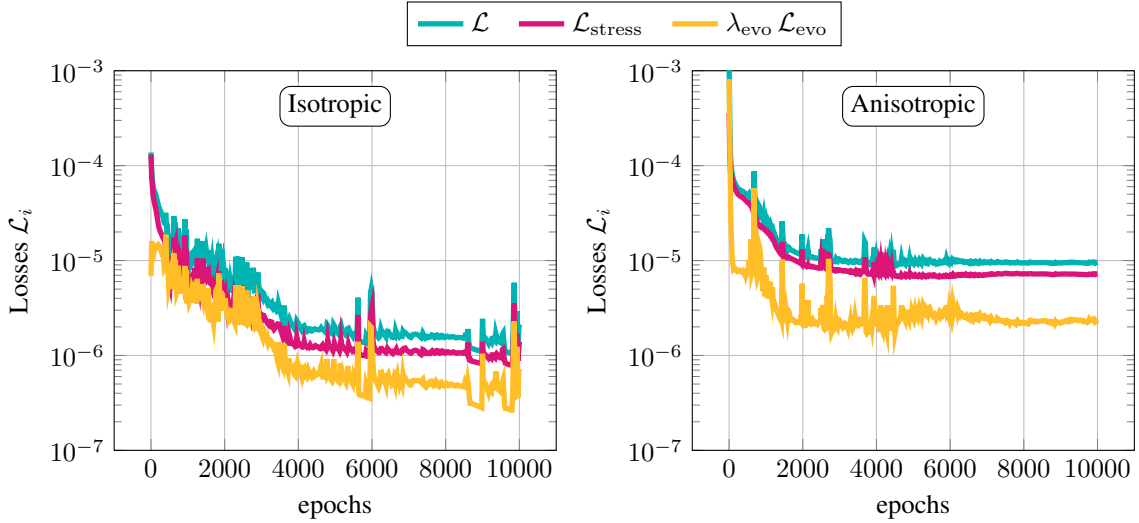


Figure 9: Training losses over 10,000 epochs for isotropic and anisotropic data sets. The penalty parameter is $\lambda_{\text{evo}} = 1000$. No additional regularization is applied beyond gradient clipping.

contribution dominates. Normalization uses maximum stresses $\underline{S} = 1406.429$ MPa (isotropic) and $\underline{S} = 1014.538$ MPa (anisotropic); cf. Equation (39).

6.2 Material point study

We next assess the performance of the discovered networks at the material point level. This study provides a controlled setting in which the predictive accuracy of the constitutive framework can be analyzed in detail, both for isotropic

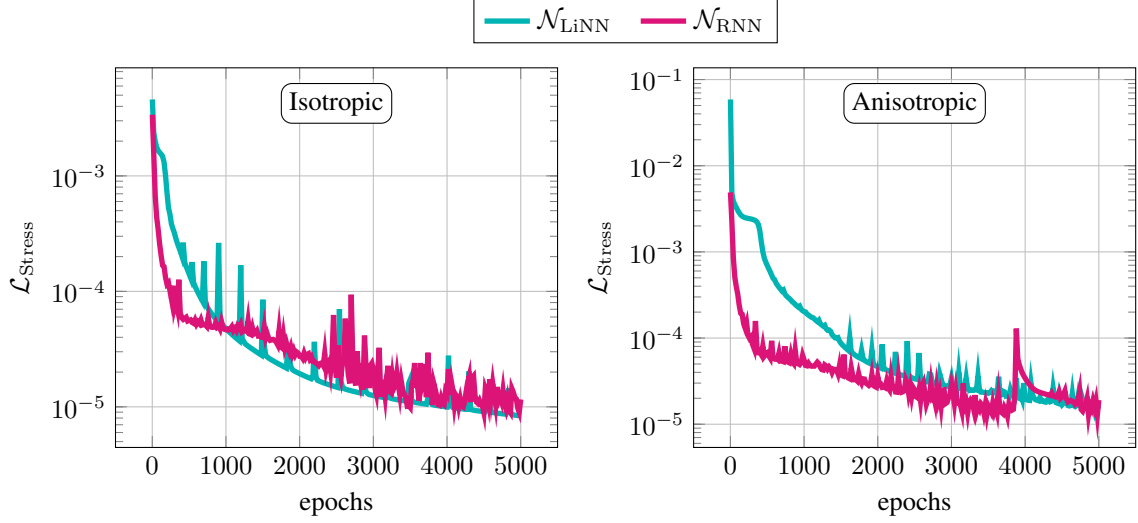


Figure 10: Training losses over the epochs for isotropic and anisotropic data sets for the plain recurrent neural networks. The total number of epochs is 5,000. The losses correspond to Equation (39), which here coincides with the total loss. Neither gradient clipping nor additional regularization is applied.

and anisotropic materials. At the same time, it allows for a systematic comparison with purely data-driven recurrent neural networks that do not incorporate constitutive knowledge, thereby highlighting the effect of embedding physical structure into the network architecture.

To this end, we consider two recurrent baselines. The first is an Elman-type recurrent neural network [108], denoted $\mathcal{N}_{\text{RNN}}$, which was already used in [36] for inelastic Constitutive Neural Networks. The second is a Liquid Neural Network, denoted $\mathcal{N}_{\text{LiNN}}$. Both serve as references representing classical recurrent approaches without physics priors. Our proposed physics-embedded models are denoted $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$, where the inelastic variables are predicted by a LiNN, and $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$, where the evolution is integrated explicitly via the exponential update rule (34). It is important to note that $\mathcal{N}_{\text{LiNN}}$, which directly predicts stresses, must not be confused with the auxiliary LiNN in $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$. Architectural details of the recurrent baselines are provided in Appendix A.6. In all recurrent models, the number of hidden states propagated through time is set to six.

For a fair comparison, the RNN and LiNN baselines are trained until the same loss value is reached after 5,000 epochs under both isotropic and anisotropic conditions. The evolution of the training losses is presented in Figure 10. In each case, the stress discrepancy decreases to magnitudes on the order of 10^{-5} , which is comparable to the performance of the physics-embedded networks shown in Figure 9. As in the previous section, stresses are normalized during preprocessing to ensure balanced training.

6.2.1 Isotropic response

We begin with the isotropic case, i.e. $\psi = \psi_{\text{iso}}$ and $\varphi = \varphi_{\text{iso}}$ according to Equation (41). The analysis proceeds in two steps: first we examine training performance on an element (ID 111) contained in the dataset, and subsequently we test the predictive capability on an element (ID 53) not used during training.

Training. Figure 11 displays the stress components predicted for one representative training element. Both physics-embedded variants, $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$ and $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$, reproduce the reference stresses with very high accuracy. The only noticeable deviation occurs in the out-of-plane shear component S_{23} , where $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$ produces an artificial stress increase. This artifact, however, disappears when the explicit exponential integrator is used. Apart from this effect, the results of both variants are almost indistinguishable, indicating that the auxiliary LiNN mainly stabilizes the update of inelastic variables while having only a marginal influence on the stress prediction itself. This observation is consistent with the findings of [33].

In contrast, the purely recurrent baselines show clear deficiencies. While $\mathcal{N}_{\text{RNN}}$ and $\mathcal{N}_{\text{LiNN}}$ capture the dominant normal stress components qualitatively, their predictions deviate significantly for shear stresses. Notably, oscillations occur in the in-plane shear and in the minor shear components, indicating a lack of robustness when stresses evolve

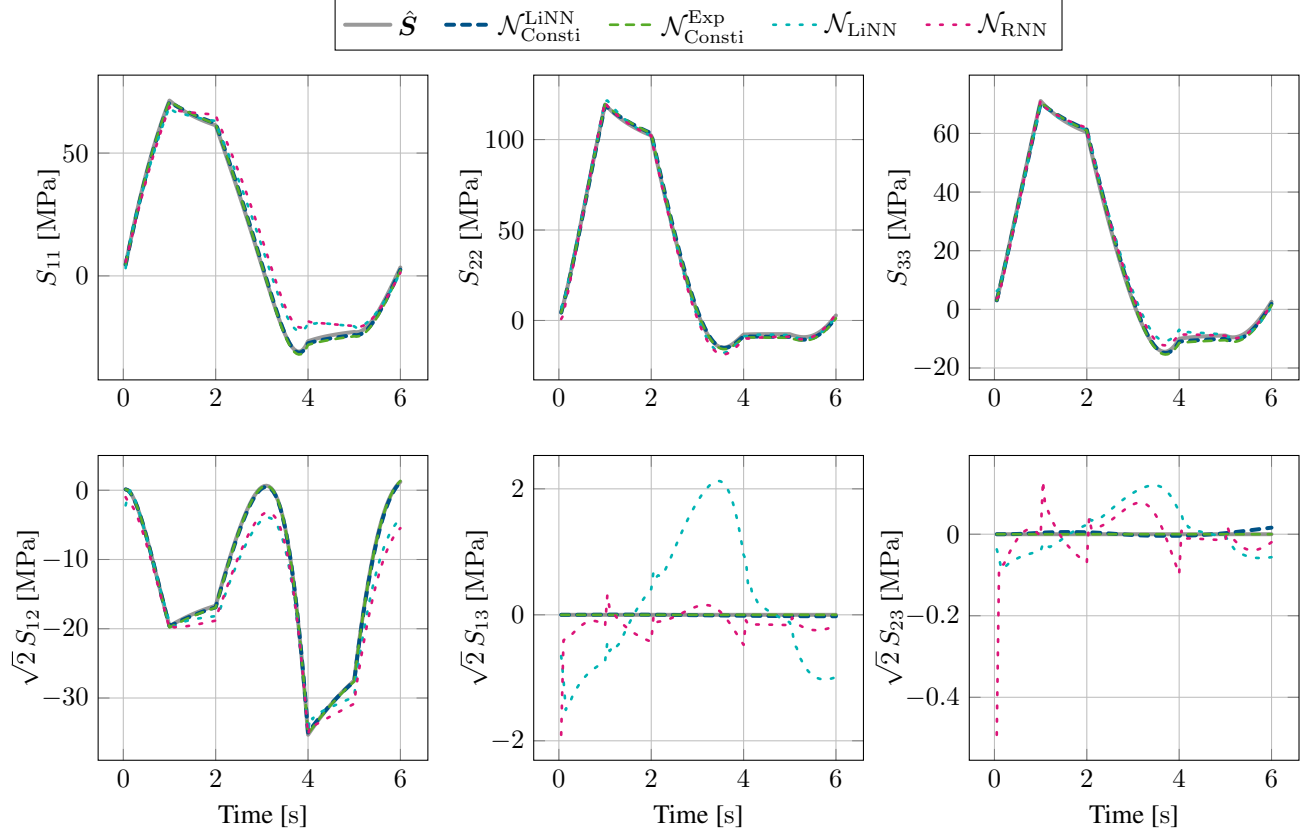


Figure 11: Training results for $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$, $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$, $\mathcal{N}_{\text{RNN}}$, and $\mathcal{N}_{\text{LiNN}}$. The reference model is isotropic, i.e., $\psi = \psi_{\text{iso}}$ and $\varphi = \varphi_{\text{iso}}$ with material parameters given in Table 1. Element ID: 111. Curves start with the first loading step at $t = 0.05$. Of all four networks, only the $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$ network correctly captures the zero shear stress in the off-plane directions.

along more complex trajectories. Although increasing the number of hidden states or employing more complex recurrent architectures may improve performance to some degree, the constitutive networks consistently achieve superior results while guaranteeing physically admissible stress states by construction.

Testing. The generalization capability of the networks is evaluated on an unseen element not included in the training data (element ID 53). The results are presented in Figure 12. Once again, the constitutive networks reproduce the reference stresses with high fidelity. Only small deviations appear, most notably in S_{11} , while the remaining stress components follow the reference solution almost perfectly. The artificial rise in the out-of-plane shear observed for $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$ during training is also present here, but is absent in $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$. The recurrent baselines capture the overall stress evolution but continue to show oscillations in the shear components and less accurate predictions in the in-plane shear. Nevertheless, no severely nonphysical stress–strain responses, such as stress decreasing with increasing strain, were observed in these tests—failure modes that have been reported for plain RNNs in the literature [36, 109]. In this regard, an advantage of more advanced networks is clearly that embedding constitutive structure yields a much more reliable predictive model.

6.2.2 Anisotropic response

We now turn to the anisotropic case, where the free energy and dual potential include anisotropic contributions, $\psi = \psi_{\text{iso}} + \psi_{\text{ani}}$ and $\varphi = \varphi_{\text{iso}} + \varphi_{\text{ani}}$; see Equation (41). As before, we analyze both training and testing performance using the same element IDs as in the isotropic study for direct comparability.

Training. The training results are shown in Figure 13. Here, the constitutive networks again outperform the recurrent baselines across most stress components. However, discrepancies are visible in the in-plane shear S_{12} , which is not

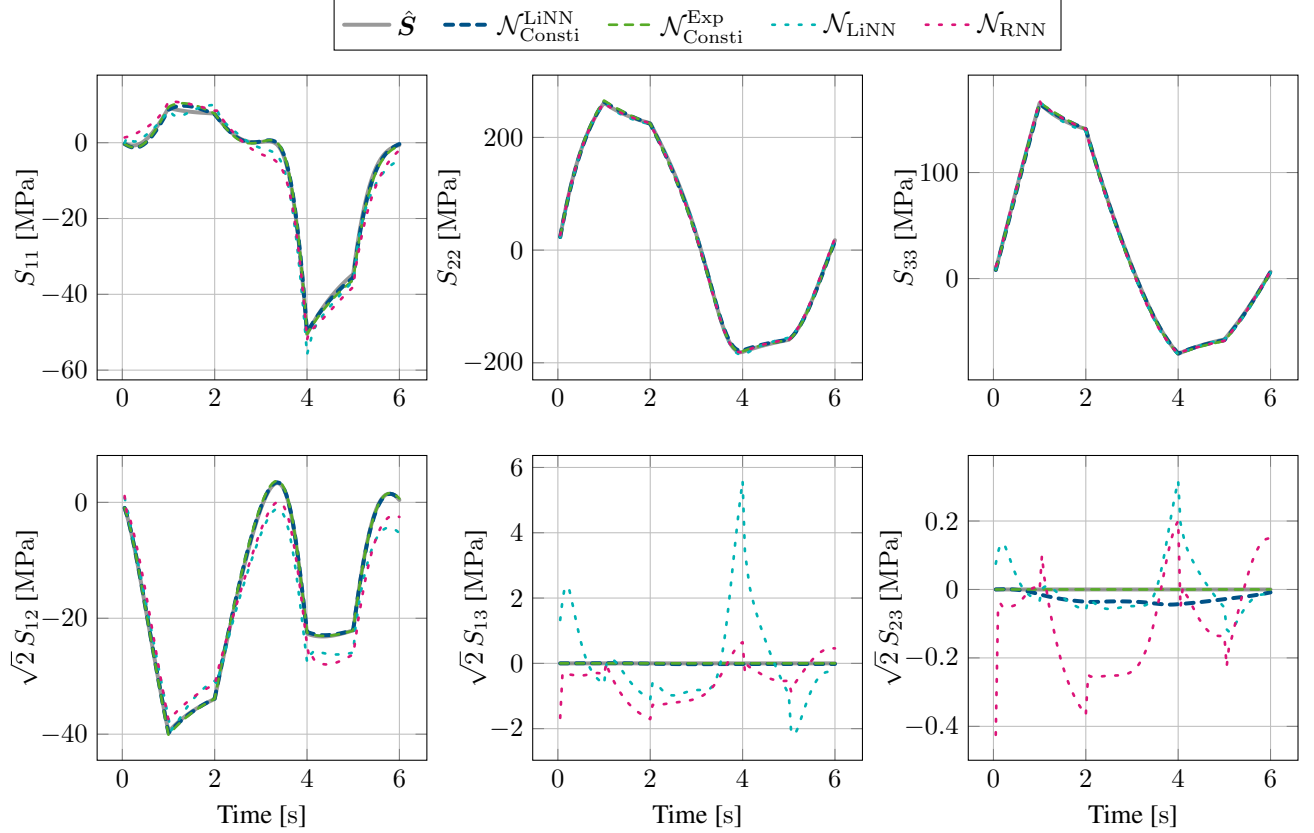


Figure 12: Testing results for $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$, $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$, $\mathcal{N}_{\text{RNN}}$, and $\mathcal{N}_{\text{LiNN}}$. The reference model is isotropic, i.e., $\psi = \psi_{\text{iso}}$ and $\varphi = \varphi_{\text{iso}}$ with material parameters given in Table 1. Element ID: 53. Curves start with the first loading step at $t = 0.05$. Of all four networks, only the $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$ network correctly captures the zero shear stress in the off-plane directions.

captured with the same accuracy as the normal stresses. This reduced accuracy can be attributed to the difference in how anisotropy is modeled: in the reference material, anisotropy is introduced via a metric tensor, whereas the neural networks rely on invariants in terms of a second-order structural tensor in their input representation. In addition, we recognize once more an artificial stress increase when employing $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$, which, however, vanishes with the explicit integration scheme. Despite this mismatch, the physics-embedded approaches retain stability and robustness, while the recurrent baselines suffer from pronounced oscillations, especially in the shear components.

Testing. The testing results for element ID 53 are displayed in Figure 14. The dominant normal stresses, in particular S_{22} and S_{33} , are predicted with high accuracy. For S_{11} , however, systematic deviations are observed across all models. The in-plane shear remains challenging, with none of the models reproducing it fully accurately. Yet, a decisive difference emerges: only the recurrent baselines exhibit oscillations in the remaining shear stresses, whereas the constitutive networks provide smooth and physically consistent responses. Interestingly, $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$ and $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$ deliver almost identical predictions, underlining that the auxiliary LiNN has little influence on accuracy but plays an important role in stabilizing training. Indeed, training $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$ alone frequently diverged to NaN losses, whereas the auxiliary LiNN enabled stable convergence.

In summary, the material point study demonstrates that the proposed physics-embedded networks provide a significant improvement over purely recurrent baselines. They achieve superior accuracy in both isotropic and anisotropic settings, eliminate oscillations in the shear response, and ensure admissible stress states by construction. The auxiliary LiNN stabilizes training, particularly in the anisotropic finite-strain regime, but has only marginal influence on the final stress predictions. These findings establish a solid basis for assessing the networks in more complex structural applications, which will be the focus of the next section.

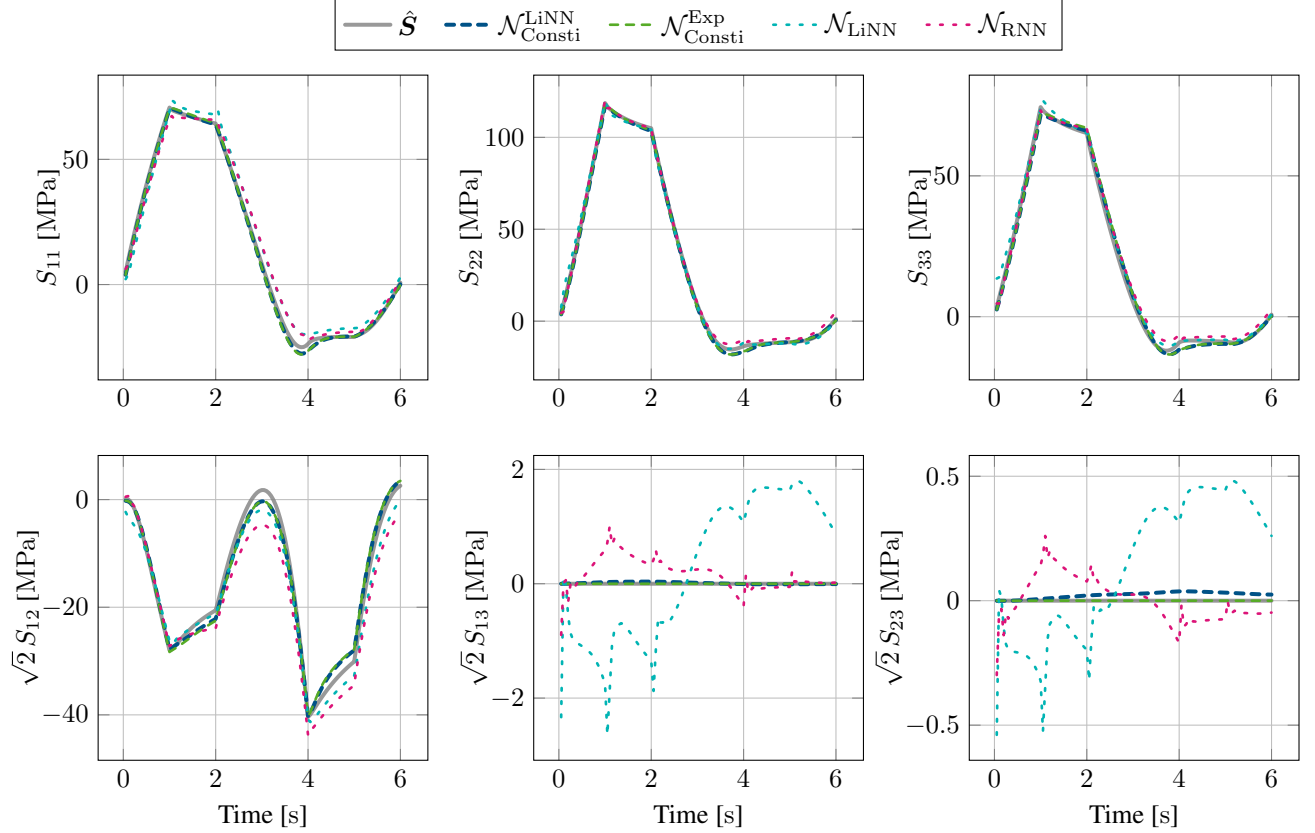


Figure 13: Training results for $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$, $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$, $\mathcal{N}_{\text{RNN}}$, and $\mathcal{N}_{\text{LiNN}}$. The reference model is anisotropic, i.e., $\psi = \psi_{\text{iso}} + \psi_{\text{ani}}$ and $\varphi = \varphi_{\text{iso}} + \varphi_{\text{ani}}$ with parameters given in Table 1. Element ID: 111. Curves start with the first loading step at $t = 0.05$. Of all four networks, only the $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$ network correctly captures the zero shear stress in the off-plane directions.

6.3 Structural level study

Having validated the proposed networks at the material point scale, we now turn to structural boundary value problems (BVPs). This step is essential, because structural simulations combine the responses of many integration points, involve complex stress redistributions, and require a consistent tangent operator to ensure convergence of the finite element method. Thus, structural studies provide a demanding test of whether the advantages observed at the local scale carry over to engineering-scale applications.

We select a variant of the classical Cook’s membrane test [110], here extended to a fully three-dimensional geometry with a central hole (Figure 15). The chosen configuration induces a bending-dominated stress state with strong stress redistributions around the hole, making it a challenging yet well-established benchmark for constitutive models. At the structural level, we focus on the network $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$, where the inelastic evolution is integrated explicitly. This choice simplifies the computation of the consistent material tangent, which can be obtained efficiently via forward-mode automatic differentiation (`jax.jacfwd`), and avoids the need to invoke the implicit function theorem when solving for the inelastic variables implicitly.

6.3.1 Isotropic response

Boundary value problem. To investigate rate dependence and relaxation effects, we consider two linear loading phases of different magnitude, followed by a hold phase at constant displacement. For program R1, the total loading time of 2 s is divided into 40 increments of $\Delta t = 0.05$ s, while for program R2 we use 50 increments of $\Delta t = 0.04$ s. This setup allows us to test whether the network correctly reproduces both the immediate rate effects and the delayed relaxation behavior. Figure 16 show the two load programs applied.

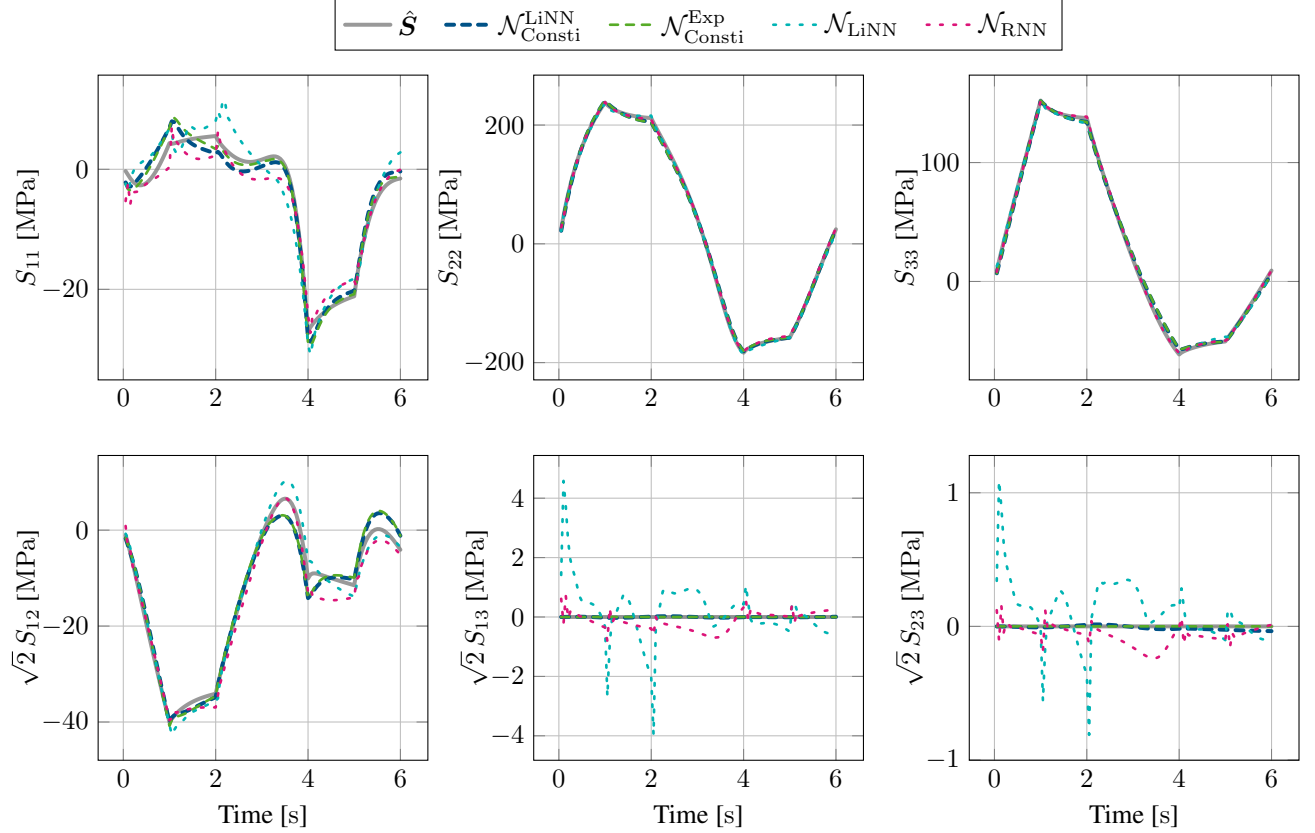


Figure 14: Testing results for $\mathcal{N}_{\text{Consti}}^{\text{LiNN}}$, $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$, $\mathcal{N}_{\text{RNN}}$, and $\mathcal{N}_{\text{LiNN}}$. The reference model is anisotropic, i.e., $\psi = \psi_{\text{iso}} + \psi_{\text{ani}}$ and $\varphi = \varphi_{\text{iso}} + \varphi_{\text{ani}}$ with parameters given in Table 1. Element ID: 53. Curves start with the first loading step at $t = 0.05$. Of all four networks, only the $\mathcal{N}_{\text{Consti}}^{\text{Exp}}$ network correctly captures the zero shear stress in the off-plane directions.

Results. The global reaction forces measured at the right edge are plotted in Figure 16. The constitutive network matches the reference solution very closely. In particular, it reproduces the increased apparent stiffness under the faster loading rate R1, as well as the gradual decrease in reaction force during the hold phase, which arises from inelastic relaxation. The corresponding force–displacement curves confirm this behavior: higher loading rates yield higher reaction forces for the same imposed displacement, and this trend is captured accurately by the network.

A closer examination of the spatial response is given in Figure 17, which presents contour plots of nodal forces and displacement errors for program R1. The nodal force distributions predicted by the network are in good agreement with the reference solution, and the outer feature edges of the deformed geometry coincide almost identical. The Euclidean norm of the displacement error $\|\hat{\mathbf{u}} - \mathbf{u}\|$ remains below 0.012 mm throughout the simulation, which is negligible relative to the prescribed displacement. The small residual errors on the right boundary arise from discrepancies in the unconstrained x - and z -directions. Overall, both global and local measures demonstrate that the discovered network provides an accurate and physically consistent prediction for isotropic materials. These observations are consistently found for the loading rate R2, which is not shown here for brevity.

6.3.2 Anisotropic response

Boundary value problem. We now repeat the Cook’s membrane study for an anisotropic material. To deliberately test generalization beyond the training conditions, the preferred direction is rotated to

$$\mathbf{n} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)^T,$$

thereby coupling all three spatial axes. The same loading programs R1 and R2 are applied, consisting of two linear loading phases followed by a hold phase.

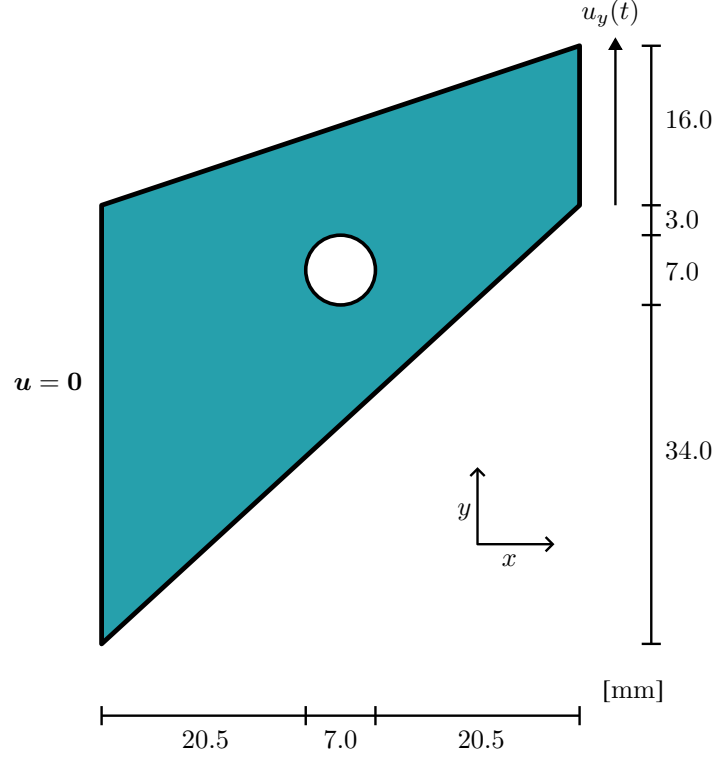


Figure 15: **Cook's membrane with circular hole.** Schematic illustration of the BVP used to evaluate predictive capability at the structural level. The left edge is fully clamped, while on the right edge a vertical displacement $u_y(t)$ is prescribed; other components are left free. A circular hole of diameter 7 mm introduces stress concentrations. The plate has a thickness of 6 mm, discretized with three hexahedral elements across the thickness and 504 elements in-plane.

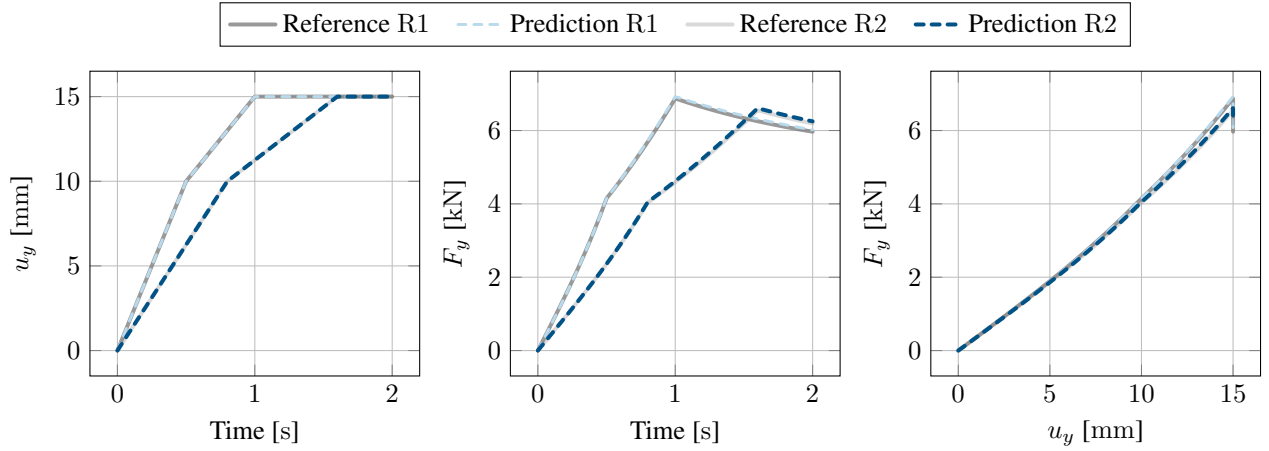


Figure 16: The loading program and resulting measured forces at the right surface (cf. Figure 15) for the isotropic material. Left: Prescribed displacement at the right surface in vertical direction at two different loading rates R1 and R2. Middle: Measured reaction force over time at both loading rates. Right: Force-displacement curves for both loading rates. Due to the inelastic effects, the stiffness between the two loading rates varies.

Results. The reaction forces are shown in Figure 18. As in the isotropic case, higher loading rates lead to higher apparent stiffness and larger reaction forces, and relaxation is visible during the hold phase. However, compared to the isotropic material, both peak forces and stiffnesses are consistently lower. This reduction reflects the physical effect of misalignment: the preferred direction must reorient toward the loading axis before it can effectively resist deformation.

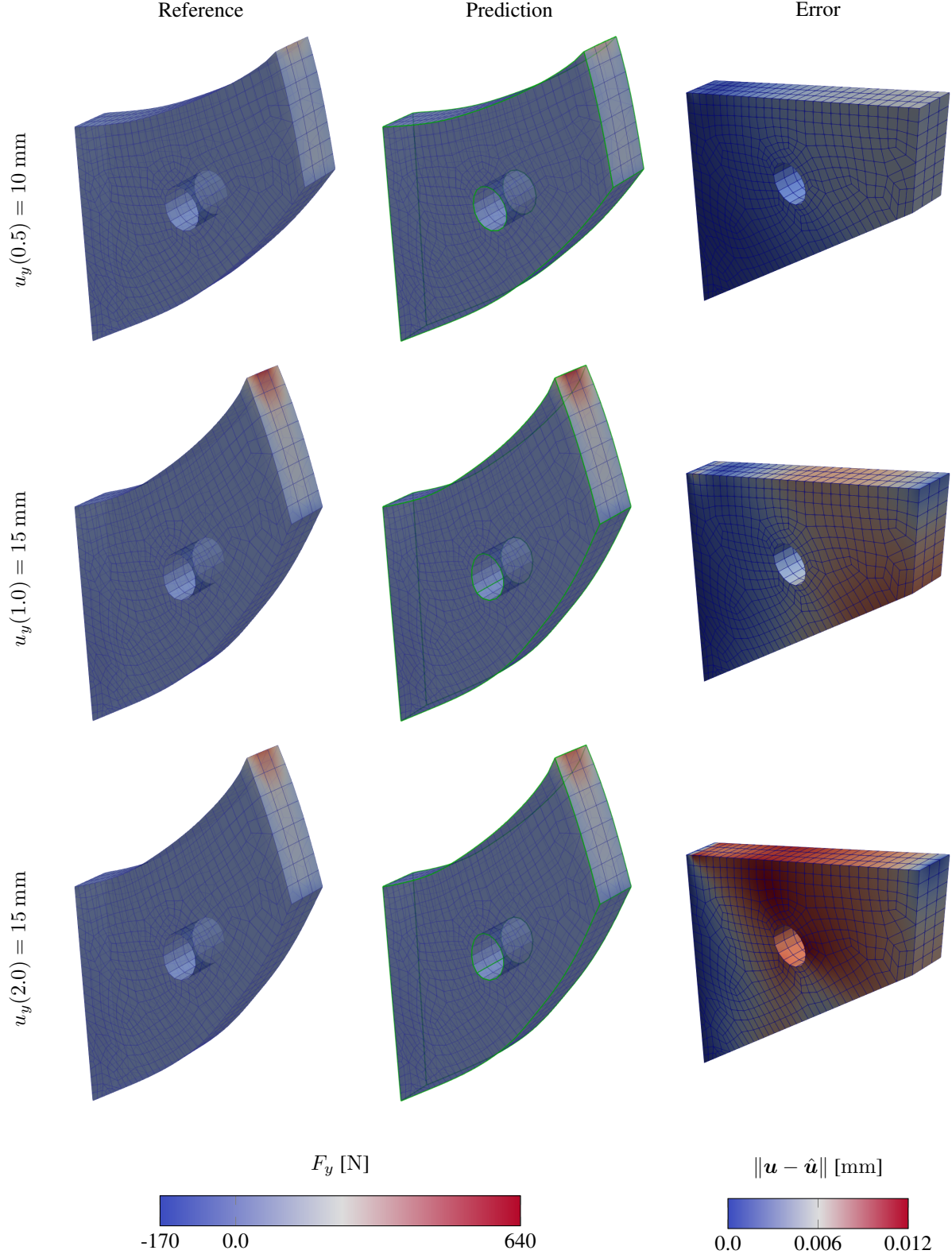


Figure 17: Contour plots of the Cook's membrane test with isotropic material for the first loading rate. Rows represent selected load steps (cf. Figure 16). Columns 1–2 show nodal force contours of the reference and neural network prediction, with green lines marking the reference deformation. The last column displays the Euclidean norm of the displacement error $\hat{u} - u$, where $\hat{u}$ corresponds to the reference and u to the prediction.

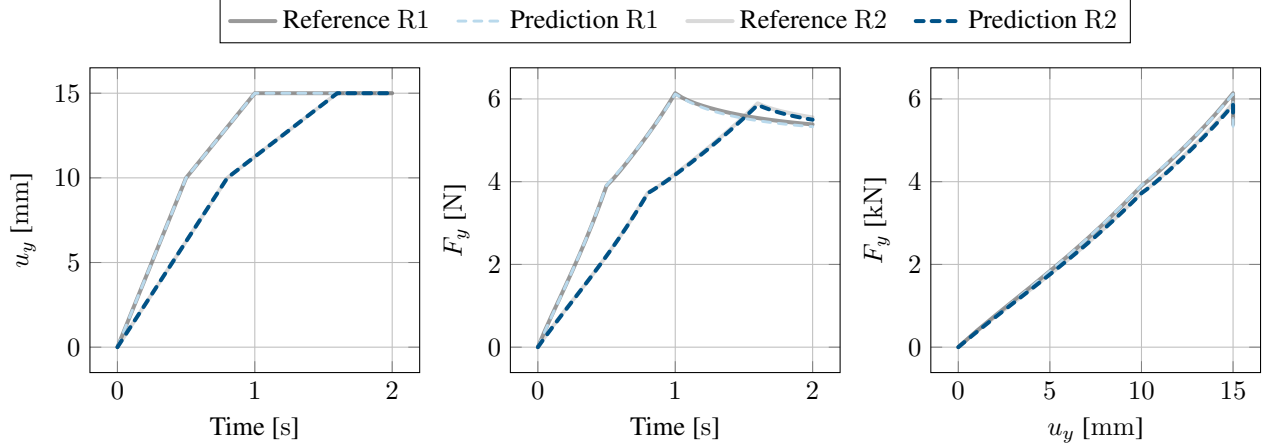


Figure 18: The loading program and resulting measured forces at the right surface (cf. Figure 15) for the anisotropic material. Left: Prescribed displacement at the right surface in vertical direction at two different loading rates R1 and R2. Middle: Measured reaction force over time at both loading rates. Right: Force-displacement curves for both loading rates. Due to the inelastic effects, the stiffness between the two loading rates varies.

Consequently, the specimen undergoes significant twisting and out-of-plane deflection, which is not present in the isotropic response.

These deformation patterns are clearly captured in the contour plots of Figure 19. The structure twists around the y -axis during the hold phase, and the network prediction follows the reference solution both qualitatively and quantitatively. Although the displacement errors are larger than in the isotropic case, reaching up to 0.35 mm at intermediate times, the overall deformation mode is reproduced faithfully. Notably, the anisotropic case also exhibits stronger relaxation in the displacement field: even at constant load, the structure continues to twist as the preferred direction gradually aligns with the principal loading direction. This time-dependent reorientation is captured convincingly by the network.

In summary, the structural study demonstrates that the proposed framework generalizes well from material point training to unseen structural BVPs. For isotropic materials, both global reaction forces and local deformation patterns are reproduced with very high accuracy. For anisotropic materials, even under rotated preferred directions and complex bending-dominated loading, the network captures the essential features of the response, including rate dependence, relaxation, and twisting deformation. These results underline the robustness of embedding constitutive structure into neural networks and confirm their potential for predictive structural simulations.

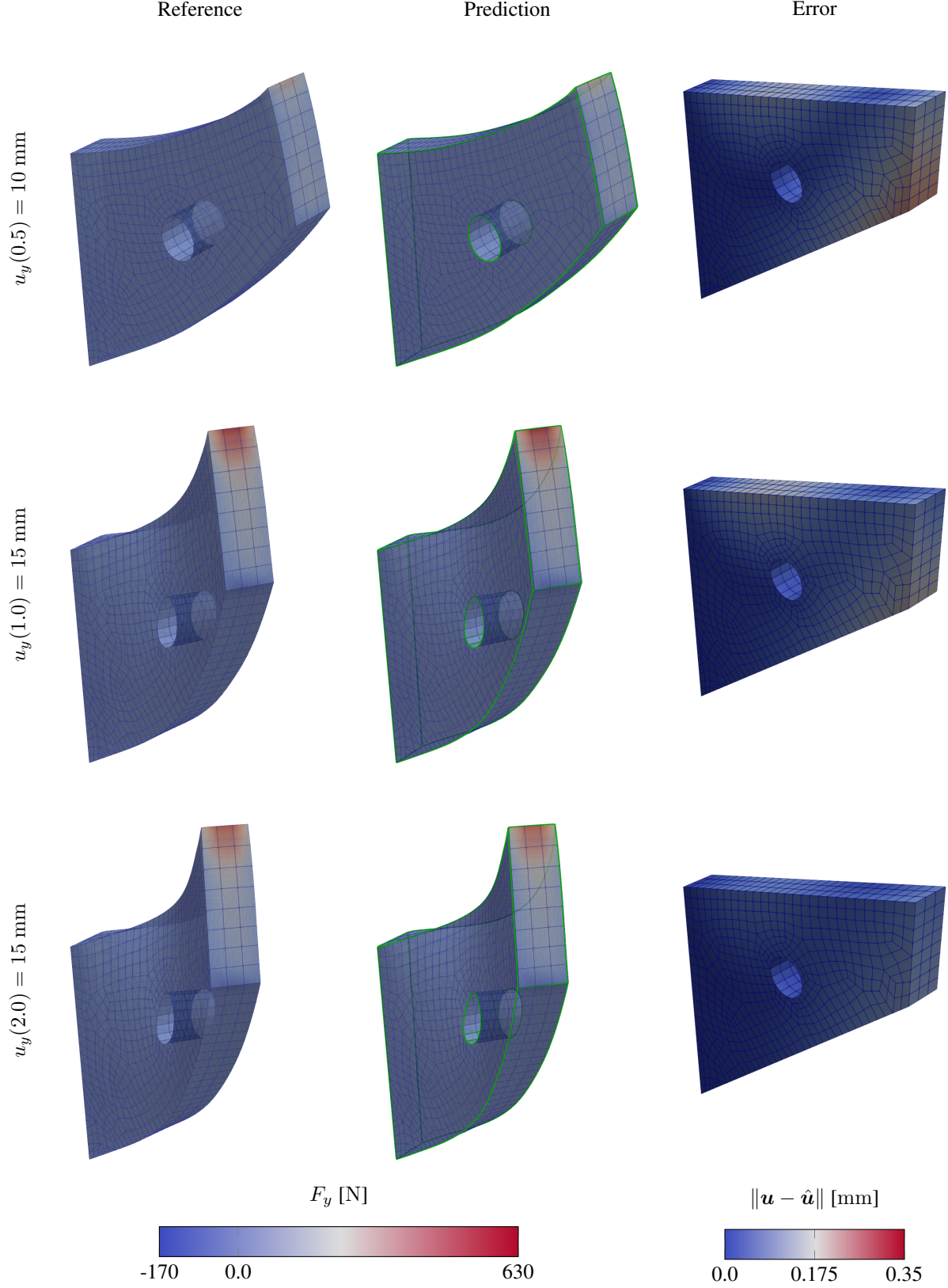


Figure 19: Contour plots of the Cook's membrane test with anisotropic material for the first loading rate. The specimen twists around the y -axis due to anisotropy, in contrast to the isotropic case (Figure 17). Rows represent selected load steps (cf. Figure 18). Columns 1–2 show nodal force contours of the reference and neural network prediction, with green lines marking the reference deformation. The last column displays the Euclidean norm of the displacement error $\hat{u} - u$, where $\hat{u}$ corresponds to the reference and u to the prediction.

7 Discussion

We have presented a framework for designing a dual potential that governs the amount of dissipation while consistently embedding anisotropy into both the Helmholtz free energy and the dual potential itself. The central theoretical extensions of our approach are twofold: (i) we propose a dual potential that is not necessarily convex; yet satisfies the dissipation inequality a priori, and (ii) we define the integrity sets of the energy and the dual potential under consideration of different consistent mappings of the structural tensor to the intermediate configuration. In this course, we discuss the entire set of invariants including those that combine elastic and inelastic tensors.

Beyond the theoretical formulation of anisotropic inelasticity at finite strains, we translated the constitutive relations into neural network architectures. Specifically, we employed a modified version of Input Convex Neural Networks (ICNNs) and introduced Input Monotonic Neural Networks (IMNNs) to extend the dual potential beyond convex formulations. For both the energy and the dual potential, we constructed suitable sets of invariants, which serve as physics-embedded inputs to the corresponding neural networks.

In line with [66, 33], we eliminate the need for classical time-integration schemes in updating the inelastic variables by using neural network prediction. To this end, we incorporated recurrent Liquid Neural Networks (LiNNs) into the overall architecture. We designed a trial-step procedure as input to the LiNN and enforced by construction that the inelastic stretches remain positive definite.

The framework was first evaluated at the material-point level, where we compared its performance and accuracy against recurrent neural architectures without physics priors. Subsequently, we validated the discovered models at the structural scale by predicting deformation and reaction forces for unseen boundary value problems. Across all investigated cases, the framework produced accurate predictions even outside the training regime and exhibited stable performance. Nevertheless, several aspects require further discussion and refinement. We address some of these aspects below.

Choice of invariants. We proposed two distinct invariant sets to represent anisotropic material behavior: one for the Helmholtz free energy (deformation invariants) and another for the dual potential (stress invariants). The choice of invariants, however, is not unique. Alternative formulations include Landau invariants [111] or generalized invariants [112], the latter being shown to capture material response with accuracy comparable to larger invariant sets while reducing redundancy. This property is particularly relevant in the neural network context [113], as smaller invariant sets can reduce network complexity and mitigate uncertainty in weight optimization. In our case, the invariants for the energy were designed to be consistent with polyconvexity. Yet, recent work by [114] introduced a singular-value-based representation capable of modeling stress plateaus in purely elastic materials, which was already implemented within a neural architecture [115]. However, the integration of anisotropy into such approaches remains unresolved. While the literature on deformation invariants is extensive, significantly fewer contributions address the invariants underlying dual potentials. Most existing designs adhere to convexity in stress space, which is sufficient but not necessary. Thus, there is considerable room to explore more expressive invariant sets. In this respect, neural networks might offer a promising tool for discovering such invariants, similar to how pattern recognition in transformer architectures has been leveraged in mathematics [116].

Data availability. Training the proposed networks required synthetic data, as high-fidelity data sets for anisotropic inelasticity at finite strains remain scarce. While the generality of the framework is an advantage, it also necessitates larger training data sets. Increasing the number of invariants directly raises the information entropy required to identify the underlying material behavior in a (near-)unique manner. Whether it is feasible to generate such extensive data sets or whether adaptive multiscale data-generation schemes [117] will be necessary remains an open question. Additionally, experimental data typically consist of force–displacement measurements rather than stress–stretch data, which we employed for training. In this context, data-driven identification methods [118], which were used in the context of physics-augmented neural network training strategies [119], may provide an alternative route. Furthermore, initialization is crucial for training neural networks, which becomes more challenging with greater non-linearity. In this regard, the material fingerprinting method [120] may help identify appropriate initial values and mitigate the need for pre-training. However, their applicability to anisotropic inelasticity has not yet been clarified.

Inelastic phenomena. Our study was restricted to viscoelasticity at finite strains, primarily because it represents the simplest inelastic phenomenon without requiring additional constraints such as yield conditions or growth laws. Nevertheless, recent efforts have also focused on plasticity [92, 23, 37, 121, 34] and growth [40], although the rigorous incorporation of anisotropy in these works remains limited, and convexity-based potentials are typically employed. Our methodology is designed to extend seamlessly to other inelastic mechanisms governed by multiplicative decomposition. Notably, the proposed non-convex dual potential can be directly adapted to other types of inelasticity such as damage models, which are also formulated via potentials. The challenge in incorporating increasingly complex inelastic

phenomena lies in ensuring their unique discovery, which directly depends on data availability and information entropy. An interesting research direction is to explore whether auxiliary neural networks for internal variables can stabilize training in such cases, particularly for plasticity.

Interpretability and regularization. In this work, we prioritized expressivity over interpretability. While our networks were designed to satisfy physical and constitutive constraints by construction, their interpretability is limited compared to, for instance, Constitutive Artificial Neural Networks [21] and their inelastic extensions [31]. We did not employ sparsity-promoting regularization such as Lasso, which could enhance interpretability, nor did we systematically investigate dropout or similar techniques that may prevent over-specialization of network pathways. To this end, clustering methods similar to those in [122] might help increase the interpretability of dense networks by clustering different pathways into a common pathway without loss of accuracy. A central question remains whether the analytical form of the entire network is essential, given that invariants already carry interpretable physical meaning as deformation or stress measures. We also did not examine whether the network architecture can reduce to a purely elastic case, which must be included as a limiting scenario. For isotropic inelasticity, this reduction was studied in [36], where Lasso regularization effectively discovered a vanishing dual potential. Interestingly, comparable behavior was observed without regularization, albeit at the expense of interpretability. Finally, emerging approaches such as neural architecture search [123] may play a crucial role in solid mechanics, both for determining whether inelasticity is relevant and for uncovering anisotropy when it is not explicitly known a priori.

8 Conclusion

We have proposed a methodology to consistently embed physics into neural networks for modeling anisotropic inelasticity at finite strains based on the design of a dual potential. The approach generalizes beyond convex formulations, ensures the dissipation inequality a priori, and integrates anisotropy consistently through invariant-based representations of the Helmholtz free energy and the dual potential. By embedding constitutive principles into adapted neural architectures, including Input Monotonic Neural Networks and Liquid Neural Networks, we achieved stable and accurate predictions both at the material-point and structural level, even outside the training regime.

While the methodology offers a flexible foundation transferable to various inelastic phenomena, its success critically depends on the availability of informative data. Future research may address the systematic interpretation of the discovered networks, different data-generation strategies, and extensions to more complex inelastic mechanisms such as plasticity.

A Additions

A.1 Input Monotonic Neural Network

In the following, we discuss the monotonicity properties of the Input Monotonic Neural Network formulation. Let $\mathbf{s} \in \mathbb{R}^n$ denote the input parameters and $\mathbf{t} \in \mathbb{R}^m$ the network output. The mapping from one layer to the next is defined as

$$\mathbf{p}_{l+1} = g_l(\mathbf{W}_l^p \mathbf{p}_l + \mathbf{W}_l^s \mathbf{s} + \mathbf{b}_l), \quad l = 0, \dots, L-1,$$

with $\mathbf{p}_0 = \mathbf{s}$. Here, $\mathbf{W}_l^p$ and $\mathbf{W}_l^s$ are weight matrices with non-negative entries, $\mathbf{b}_l$ are bias vectors, and the activation functions g_l are chosen to be monotonically increasing functions (such as ReLU, softplus, or sigmoid).

This structure directly implies that the overall network mapping $\mathbf{s} \mapsto \mathbf{t}$ is monotonically non-decreasing with respect to each input component. The reasoning is straightforward: if one increases any component of the input vector $\mathbf{s}$, the affine terms $\mathbf{W}_l^p \mathbf{p}_l + \mathbf{W}_l^s \mathbf{s} + \mathbf{b}_l$ in each layer cannot decrease, because all entries of $\mathbf{W}_l^p$ and $\mathbf{W}_l^s$ are non-negative. The subsequent application of a monotonically increasing activation function preserves this order. Consequently, the intermediate states $\mathbf{p}_l$ remain ordered in the same way for all layers. The final output $\mathbf{t} = \mathbf{p}_L$ therefore inherits this property.

From a differential perspective, whenever the activation functions are differentiable, the partial derivatives of the outputs with respect to the inputs satisfy

$$\frac{\partial t_i}{\partial s_j} \geq 0 \quad \text{for all } i, j.$$

This means that no output component decreases when any input parameter increases. In other words, the network as defined above provides a mapping that is monotone: the response $\mathbf{t}$ cannot decrease under an increase of any component of the parameter vector $\mathbf{s}$.

A.2 Numerical details

Computation of roots. In the computation of the square and cubic roots of the stress invariants, we are particularly concerned with their derivatives. However, these are not defined at zero. To circumvent this issue, following [36], we slightly modify the definitions as

$$\sqrt{x} := \frac{x}{(x + \epsilon)^{1/2}}, \quad \sqrt[3]{x} := \frac{x}{(|x| + \epsilon)^{2/3}}.$$

Throughout this work, we set $\epsilon = 0.01$. With this choice, one obtains

$$\left. \frac{d}{dx} \sqrt{x} \right|_{x=0} = 10.0, \quad \left. \frac{d}{dx} \sqrt[3]{x} \right|_{x=0} \approx 21.54435.$$

Cholesky decomposition. Within the time discretization of the inelastic variable, we employ a recurrent Liquid Neural Network (LiNN) architecture to predict the inelastic stretches $\mathbf{U}_i \in \text{Sym}_+^{3 \times 3}$ at t_{n+1} . Thus, we must ensure that the hidden states $\mathbf{h}$ propagated by the LiNN are consistently mapped to a symmetric positive definite tensor. Since every $\mathbf{U}_i$ admits a Cholesky factorization,

$$\mathbf{U}_i = \mathbf{L} \mathbf{L}^T,$$

with $\mathbf{L}$ a real lower triangular matrix with strictly positive diagonal entries $L_{ii} > 0$, we parameterize $\mathbf{L}$ through six independent states $\mathbf{h} = (h_1, \dots, h_6)$

$$\mathbf{L} = \begin{pmatrix} f(h_1) & 0 & 0 \\ h_4 & f(h_2) & 0 \\ h_6 & h_5 & f(h_3) \end{pmatrix},$$

where

$$f(x) = \frac{1}{\ln(2)} \ln(1 + \exp(\ln(2) x))$$

is a scaled softplus activation ensuring $f(x) > 0$ for all x and $f(0) = 1$. With the initialization $\mathbf{h} = \mathbf{0}$, we obtain $\mathbf{L} = \mathbf{I}$ and hence $\mathbf{U}_i = \mathbf{I}$.

A.3 Thermodynamic consistency

Isotropic materials. We define the stress invariants

$$S_1 := \text{tr } \boldsymbol{\Sigma}, \quad S_2' := \frac{1}{2} \text{dev } \boldsymbol{\Sigma} : \text{dev } \boldsymbol{\Sigma}, \quad S_3' := \frac{1}{3} \text{tr}((\text{dev } \boldsymbol{\Sigma})^3),$$

together with $S_2 := \sqrt{S_2'}$ and $S_3 := \sqrt[3]{S_3'}$. Set

$$\mathcal{S}_\varphi(\boldsymbol{\Sigma}) := (S_1(\boldsymbol{\Sigma}), S_2(\boldsymbol{\Sigma}), S_3(\boldsymbol{\Sigma})), \quad \varphi(\boldsymbol{\Sigma}) := \omega(\mathcal{S}_\varphi(\boldsymbol{\Sigma})),$$

where $\omega : \mathbb{R}^3 \rightarrow \mathbb{R}$ is convex, non-negative, and satisfies $\omega(\mathbf{0}) = 0$. The derivatives are

$$\frac{\partial S_1}{\partial \boldsymbol{\Sigma}} = \mathbf{I}, \quad \frac{\partial S_2}{\partial \boldsymbol{\Sigma}} = \frac{1}{2\sqrt{S_2'}} \text{dev } \boldsymbol{\Sigma}, \quad \frac{\partial S_3}{\partial \boldsymbol{\Sigma}} = \frac{1}{3S_3'^{2/3}} \text{dev}((\text{dev } \boldsymbol{\Sigma})^2).$$

Since S_1, S_2, S_3 are homogeneous of degree 1, it follows that

$$\frac{\partial S_k}{\partial \boldsymbol{\Sigma}} : \boldsymbol{\Sigma} = S_k \quad (k = 1, 2, 3).$$

By the chain rule,

$$\partial_{\boldsymbol{\Sigma}} \varphi = \sum_k \frac{\partial \omega}{\partial S_k} \frac{\partial S_k}{\partial \boldsymbol{\Sigma}}, \quad \Rightarrow \quad \boldsymbol{\Sigma} : \partial_{\boldsymbol{\Sigma}} \varphi = \langle \mathcal{S}_\varphi, \partial_{\mathcal{S}_\varphi} \omega \rangle.$$

With $\mathbf{D}_i \in \partial_{\boldsymbol{\Sigma}} \varphi$, the reduced dissipation inequality reads

$$\boldsymbol{\Sigma} : \mathbf{D}_i = \langle \mathcal{S}_\varphi, \partial_{\mathcal{S}_\varphi} \omega \rangle \geq \omega(\mathcal{S}_\varphi) - \omega(\mathbf{0}) = \omega(\mathcal{S}_\varphi) \geq 0,$$

ensuring non-negative dissipation. For a more detailed derivation, see e.g. [36]. If $\varphi = (\zeta \circ \omega)(\mathcal{S}_\varphi)$ with $\zeta : \mathbb{R} \rightarrow \mathbb{R}$ monotone increasing, then

$$\boldsymbol{\Sigma} : \partial_{\boldsymbol{\Sigma}} \varphi = \frac{\partial \zeta}{\partial \omega} \langle \mathcal{S}_\varphi, \partial_{\mathcal{S}_\varphi} \omega \rangle \geq 0,$$

so thermodynamic consistency is still preserved.

Anisotropic materials. The push-forward $\tilde{M} = F_i \star M$ represents either a covariant or contravariant mapping of the structural vector; see Equation (23). Following [93], we may form the irreducible basis in terms of the mixed invariants of the stress and the structural vectors

$$\Sigma : \tilde{M}_j = \text{dev} \Sigma : \tilde{M}_j + \frac{S_1}{3}, \quad \Sigma^2 : \tilde{M}_j = (\text{dev} \Sigma)^2 : \tilde{M}_j + 2 \frac{S_1}{3} \text{dev} \Sigma : \tilde{M}_j + \frac{S_1^2}{9},$$

$$\tilde{n}_1 \cdot \Sigma \tilde{n}_2 = \Sigma : \text{sym}(\tilde{n}_1 \otimes \tilde{n}_2), \quad \tilde{n}_1 \cdot \Sigma^2 \tilde{n}_2 = \Sigma^2 : \text{sym}(\tilde{n}_1 \otimes \tilde{n}_2),$$

where $\tilde{M}_j := \tilde{n}_j \otimes \tilde{n}_j$ with $k = 1, 2$. We introduce

$$\mathcal{A}_\varphi(\Sigma, \tilde{n}_1, \tilde{n}_2) := (A_{1j}, A_{2j}, A_3, A_4)_{j=1,2},$$

with

$$A_{1j} := \text{dev} \Sigma : \tilde{M}_j, \quad A_{2j} := \sqrt{\frac{1}{2}(\text{dev} \Sigma)^2 : \tilde{M}_j}, \quad A_3 := \Sigma : \text{sym}(\tilde{n}_1 \otimes \tilde{n}_2), \quad A_4 := \sqrt{\frac{1}{2} \Sigma^2 : \text{sym}(\tilde{n}_1 \otimes \tilde{n}_2)}.$$

Each of these invariants is homogeneous of degree 1. We set

$$\varphi(\Sigma, \tilde{n}_1, \tilde{n}_2) := (\zeta \circ \omega)(\mathcal{T}_\varphi(\Sigma, \tilde{n}_1, \tilde{n}_2)), \quad \mathcal{T}_\varphi(\Sigma, \tilde{n}_1, \tilde{n}_2) := (\mathcal{S}_\varphi(\Sigma), \mathcal{A}_\varphi(\Sigma, \tilde{n}_1, \tilde{n}_2)),$$

with $\omega : \mathbb{R}^9 \rightarrow \mathbb{R}$ convex, non-negative, and $\omega(0) = 0$. The chain rule gives

$$\Sigma : \partial_\Sigma \varphi = \frac{\partial \zeta}{\partial \omega} \langle \mathcal{T}_\varphi, \partial \omega \mathcal{T}_\varphi \rangle \geq 0,$$

ensuring thermodynamic consistency as in the isotropic case.

A.4 Helmholtz free energy

Mixed-variant mappings. In Equation (24), for $k = r = 0$ and $s = t = 1$,

$$\tilde{M}_1 \tilde{M}_2 = \frac{1}{\gamma} F_i M F_i^{-1},$$

i.e., proportional to the mixed-variant map $F_i \star M = F_i M F_i^{-1}$, where

$$\gamma := (C_i : M) (C_i^{-1} : M) = (\mathbf{n} \cdot C_i \mathbf{n}) (\mathbf{n} \cdot C_i^{-1} \mathbf{n}) \geq 1$$

by the Cauchy–Schwarz inequality, with equality if $\mathbf{n}$ is an eigenvector of C_i . If one wishes to mimic $s = t = -1$ by definition (not inversion), set

$$\left(\frac{C_i M}{C_i : M} \right)^{-1} := (C_i : M) M C_i^{-1}, \quad \left(\frac{M C_i^{-1}}{C_i^{-1} : M} \right)^{-1} := (C_i^{-1} : M) C_i M,$$

which yields, for $k = r = 0$ and $s = t = -1$,

$$\tilde{M}_1^{-1} \tilde{M}_2^{-1} = \gamma F_i^{-T} M F_i^T,$$

i.e., proportional to the alternative mixed-variant map.

Polyconvexity. To construct a polyconvex Helmholtz free energy, one must employ invariants that are convex in $(F, \text{cof } F, J)$. As an example, consider

$$f(F, G) := \text{tr}(F^T F G).$$

Its second Gâteaux derivative in direction $H \in \mathbb{M}^{3 \times 3}$ is

$$\mathcal{D}_F^2[f(F, G)](H, H) = \frac{d^2}{d\epsilon^2} \text{tr}((F + \epsilon H)^T (F + \epsilon H) G) \Big|_{\epsilon=0} = 2 G : H^T H.$$

Thus, f is convex in F whenever $G \succeq 0$ (positive semidefinite) [105]. Analogously, replacing F by $\text{cof } F$ shows the same condition for convexity in the cofactor.

Polyconvexity of mixed-variant mappings. Let $F_i \star M$ denote a mixed-variant mapping consistent with the principle of indifference to the choice of intermediate configuration, e.g., $F_i \star M = F_i M F_i^{-1}$, $F_i \star M = F_i^{-T} M F_i^T$ or any analogous form. Then the function $f(F, F_i \star M) = \text{tr}(F^T F F_i \star M) = \text{tr}(F^T F \text{sym}(F_i \star M))$ cannot be guaranteed to be polyconvex *a priori* as $\text{sym}(F_i \star M)$ is not guaranteed to be positive (semi-)definite.

Outer tensor product of second order tensors. Let $\#$ denote the outer tensor product of two second-order tensors,

$$(\mathbf{a} \otimes \mathbf{b}) \# (\mathbf{c} \otimes \mathbf{d}) := (\mathbf{a} \times \mathbf{c}) \otimes (\mathbf{b} \times \mathbf{d}),$$

where $\times$ denotes the vector cross product. Let $\mathbf{B} \in \text{Sym}_+^{3 \times 3}$ and $\mathbf{G} = \mathbf{g} \otimes \mathbf{g}$. Then their outer product can be written as

$$\mathbf{B} \# \mathbf{G} = \sum_{i=1}^3 B_i (\mathbf{p}_i \otimes \mathbf{p}_i) \# \mathbf{G} = \sum_{i=1}^3 B_i (\mathbf{p}_i \times \mathbf{g}) \otimes (\mathbf{p}_i \times \mathbf{g}),$$

which is positive semi-definite, with B_i and $\mathbf{p}_i$ denoting the eigenvalues and eigenvectors of $\mathbf{B}$, respectively. More generally, the outer tensor product of any two tensors $\mathbf{A}, \mathbf{B} \in \mathbb{M}^{3 \times 3}$ is given by

$$\mathbf{A} \# \mathbf{B} = (\text{tr } \mathbf{A} \text{ tr } \mathbf{B} - \text{tr}(\mathbf{A}\mathbf{B})) \mathbf{I} + \mathbf{B}^T \mathbf{A}^T + \mathbf{A}^T \mathbf{B}^T - (\text{tr } \mathbf{A}) \mathbf{B}^T - (\text{tr } \mathbf{B}) \mathbf{A}^T,$$

which yields a symmetric tensor whenever both arguments are symmetric. In addition, for any $\mathbf{P} \in \text{GL}(3, \mathbb{R})$ one obtains the covariance property

$$(\mathbf{P}\mathbf{A}\mathbf{P}^{-1}) \# (\mathbf{P}\mathbf{B}\mathbf{P}^{-1}) = \mathbf{P}^{-T} (\mathbf{A} \# \mathbf{B}) \mathbf{P}^T.$$

In view of these properties and of the Helmholtz free energy in Section 4.1, the invariant

$$\text{tr} \left(\mathbf{C}_e (\mathbf{B}_i \# \tilde{\mathbf{M}}) \right) = \text{tr} \left(\mathbf{C} \mathbf{F}_i^{-1} (\mathbf{B}_i \# \tilde{\mathbf{M}}) \mathbf{F}_i^{-T} \right)$$

is polyconvex, since $\mathbf{F}_i^{-1} (\mathbf{B}_i \# \tilde{\mathbf{M}}) \mathbf{F}_i^{-T}$ is positive semi-definite. Finally, using Equation (24), we obtain

$$\text{tr} \left(\mathbf{C} \mathbf{F}_i^{-1} (\mathbf{B}_i \# \tilde{\mathbf{M}}) \mathbf{F}_i^{-T} \right) = \text{tr} (\mathbf{C}_i^{-1} \mathbf{C} (\mathbf{C}_i \# \mathbf{C}_i \mathbf{M})) = \text{tr} (\mathbf{C} \mathbf{C}_i^{-1} (\mathbf{C}_i \# \mathbf{M} \mathbf{C}_i)).$$

A.5 Reference material model

As reference material model, we adopt a rheological formulation with three elements. The model consists of an equilibrium branch, represented by a spring element, arranged in parallel to a non-equilibrium branch, consisting of a spring–dashpot series connection. Accordingly, the Helmholtz free energy ψ and the dual potential φ are additively decomposed into isotropic and anisotropic contributions:

$$\psi = \underbrace{\psi_{\text{iso}}^{\text{eq}}(\mathbf{C}; a^{\text{eq}}, b^{\text{eq}}, c^{\text{eq}}) + \psi_{\text{iso}}^{\text{neq}}(\mathbf{C}_e; a^{\text{neq}}, b^{\text{neq}}, c^{\text{neq}})}_{=: \psi_{\text{iso}}} + \underbrace{\psi_{\text{ani}}^{\text{eq}}(\mathbf{C}, \mathbf{G}; \alpha^{\text{eq}}, \eta^{\text{eq}}) + \psi_{\text{ani}}^{\text{neq}}(\mathbf{C}_e, \tilde{\mathbf{G}}; \alpha^{\text{neq}}, \eta^{\text{neq}})}_{=: \psi_{\text{ani}}},$$

$$\varphi = \varphi_{\text{iso}}(\boldsymbol{\Sigma}; a^{\text{neq}}, b^{\text{neq}}, c^{\text{neq}}, \tau) + \varphi_{\text{ani}}(\boldsymbol{\Sigma}, \tilde{\mathbf{M}}_3; \eta^{\text{neq}}, \tau).$$

Here, a, b, c, α , and η denote material parameters, while $\mathbf{G}$ represents a metric tensor that accounts for the preferred direction [105], and must not be confused with a second-order structural tensor. Assuming the preferred direction aligns with $\mathbf{e}_1$, the reference metric reads

$$\mathbf{G}_0 = \begin{pmatrix} \beta^2 & 0 & 0 \\ 0 & \frac{1}{\beta} & 0 \\ 0 & 0 & \frac{1}{\beta} \end{pmatrix},$$

where β is an anisotropy parameter. To obtain $\mathbf{G}$ aligned with an arbitrary preferred direction $\mathbf{n} \in \mathbb{S}^2$, we employ Rodrigues' rotation formula. A rotation matrix $\mathbf{R} \in \text{SO}(3)$ is constructed such that $\mathbf{e}_1$ is mapped onto $\mathbf{n}$, with axis and angle defined by

$$\mathbf{k} = \mathbf{e}_1 \times \mathbf{n}, \quad \hat{\mathbf{k}} = \frac{\mathbf{k}}{\|\mathbf{k}\|}, \quad \theta = \arccos(\mathbf{e}_1 \cdot \mathbf{n}).$$

The corresponding rotation matrix is

$$\mathbf{R} = \mathbf{I} + \sin \theta [\hat{\mathbf{k}}]_{\times} + (1 - \cos \theta) [\hat{\mathbf{k}}]_{\times}^2,$$

with $[\hat{\mathbf{k}}]_{\times}$ denoting the skew-symmetric cross-product operator. The metric tensor is then given by

$$\mathbf{G} = \mathbf{R} \mathbf{G}_0 \mathbf{R}^T.$$

For the push-forward to the intermediate configuration, we adopt

$$\tilde{\mathbf{G}} = \mathbf{F}_i \mathbf{G} \mathbf{F}_i^T, \quad \tilde{\mathbf{M}}_3 = \mathbf{F}_i (\mathbf{n} \otimes \mathbf{n}) \mathbf{F}_i^{-1}.$$

For the isotropic Helmholtz free energy, we employ the form proposed in [105]:

$$\psi_{\text{iso}}^{\text{eq/neq}}(\mathbf{A}) = a \operatorname{tr} \mathbf{A} + b \operatorname{tr} \operatorname{cof} \mathbf{A} + c \det \mathbf{A} - \frac{d_1}{2} \ln \det \mathbf{A}, \quad d_1 = 2a + 4b + 2c.$$

Further, for the anisotropic Helmholtz free energy, we adopt

$$\psi_{\text{ani}}^{\text{eq/neq}}(\mathbf{A}, \mathbf{B}) = \frac{\eta}{\alpha (\operatorname{tr} \mathbf{B})^\alpha} ((\operatorname{tr} \mathbf{A} \mathbf{B})^\alpha + (\operatorname{tr} \operatorname{cof} \mathbf{A} \mathbf{B})^\alpha) - \frac{d_2}{2} \ln \det \mathbf{A}, \quad d_2 = 2\eta.$$

In analogy to [106], the isotropic part of the dual potential is chosen as

$$\varphi_{\text{iso}}(\boldsymbol{\Sigma}) = \frac{1}{\tau} \left(\frac{1}{18} \frac{(\operatorname{tr} \boldsymbol{\Sigma})^2}{c} + \frac{2}{a+b} \left(\frac{1}{2} \operatorname{tr} \operatorname{dev} \boldsymbol{\Sigma} \operatorname{dev} \boldsymbol{\Sigma} \right) \right),$$

while the anisotropic contribution is given by

$$\varphi_{\text{ani}}(\boldsymbol{\Sigma}, \tilde{\mathbf{M}}_3) = \frac{1}{\tau} \left(\frac{1}{2\eta} \left((\operatorname{tr} \boldsymbol{\Sigma} \tilde{\mathbf{M}}_3)^2 + \operatorname{tr} (\boldsymbol{\Sigma}^2 \tilde{\mathbf{M}}_3) \right) \right).$$

A.6 Plain recurrent neural networks

Recurrent Neural Network (RNN). The recurrent neural network employed in this work is formulated in discrete time with input $\mathbf{x}_t \in \mathbb{R}^{n_x}$, hidden state $\mathbf{h}_t \in \mathbb{R}^{n_h}$, and output $\mathbf{y}_t \in \mathbb{R}^{n_y}$. At each time step t , the hidden state is updated by concatenating the input and the previous hidden state,

$$\mathbf{z}_t^{(0)} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{h}_{t-1} \end{bmatrix} \in \mathbb{R}^{n_x+n_h}.$$

This vector is processed by a sequence of dense layers with nonlinear activations,

$$\mathbf{z}_t^{(\ell)} = \phi_\ell \left(\mathbf{W}_\ell \mathbf{z}_t^{(\ell-1)} + \mathbf{b}_\ell \right), \quad \ell = 1, \dots, L,$$

with $\mathbf{W}_\ell \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$, $\mathbf{b}_\ell \in \mathbb{R}^{n_\ell}$, and nonlinear activation $\phi_\ell(\cdot)$. The final transformation maps onto the hidden state space of size n_h ,

$$\mathbf{h}_t = \mathbf{W}_h \mathbf{z}_t^{(L)} + \mathbf{b}_h, \quad \mathbf{W}_h \in \mathbb{R}^{n_h \times n_L}.$$

The output is computed as a linear mapping of the hidden state,

$$\mathbf{y}_t = \mathbf{W}_{hy} \mathbf{h}_t + \mathbf{b}_y, \quad \mathbf{W}_{hy} \in \mathbb{R}^{n_y \times n_h}.$$

This flexible cell structure allows one to define arbitrary numbers of hidden layers and activation functions inside the recurrent update.

In compact form, the chosen RNN topology reads

$$\mathcal{N}_{\text{RNN}} : \mathbb{R}^{n_x+n_h} \rightarrow \mathbb{R}^{n_h}, \quad (\mathbf{x}_t, \mathbf{h}_{t-1}) \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{Linear}} \mathbf{h}_t,$$

with $n_x = 7$ (strain-like features $\mathbf{C}$ and time increment Δt) and $n_h = 6$. The output mapping is

$$\mathbf{y}_t = \mathbf{W}_{hy} \mathbf{h}_t + \mathbf{b}_y \in \mathbb{R}^6.$$

Liquid Neural Network (LiNN). The Liquid Neural Network builds upon the same input–hidden–output structure with $\mathbf{x}_t \in \mathbb{R}^{n_x}$, $\mathbf{h}_t \in \mathbb{R}^{n_h}$, and $\mathbf{y}_t \in \mathbb{R}^{n_y}$. The key difference lies in the hidden state update, which introduces adaptive dynamics through two subnetworks: a *source network* $\mathbf{f}(\mathbf{x}_t, \mathbf{h}_{t-1}) \in \mathbb{R}^{n_h}$ and a *modulation network* $\boldsymbol{\alpha}(\mathbf{x}_t, \mathbf{h}_{t-1}) \in \mathbb{R}^{n_h}$. Both subnetworks take the concatenated input–state vector

$$\mathbf{z}_t^{(0)} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{h}_{t-1} \end{bmatrix} \in \mathbb{R}^{n_x+n_h},$$

as input and are constructed entirely from *centered layers*, i.e.

$$\text{CL}(z) = \phi(\mathbf{W}z + \mathbf{b}) - \phi(\mathbf{b}),$$

which stabilize training by removing trivial bias-induced offsets. The hidden state update follows an explicit Euler integration scheme,

$$\mathbf{h}_t = (1 - \boldsymbol{\alpha}) \odot \mathbf{h}_{t-1} + \Delta t \mathbf{f},$$

where $\odot$ denotes the Hadamard product and Δt is a prescribed time step. The output is obtained by

$$\mathbf{y}_t = \mathbf{W}_{hy} \mathbf{h}_t + \mathbf{b}_y \in \mathbb{R}^6.$$

In compact form, the chosen LiNN topologies are

$$\begin{aligned} \mathcal{N}_f : \mathbb{R}^{n_x+n_h} &\rightarrow \mathbb{R}^{n_h}, \quad (\mathbf{x}_t, \mathbf{h}_{t-1}) \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{Linear}} \mathbf{f}, \\ \mathcal{N}_\alpha : \mathbb{R}^{n_x+n_h} &\rightarrow \mathbb{R}^{n_h}, \quad (\mathbf{x}_t, \mathbf{h}_{t-1}) \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{GELU}} 64 \xrightarrow{\text{ReLU}} \alpha. \end{aligned}$$

Here, $n_x = 7$ (strain-like features $\mathbf{C}$ and Δt) and $n_h = 6$. Although Δt already appears in the update scheme, we include it also as network input for consistency with the RNN case.

B Declarations

B.1 Acknowledgements

First of all, the authors would like to thank Ahmad Awad for his help with implementing the finite element framework in JAX. Hagen Holthausen gratefully acknowledge financial support of the projects 495926269 and 417002380 by the Deutsche Forschungsgemeinschaft. Further, this work was supported by the NSF CMMI Award 2320933 Automated Model Discovery for Soft Matter and by the ERC Advanced Grant 101141626 DISCOVER to Ellen Kuhl.

B.2 Conflict of interest

The authors of this work certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

B.3 Availability of data, source code and material

Our data, source code and examples of the JAX/Flax implementations of the neural networks and the finite element framework are accessible to the public at <https://doi.org/10.5281/zenodo.17199965>.

B.4 Contributions by the authors

Hagen Holthausen: Conceptualization, Methodology, Software, Validation, Investigation, Formal Analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Funding acquisition

Ellen Kuhl: Writing - Review & Editing, Supervision, Funding acquisition

B.5 Statement of AI-assisted tools usage

The authors acknowledge the use of OpenAI's ChatGPT, an AI language model, for assistance in generating and refining text. The authors reviewed, edited, and take full responsibility for the content and conclusions of this work

References

- [1] Lu Lu, Raphael Pestourie, Wenjie Yao, Zhicheng Wang, Francesc Verdugo, and Steven G Johnson. Physics-informed neural networks with hard constraints for inverse design. *SIAM Journal on Scientific Computing*, 43(6):B1105–B1132, 2021.
- [2] Asghar A. Jadoon, Karl A. Kalina, Manuel K. Rausch, Reese Jones, and Jan Niklas Fuhg. Inverse design of anisotropic microstructures using physics-augmented neural networks. *Journal of the Mechanics and Physics of Solids*, 203:106161, 2025.
- [3] Saeid Ghoul, Moritz Flaschel, Siddhant Kumar, and Laura De Lorenzis. A topology optimisation framework to design test specimens for one-shot identification or discovery of material models. *Journal of the Mechanics and Physics of Solids*, 203:106210, 2025.

- [4] Jonathan Stollberg, Tarun Gangwar, Oliver Weeger, and Dominik Schillinger. Multiscale topology optimization of functionally graded lattice structures based on physics-augmented neural network material models. *Computer Methods in Applied Mechanics and Engineering*, 438:117808, 2025.
- [5] Aymen Danoun, Etienne Prulière, and Yves Chemisky. Fe-Istm: A hybrid approach to accelerate multiscale simulations of architected materials using recurrent neural networks and finite element analysis. *Computer Methods in Applied Mechanics and Engineering*, 429:117192, 2024.
- [6] Karl A. Kalina, Lennart Linden, Jörg Brummund, and Markus Kästner. Fe^{ANN}: an efficient data-driven multiscale approach based on physics-constrained neural networks and automated data mining. *Computational Mechanics*, 71(5):827–851, February 2023.
- [7] Diab W. Abueidda and Mostafa E. Mobasher. I-fenn for thermoelasticity based on physics-informed temporal convolutional network (pi-tcn). *Computational Mechanics*, 74(6):1229–1259, April 2024.
- [8] Jan N. Fuhg, Asghar Jadoon, Oliver Weeger, D. Thomas Seidl, and Reese E. Jones. Polyconvex neural network models of thermoelasticity, 2024.
- [9] Jannick Kehls, Ellen Kuhl, Tim Brepols, Kevin Linka, and Hagen Holthusen. Autoencoder-based non-intrusive model order reduction in continuum mechanics, 2025.
- [10] Jan S Hesthaven and Stefano Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363:55–78, 2018.
- [11] Moritz Flaschel, Siddhant Kumar, and Laura De Lorenzis. Unsupervised discovery of interpretable hyperelastic constitutive laws. *Computer Methods in Applied Mechanics and Engineering*, 381:113852, 2021.
- [12] Kevin Linka, Markus Hillgärtner, Kian P. Abdolazizi, Roland C. Aydin, Mikhail Itskov, and Christian J. Cyron. Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning. *Journal of Computational Physics*, 429:110010, 2021.
- [13] Dominik K. Klein, Mauricio Fernández, Robert J. Martin, Patrizio Neff, and Oliver Weeger. Polyconvex anisotropic hyperelasticity with neural networks. *Journal of the Mechanics and Physics of Solids*, 159:104703, 2022.
- [14] Manuel Chiachío, María Megía, Juan Chiachío, Juan Fernandez, and María L. Jalón. Structural digital twin framework: Formulation and technology integration. *Automation in Construction*, 140:104333, 2022.
- [15] Eamon Duede, William Dolan, André Bauer, Ian Foster, and Karim Lakhani. Oil & water? diffusion of ai within and across scientific fields, 2024.
- [16] Stefan Hajkowicz, Conrad Sanderson, Sarvnaz Karimi, Alexandra Bratanova, and Claire Naughtin. Artificial intelligence adoption in the physical sciences, natural sciences, life sciences, social sciences and the arts and humanities: A bibliometric analysis of research publications from 1960-2021. *Technology in Society*, 74, August 2023.
- [17] J. Ghaboussi, J. H. Garrett, and X. Wu. Knowledge-based modeling of material behavior with neural networks. *Journal of Engineering Mechanics*, 117(1):132–153, 1991.
- [18] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- [19] Filippo Masi, Ioannis Stefanou, Paolo Vannucci, and Victor Maffi-Berthier. Thermodynamics-based artificial neural networks for constitutive modeling. *Journal of the Mechanics and Physics of Solids*, 147:104277, 2021.
- [20] Filippo Masi and Ioannis Stefanou. Multiscale modeling of inelastic materials with thermodynamics-based artificial neural networks (tann). *Computer Methods in Applied Mechanics and Engineering*, 398:115190, 2022.
- [21] Kevin Linka and Ellen Kuhl. A new family of constitutive artificial neural networks towards automated model discovery. *Computer Methods in Applied Mechanics and Engineering*, 403:115731, 2023.
- [22] Lennart Linden, Dominik K. Klein, Karl A. Kalina, Jörg Brummund, Oliver Weeger, and Markus Kästner. Neural networks meet hyperelasticity: A guide to enforcing physics. *Journal of the Mechanics and Physics of Solids*, 179:105363, October 2023.
- [23] Moritz Flaschel, Siddhant Kumar, and Laura De Lorenzis. Discovering plasticity models without stress data. *npj Computational Materials*, 8(1), April 2022.
- [24] Haotian Xu, Moritz Flaschel, and Laura De Lorenzis. Discovering non-associated pressure-sensitive plasticity models with euclid. *Advanced Modeling and Simulation in Engineering Sciences*, 12(1), January 2025.

- [25] Enzo Marino, Moritz Flaschel, Siddhant Kumar, and Laura De Lorenzis. Automated identification of linear viscoelastic constitutive laws with euclid. *Mechanics of Materials*, 181:104643, 2023.
- [26] Moritz Flaschel, Siddhant Kumar, and Laura De Lorenzis. Automated discovery of generalized standard material models with euclid. *Computer Methods in Applied Mechanics and Engineering*, 405:115867, 2023.
- [27] Bernard Halphen and Quoc Son Nguyen. Sur les matériaux standard généralisés. *Journal de mécanique*, 14(1):39–63, 1975.
- [28] Klaus Hackl. Generalized standard media and variational principles in classical and finite strain elastoplasticity. *Journal of the Mechanics and Physics of Solids*, 45(5):667–688, 1997.
- [29] Moritz Flaschel, Paul Steinmann, Laura De Lorenzis, and Ellen Kuhl. Convex neural networks learn generalized standard material models. *Journal of the Mechanics and Physics of Solids*, 200:106103, 2025.
- [30] Shenglin Huang, Zequn He, Bryan Chem, and Celia Reina. Variational onsager neural networks (vonns): A thermodynamics-based variational learning strategy for non-equilibrium pdes. *Journal of the Mechanics and Physics of Solids*, 163:104856, 2022.
- [31] Hagen Holthusen, Lukas Lamm, Tim Brepols, Stefanie Reese, and Ellen Kuhl. Theory and implementation of inelastic constitutive artificial neural networks. *Computer Methods in Applied Mechanics and Engineering*, 428:117063, 2024.
- [32] Antoine Benady, Emmanuel Baranger, and Ludovic Chamoin. Unsupervised learning of history-dependent constitutive material laws with thermodynamically-consistent neural networks in the modified constitutive relation error framework. *Computer Methods in Applied Mechanics and Engineering*, 425:116967, 2024.
- [33] Max Rosenkranz, Karl A. Kalina, Jörg Brummund, WaiChing Sun, and Markus Kästner. Viscoelasticity with physics-augmented neural networks: model formulation and training methods without prescribed internal variables. *Computational Mechanics*, 74(6):1279–1301, Dec 2024.
- [34] Wulf G. Dettmer, Eugenio J. Muttio, Reem Alhayki, and Djordje Perić. A framework for neural network based constitutive modelling of inelastic materials. *Computer Methods in Applied Mechanics and Engineering*, 420:116672, 2024.
- [35] Faisal As’ad and Charbel Farhat. A mechanics-informed deep learning framework for data-driven nonlinear viscoelasticity. *Computer Methods in Applied Mechanics and Engineering*, 417:116463, 2023.
- [36] Hagen Holthusen, Kevin Linka, Ellen Kuhl, and Tim Brepols. A generalized dual potential for inelastic constitutive artificial neural networks: A jax implementation at finite strains. *Journal of the Mechanics and Physics of Solids*, page 106337, 2025.
- [37] Asghar Arshad Jadoon, Knut Andreas Meyer, and Jan Niklas Fuhg. Automated model discovery of finite strain elastoplasticity from uniaxial experiments. *Computer Methods in Applied Mechanics and Engineering*, 435:117653, 2025.
- [38] Birte Boes, Jaan-Willem Simon, and Hagen Holthusen. Accounting for plasticity: An extension of inelastic constitutive artificial neural networks, 2025.
- [39] Franz Dammaß, Karl A. Kalina, and Markus Kästner. Neural networks meet phase-field: A hybrid fracture model. *Computer Methods in Applied Mechanics and Engineering*, 440:117937, 2025.
- [40] Hagen Holthusen, Tim Brepols, Kevin Linka, and Ellen Kuhl. Automated model discovery for tensional homeostasis: Constitutive machine learning in growth and remodeling. *Computers in Biology and Medicine*, 186:109691, March 2025.
- [41] Asghar A. Jadoon, Ravi G. Patel, Brian N. Granzow, Reese E. Jones, D. Thomas Seidl, and Jan N. Fuhg. Thermodynamically consistent hybrid and permutation-invariant neural yield functions for anisotropic plasticity, 2025.
- [42] Brandon Amos, Lei Xu, and J. Zico Kolter. Input convex neural networks, 2017.
- [43] Sanjeev Kumar, D. Thomas Seidl, Brian N. Granzow, Jin Yang, and Jan Niklas Fuhg. A comparative study of calibration techniques for finite strain elastoplasticity: Numerically-exact sensitivities for femu and vfm. *Computer Methods in Applied Mechanics and Engineering*, 444:118159, 2025.
- [44] Reese E. Jones, Ari L. Frankel, and K. L. Johnson. A neural ordinary differential equation framework for modeling inelastic stress response via internal state variables. *Journal of Machine Learning for Modeling and Computing*, 3(3):1–35, 2022.
- [45] Reese E. Jones and Jan N. Fuhg. An attention-based neural ordinary differential equation framework for modeling inelastic processes, 2025.

- [46] L. Borkowski, C. Sorini, and A. Chattopadhyay. Recurrent neural network-based multiaxial plasticity model with regularization for physics-informed constraints. *Computers & Structures*, 258:106678, 2022.
- [47] Jin-Zhao Li, Zhi-Ping Guan, Jiong-Rui Chen, and Hui-Chao Jin. A long short-term memory-based constitutive modeling framework for capturing strain path dependence in plastic deformation. *Mechanics of Materials*, 205:105325, 2025.
- [48] Binyao Guo, Zihan Lin, and QiZhi He. History-aware neural operator: Robust data-driven constitutive modeling of path-dependent materials. *Computer Methods in Applied Mechanics and Engineering*, 447:118358, 2025.
- [49] Marco Maurizi, Chao Gao, and Filippo Berto. Predicting stress, strain and deformation fields in materials and structures with graph neural networks. *Scientific Reports*, 12(1), December 2022.
- [50] Ravi G. Patel, Reese E. Jones, D. Thomas Seidl, Brian N. Granzow, and Jan N. Fuhg. A general, automated method for building structural tensors of arbitrary order for anisotropic function representations, 2025.
- [51] Vahidullah Tac, Francisco Sahli Costabal, and Adrian B. Tepole. Data-driven tissue mechanics with polyconvex neural ordinary differential equations. *Computer Methods in Applied Mechanics and Engineering*, 398:115248, 2022.
- [52] Jan N. Fuhg and Nikolaos Bouklas. On physics-informed data-driven isotropic and anisotropic constitutive models through probabilistic machine learning and space-filling sampling. *Computer Methods in Applied Mechanics and Engineering*, 394:114915, 2022.
- [53] Karl A. Kalina, Jörg Brummund, WaiChing Sun, and Markus Kästner. Neural networks meet anisotropic hyperelasticity: A framework based on generalized structure tensors and isotropic tensor functions. *Computer Methods in Applied Mechanics and Engineering*, 437:117725, 2025.
- [54] Jeremy A. McCulloch and Ellen Kuhl. Automated model discovery for textile structures: The unique mechanical signature of warp knitted fabrics. *Acta Biomaterialia*, 189:461–477, 2024.
- [55] Vahidullah Tac, Vivek D. Sree, Manuel K. Rausch, and Adrian B. Tepole. Data-driven modeling of the mechanical behavior of anisotropic soft biological tissue. *Engineering with Computers*, 38(5):4167–4182, September 2022.
- [56] Osman Gültekin, Ahmad Moeineddin, Barış Cansız, Krunoslav Sveric, Axel Linke, and Michael Kaliske. A physics-informed neural network model for the anisotropic hyperelasticity of the human passive myocardium. *International Journal for Numerical Methods in Engineering*, 126(14):e70067, 2025.
- [57] Denisa Martonová, Sigrid Leyendecker, Gerhard A. Holzapfel, and Ellen Kuhl. Discovering dispersion: How robust is automated model discovery for human myocardial tissue? *Biomechanics and Modeling in Mechanobiology*, August 2025.
- [58] Thibault Vervenne, Mathias Peirlinck, Nele Famaey, and Ellen Kuhl. Constitutive neural networks for main pulmonary arteries: discovering the undiscovered. *Biomechanics and Modeling in Mechanobiology*, 24(2):615–634, February 2025.
- [59] Jan Niklas Fuhg, Govinda Anantha Padmanabha, Nikolaos Bouklas, Bahador Bahmani, WaiChing Sun, Nikolaos N. Vlassis, Moritz Flaschel, Pietro Carrara, and Laura De Lorenzis. A review on data-driven constitutive laws for solids, 2024.
- [60] Joe Watson, Chen Song, Oliver Weeger, Theo Gruner, An T. Le, Kay Pompetzki, Ahmed Hendawy, Oleg Arenz, Will Trojak, Miles Cranmer, Carlo D’Eramo, Fabian Bülow, Tanmay Goyal, Jan Peters, and Martin W. Hoffman. Machine learning with physics knowledge for prediction: A survey, 2025.
- [61] P. Germain, Q. S. Nguyen, and P. Suquet. Continuum thermodynamics. *Journal of Applied Mechanics*, 50(4b):1010–1020, December 1983.
- [62] Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
- [63] Joseph Sill. Monotonic networks. In *Neural Information Processing Systems*, 1997.
- [64] Hennie Daniels and Marina Velikova. Monotone and partially monotone neural networks. *IEEE Transactions on Neural Networks*, 21(6):906–917, June 2010.
- [65] Seungil You, David Ding, Kevin Canini, Jan Pfeifer, and Maya Gupta. Deep lattice networks and partial monotonic functions, 2017.
- [66] Faisal As’ad and Charbel Farhat. A mechanics-informed neural network framework for data-driven nonlinear viscoelasticity. In *AIAA SCITECH 2023 Forum*. American Institute of Aeronautics and Astronautics, January 2023.
- [67] Ramin Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. Liquid time-constant networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(9):7657–7666, May 2021.

- [68] S.M. Cox and P.C. Matthews. Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2):430–455, March 2002.
- [69] Mikhail Itskov. On the application of the additive decomposition of generalized strain measures in large strain plasticity. *Mechanics Research Communications*, 31(5):507–517, 2004.
- [70] Carl Eckart. The thermodynamics of irreversible processes. iv. the theory of elasticity and anelasticity. *Physical Review*, 73(4):373–382, February 1948.
- [71] Ekkehart Kröner. Allgemeine kontinuumstheorie der versetzungen und eigenspannungen. *Archive for Rational Mechanics and Analysis*, 4(1):273–334, January 1959.
- [72] François Sidoroff. Un modèle viscoélastique non linéaire avec configuration intermédiaire. *J. Mécanique*, 13:679–713, 1974.
- [73] Edward K. Rodriguez, Anne Hoger, and Andrew D. McCulloch. Stress-dependent finite growth in soft elastic tissues. *Journal of Biomechanics*, 27(4):455–467, April 1994.
- [74] Daniel Christ and Stefanie Reese. A finite element model for shape memory alloys considering thermomechanical couplings at large strains. *International Journal of Solids and Structures*, 46(20):3694–3709, October 2009.
- [75] James Casey. A convenient form of the multiplicative decomposition of the deformation gradient. *Mathematics and Mechanics of Solids*, 22(3):528–537, August 2016.
- [76] Jean-Paul Boehler. A simple derivation of representations for non-polynomial constitutive equations in some cases of anisotropy. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 59(4):157–167, 1979.
- [77] A. J. M. Spencer. *Constitutive Theory for Strongly Anisotropic Solids*, page 1–32. Springer Vienna, 1984.
- [78] Jörg Schröder and Patrizio Neff. Invariant formulation of hyperelastic transverse isotropy based on polyconvex free energy functions. *International Journal of Solids and Structures*, 40(2):401–445, January 2003.
- [79] John M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Archive for Rational Mechanics and Analysis*, 63(4):337–403, December 1976.
- [80] John M Ball. Constitutive inequalities and existence theorems in nonlinear elastostatics. In *Nonlinear analysis and mechanics: Heriot-Watt symposium*, volume 1, pages 187–241. Pitman London, 1977.
- [81] Jerrold E Marsden and Thomas JR Hughes. *Mathematical foundations of elasticity*. Courier Corporation, 1994.
- [82] Bob Svendsen. On the modelling of anisotropic elastic and inelastic material behaviour at large deformation. *International Journal of Solids and Structures*, 38(52):9579–9599, December 2001.
- [83] Bernard D. Coleman and Walter Noll. Foundations of linear viscoelasticity. *Reviews of Modern Physics*, 33(2):239–249, April 1961.
- [84] Bernard D. Coleman and Walter Noll. The thermodynamics of elastic materials with heat conduction and viscosity. *Archive for Rational Mechanics and Analysis*, 13(1):167–178, December 1963.
- [85] Bernard D. Coleman and Morton E. Gurtin. Thermodynamics with internal state variables. *The Journal of Chemical Physics*, 47(2):597–613, July 1967.
- [86] Wulf Dettmer and Stefanie Reese. On the theoretical and numerical modelling of armstrong–frederick kinematic hardening in the finite strain regime. *Computer Methods in Applied Mechanics and Engineering*, 193(1–2):87–116, January 2004.
- [87] S. Reese, T. Brepols, M. Fassin, L. Poggenpohl, and S. Wulfinghoff. Using structural tensors for inelastic material modeling in the finite strain regime – a novel approach to anisotropic damage. *Journal of the Mechanics and Physics of Solids*, 146:104174, January 2021.
- [88] J R Rice. Inelastic constitutive relations for solids: An internal-variable theory and its application to metal plasticity. *J. Mech. Phys. Solids*, 19(6):433–455, November 1971.
- [89] Paul Germain. Functional concepts in continuum mechanics. *Meccanica*, 33(5):433–444, October 1998.
- [90] Y. F. Dafalias. Issues on the constitutive formulation at large elastoplastic deformations, part 1: Kinematics. *Acta Mechanica*, 69(1–4):119–138, December 1987.
- [91] Stefanie Reese. Meso-macro modelling of fibre-reinforced rubber-like composites exhibiting large elastoplastic deformation. *International Journal of Solids and Structures*, 40(4):951–980, February 2003.
- [92] Birte Boes, Jaan-Willem Simon, Stefanie Reese, and Hagen Holthausen. A novel continuum mechanical framework for decoupled material behavior in thickness and in-plane directions. *Computer Methods in Applied Mechanics and Engineering*, 415:116192, October 2023.

- [93] Q.-S. Zheng. Theory of representations for tensor functions—a unified invariant approach to constitutive equations. *Applied Mechanics Reviews*, 47(11):545–587, November 1994.
- [94] Rainer Glüge and Sara Bucci. Does convexity of yield surfaces in plasticity have a physical significance? *Mathematics and Mechanics of Solids*, 23(9):1364–1373, August 2017.
- [95] Adam Janečka and Michal Pavelka. Non-convex dissipation potentials in multiscale non-equilibrium thermodynamics. *Continuum Mechanics and Thermodynamics*, 30(4):917–941, April 2018.
- [96] J.H. Argyris, G. Faust, J. Szimmat, E.P. Warnke, and K.J. Willam. Recent developments in the finite element analysis of prestressed concrete reactor vessels. *Nuclear Engineering and Design*, 28(1):42–75, July 1974.
- [97] Hagen Holthusen, Christiane Rothkranz, Lukas Lamm, Tim Brepols, and Stefanie Reese. Inelastic material formulations based on a co-rotated intermediate configuration—application to bioengineered tissues. *Journal of the Mechanics and Physics of Solids*, 172:105174, March 2023.
- [98] Zeng Liu, Rogelio Ortigosa, Antonio J. Gil, and Javier Bonet. Large strain constitutive modelling of soft compressible and incompressible solids: Generalised isotropic and anisotropic viscoelasticity. *Journal of the Mechanics and Physics of Solids*, 203:106194, 2025.
- [99] Dominik K. Klein, Rogelio Ortigosa, Jesús Martínez-Frutos, and Oliver Weeger. Finite electro-elasticity with physics-augmented neural networks. *Computer Methods in Applied Mechanics and Engineering*, 400:115501, 2022.
- [100] Iwaylo N. Vladimirov, Michael P. Pietryga, and Stefanie Reese. On the modelling of non-linear kinematic hardening at finite strains with application to springback—comparison of time integration algorithms. *International Journal for Numerical Methods in Engineering*, 75(1):1–28, 2008.
- [101] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus), 2023.
- [102] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necoara, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. Jax: composable transformations of Python+NumPy programs. *GitHub repository*, 2018.
- [103] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79(11):1309–1331, 2009.
- [104] Utkarsh Ayachit. The paraview guide: A parallel visualization application. 2015.
- [105] J. Schröder, P. Neff, and V. Ebbing. Anisotropic polyconvex energies on the basis of crystallographic motivated structural tensors. *Journal of the Mechanics and Physics of Solids*, 56(12):3486–3506, 2008.
- [106] Stefanie Reese and Sanjay Govindjee. A theory of finite viscoelasticity and numerical aspects. *International Journal of Solids and Structures*, 35(26):3455–3482, 1998.
- [107] Karl A. Kalina, Lennart Linden, Jörg Brummund, Philipp Metsch, and Markus Kästner. Automated constitutive modeling of isotropic hyperelasticity based on artificial neural networks. *Computational Mechanics*, 69(1):213–232, October 2021.
- [108] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.
- [109] Lucy M. Wang, Kevin Linka, and Ellen Kuhl. Automated model discovery for muscle using constitutive recurrent neural networks. *Journal of the Mechanical Behavior of Biomedical Materials*, 145:106021, 2023.
- [110] Robert D. Cook, David S. Malkus, and Michael E. Plesha. *Concepts and Applications of Finite Element Analysis*. John Wiley & Sons, 2nd edition, 1981.
- [111] Jiashen Guan, Xin Li, Hongyan Yuan, and Ju Liu. Hyperelastic modeling based on generalized landau invariants and multi-stage calibration. *Journal of the Mechanics and Physics of Solids*, 206:106338, 2026.
- [112] Afshin Anssari-Benam, Alain Goriely, and Giuseppe Saccomandi. Generalised invariants and pseudo-universal relationships for hyperelastic materials: A new approach to constitutive modelling. *Journal of the Mechanics and Physics of Solids*, 193:105883, 2024.
- [113] Denisa Martonová, Alain Goriely, and Ellen Kuhl. Generalized invariants meet constitutive neural networks: A novel framework for hyperelastic materials. *Journal of the Mechanics and Physics of Solids*, page 106352, 2025.
- [114] David Wiedemann and Malte A. Peter. Characterization of polyconvex isotropic functions, 2023.
- [115] Gian-Luca Geuken, Patrick Kurzeja, David Wiedemann, and Jörn Mosler. A novel neural network for isotropic polyconvex hyperelasticity satisfying the universal approximation theorem. *Journal of the Mechanics and Physics of Solids*, 203:106209, 2025.

- [116] François Charton, Jordan S. Ellenberg, Adam Zsolt Wagner, and Geordie Williamson. Patternboost: Constructions in mathematics with a little help from ai, 2024.
- [117] E. Prume, C. Gierden, M. Ortiz, and S. Reese. Direct data-driven algorithms for multiscale mechanics. *Computer Methods in Applied Mechanics and Engineering*, 433:117525, 2025.
- [118] Laurent Stainier, Adrien Leygue, and Michael Ortiz. Model-free data-driven methods in mechanics: material data identification and solvers. *Computational Mechanics*, 64(2):381–393, June 2019.
- [119] Lennart Linden, Karl A. Kalina, Jörg Brummund, Brain Riemer, and Markus Kästner. A dual-stage constitutive modeling framework based on finite strain data-driven identification and physics-augmented neural networks, 2025.
- [120] Moritz Flaschel, Denisa Martonová, Carina Veil, and Ellen Kuhl. Material fingerprinting: A shortcut to material model discovery without solving optimization problems, 2025.
- [121] Adnan Egthesad, Jingye Tan, Jan Niklas Fuhg, and Nikolaos Bouklas. Nn-evp: A physics informed neural network-based elasto-viscoplastic framework for predictions of grain size-aware flow response. *International Journal of Plasticity*, 181:104072, 2024.
- [122] Moritz Flaschel, Huitian Yu, Nina Reiter, Jan Hinrichsen, Silvia Budday, Paul Steinmann, Siddhant Kumar, and Laura De Lorenzis. Automated discovery of interpretable hyperelastic material models for human brain tissue with euclid. *Journal of the Mechanics and Physics of Solids*, 180:105404, 2023.
- [123] Matt Poyser and Toby P. Breckon. Neural architecture search: A contemporary literature review for computer vision applications. *Pattern Recognition*, 147:110052, 2024.