Computer Methods in Applied Mechanics and Engineering 450 (2026) 118612

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

journal homepage: www.elsevier.com/locate/cma

A complement to neural networks for anisotropic inelasticity at finite strains

Hagen Holthusen [a,*], Ellen Kuhl [b]

[a] Institute of Applied Mechanics, University of Erlangen-Nuremberg, Erlangen, 91058, Germany
[b] Department of Mechanical Engineering, Stanford University, Stanford, CA, 94305, United States

ARTICLE INFO

Keywords:
Neural networks
Inelasticity
Anisotropy
Finite strains
Automated model discovery
Finite element simulation

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 circumvent the use of exponential-map time integration during training–which often leads to numerical instabilities–we employ recurrent Liquid Neural Networks as an auxiliary architecture. During inference, however, the exponential-map update is reinstated to ensure admissibility of the inelastic variables.

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.

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 [4–6], and multiphysics problems [7,8], to model order reduction [9,10], the identification of material behavior [11–13], 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.

* Corresponding author.
E-mail addresses: hagen.holthusen@fau.de (H. Holthusen), ekuhl@stanford.edu (E. Kuhl).

https://doi.org/10.1016/j.cma.2025.118612
Received 5 October 2025; Received in revised form 17 November 2025; Accepted 25 November 2025

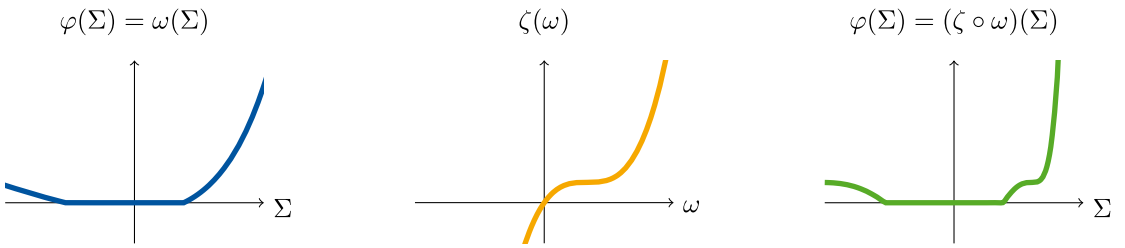

Fig. 1. Schematic illustration of possible constructions for a function $\varphi(\Sigma)$ satisfying InequalityEq. (3). The function ω is *convex*, *non-negative*, and *zero-valued at the origin*, whereas ζ is *monotonically increasing* and *zero-valued* with respect to ω. Their composition $(\zeta \circ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

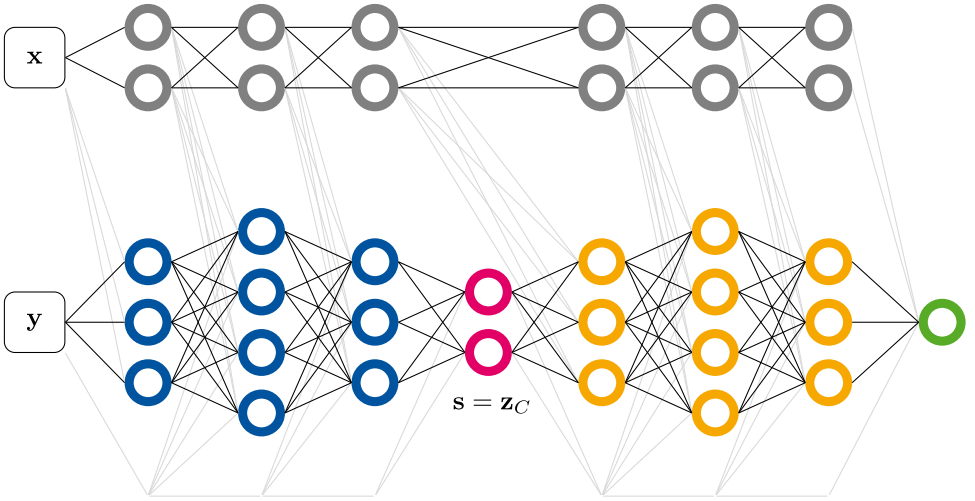

Fig. 2. Schematic illustration of the composition network $\mathcal{N}_\circ$. The blue part represents the Input Convex Neural Network $\mathcal{N}_c$, which is convex in **y** but not in **x**. Each layer takes as input the output of the previous layer and of the parallel network in gray carrying **x** as well as **y** itself. Its final output $\mathbf{z}_C$ highlighted in red serves as the input **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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

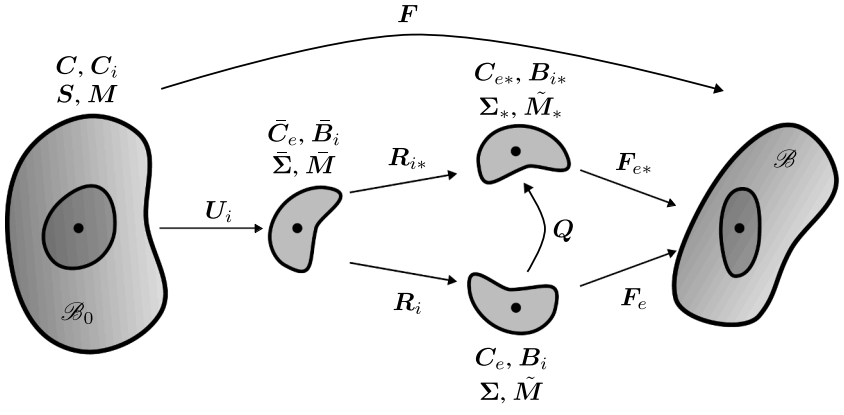

Fig. 3. Motion from the reference configuration $\mathscr{B}_0$ to the current configuration $\mathscr{B}$. An infinitesimal material element is described by the deformation gradient $\boldsymbol{F}$, which admits a multiplicative decomposition into elastic $\boldsymbol{F}_e$ and inelastic $\boldsymbol{F}_i$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

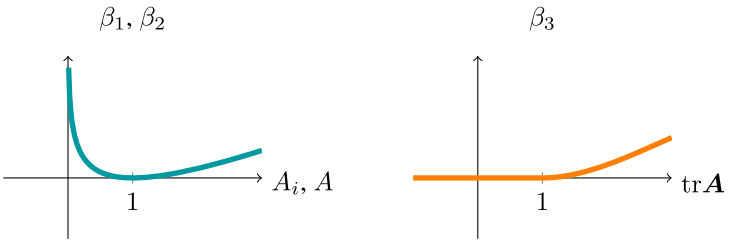

Fig. 4. Illustration of the normalization functions for the isotropic $\mathcal{S}_\psi$ and anisotropic $\mathcal{A}_\psi$ sets for the Helmholtz free energy ψ.

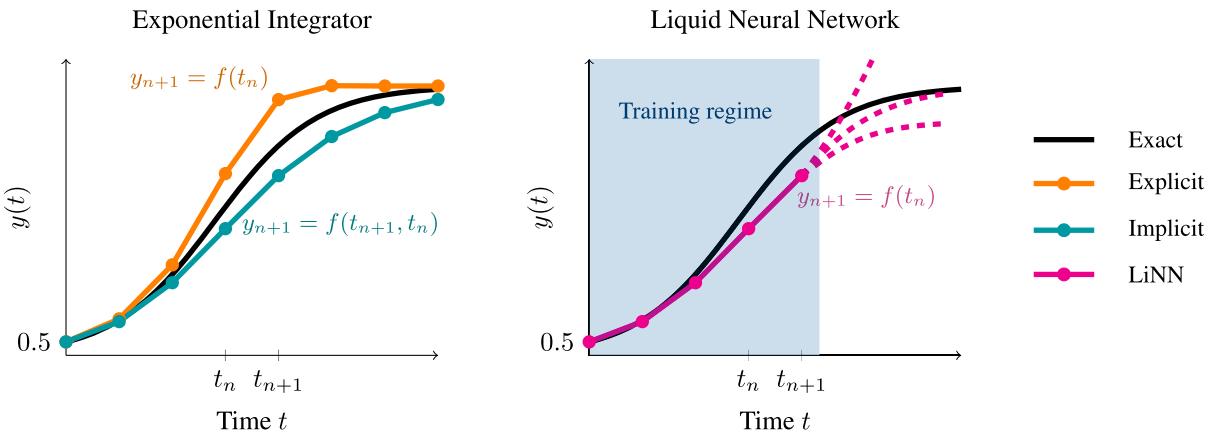

Fig. 5. Schematic of time discretizations of the ordinary differential equation $\dot{y} = A(y)\, y$ with $A(y) = r(1 - y/k)$, where $r = 0.7$ and $k = 10$. The exponential integrator with time step $\Delta t = 1.5$ shows the explicit scheme $y_{n+1} = \exp(\Delta t\, A(y_n))\, y_n$ and the implicit 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.

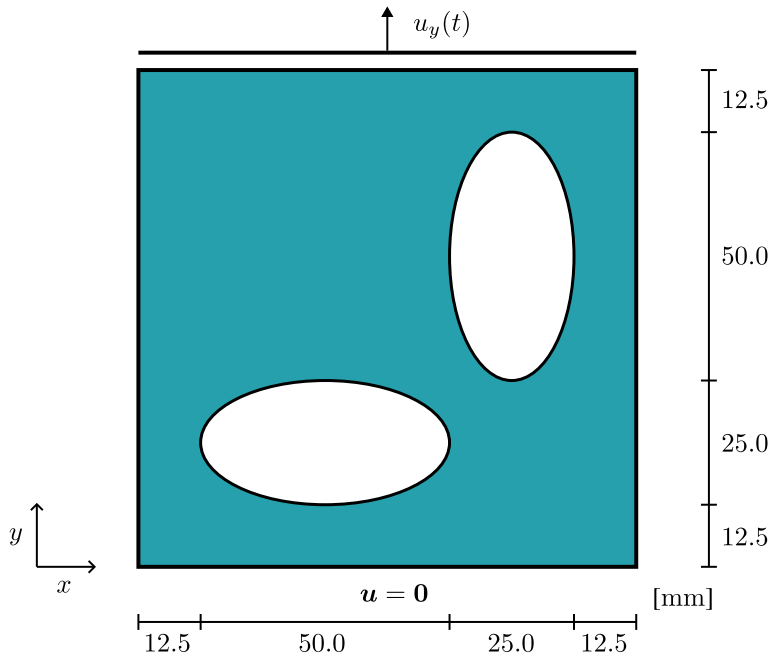

Fig. 6. Plate with two elliptic holes. Geometry and boundary conditions of the BVP used to generate the training data set, adapted from [23]. The bottom edge is clamped, while on the top edge a vertical displacement $u_y(t)$ is prescribed. The displacements in x and z directions are free on the top, 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.

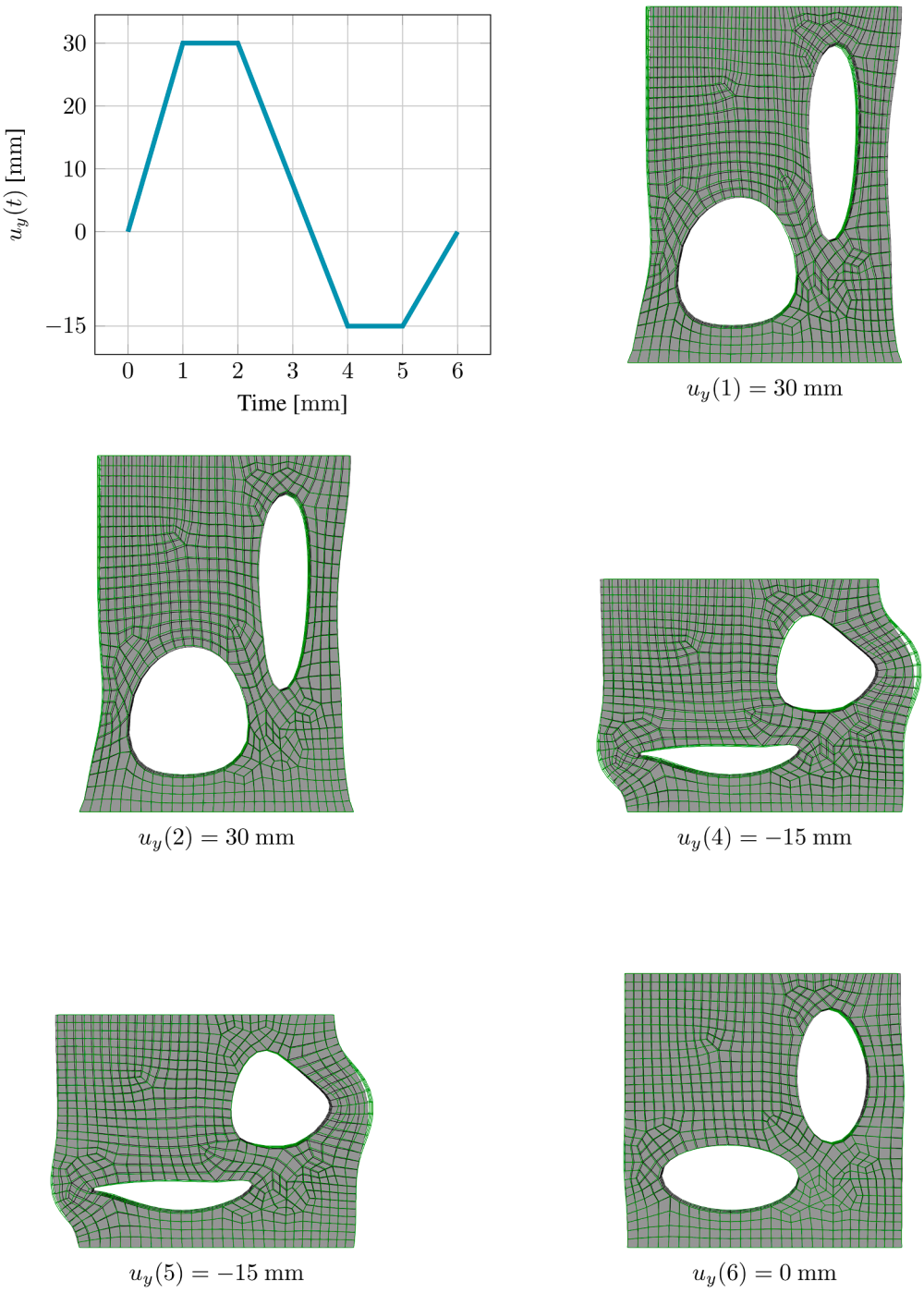

Fig. 7. Prescribed loading program and representative deformed configurations for isotropic (green mesh) and anisotropic (gray body/black mesh) materials. The comparison highlights the influence of anisotropy under otherwise identical loading. At the end of the program, although the external displacement returns to its initial value, residual strains remain due to inelasticity and the unloaded configurations are not identical; see also Fig. 8. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

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. Thermodynamics-based Artificial Neural Networks (TANNs) [19,20] might be considered counterparts at the material point level that integrate physical constraints, such as positivity of dissipation, into the loss function. An alternative paradigm designs neural network

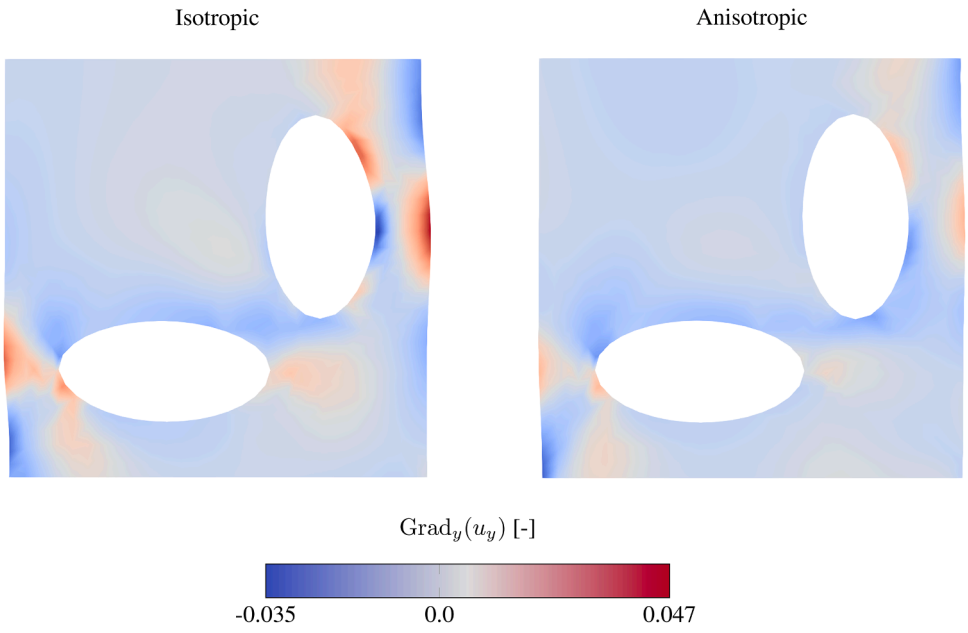

Fig. 8. Gradient of the vertical displacement u_y along the y-direction at the end of the loading program with $u_y(6) = 0$ mm. For a purely elastic material, the gradient vanishes throughout the specimen. In contrast, due to inelastic effects, a residual deformation remains for both the isotropic and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

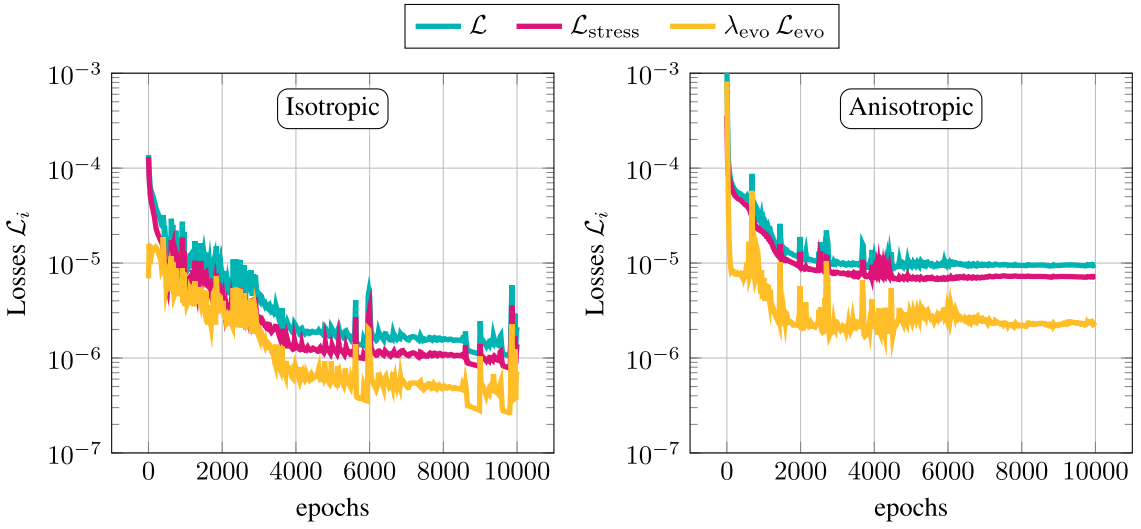

Fig. 9. Training losses over 10,000 epochs for isotropic and anisotropic data sets. The penalty parameter is $\lambda_{evo} = 1000$. No additional regularization is applied beyond gradient clipping.

architectures that satisfy physical principles *a priori*, thereby guaranteeing that predictions adhere to the underlying laws. Prominent approaches include the unsupervised EUCLID framework [11], 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.

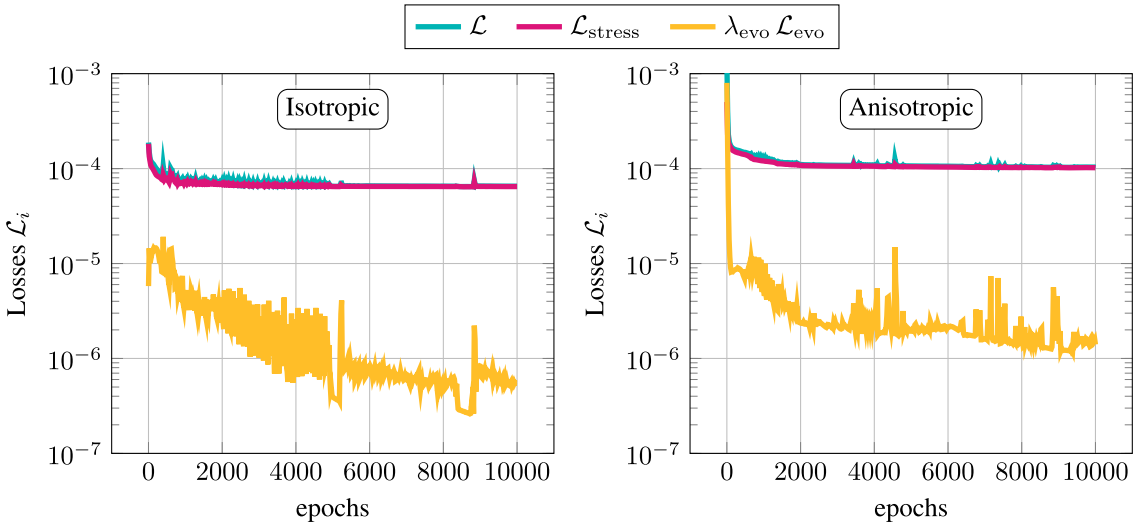

Fig. 10. Training losses over 10,000 epochs for isotropic and anisotropic data sets. The data set is corrupted by synthetic noise. The penalty parameter is $\lambda_{evo} = 1000$. No additional regularization is applied beyond gradient clipping.

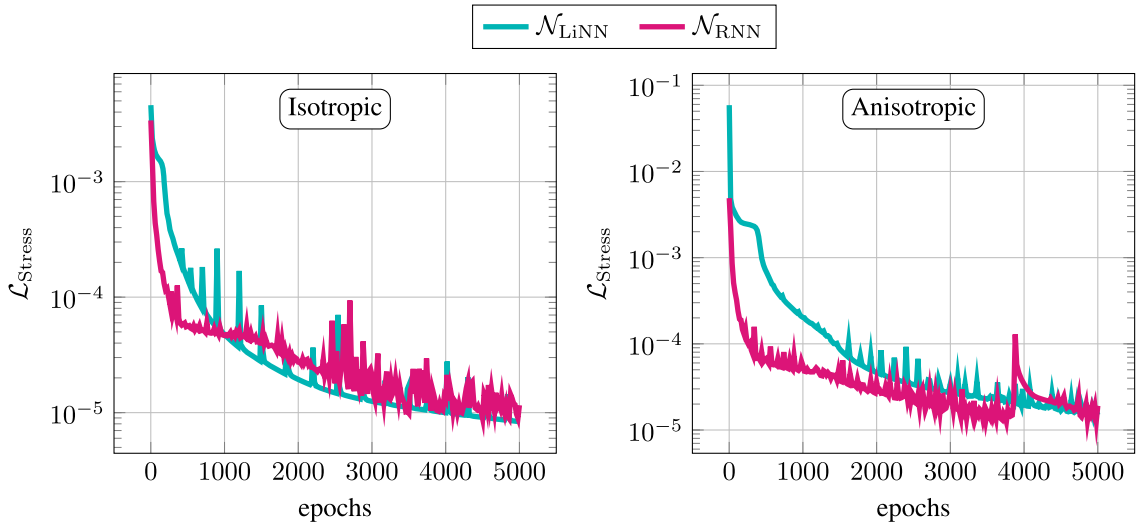

Fig. 11. Training losses over the epochs for isotropic and anisotropic data sets for the plain recurrent neural networks. The total number of epochs is 5000. The losses correspond to Eq. (42), which here coincides with the total loss. Neither gradient clipping nor additional regularization is applied.

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, Flaschel et al. [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, Flaschel et al. [29], Huang et al. [30], Holthusen et al. [31], Benady et al. [32], Rosenkranz et al. [33], Dettmer et al. [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],

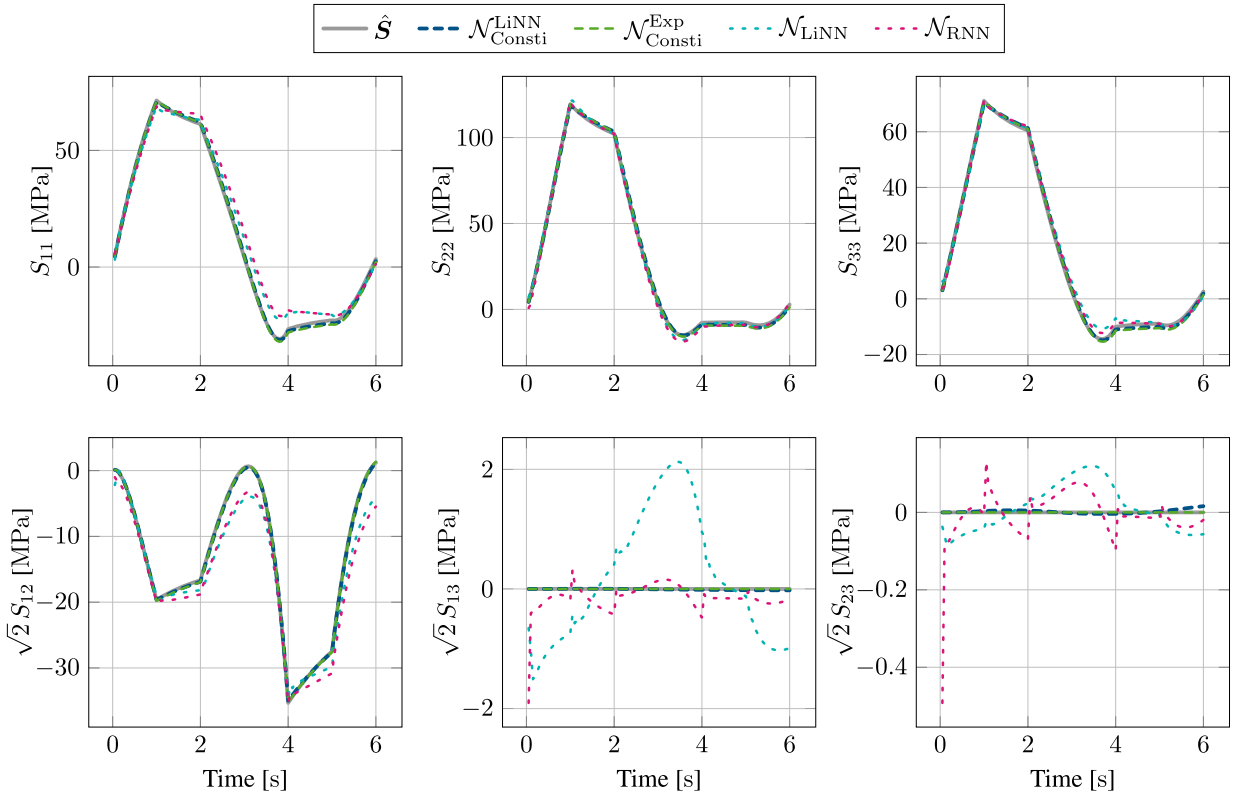

Fig. 12. Training results for $\mathcal{N}^{LiNN}_{Consti}$, $\mathcal{N}^{Exp}_{Consti}$, $\mathcal{N}_{RNN}$, and $\mathcal{N}_{LiNN}$. The reference model is isotropic, i.e., $\psi = \psi_{iso}$ and $\varphi = \varphi_{iso}$ with material parameters given in Table 1. Element ID: 111. Curves start with the first loading step at $t = 0.05$. Of all four networks, only the $\mathcal{N}^{Exp}_{Consti}$ network correctly captures the zero shear stress in the off-plane directions.

elasto-plasticity [37–39], fracture mechanics [40], and growth and remodeling phenomena [41]. 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 [42]: the first augments the classical Hill-48 yield criterion with an Input Convex Neural Network (ICNN) [43], 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. Moreover, the viscoelastic behavior of incompressible solids at finite strains was very recently investigated within the PANN framework in [44]. The study employs an implicit time-integration scheme based on an exponential map integrator and introduces a finite number of parallel decompositions of the deformation gradient into elastic and inelastic parts. Interestingly, during training, the initial number of parallel decompositions is automatically reduced to the minimum required through an additional sparsity-promoting loss term.

In the training of neural networks for inelastic material modeling, the availability of high-quality data is a central challenge. While most physics-embedded 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 [45].

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: Jones et al. [46] 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, Jones and Fuhg [47] 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 [48], 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 [49] 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.

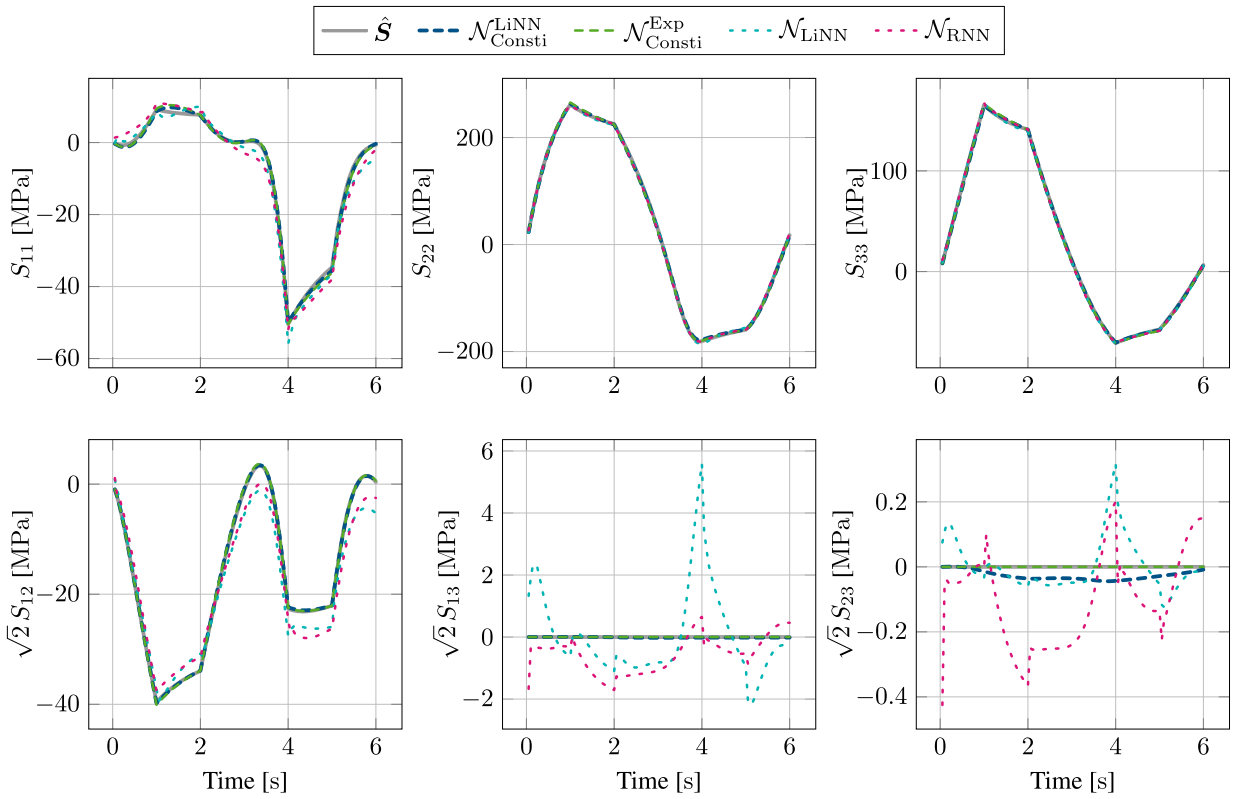

Fig. 13. Testing results for $\mathcal{N}^{LiNN}_{Consti}$, $\mathcal{N}^{Exp}_{Consti}$, $\mathcal{N}_{RNN}$, and $\mathcal{N}_{LiNN}$. The reference model is isotropic, i.e., $\psi = \psi_{iso}$ and $\varphi = \varphi_{iso}$ with material parameters given in Table 1. Element ID: 53. Curves start with the first loading step at $t = 0.05$. Of all four networks, only the $\mathcal{N}^{Exp}_{Consti}$ network correctly captures the zero shear stress in the off-plane directions.

A second direction emphasizes operator learning. The work of [50] 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: Maurizi et al. [51] 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 [52]. 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, Tac et al. [53] 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 [54], who employed Gaussian process regression and physics-informed sampling techniques to build anisotropic hyperelastic models that respect material symmetries and thermodynamic consistency. More recently, Kalina et al. [55] 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 [56], 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 [57], 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 [58,59] as well as the constitutive behavior of pulmonary arteries [60]. 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.

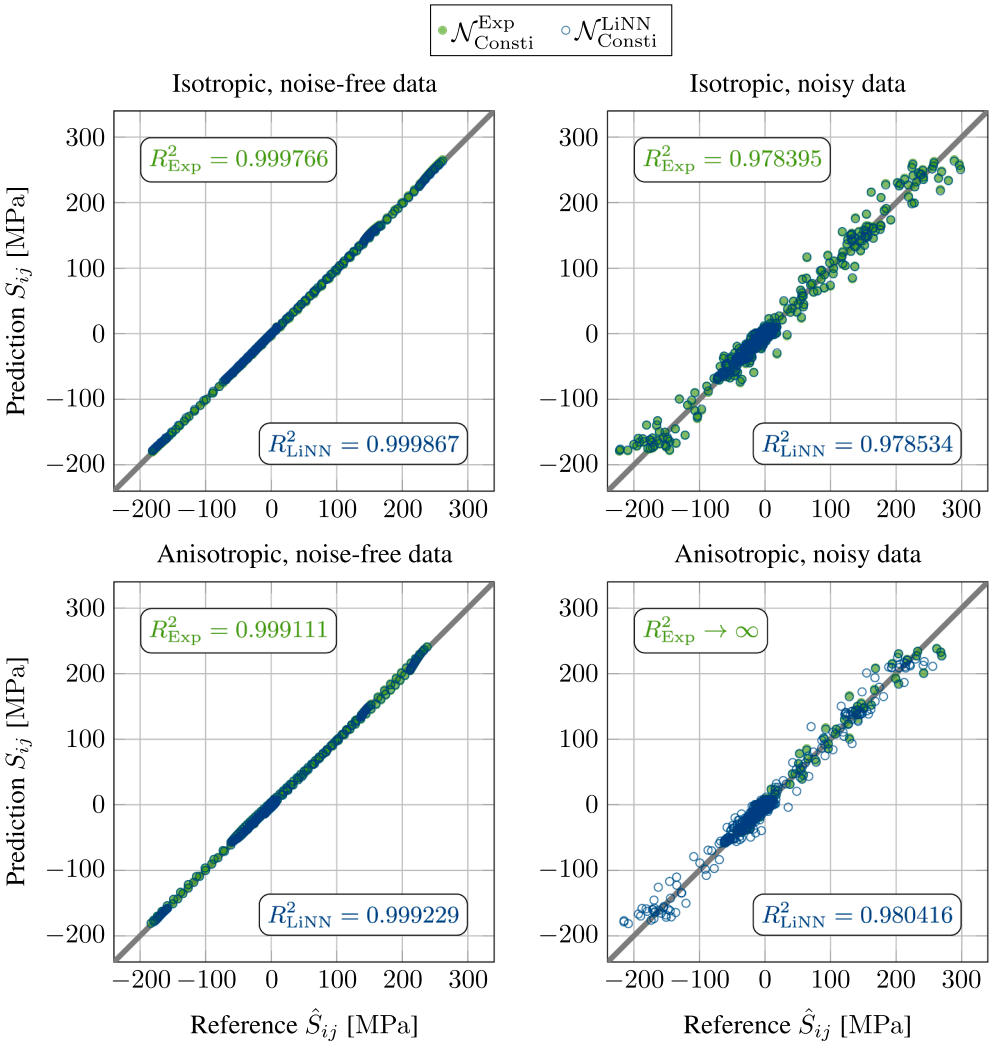

Fig. 14. Comparison between predicted S_{ij} and reference $\hat{S}_{ij}$ stress components across all data sets for the respective testing regimes. The shown data correspond to the Figs. 13, 16, 18, and 20. The solid gray line represents the ideal mapping $S_{ij} = \hat{S}_{ij}$. Results are shown for constitutive neural networks, $\mathcal{N}^{Exp}_{Consti}$ and $\mathcal{N}^{LiNN}_{Consti}$, respectively. The coefficients of determination R^2 indicate the predictive accuracy for each configuration. Note that, due to stress divergence in the anisotropic noisy data set, R^2 for $\mathcal{N}^{Exp}_{Consti}$ could not be obtained.

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 [61,62], 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.

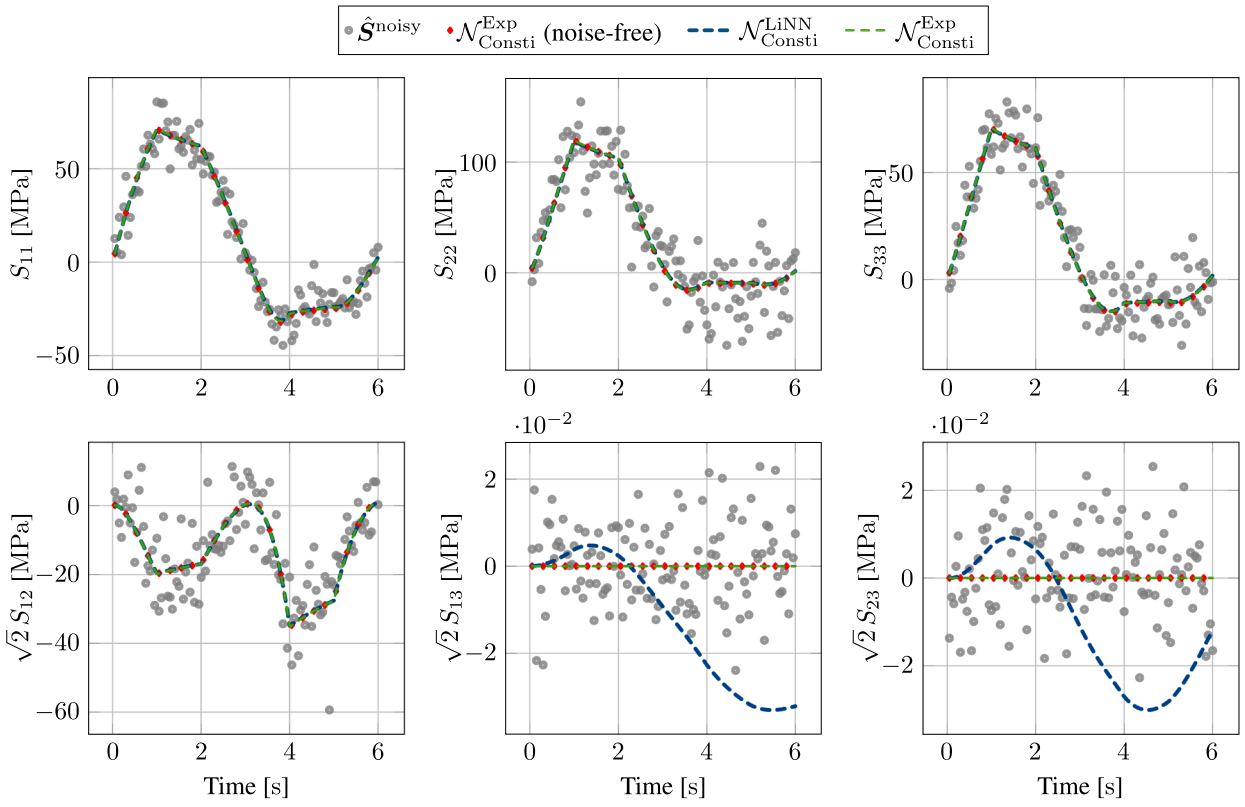

Fig. 15. Training results for the data set corrupted by noise for both $\mathcal{N}^{LiNN}_{Consti}$ and $\mathcal{N}^{Exp}_{Consti}$. The reference model is isotropic, i.e., $\psi = \psi_{iso}$ and $\varphi = \varphi_{iso}$ with material parameters given in Table 1. For comparison, the results of the noise-free training case are given by $\mathcal{N}^{Exp}_{Consti}$ (noise-free). Element ID: 111. Curves start with the first loading step at $t = 0.05$. The noise-free counterpart is depicted in Fig. 12.

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 architectured 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.

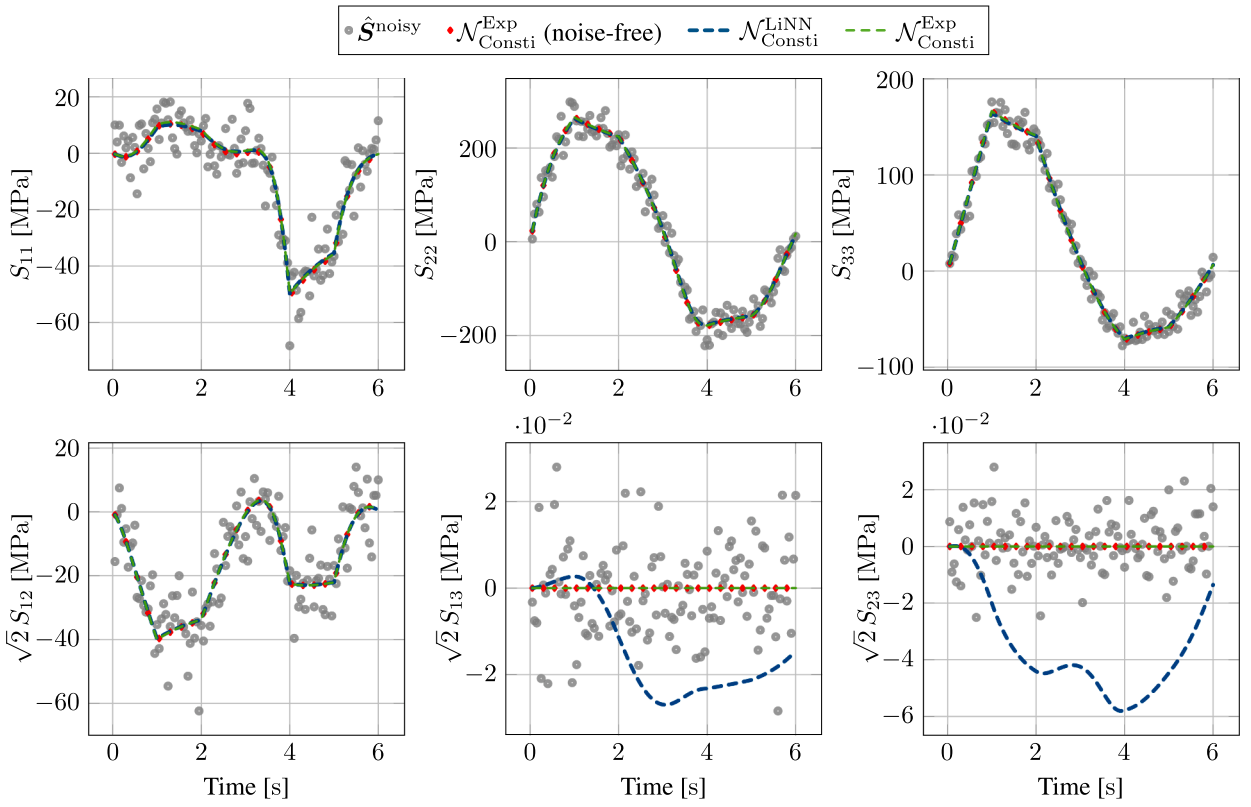

Fig. 16. Testing results for the data set corrupted by noise for both $\mathcal{N}^{\text{LiNN}}_{\text{Consti}}$ and $\mathcal{N}^{\text{Exp}}_{\text{Consti}}$. The reference model is isotropic, i.e., $\psi = \psi_{\text{iso}}$ and $\varphi = \varphi_{\text{iso}}$ with material parameters given in Table 1. For comparison, the results of the noise-free training case are given by $\mathcal{N}^{\text{Exp}}_{\text{Consti}}$ (noise-free). Element ID: 53. Curves start with the first loading step at $t = 0.05$. The noise-free counterpart is depicted in Fig. 13.

2. Preliminary problem description

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, $\boldsymbol{a}$, and $\boldsymbol{A}$ 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 $\mathbb{S}^{n-1} := \{\, \boldsymbol{v} \in \mathbb{R}^n : \|\boldsymbol{v}\| = 1 \,\}$ denotes the set of unit vectors. 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 $\mathrm{G}L(n,\mathbb{R}) = \{\, \boldsymbol{A} \in \mathbb{M}^{n\times n} : \det \boldsymbol{A} \neq 0 \,\}$, with the subgroup of orientation-preserving transformations $\mathrm{G}L^{+}(n,\mathbb{R}) = \{\, \boldsymbol{A} \in \mathrm{G}L(n,\mathbb{R}) : \det \boldsymbol{A} > 0 \,\}$. The set of symmetric matrices is denoted by $\mathrm{S}ym^{n\times n} = \{\, \boldsymbol{A} \in \mathbb{M}^{n\times n} : \boldsymbol{A}^T = \boldsymbol{A} \,\}$, with the subset of symmetric positive definite matrices $\mathrm{S}ym^{n\times n}_{+} = \{\, \boldsymbol{A} \in \mathrm{S}ym^{n\times n} : \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} > 0 \;\forall \boldsymbol{x} \neq \boldsymbol{0} \,\}$. Any two tensors $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{M}^{n\times n}$ related by $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{B}\boldsymbol{P}^{-1}$ for $\boldsymbol{P} \in \mathrm{G}L(n,\mathbb{R})$ are called *similar*, i.e., they share the same eigenvalues. The cofactor of a tensor $\boldsymbol{A} \in \mathrm{G}L(n,\mathbb{R})$ is defined as $\operatorname{cof} \boldsymbol{A} = \det(\boldsymbol{A})\, \boldsymbol{A}^{-T}$, and $\operatorname{cof} \boldsymbol{A} = \det(\boldsymbol{A})\, \boldsymbol{A}^{-1}$ if $\boldsymbol{A} \in \mathrm{S}ym^{n\times n}_{+}$. The set of orthogonal tensors is given by $\mathrm{O}(n) = \{\, \boldsymbol{Q} \in \mathbb{M}^{n\times n} : \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I},\ \det \boldsymbol{Q} = \pm 1 \,\}$, while the subgroup of proper orthogonal tensors (rotations) is denoted by $\mathrm{S}O(n) = \{\, \boldsymbol{Q} \in \mathrm{O}(n) : \det \boldsymbol{Q} = 1 \,\}$. The second-order identity tensor is expressed as $\boldsymbol{I} = \delta_{ij}\, \boldsymbol{e}_i \otimes \boldsymbol{e}_j$ with $\{\boldsymbol{e}_i\}$ 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 $\mathrm{d}ev$ is defined as $\mathrm{d}ev\boldsymbol{A} = \boldsymbol{A} - \frac{1}{3}\operatorname{tr}(\boldsymbol{A})\, \boldsymbol{I}$. The inner product of two objects is denoted by $\langle \cdot, \cdot \rangle$, and a superposed dot $(\dot{\bullet})$ 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.

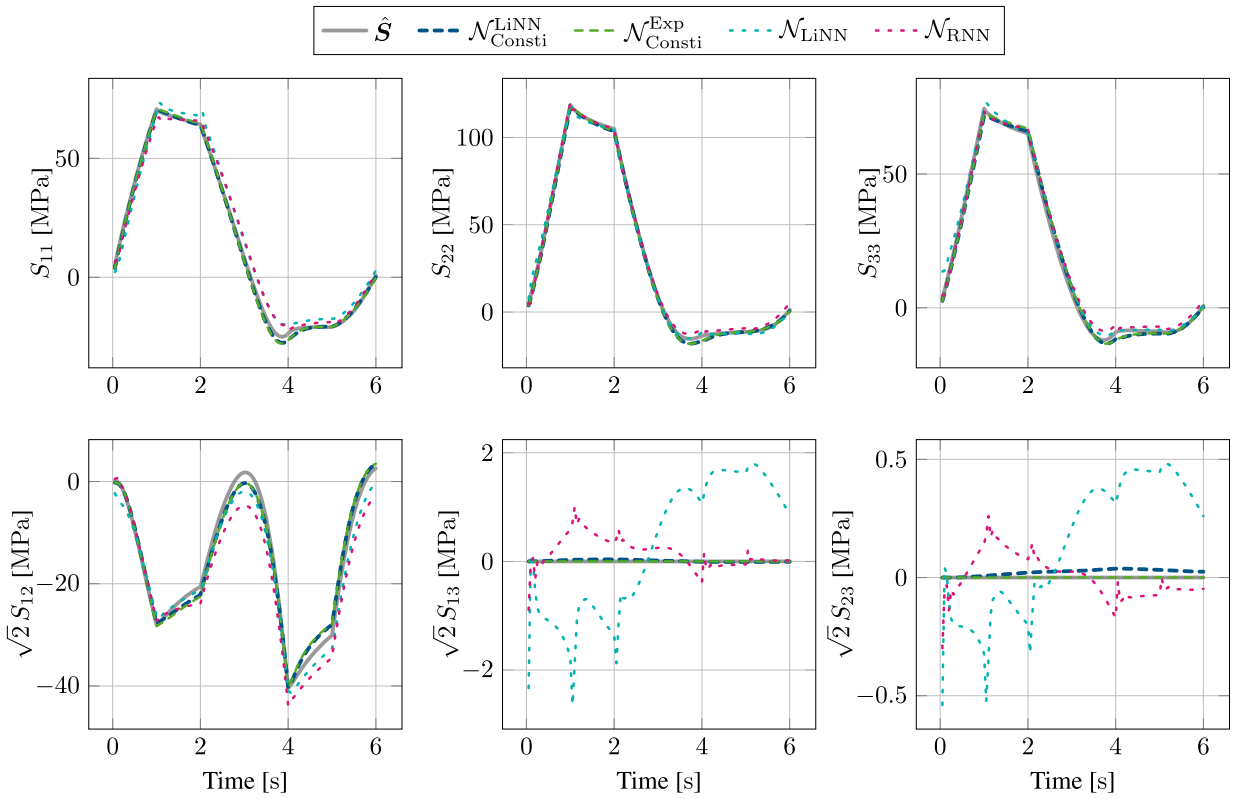

Fig. 17. Training results for $\mathcal{N}^{LiNN}_{Consti}$, $\mathcal{N}^{Exp}_{Consti}$, $\mathcal{N}_{RNN}$, and $\mathcal{N}_{LiNN}$. The reference model is anisotropic, i.e., $\psi = \psi_{iso} + \psi_{ani}$ and $\varphi = \varphi_{iso} + \varphi_{ani}$ with parameters given in Table 1. Element ID: 111. Curves start with the first loading step at $t = 0.05$. Of all four networks, only the $\mathcal{N}^{Exp}_{Consti}$ network correctly captures the zero shear stress in the off-plane directions.

2.2. One-dimensional illustrative example

We begin by defining two scalar-valued functions

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

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

$$\sigma = \frac{\partial \psi}{\partial \varepsilon}, \qquad \Sigma = -\frac{\partial \psi}{\partial \varepsilon_i}, \tag{1}$$

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

$$\Sigma\, \dot{\varepsilon}_i \geq 0. \tag{2}$$

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.,

$$\Sigma\, \dot{\varepsilon}_i = \Sigma\, \partial_\Sigma \varphi \geq 0. \tag{3}$$

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 inequalityEq. (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* [63]. Such a function ensures that the dissipation inequality is satisfied automatically, due to the properties of subgradients [64], namely

$$\omega(0) \geq \omega(\Sigma) - \langle \partial_\Sigma \omega, \Sigma \rangle, \tag{4}$$

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

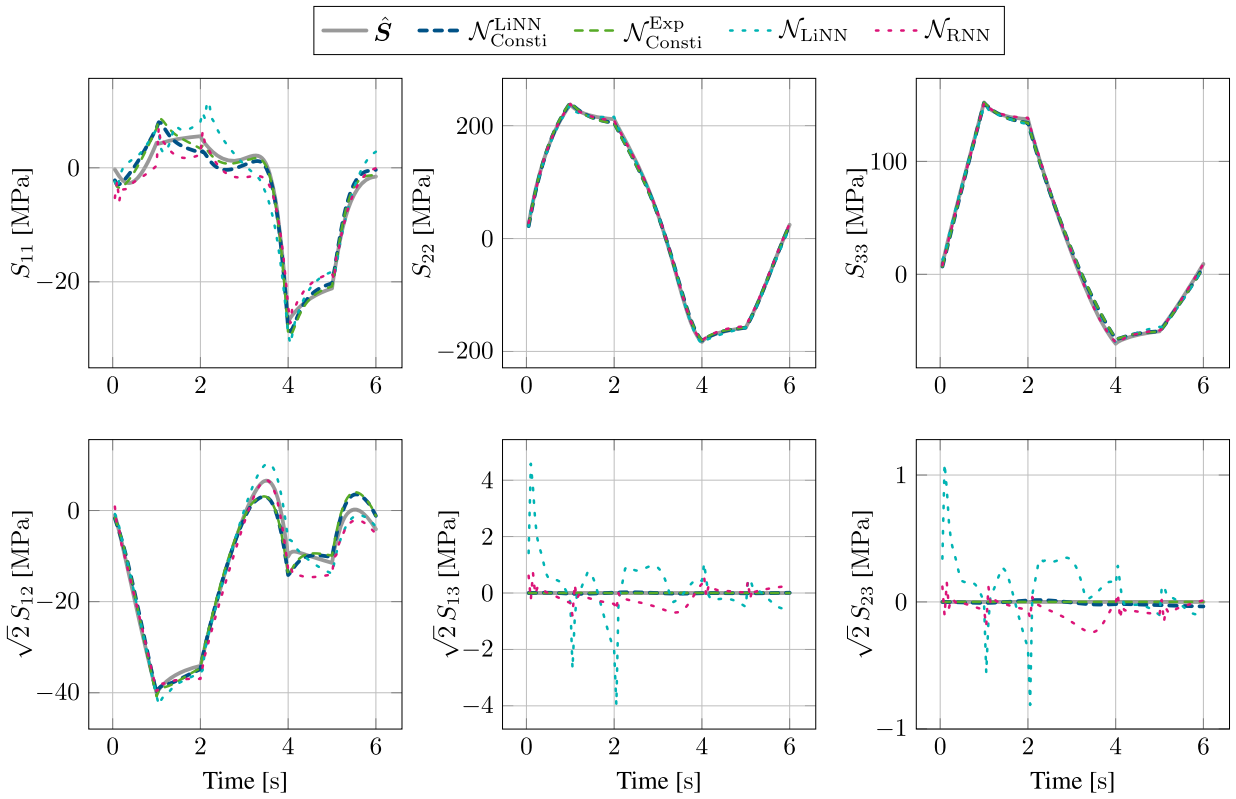

Fig. 18. Testing results for $\mathcal{N}^{LiNN}_{Consti}$, $\mathcal{N}^{Exp}_{Consti}$, $\mathcal{N}_{RNN}$, and $\mathcal{N}_{LiNN}$. The reference model is anisotropic, i.e., $\psi = \psi_{iso} + \psi_{ani}$ and $\varphi = \varphi_{iso} + \varphi_{ani}$ with parameters given in Table 1. Element ID: 53. Curves start with the first loading step at $t = 0.05$. Of all four networks, only the $\mathcal{N}^{Exp}_{Consti}$ network correctly captures the zero shear stress in the off-plane directions.

Monotonic functions. While convexity is *sufficient*[1], it is not *necessary* to guarantee positive dissipation. Indeed, Eq. (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 $\zeta : \mathbb{R} \to \mathbb{R}$ be a monotonically increasing, zero-valued function. We can then define

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

By the chain rule, the dissipation reads

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

which is guaranteed by the monotonicity of ζ with respect to ω and the convexity propertyEq. (4). This generalized construction allows greater flexibility in designing constitutive potentials while maintaining thermodynamic consistency. Fig. 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*.

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

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

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 [43]. 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 [43] reads

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

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 architectureEq. (6), yielding

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

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

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

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 formEq. (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:

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

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 Eq. (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 [65] and later extended in [66] to partially input specific monotonic networks. More recently, You et al. [67] 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 Fig. 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 $\mathcal{N}_\circ$ by composing ICNNs and IMNNs:

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

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 $\mathbf{z}_C$. Fig. 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 [68].

In this work, we employ *Liquid Neural Networks* (LiNNs) [69], 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

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

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

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

or by a (semi-)explicit exponential time integration scheme [70]

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

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 Eq. (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.

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 $\boldsymbol{n}^i$ perpendicular to a plane. For brevity, we derive the equations for one family of vectors $\boldsymbol{n}$ (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 $\boldsymbol{\chi} : \mathscr{B}_0 \subset \mathbb{R}^3 \to \mathscr{B}$ denote the motion that maps the reference configuration $\mathscr{B}_0$ onto the current configuration $\mathscr{B}$ in Euclidean space. The deformation is characterized by the deformation gradient

$$\boldsymbol{F} = \mathrm{Grad}\, \boldsymbol{\chi}, \quad J := \det \boldsymbol{F}.$$

To ensure an admissible, orientation-preserving motion one requires $J > 0$ in $\mathscr{B}_0$, such that $\boldsymbol{F} \in \mathrm{GL}^+(3, \mathbb{R})$. In the finite strain regime, inelastic effects are commonly described by postulating a multiplicative decomposition[3] of the deformation gradient [71–75], irrespective of the specific inelastic mechanism,

$$\boldsymbol{F} = \boldsymbol{F}_e \boldsymbol{F}_i = \boldsymbol{F}_{e*} \boldsymbol{F}_{i*}. \tag{14}$$

Here, $\boldsymbol{F}_e$ and $\boldsymbol{F}_i$ denote the elastic and inelastic parts, respectively. Constitutive admissibility requires $J_e := \det \boldsymbol{F}_e > 0$ and $J_i := \det \boldsymbol{F}_i > 0$. The decomposition is not unique, since any rotation $\boldsymbol{Q} \in \mathrm{SO}(3)$ yields an equivalent representation with $\boldsymbol{F}_{e*} = \boldsymbol{F}_e \boldsymbol{Q}^T$ and $\boldsymbol{F}_{i*} = \boldsymbol{Q} \boldsymbol{F}_i$ (see, e.g., Casey [76]). Since $\boldsymbol{F}_e, \boldsymbol{F}_i \in \mathrm{GL}^+(3, \mathbb{R})$, both admit polar decompositions, $\boldsymbol{F}_e = \boldsymbol{R}_e \boldsymbol{U}_e$ and $\boldsymbol{F}_i = \boldsymbol{R}_i \boldsymbol{U}_i$, with $\boldsymbol{R}_e, \boldsymbol{R}_i \in \mathrm{SO}(3)$ and $\boldsymbol{U}_e, \boldsymbol{U}_i \in \mathrm{Sym}^{3\times3}_+$. It follows that the elastic right stretch tensor in the rotated configuration, $\boldsymbol{U}_{e*} = \boldsymbol{Q} \boldsymbol{U}_e \boldsymbol{Q}^T$, is orthogonally similar to $\boldsymbol{U}_e$, while the inelastic stretch remains unchanged, $\boldsymbol{U}_{i*} = \boldsymbol{U}_i$. Fig. 3 shows the various mappings involved.

Material principles. For anisotropic materials, we introduce a symmetry class $\mathcal{S} \subset \mathrm{O}(3)$ consisting of all orthogonal tensors $\boldsymbol{Q}$ relative to which the material response remains invariant. In this work, following the structural-tensor approach (cf. [77,78]), we focus on a single preferred direction $\boldsymbol{n} \in \mathbb{S}^2$ in the reference configuration and set the structural tensor $\boldsymbol{M} = \boldsymbol{n} \otimes \boldsymbol{n}$. In analogy to [79], we introduce the group $\mathcal{G} = \{\boldsymbol{Q} \in \mathrm{SO}(3) : \boldsymbol{Q}\boldsymbol{n} = \boldsymbol{n}\}$, i.e., the set of proper rotations leaving $\boldsymbol{n}$ invariant. As mentioned at the beginning, we set $\mathcal{S} = \mathcal{G}$, and thus, the material is invariant about $\boldsymbol{n}$. For isotropic materials, in contrast, the symmetry class is $\mathcal{S} = \mathrm{SO}(3)$ (often $\mathrm{O}(3)$ is admitted as well).

[3] As shown in [126], the multiplicative decomposition is superior to the additive decomposition in case of finite elasto-plasticity regardless of the elastic material law.

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

$$\psi \,:\, \mathrm{G}L^{+}(3,\mathbb{R}) \times \mathrm{G}L^{+}(3,\mathbb{R}) \times \mathcal{M} \;\to\; \mathbb{R}, \qquad (\boldsymbol{F}, \boldsymbol{F}_i, \boldsymbol{M}) \;\mapsto\; \psi(\boldsymbol{F}, \boldsymbol{F}_i, \boldsymbol{M}),$$

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

- Objectivity (frame indifference)

$$\psi(\boldsymbol{F}, \boldsymbol{F}_i, \boldsymbol{M}) \;=\; \psi(\boldsymbol{Q}\boldsymbol{F}, \boldsymbol{F}_i, \boldsymbol{M}) \qquad \forall\, \boldsymbol{Q} \in \mathrm{SO}(3).$$

- Material symmetry

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

- Indifference to the choice of intermediate configuration

$$\psi(\boldsymbol{F}, \boldsymbol{F}_i, \boldsymbol{M}) \;=\; \psi(\boldsymbol{F},\; \boldsymbol{Q}\boldsymbol{F}_i,\; \boldsymbol{M}) \qquad \forall\, \boldsymbol{Q} \in \mathrm{SO}(3).$$

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

$$\psi(\boldsymbol{F}, \boldsymbol{F}_i, \boldsymbol{M}) \;=\; \psi(\boldsymbol{C}, \boldsymbol{C}_i, \boldsymbol{M}) \;=\; \psi\big(\boldsymbol{Q}\boldsymbol{C}\boldsymbol{Q}^T,\; \boldsymbol{Q}\boldsymbol{C}_i\boldsymbol{Q}^T,\; \boldsymbol{Q}\boldsymbol{M}\boldsymbol{Q}^T\big) \quad \forall\, \boldsymbol{Q} \in \mathrm{SO}(3). \tag{15}$$

Further, we impose volumetric growth (coercivity) conditions

$$\psi \to \infty \quad \text{as} \quad J \searrow 0 \text{ or } J \to \infty, \tag{16}$$

$$\psi \to \infty \quad \text{as} \quad J_e \searrow 0 \text{ or } J_e \to \infty. \tag{17}$$

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

To ensure existence of minimizers, we design ψ to be polyconvex in $\boldsymbol{F}$ [80,81], i.e., there exists a function $W = W\big(\boldsymbol{F}, \operatorname{cof}\boldsymbol{F}, J;\, \boldsymbol{F}_i, \boldsymbol{M}\big)$ that is convex in $(\boldsymbol{F}, \operatorname{cof}\boldsymbol{F}, J)$ (with $\boldsymbol{F}_i$ and $\boldsymbol{M}$ acting as parameters) and satisfies

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

For background on polyconvex hyperelastic formulations in terms of the deformation gradient for anisotropic solids we refer the reader to [79]; see also [82] for classical symmetry principles and [83] for a discussion of isomorphism concepts in inelastic anisotropy.

Thermodynamic consistency for isotropic materials. Considering Eq. (15), we evaluate the Clausius–Planck inequality $-\dot{\psi} + \frac{1}{2}\,\boldsymbol{S} : \dot{\boldsymbol{C}} \geq 0$, where $\boldsymbol{S}$ denotes the second Piola–Kirchhoff stress tensor. Following the arguments of [84–86], this yields the state law $\boldsymbol{S} = 2\,\partial_{\boldsymbol{C}}\psi$ as well as the reduced dissipation inequality

$$\mathscr{D} := \boldsymbol{\Sigma} : \boldsymbol{D}_i \;\geq\; 0, \qquad \boldsymbol{\Sigma} := -2\,\boldsymbol{F}_i\,\frac{\partial \psi}{\partial \boldsymbol{C}_i}\,\boldsymbol{F}_i^T, \tag{18}$$

where $\boldsymbol{D}_i$ denotes the symmetric part of $\dot{\boldsymbol{F}}_i\boldsymbol{F}_i^{-1}$ and is related to $\boldsymbol{C}_i$ by $\dot{\boldsymbol{C}}_i = 2\,\boldsymbol{F}_i^T\boldsymbol{D}_i\boldsymbol{F}_i$. As discussed in [87], $\boldsymbol{\Sigma}$ shares the same eigenvalues as the Kirchhoff stress $\boldsymbol{F}\boldsymbol{S}\boldsymbol{F}^T$ if ψ depends solely on the elastic part of the deformation[4]. Note that neither $\boldsymbol{D}_i$ nor $\boldsymbol{\Sigma}$ is invariant under the choice of the intermediate configuration; indeed, $\boldsymbol{D}_{i*} = \boldsymbol{Q}\boldsymbol{D}_i\boldsymbol{Q}^T$, $\boldsymbol{\Sigma}_* = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^T$ for any $\boldsymbol{Q} \in \mathrm{SO}(3)$. However, the resulting dissipation is invariant with respect to the intermediate configuration, i.e., $\boldsymbol{\Sigma}_* : \boldsymbol{D}_{i*} = \boldsymbol{\Sigma} : \boldsymbol{D}_i$.

We postulate the existence of a dual potential [88]

$$\varphi \,:\, \mathrm{Sym}^{3\times 3} \;\longrightarrow\; \mathbb{R}, \qquad \boldsymbol{\Sigma} \;\mapsto\; \varphi(\boldsymbol{\Sigma}).$$

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

$$\boldsymbol{D}_i \;\in\; \partial_{\boldsymbol{\Sigma}}\varphi, \tag{19}$$

which in general is nonsmooth [89]. Contrary to the one-dimensional example presented in Section 2.2, the argument of the dual potential is now a second-order tensor. Hence, it is natural to ask how our previous findings can be consistently extended to the three-dimensional case. A standard approach is to assume that the potential is convex in $\boldsymbol{\Sigma}$, which ensures positive dissipation, analogous to InequalityEq. (4). Such an assumption falls within the framework of Generalized Standard Materials [27]. However, for certain materials, this convexity requirement may be overly restrictive. As shown in [36], we may relax this requirement by formulating a

[4] Alternatively, one may derive the reduced dissipation inequality with respect to $\dot{\boldsymbol{C}}_i$ and its conjugated force (see [127]). However, this approach loses the physical interpretation of $\boldsymbol{\Sigma}$, which is considered a disadvantage here

potential ω that is convex, non-negative, and vanishes at the origin, expressed in terms of the invariants $S_\varphi(\boldsymbol{\Sigma}) := (S_1, S_2, S_3)$, and setting $\varphi = \omega(S_\varphi)$. Importantly, these invariants must be *positively homogeneous of degree one*, i.e.,

$$\frac{\partial S_k}{\partial \boldsymbol{\Sigma}} : \boldsymbol{\Sigma} = S_k, \qquad k = 1, \dots, 3. \tag{20}$$

With these properties, the reduced dissipationEq. (18) simplifies to $\partial_{\boldsymbol{\Sigma}}\varphi : \boldsymbol{\Sigma} = \sum_i (\partial_{S_i}\varphi)\, S_i$. Analogous to InequalityEq. (4), this formulation guarantees positive dissipation–yet without requiring convexity with respect to $\boldsymbol{\Sigma}$ itself. A suitable choice for the invariants S_1, S_2, and S_3 was proposed in [36] as

$$S_1 := \operatorname{tr} \boldsymbol{\Sigma}, \qquad S_2' := \tfrac{1}{2} \operatorname{tr}\big((\mathrm{dev}\, \boldsymbol{\Sigma})^2\big), \qquad S_3' := \tfrac{1}{3} \operatorname{tr}\big((\mathrm{dev}\, \boldsymbol{\Sigma})^3\big), \tag{21}$$

where $S_2 := \sqrt{S_2'}$ and $S_3 := \sqrt[3]{S_3'}$; see Appendix A.2 for details on the computation of the roots and Appendix A.3 for a more comprehensive proof of thermodynamic consistency.

Although the convexity requirement with respect to the stress tensor has now been relaxed, convexity with respect to the invariants is still assumed. Recalling the insights from Section 2 and Fig. 1, we may further extend the concept of monotonicity to the three-dimensional case. This allows us to generalize the potential by introducing a monotonically increasing function ζ and defining

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

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

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

$$\varphi : \mathrm{Sym}^{3\times3} \times \mathbb{M}^{3\times3} \longrightarrow \mathbb{R}, \qquad (\boldsymbol{\Sigma}, \tilde{\boldsymbol{M}}) \mapsto \varphi(\boldsymbol{\Sigma}, \tilde{\boldsymbol{M}}),$$

where $\tilde{\boldsymbol{M}} = \boldsymbol{F}_i \star \boldsymbol{M}$ denotes an appropriate push-forward of $\boldsymbol{M}$ onto the intermediate configuration such that $\boldsymbol{Q}\boldsymbol{F}_i \star \boldsymbol{M} = \boldsymbol{Q}\, \tilde{\boldsymbol{M}}\, \boldsymbol{Q}^T \quad \forall\, \boldsymbol{Q} \in \mathrm{SO}(3)$, 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

$$\mathcal{I}_\varphi(\boldsymbol{\Sigma}, \tilde{\boldsymbol{M}}) := \big(S_\varphi(\boldsymbol{\Sigma}),\, \mathcal{A}_\varphi(\boldsymbol{\Sigma}, \tilde{\boldsymbol{M}})\big),$$

where $\mathcal{A}_\varphi(\boldsymbol{\Sigma}, \tilde{\boldsymbol{M}}) := \big(A_1(\boldsymbol{\Sigma}, \tilde{\boldsymbol{M}}), \dots, A_m(\boldsymbol{\Sigma}, \tilde{\boldsymbol{M}})\big)$ accounts for anisotropy. Similarly to S_φ, the anisotropic invariants A_k must be *positively homogeneous of degree one*, i.e.

$$\frac{\partial A_k}{\partial \boldsymbol{\Sigma}} : \boldsymbol{\Sigma} = A_k, \qquad k = 1, \dots, m. \tag{22}$$

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

$$\varphi(\mathcal{I}_\varphi) = (\zeta \circ \omega)(\mathcal{I}_\varphi). \tag{23}$$

A suitable choice of invariants $\mathcal{A}_\varphi$ will be introduced in Section 4.2, while a more detailed proof of thermodynamic consistency is given in Appendix A.3.

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

$$\tilde{\boldsymbol{n}} \in \left\{ \frac{\boldsymbol{F}_i\, \boldsymbol{n}}{\|\boldsymbol{F}_i\, \boldsymbol{n}\|}, \quad \frac{\operatorname{cof} \boldsymbol{F}_i\, \boldsymbol{n}}{\|\operatorname{cof} \boldsymbol{F}_i\, \boldsymbol{n}\|} \right\} =: \{\tilde{\boldsymbol{n}}_1, \tilde{\boldsymbol{n}}_2\}, \tag{24}$$

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

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

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

$$\boldsymbol{C}_e^k\, \boldsymbol{B}_i^r\, \tilde{\boldsymbol{M}}_1^s\, \tilde{\boldsymbol{M}}_2^t = \boldsymbol{F}_i^{-T} \left(\boldsymbol{C}\boldsymbol{C}_i^{-1}\right)^k \boldsymbol{C}_i^r \left(\frac{\boldsymbol{C}_i \boldsymbol{M}}{\boldsymbol{C}_i : \boldsymbol{M}}\right)^s \left(\frac{\boldsymbol{M}\boldsymbol{C}_i^{-1}}{\boldsymbol{C}_i^{-1} : \boldsymbol{M}}\right)^t \boldsymbol{F}_i^T, \quad k, r \in \mathbb{Z}, \quad s, t \in \mathbb{N}_0. \tag{25}$$

Consequently, the mechanically relevant eigenvalues in the intermediate configuration can be determined solely from tensorial quantities defined with respect to the reference configuration. This represents an important result, as it shows that although the intermediate

configuration is arbitrary, the eigenvalues can still be uniquely obtained. Furthermore, since $\tilde{\boldsymbol{M}}_1$ and $\tilde{\boldsymbol{M}}_2$ are idempotent, we may, without loss of generality, restrict $s,t \in \{0,1\}$.[5]

Similarly to the dual potential, we are aiming for an isotropic set of invariants $\mathcal{S}_\psi$ and an anisotropic set $\mathcal{A}_\psi$ to express the Helmholtz free energy as a scalar-valued isotropic function. With the relation given in Eq. (25), we are enabled to define a suitable set of invariants in terms of $\boldsymbol{C}_e$, $\boldsymbol{B}_i$, and $\tilde{\boldsymbol{M}}$ for both sets and subsequently express them in terms of $\boldsymbol{C}$, $\boldsymbol{C}_i$, and $\boldsymbol{M}$. We start with the isotropic set. In line with [93], the corresponding integrity basis is given by the following ten invariants

$$\begin{aligned}
&\operatorname{tr}(\boldsymbol{C}_e), \quad \operatorname{tr}(\operatorname{cof}\boldsymbol{C}_e), \quad J_e,\\
&\operatorname{tr}(\boldsymbol{B}_i), \quad \operatorname{tr}(\operatorname{cof}\boldsymbol{B}_i), \quad J_i,\\
&\operatorname{tr}(\boldsymbol{C}_e\boldsymbol{B}_i), \quad \operatorname{tr}(\operatorname{cof}\boldsymbol{C}_e\operatorname{cof}\boldsymbol{B}_i), \quad \operatorname{tr}(\operatorname{cof}\boldsymbol{C}_e\,\boldsymbol{B}_i), \quad \operatorname{tr}(\boldsymbol{C}_e\operatorname{cof}\boldsymbol{B}_i).
\end{aligned}$$

Making use of Eq. (25), we express these invariants as

$$\begin{aligned}
&\operatorname{tr}(\boldsymbol{C}\boldsymbol{C}_i^{-1}), \quad \tfrac{1}{J_i^2}\operatorname{tr}(\boldsymbol{C}_i\operatorname{cof}\boldsymbol{C}), \quad \tfrac{J}{J_i},\\
&\operatorname{tr}(\boldsymbol{C}_i), \quad \operatorname{tr}(\operatorname{cof}\boldsymbol{C}_i), \quad J,\\
&\operatorname{tr}(\boldsymbol{C}), \quad \operatorname{tr}(\operatorname{cof}\boldsymbol{C}), \quad \tfrac{1}{J_i^2}\operatorname{tr}(\boldsymbol{C}_i), \quad \operatorname{tr}(\operatorname{cof}\boldsymbol{C}\,\boldsymbol{C}_i), \quad J_i^2\operatorname{tr}(\boldsymbol{C}\boldsymbol{C}_i^{-2}),
\end{aligned}$$

where J_i has been replaced by J to directly account for the volumetric growth condition (16). Notably, it is interesting to recognize that the seventh and eighth mixed invariants of elastic and inelastic arguments simplify to total deformation, which gives them a clear physical interpretation. Further, we observe that each trace-like invariant can be expressed as $\operatorname{tr}(\boldsymbol{F}^T\boldsymbol{F}\boldsymbol{G})$ with $\boldsymbol{G} \succeq 0$ and analogously when replacing $\boldsymbol{F}$ with its cofactor. This particular type of invariant can be shown to be polyconvex; see Appendix A.4. Thus, each of these ten invariants is polyconvex as they are convex in either $\boldsymbol{F}$, $\operatorname{cof}\boldsymbol{F}$ or J. It remains to explicitly define the isotropic set $\mathcal{S}_\psi := (S_1,\dots,S_{10})$. Unfortunately, none of the above mentioned invariants satisfies to normalize the energy and stresses. To this end, we introduce the following convex functions[6]

$$\beta_1(\boldsymbol{A}) := \operatorname{tr}\boldsymbol{A} - 3 - \ln\det\boldsymbol{A}, \quad \beta_2(A) := A - 1 - \ln A, \quad \beta_3(\boldsymbol{A}) := \ln\cosh(\max(\operatorname{tr}\boldsymbol{A} - 1, 0)), \tag{26}$$

which are illustrated in Fig. 4. With these functions at hand, we define the isotropic set as follows

$$\begin{aligned}
&S_1 = \beta_1(\boldsymbol{C}\boldsymbol{C}_i^{-1}), && S_2 = \beta_1\Big(\tfrac{1}{J_i^2}\boldsymbol{C}_i\operatorname{cof}\boldsymbol{C}\Big), && S_3 = \beta_2\Big(\tfrac{J}{J_i}\Big), && S_4 = \beta_1(\boldsymbol{C}_i),\\
&S_5 = \beta_1(\operatorname{cof}\boldsymbol{C}_i), && S_6 = \beta_2(J), && S_7 = \beta_1(\boldsymbol{C}), && S_8 = \beta_1(\operatorname{cof}\boldsymbol{C}),\\
&S_9 = \beta_3\Big(\tfrac{1}{J_i^2}\boldsymbol{C}_i\operatorname{cof}\boldsymbol{C}\,\boldsymbol{C}_i\Big), && S_{10} = \beta_3(J_i^2\boldsymbol{C}\boldsymbol{C}_i^{-2}),
\end{aligned} \tag{27}$$

where each S_k represents a polyconvexity-preserving combination of our above introduced invariants.[7]

Next, we turn our attention to the anisotropic set $\mathcal{A}_\psi$. We restrict ourselves to the mapping $\tilde{\boldsymbol{M}}_1$ for simplicity, though both can be included (e.g. letting the neural network select the more appropriate one). Once again, in line with [93], we introduce the integrity basis

$$\begin{aligned}
&\operatorname{tr}(\boldsymbol{C}_e\,\tilde{\boldsymbol{M}}_1), \quad \operatorname{tr}(\operatorname{cof}\boldsymbol{C}_e\,\tilde{\boldsymbol{M}}_1), \quad \operatorname{tr}(\boldsymbol{B}_i\,\tilde{\boldsymbol{M}}_1),\\
&\operatorname{tr}(\operatorname{cof}\boldsymbol{B}_i\,\tilde{\boldsymbol{M}}_1), \quad \operatorname{tr}\big(\boldsymbol{C}_e\,(\boldsymbol{B}_i\,\#\,\tilde{\boldsymbol{M}}_1)\big),
\end{aligned}$$

which map to their referential counterparts as

$$\begin{aligned}
&\operatorname{tr}\Big(\tfrac{\boldsymbol{C}\boldsymbol{M}}{\boldsymbol{C}_i:\boldsymbol{M}}\Big), \quad \tfrac{1}{J_i^2}\operatorname{tr}\Big(\boldsymbol{C}_i\operatorname{cof}\boldsymbol{C}\,\tfrac{\boldsymbol{C}_i\boldsymbol{M}}{\boldsymbol{C}_i:\boldsymbol{M}}\Big), \quad \operatorname{tr}\Big(\tfrac{\boldsymbol{C}_i^2\boldsymbol{M}}{\boldsymbol{C}_i:\boldsymbol{M}}\Big),\\
&J_i^2\operatorname{tr}\Big(\tfrac{\boldsymbol{M}}{\boldsymbol{C}_i:\boldsymbol{M}}\Big), \quad \operatorname{tr}\Big(\boldsymbol{C}\boldsymbol{C}_i^{-1}\Big(\boldsymbol{C}_i\,\#\,\tfrac{\boldsymbol{M}\boldsymbol{C}_i}{\boldsymbol{C}_i:\boldsymbol{M}}\Big)\Big),
\end{aligned}$$

with the outer product # of second-order tensors. This unusual choice[8] replaces the classical invariant $\boldsymbol{C}_e\boldsymbol{B}_i\tilde{\boldsymbol{M}}_1$, which is generally not polyconvex; see Appendix A.4 for the definition of the outer product as well as the corresponding discussion of polyconvexity. For the same reasons as in the isotropic case, we observe that each of these five invariants is polyconvex. In fact, the third and fourth invariants are independent of the deformation gradient; hence, polyconvexity with respect to these invariants is trivially satisfied.

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

[6] For β_1, one has $\beta_1 = \sum_{i=1}^{3}(A_i - 1 - \ln A_i)$ if $\boldsymbol{A} \in \mathrm{Sym}_+^{3\times3}$ with the eigenvalues A_i. All β_j are non-negative, convex, and vanish at the normalized state: $\boldsymbol{A} = \boldsymbol{I}$ for β_1, $A = 1$ for β_2, and $\operatorname{tr}\boldsymbol{A} = 1$ for β_3. Further, β_3 is non-decreasing.

[7] For $f(J) = -\ln(aJ^b) = -\ln a - b\ln J$ 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 $f(\boldsymbol{C},\boldsymbol{G}) = \operatorname{tr}(\boldsymbol{C}\boldsymbol{G})$ is polyconvex if $\boldsymbol{G} \succeq 0$, same holds when $\boldsymbol{C}$ is replaced by its cofactor; see Appendix A.4.

[8] For 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]

However, as in the isotropic case, none of them ensures proper normalization of the energy and the stresses. Therefore, we make use of the convex function β_3 introduced in Eq. (26) and define the anisotropic set $\mathcal{A}_\psi := (A_1, \dots, A_5)$

$$A_1 = \beta_3\Big(\frac{\boldsymbol{C}\boldsymbol{M}}{\boldsymbol{C}_i : \boldsymbol{M}}\Big), \qquad A_2 = \beta_3\Big(\frac{1}{J_i^2}\,\boldsymbol{C}_i \operatorname{cof}\boldsymbol{C}\,\frac{\boldsymbol{C}_i\boldsymbol{M}}{\boldsymbol{C}_i : \boldsymbol{M}}\Big), \qquad A_3 = \beta_3\Big(\frac{\boldsymbol{C}_i^2\boldsymbol{M}}{\boldsymbol{C}_i : \boldsymbol{M}}\Big),$$
$$A_4 = \beta_3\Big(\frac{J_i^2\boldsymbol{M}}{\boldsymbol{C}_i : \boldsymbol{M}}\Big), \qquad A_5 = \beta_3\Big(\boldsymbol{C}\boldsymbol{C}_i^{-1}\Big(\boldsymbol{C}_i \,\#\, \frac{\boldsymbol{M}\boldsymbol{C}_i}{\boldsymbol{C}_i : \boldsymbol{M}}\Big)\Big), \tag{28}$$

where each A_k is polyconvex, as it arises from composing the convex and non-decreasing function β_3 with a polyconvex invariant.

Finally, the Helmholtz free energy is expressed as

$$\psi = \psi\big(\mathcal{T}_\psi\big), \quad \mathcal{T}_\psi(\boldsymbol{C}, \boldsymbol{C}_i, \boldsymbol{M}) := \big(\mathcal{S}_\psi(\boldsymbol{C}, \boldsymbol{C}_i),\, \mathcal{A}_\psi(\boldsymbol{C}, \boldsymbol{C}_i, \boldsymbol{M})\big). \tag{29}$$

Noteworthy, within the above introduced definitions, we followed the energy normalization condition, but slightly relaxed the stress normalization condition. That is, we allowed 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 $\mathcal{T}_\varphi$ Eq. (23) 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 $(\boldsymbol{F}, \operatorname{cof}\boldsymbol{F}, J)$ (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 $\mathcal{S}_\varphi$ are already given in Eq. (21). For the anisotropic contribution, we permit both push-forwards of the structural vector in Eq. (24). Following [93], this leads to

$$A_{11} = \operatorname{dev}\boldsymbol{\Sigma} : \tilde{\boldsymbol{M}}_1, \qquad A_{12} = \operatorname{dev}\boldsymbol{\Sigma} : \tilde{\boldsymbol{M}}_2, \qquad A_{21} = \sqrt{\tfrac{1}{2}\,(\operatorname{dev}\boldsymbol{\Sigma})^2 : \tilde{\boldsymbol{M}}_1},$$
$$A_{22} = \sqrt{\tfrac{1}{2}\,(\operatorname{dev}\boldsymbol{\Sigma})^2 : \tilde{\boldsymbol{M}}_2}, \qquad A_3 = \boldsymbol{\Sigma} : \operatorname{sym}(\tilde{\boldsymbol{n}}_1 \otimes \tilde{\boldsymbol{n}}_2), \qquad A_4 = \sqrt{\tfrac{1}{2}\,\boldsymbol{\Sigma}^2 : \operatorname{sym}(\tilde{\boldsymbol{n}}_1 \otimes \tilde{\boldsymbol{n}}_2)}, \tag{30}$$

and we set $\mathcal{A}_\varphi := \big(A_{11}, A_{12}, A_{21}, A_{22}, A_3, A_4\big)$. All these invariants are positively homogeneous of degree one 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 and being related to the Kirchhoff stress by the determinant J. Hence, it seems reasonable to use the Mandel stress divided by J as the argument of the dual potential. Using the positive homogeneity of degree one of $\mathcal{T}_\varphi := (T_1, \dots, T_9) = (\mathcal{S}_\varphi, \mathcal{A}_\varphi)$ with respect to $\boldsymbol{\Sigma}$, the evolution Eq. (19) can be written as

$$\boldsymbol{D}_i \in \partial_{\boldsymbol{\Sigma}}\,\varphi\Big(\frac{\mathcal{T}_\varphi}{J}\Big) = \frac{1}{J}\,\Big[\partial_{\mathbf{z}}\varphi(\mathbf{z})\Big]_{\mathbf{z}=\mathcal{T}_\varphi/J} : \partial_{\boldsymbol{\Sigma}}\mathcal{T}_\varphi, \tag{31}$$

where the gradient of φ is evaluated at $\mathcal{T}_\varphi/J$. Having in mind that each stress invariant satisfies $(\partial_{\boldsymbol{\Sigma}} T_k) : \boldsymbol{\Sigma} = T_k$, we observe that the reduced dissipation Eq. (18)

$$\boldsymbol{\Sigma} : \boldsymbol{D}_i = \Big\langle \boldsymbol{\Sigma}, \partial_{\boldsymbol{\Sigma}}\,\varphi\Big(\frac{\mathcal{T}_\varphi}{J}\Big)\Big\rangle = \Big\langle \frac{\mathcal{T}_\varphi}{J}, \Big[\partial_{\mathbf{z}}\varphi(\mathbf{z})\Big]_{\mathbf{z}=\mathcal{T}_\varphi/J}\Big\rangle \tag{32}$$

is satisfied, since φ is composed of a zero-valued, monotonically increasing function ζ and a convex, zero-valued, and non-negative function ω, together with the property of subgradients Eq. (4); see Appendix A.3 for a detailed proof of thermodynamic consistency.

Co-rotated intermediate configuration. Direct computation of $\boldsymbol{\Sigma}$, $\tilde{\boldsymbol{n}}_1$, and $\tilde{\boldsymbol{n}}_2$ would require the inelastic part $\boldsymbol{F}_i$ of the deformation gradient, which is non-unique. One may either establish a mapping from intermediate to referential invariants (analogous to Eq. (25)) or adopt a co-rotated intermediate configuration [97] (similarly exploited in [98]) that maps all non-unique quantities to unique counterparts via (cf. Fig. 3)

$$\bar{(\bullet)} := \boldsymbol{R}_i^T \star (\bullet) = \boldsymbol{R}_{i*}^T \star (\bullet)_*,$$

where $\boldsymbol{R}_i$ and $\boldsymbol{R}_{i*}$ denote the rotational parts associated with $\boldsymbol{F}_i$ and $\boldsymbol{F}_{i*}$, respectively. This yields

$$\bar{\boldsymbol{\Sigma}} = -2\,\boldsymbol{U}_i\,\frac{\partial\psi}{\partial\boldsymbol{C}_i}\,\boldsymbol{U}_i, \quad \bar{\boldsymbol{D}}_i = \operatorname{sym}(\dot{\boldsymbol{U}}_i\boldsymbol{U}_i^{-1}) = \tfrac{1}{2}\,\boldsymbol{U}_i^{-1}\,\dot{\boldsymbol{C}}_i\,\boldsymbol{U}_i^{-1}, \quad \bar{\boldsymbol{n}}_1 = \frac{\boldsymbol{U}_i\,\boldsymbol{n}}{\|\boldsymbol{U}_i\,\boldsymbol{n}\|}, \quad \bar{\boldsymbol{n}}_2 = \frac{\operatorname{cof}\boldsymbol{U}_i\,\boldsymbol{n}}{\|\operatorname{cof}\boldsymbol{U}_i\,\boldsymbol{n}\|}, \tag{33}$$

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 $\mathcal{I}_\psi$ and $\mathcal{I}_\varphi$ 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 [13,22,99]. Furthermore, we seek to replace the solution of the implicit time discretization scheme by a recurrent architecture, in analogy to [33,68]. While the small strain regime is investigated in [33] and the finite strain regime is studied in [68], both assume an additive decomposition of the total strain and internal variables. We refer the interested reader to [33] for a graphical representation of this procedure and the connections between the individual neural networks.

Helmholtz free energy. The Helmholtz free energy is represented by an Input Convex Neural Network (ICNN) that is polyconvex with respect to $(\boldsymbol{F}, \operatorname{cof}\boldsymbol{F}, J)$. However, we pass $\mathcal{I}_\psi$ to the ICNN, which already provides a convex representation of $(\boldsymbol{F}, \operatorname{cof}\boldsymbol{F}, J)$. This motivates the slight adaptation of the ICNN architecture in Eq. (7) compared to the original (6). To enforce the volumetric growth conditionsEqs. (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

$$\psi = \mathcal{N}_c(\bullet,\, \mathcal{I}_\psi) + \psi_{\text{gr}}(J, \tfrac{J}{J_i}), \tag{34}$$

with

$$\psi_{\text{gr}}(J, \tfrac{J}{J_i}) = \lambda_{\text{gr}}\Big(\beta_2(J) + \beta_2(\tfrac{J}{J_i})\Big), \tag{35}$$

where λ_{gr} is a penalty parameter, fixed to 10^{-4} in all simulations. In Eq. (34), $\bullet$ 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):

$$\varphi = \mathcal{N}_\circ\left(\bullet,\, \left(\frac{\mathcal{I}_\varphi}{J}, -\frac{\mathcal{I}_\varphi}{J}\right)\right), \tag{36}$$

where $\bullet$ again denotes the dependence on non-convex quantities. Unlike the free energy, the inner function ω of the composition $(\zeta \circ \omega)$ must only be convex with respect to the invariants themselves, not with respect to their arguments, i.e., the thermodynamic driving force $\boldsymbol{\Sigma}$. To recover the structure of the original ICNN (6), we pass both $\mathcal{I}_\varphi$ and $-\mathcal{I}_\varphi$ through the convex branch of the network. These are scaled by the determinant J of the deformation gradient, which was motivated by the relation between the Kirchhoff stress tensor and the Cauchy stress tensor. 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 $\boldsymbol{U}_i$ [41]. The *explicit* integration step reads

$$\boldsymbol{C}_{i_{n+1}} = \boldsymbol{U}_{i_n} \exp(2\,\Delta t\, \boldsymbol{D}_{i_n})\, \boldsymbol{U}_{i_n}, \qquad \boldsymbol{U}_{i_{n+1}} = \sqrt{\boldsymbol{C}_{i_{n+1}}}, \tag{37}$$

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

$$\mathbf{r} := \boldsymbol{C}_{i_n} - \boldsymbol{U}_{i_{n+1}} \exp(-2\,\Delta t\, \boldsymbol{D}_{i_{n+1}})\, \boldsymbol{U}_{i_{n+1}} \overset{!}{=} \mathbf{0}, \tag{38}$$

where $\boldsymbol{D}_{i_{n+1}}$ depends nonlinearly on $\boldsymbol{U}_{i_{n+1}}$.

In our experience, the exponential time-integration scheme–irrespective of whether it is implemented explicitly or implicitly–tends to destabilize the training process.[9] This instability can ultimately cause the training to terminate prematurely at unpredictable stages. To address this issue, we follow the approach of [33,68] and predict the evolution of the internal variables during training through an auxiliary neural network. Importantly, after training, the exponential-map update is reinstated during inference to ensure admissible internal variables. This strategy enables stable training of the physically motivated networks associated with ψ and φ, while preventing unphysical extrapolations of the internal variables beyond the training regime. In this contribution, we adopt an architecture based on LiNNs (12) to predict the evolution of internal variables; Fig. 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 $\boldsymbol{\alpha} = \mathcal{N}_\alpha$ and $\mathbf{f} = \mathcal{N}_f$ via

$$(\mathbf{h}_n, (\boldsymbol{E}_{n+1}, \boldsymbol{D}_i^{trial})) \mapsto \mathcal{N}_\alpha(\mathbf{h}_n, (\boldsymbol{E}_{n+1}, \boldsymbol{D}_i^{trial})), \quad (\mathbf{h}_n, (\boldsymbol{E}_{n+1}, \boldsymbol{D}_i^{trial})) \mapsto \mathcal{N}_f(\mathbf{h}_n, (\boldsymbol{E}_{n+1}, \boldsymbol{D}_i^{trial})), \tag{39}$$

[9] However, as mentioned in the introduction, an implicit time integration scheme based on the exponential integrator was recently employed for finite viscoelasticity in [44].

where $\boldsymbol{E}_{n+1}$ denotes the Green–Lagrange strain tensor and $\boldsymbol{D}_i^{trial}$ is the trial value of $\boldsymbol{D}_i$, i.e., $\boldsymbol{D}_i$ evaluated with the current $\boldsymbol{C}_{n+1}$ but freezing $\boldsymbol{U}_{i_{n+1}} = \boldsymbol{U}_{i_n}$. 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 Eq. (39) allows us to update the hidden state according to Eq. (12). Since these states do not generally coincide with the independent components of $\boldsymbol{U}_i \in \mathrm{Sym}_+^{3\times3}$, positive definiteness must be enforced. For this purpose, we employ the Cholesky decomposition of $\boldsymbol{U}_i$ and map $\mathbf{h}$ to the lower-triangular factor of $\boldsymbol{U}_i$; see Appendix A.2.

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

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \; \mathcal{L}(\boldsymbol{\theta}), \tag{40}$$

where the total loss is additively split into[10]

$$\mathcal{L}(\boldsymbol{\theta}) = \mathcal{L}_{stress}(\boldsymbol{\theta}) + \lambda_{evo}\, \mathcal{L}_{evo}(\boldsymbol{\theta}). \tag{41}$$

Here, $\mathcal{L}_{stress}$ measures the discrepancy between predicted and experimentally observed stresses, while $\mathcal{L}_{evo}$ quantifies the residual of the implicit evolution Eq. (38). 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

$$\mathcal{L}_{stress}(\boldsymbol{\theta}) = \frac{1}{B\,T\,O} \sum_{b=1}^{B} \sum_{t=1}^{T} \sum_{o=1}^{O} \left(\underline{S}_{b,t,o}(\boldsymbol{\theta}) - \underline{\hat{S}}_{b,t,o} \right)^2, \tag{42}$$

with the number of batches B, time steps T, independent stress components $O = 6$, the predicted normalized second Piola–Kirchhoff stress components $\underline{S}(\boldsymbol{\theta})$, and the normalized experimental stresses $\underline{\hat{S}}$. Normalization is performed by dividing the raw experimental stress $\hat{\boldsymbol{S}}$ by the absolute maximum $\underline{S}$ across all batches and time steps, i.e., $\underline{\hat{\boldsymbol{S}}} = \hat{\boldsymbol{S}}/\underline{S}$. Thus, after training we have to multiply $\underline{\boldsymbol{S}}$ by $\underline{S}$ to obtain $\boldsymbol{S}$. Analogously, the evolution loss is defined as

$$\mathcal{L}_{evo}(\boldsymbol{\theta}) = \frac{1}{B\,T\,O} \sum_{b=1}^{B} \sum_{t=1}^{T} \sum_{o=1}^{O} \left(r_{b,t,o}(\boldsymbol{\theta}) \right)^2, \tag{43}$$

where $r(\boldsymbol{\theta})$ denotes the independent components of the residual of the implicit evolution Eq. (38).

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 (34) is defined as

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

while the network for the dual potential Eq. (36) is designed as (cf. Fig. 2 for the color code)

$$\mathcal{N}_\circ : \mathbb{R}^{18} \to \mathbb{R}, \quad \left(\frac{\mathcal{I}_\varphi}{J}, -\frac{\mathcal{I}_\varphi}{J} \right) \xrightarrow{\exp} 16 \xrightarrow{ReLU} 16 \xrightarrow{ReLU} 16 \xrightarrow{ReLU} 4 \xrightarrow{\tanh} 16 \xrightarrow{\tanh} 16 \xrightarrow{Softplus} 16 \xrightarrow{Linear} \varphi.$$

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

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

and

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

where $GELU$ refers to the Gaussian Error Linear Unit activation function [101]. Note that the above denoted input dimensions of the energy and the potential correlate to the anisotropic case. In case of isotropic materials, the input dimensions are reduced from 15 to 10 for ψ and from 18 to 6 for φ; however, as mentioned we keep the number and size of hidden layers per network.

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].

[10] A regularization term can be added if desired.

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			
a	b	c	a	b	c	α	η	α	η	β	τ
MPa	MPa	MPa	MPa	MPa	MPa	–	MPa	–	MPa	–	s
80	100	100	40	50	50	2	10	2	10	2	12

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

$$\psi = \psi_{iso} + \psi_{ani}, \quad \varphi = \varphi_{iso} + \varphi_{ani}. \tag{44}$$

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 Eq. (37). Additionally, in the anisotropic case, the preferred direction is

$$\boldsymbol{n} = \left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0 \right)^T,$$

Boundary value problem and loading. The simulation setup is illustrated in Fig. 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 Fig. 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 Fig. 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 Fig. 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 $\lambda_{evo} = 1000$, neither contribution dominates. Normalization uses maximum stresses $\underline{S} = 1406.429$ MPa (isotropic) and $\underline{S} = 1014.538$ MPa (anisotropic); cf. Eq. (42).

Training on noisy data. In practical applications, experimental data are typically corrupted by measurement noise, which can distort stress-strain relationships and consequently impair the training of neural networks. To evaluate the robustness of the proposed framework under such conditions, we perform additional training runs based on synthetically generated noisy data sets. The perturbations are introduced directly at the level of the target stresses by superimposing Gaussian noise whose amplitude depends on both the statistical variability of the data and a small constant offset. This combination ensures that all stress components are perturbed, including those with small or vanishing magnitudes such as the out-of-plane shear stresses. Formally, the noisy stress targets are obtained by

$$\hat{\boldsymbol{S}}^{noisy} = \hat{\boldsymbol{S}} + (\alpha_1 \, \boldsymbol{s} + \alpha_2 \, \boldsymbol{1}) \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{1}),$$

where $\boldsymbol{s}$ denotes the empirical standard deviation of each stress component computed over all temporal and batch dimensions and $\mathcal{N}$ refers to the normal distribution. The parameters $\alpha_1 = 0.15$ and $\alpha_2 = 0.01$ specify the relative and absolute noise magnitudes, respectively.

Training follows exactly the same two-stage strategy as in the noise-free case. Both isotropic and anisotropic networks are first pre-trained for 500 epochs using data from a single material point to obtain stable initial parameters. Subsequently, full-data training is performed for 10,000 epochs using the complete data set. The evolution of the corresponding loss functions is presented in Fig. 10. Unlike the noise-free scenario, the total losses no longer decrease monotonically but fluctuate around a mean value after an initial decay phase, reflecting the random variability introduced by the noise. Only the auxiliary loss associated with the evolution equation exhibits a steady convergence, reaching magnitudes on the order of 10^{-6} in both isotropic and anisotropic models. The overall higher loss levels may indicate that the current choice of the penalty parameter λ_{evo} is too small to fully balance the competing terms in the objective function. Nevertheless, a detailed hyperparameter sensitivity analysis is beyond the scope of the present work. For consistency, the same normalization scheme is applied, using maximum stresses of $\underline{S} = 1420.851$ MPa (isotropic) and $\underline{S} = 1012.571$ MPa (anisotropic).

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 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 trained exclusively on the noise-free data sets. The first is an Elman-type recurrent neural network [108], denoted $\mathcal{N}_{RNN}$, which was already used in [36] for inelastic Constitutive Neural Networks. The second is a Liquid Neural Network, denoted $\mathcal{N}_{LiNN}$. Both serve as references representing classical recurrent approaches without physics priors. Our proposed physics-embedded models are denoted $\mathcal{N}^{LiNN}_{Consti}$, where the inelastic variables are predicted by a LiNN, and $\mathcal{N}^{Exp}_{Consti}$, where the evolution is integrated explicitly via the exponential update rule (37). It is important to note that $\mathcal{N}_{LiNN}$, which directly predicts stresses, must not be confused with the auxiliary LiNN in $\mathcal{N}^{LiNN}_{Consti}$. 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 Fig. 11. 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 Fig. 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. $\psi = \psi_{iso}$ and $\varphi = \varphi_{iso}$ according to Eq. (44). 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. Fig. 12 displays the stress components predicted for one representative training element. Both physics-embedded variants, $\mathcal{N}^{LiNN}_{Consti}$ and $\mathcal{N}^{Exp}_{Consti}$, reproduce the reference stresses with very high accuracy. It should be emphasized once again that $\mathcal{N}^{Exp}_{Consti}$ is not really trained, but rather evaluated using the weights and biases of the Helmholtz free energy and the dual potential discovered during training of $\mathcal{N}^{LiNN}_{Consti}$. The only noticeable deviation occurs in the out-of-plane shear component S_{23}, where $\mathcal{N}^{LiNN}_{Consti}$ 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 $\mathcal{N}_{RNN}$ and $\mathcal{N}_{LiNN}$ 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 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 Fig. 13 as well as the corresponding parity plot in Fig. 14. 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 $\mathcal{N}^{LiNN}_{Consti}$ during training is also present here, but is absent in $\mathcal{N}^{Exp}_{Consti}$. 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.

Noisy data. Next, we examine how well the proposed framework performs when trained on data sets corrupted by synthetic noise. Fig. 15 presents the training results (element ID 111), while the testing results are shown in Fig. 16. In addition, Fig. 14 shows the corresponding parity plot. Here, we observe that $\mathcal{N}^{LiNN}_{Consti}$ and $\mathcal{N}^{Exp}_{Consti}$ continue to deliver stable and accurate predictions across all

stress components, despite the presence of noise in the training data. For comparison, we also include the results of the noise-free training case obtained with $\mathcal{N}^{Exp}_{Consti}$ (noise-free). Interestingly, the network trained on noisy data discovers almost the same stress-strain relationships as in the noise-free scenario. In addition, as in the noise-free case, we do not use any regularization techniques beyond gradient clipping, yet overfitting is not observed although the depth and width of the involved networks are relatively large. We attribute this favorable behavior to the physics-embedded approach, which constrained the architectures of the involved neural networks. In simple terms, for example, a convex energy and dissipation potential limit the space of possible stress-strain relationships and suppress oscillating behavior, thereby acting as an implicit regularizer. Of course, this favorable effect strongly relies on noise-free input data, i.e., the deformation measures.

6.2.2. Anisotropic response

We now turn to the anisotropic case, where the free energy and dual potential include anisotropic contributions, $\psi = \psi_{iso} + \psi_{ani}$ and $\varphi = \varphi_{iso} + \varphi_{ani}$; see Eq. (44). 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 Fig. 17. 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 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 $\mathcal{N}^{LiNN}_{Consti}$, 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 Fig. 18 with its corresponding parity plot shown in Fig. 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, $\mathcal{N}^{LiNN}_{Consti}$ and $\mathcal{N}^{Exp}_{Consti}$ deliver almost identical predictions, underlining that the auxiliary LiNN has little influence on accuracy but plays an important role in stabilizing training. Indeed, training $\mathcal{N}^{Exp}_{Consti}$ alone[11] frequently diverged to NaN losses, whereas the auxiliary LiNN enabled stable convergence.

Noisy data. Finally, we assess the capability of the proposed framework to handle noisy data in the anisotropic setting. Fig. 19 presents the training results, while the testing results are shown in Fig. 20. Fig. 14 depicts the respective parity plot. As in the isotropic case, training on noisy data does not suffer from overfitting, and further, the discovered stress-strain relationships closely resemble those obtained from noise-free training. Noteworthy, this would likely change if the input data were no longer noise-free. Although the training results are satisfactory, we observe an unstable behavior during testing when using the explicit exponential integrator. Specifically, as shown in Fig. 20, the stresses predicted by $\mathcal{N}^{Exp}_{Consti}$ diverge after $t = 1.15\,\mathrm{s}$. We attribute this instability to the chosen time integration scheme rather than to the network architecture itself. Explicit schemes are known to be conditionally stable; probably the size of the networks–which is higher compared to the isotropic case due to the change in the input dimensions–may lead to a stronger accumulation of numerical errors during time integration. One may tend to use an implicit version as recently utilized in [44]. An explicit Runge-Kutta-like exponential scheme of higher order could be a viable alternative to improve stability without the need for implicit solvers; however, this is beyond the scope of the present work. Interestingly, $\mathcal{N}^{LiNN}_{Consti}$ still performs well in the testing regime and closely matches the noise-free reference obtained using $\mathcal{N}^{Exp}_{Consti}$.

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.

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 (Fig. 21). 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

[11] As mentioned earlier, we do not train $\mathcal{N}^{Exp}_{Consti}$ in this contribution. Rather, it is evaluated using the weights and biases discovered during the training of its LiNN counterpart.

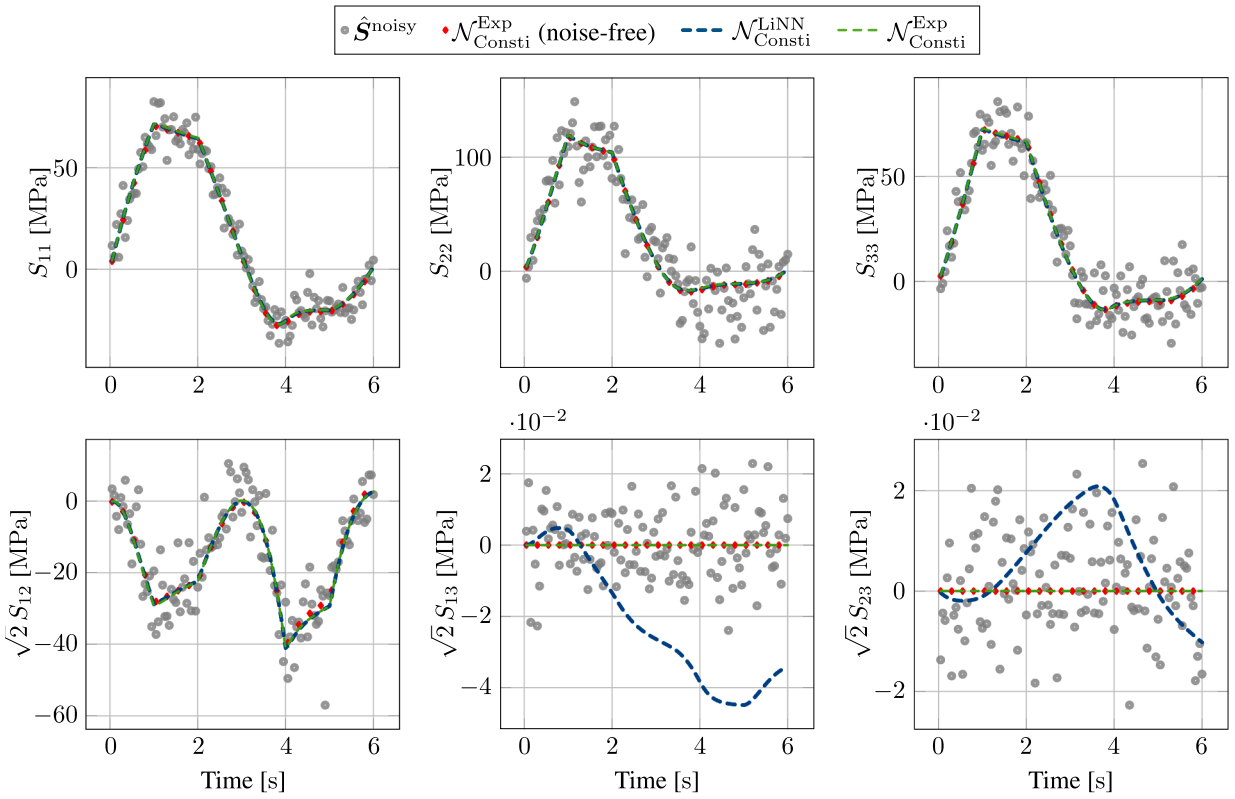

Fig. 19. Training results for the data set corrupted by noise for both $\mathcal{N}^{LiNN}_{Consti}$ and $\mathcal{N}^{Exp}_{Consti}$. The reference model is anisotropic, i.e., $\psi = \psi_{iso} + \psi_{ani}$ and $\varphi = \varphi_{iso} + \varphi_{ani}$ with material parameters given in Table 1. For comparison, the results of the noise-free training case are given by $\mathcal{N}^{Exp}_{Consti}$ (noise-free). Element ID: 111. Curves start with the first loading step at $t = 0.05$. The noise-free counterpart is depicted in Fig. 17.

network $\mathcal{N}^{Exp}_{Consti}$, 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. For all studies, we use the weights and biases determined which belong to the noise-free data.

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. Fig. 22 show the two load programs applied.

Results. The global reaction forces measured at the right edge are plotted in Fig. 22. 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 Fig. 23, 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 $\|\hat{\boldsymbol{u}} - \boldsymbol{u}\|$ 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.

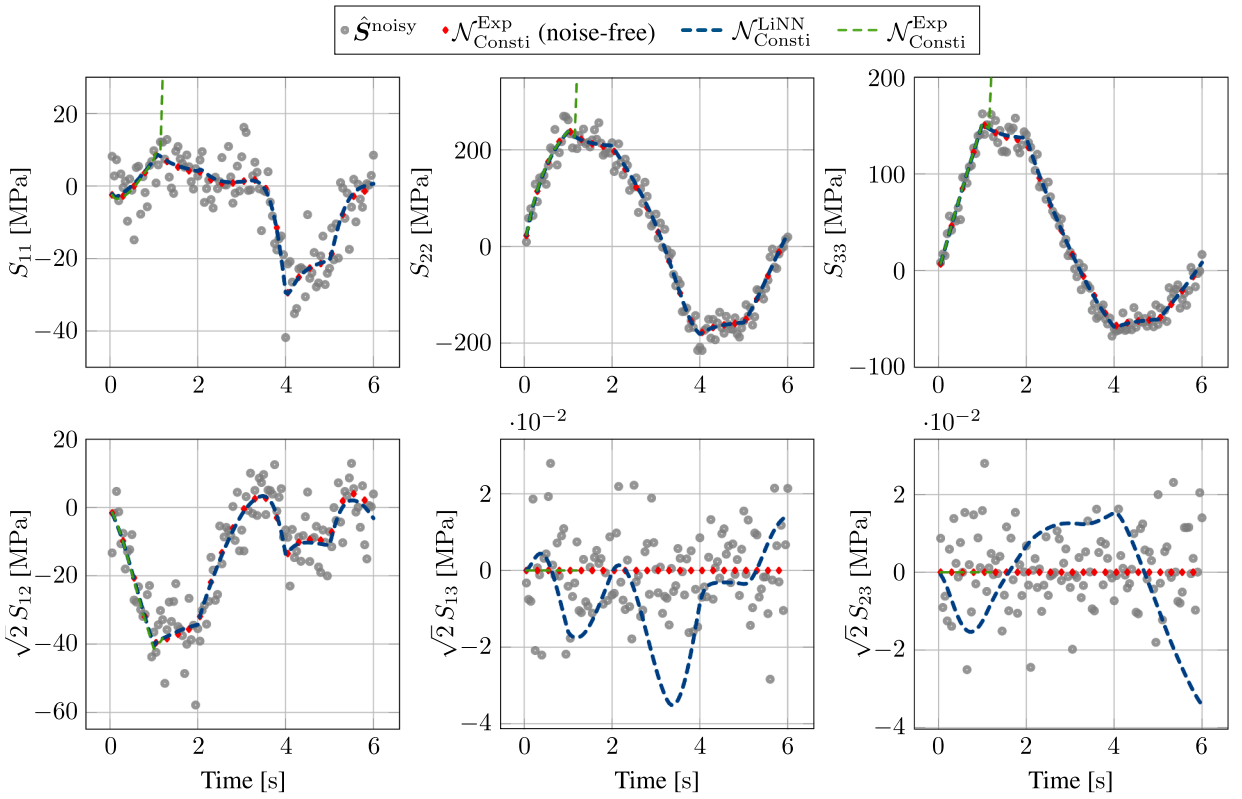

Fig. 20. Testing results for the data set corrupted by noise for both $\mathcal{N}^{LiNN}_{Consti}$ and $\mathcal{N}^{Exp}_{Consti}$. The reference model is anisotropic, i.e., $\psi = \psi_{iso} + \psi_{ani}$ and $\varphi = \varphi_{iso} + \varphi_{ani}$ with material parameters given in Table 1. For comparison, the results of the noise-free training case are given by $\mathcal{N}^{Exp}_{Consti}$ (noise-free). Element ID: 53. Curves start with the first loading step at $t = 0.05$. The noise-free counterpart is depicted in Fig. 18. When using the explicit exponential integrator, the stresses diverge after $t = 1.15$ s.

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

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

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.

Results. The reaction forces are shown in Fig. 24. 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. 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 Fig. 25. 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 align 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.

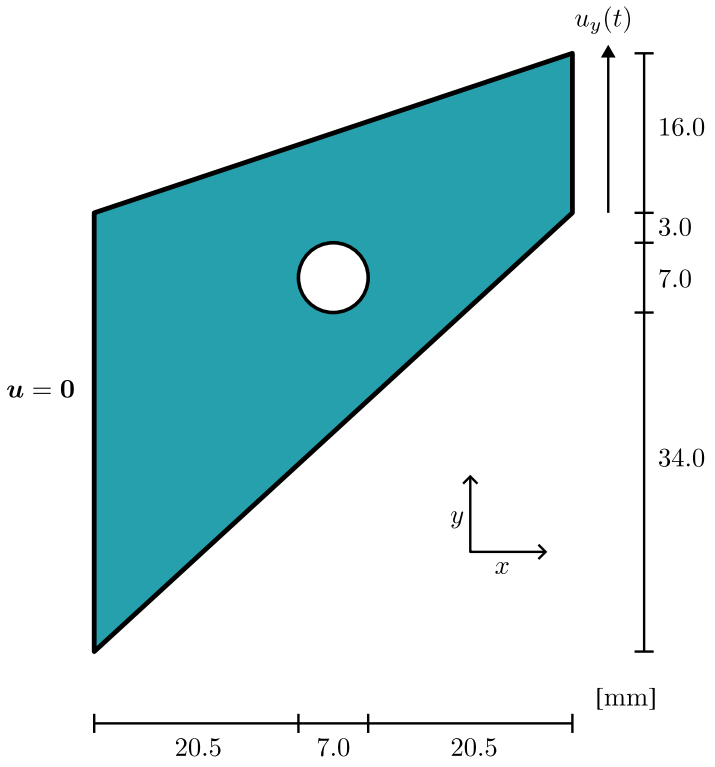

Fig. 21. Cook's membrane with circular hole. Schematic illustration of the BVP used to evaluate predictive capability at the structural level. The left edge is fully clamped, while on the right edge a vertical displacement $u_y(t)$ is prescribed; other components are left free. A circular hole of diameter 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.

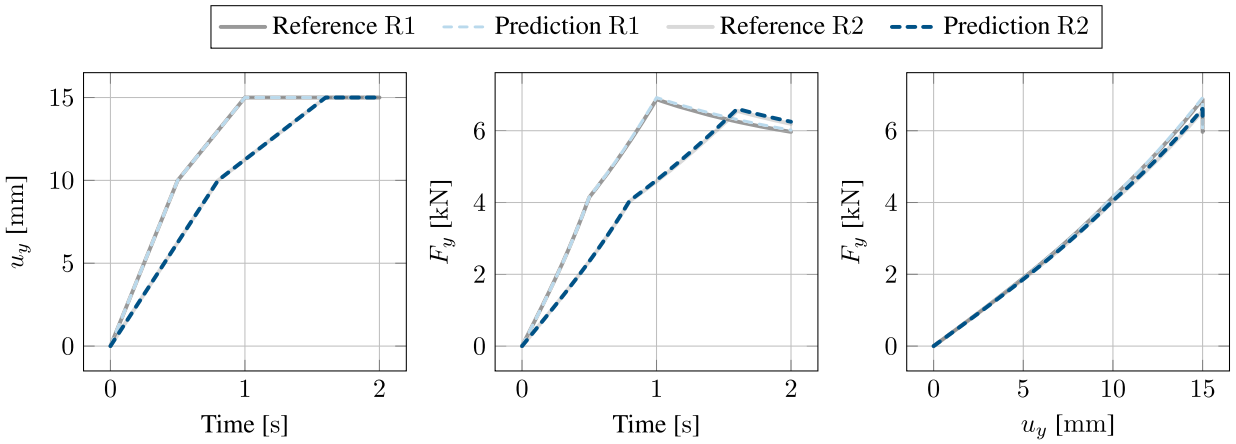

Fig. 22. The loading program and resulting measured forces at the right surface (cf. Fig. 21) for the isotropic material. Left: Prescribed displacement at the right surface in vertical direction at two different loading rates R1 and R2. Middle: Measured reaction force over time at both loading rates. Right: Force-displacement curves for both loading rates. Due to the inelastic effects, the stiffness between the two loading rates varies.

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

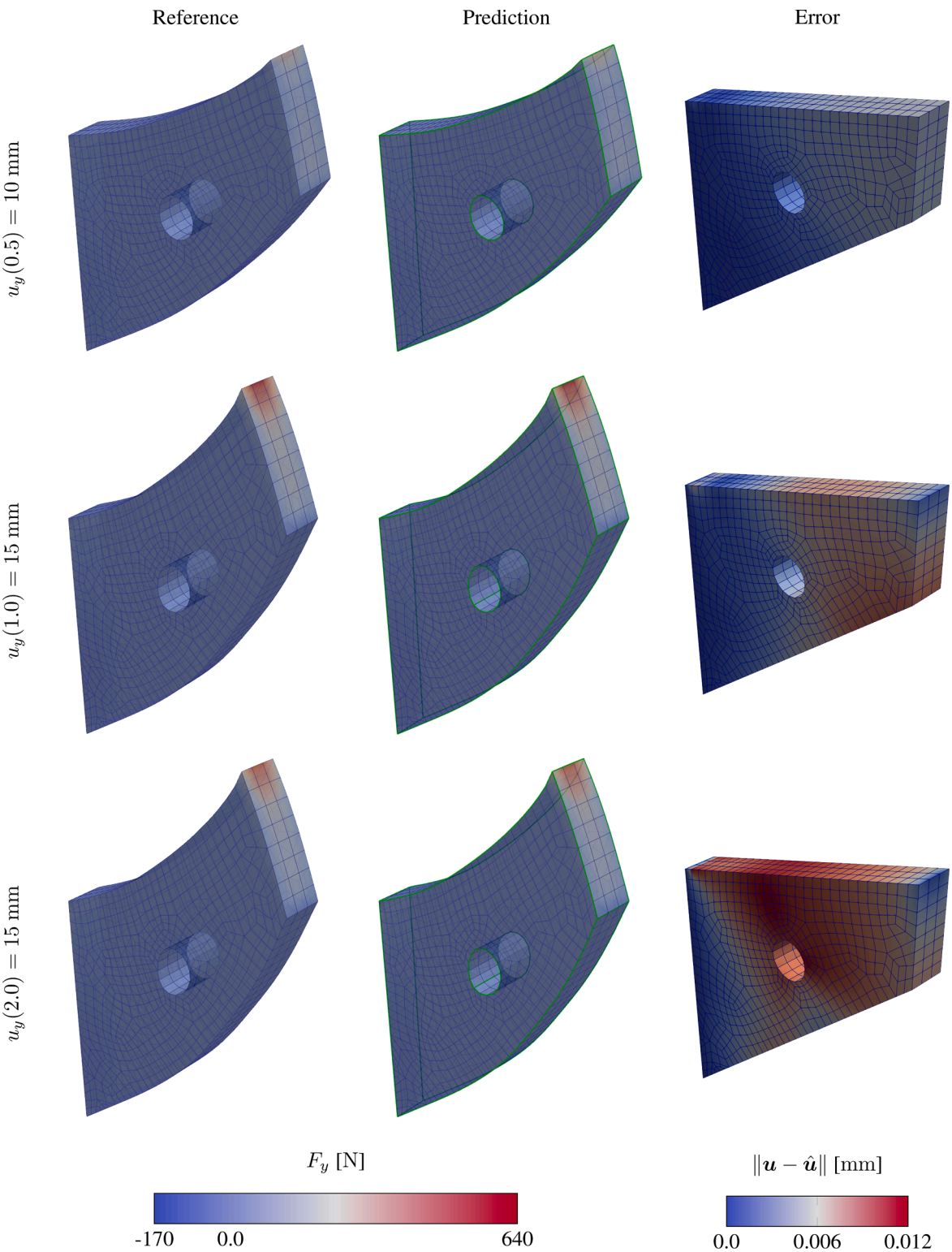

Fig. 23. Contour plots of the Cook's membrane test with isotropic material for the first loading rate. Rows represent selected load steps (cf. Fig. 22). Columns 1-2 show nodal force contours of the reference and neural network prediction, with green lines marking the reference deformation. The last column displays the Euclidean norm of the displacement error $\hat{\boldsymbol{u}} - \boldsymbol{u}$, where $\hat{\boldsymbol{u}}$ corresponds to the reference and $\boldsymbol{u}$ to the prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

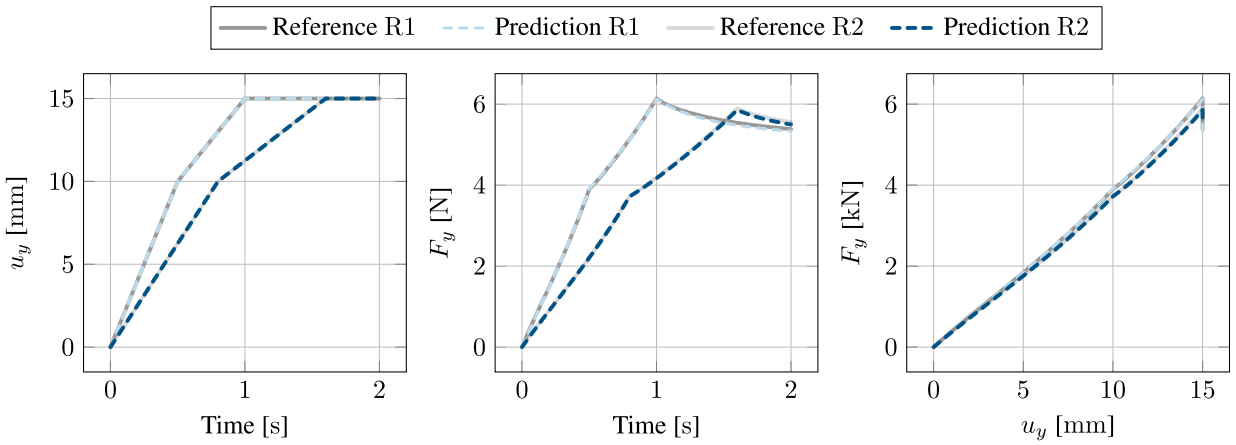

Fig. 24. The loading program and resulting measured forces at the right surface (cf. Fig. 21) for the anisotropic material. Left: Prescribed displacement at the right surface in vertical direction at two different loading rates R1 and R2. Middle: Measured reaction force over time at both loading rates. Right: Force-displacement curves for both loading rates. Due to the inelastic effects, the stiffness between the two loading rates varies.

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 [33,68], 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.

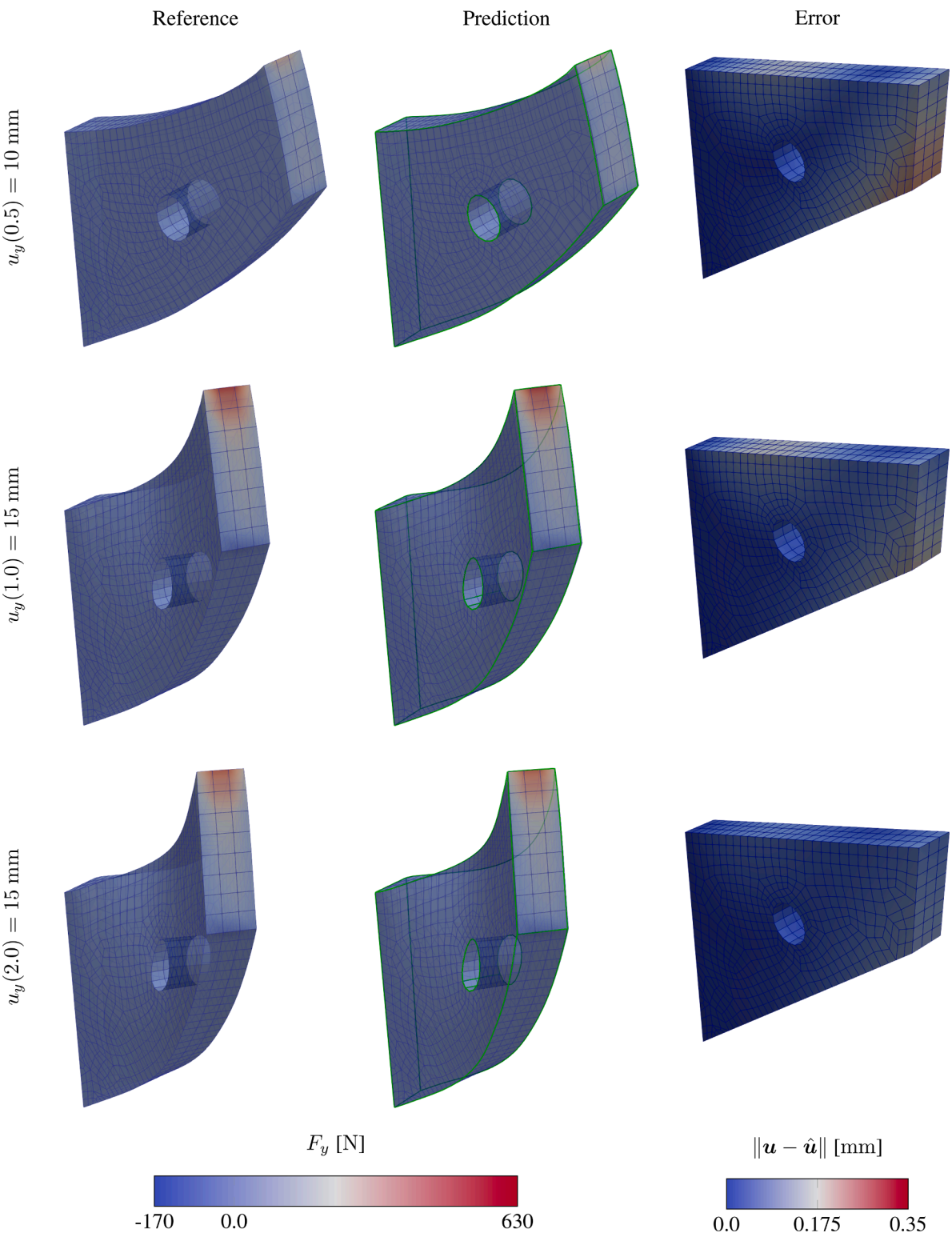

Fig. 25. Contour plots of the Cook's membrane test with anisotropic material for the first loading rate. The specimen twists around the y-axis due to anisotropy, in contrast to the isotropic case (Fig. 23). Rows represent selected load steps (cf. Fig. 24). Columns 1-2 show nodal force contours of the reference and neural network prediction, with green lines marking the reference deformation. The last column displays the Euclidean norm of the displacement error $\hat{\boldsymbol{u}} - \boldsymbol{u}$, where $\hat{\boldsymbol{u}}$ corresponds to the reference and $\boldsymbol{u}$ to the prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Architectural hyperparameters. In this contribution, we made deliberate choices regarding the architectural hyperparameters of the neural networks, including the depth, width, and layer-wise activation functions. These selections were guided by empirical evidence and prior experience in comparable contexts. For instance, we observed that employing rapidly increasing activation functions, such as the exponential function $\exp(\bullet)$, in the earlier layers tended to stabilize the training process, whereas their use in later layers often led to instability. Furthermore, we adopted identical hyperparameter configurations across both the isotropic and anisotropic studies, despite the differing complexity of the underlying material responses. Although the chosen hyperparameters produced satisfactory results in all experiments, they may not represent the optimal trade-off between expressive capacity and architectural simplicity in every case. A systematic analysis of the sensitivity of model performance, convergence characteristics, and generalization behavior with respect to these hyperparameters remains an important direction for future work. Beyond these architectural aspects, we did not explore the influence of training-related parameters such as learning rate, batch size, or the choice of optimization algorithm.

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 [23,34,37,92,121] and growth [41], 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. In case of Physics-Augmented Neural Networks (PANNs), i.e. highly densely connected neural networks, extreme sparsification methods are investigated in [123] which drastically reduce the number effective weights and biases within the network, and thus, enable to interpret the discovered energy. 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. For instance, from a mechanical point of view, examining whether the partial derivative $\partial_{S_{\varphi,1}} \varphi$ vanishes is highly insightful. If this derivative of the dual potential with respect to the hydrostatic pressure of $\boldsymbol{\Sigma}$ is zero, the inelastic evolution is volume-preserving, as in von Mises-type plasticity, irrespective of the discovered neural networks. As shown in [41], similar observations can be made for the deviatoric part of the stress tensor and its influence on isochoric inelastic evolution.

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 [124] 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. In this regard, gating mechanisms, such as those investigated in [125], may be of particular interest for identifying the relevant invariant subsets.

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 for noise-free data sets, even outside the training regime. We encountered issues during the inference process using the explicit exponential integrator scheme when dealing with data sets corrupted by noise in case of anisotropy. Nevertheless, the training procedure employing Liquid Neural Networks remained stable throughout all settings.

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.

CRediT authorship contribution statement

Hagen Holthusen: Writing - review & editing, Writing - original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization; **Ellen Kuhl:** Writing - review & editing, Writing - original draft, Supervision, Funding acquisition.

Data availability

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.17199964. We shared a link via Zenodo.org

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

First of all, the authors would like to thank Ahmad Awad for his help with implementing the finite element framework in `JAX`. Hagen Holthusen 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.

Appendix 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

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

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

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

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

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

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

$$\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{x}\Big|_{x=0} = 10.0, \qquad \frac{\mathrm{d}}{\mathrm{d}x}\sqrt[3]{x}\Big|_{x=0} \approx 21.54435.$$

Cholesky decomposition. Within the time discretization of the inelastic variable, we employ a recurrent Liquid Neural Network (LiNN) architecture to predict the inelastic stretches $\boldsymbol{U}_i \in Sym_+^{3\times3}$ 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 $\boldsymbol{U}_i$ admits a Cholesky factorization,

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

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

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

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 $\boldsymbol{L} = \boldsymbol{I}$ and hence $\boldsymbol{U}_i = \boldsymbol{I}$.

A.3. Thermodynamic consistency

Isotropic materials. We define the stress invariants

$$S_1 := \mathrm{tr}\boldsymbol{\Sigma}, \qquad S_2' := \tfrac{1}{2}\, \mathrm{dev}\boldsymbol{\Sigma} : \mathrm{dev}\boldsymbol{\Sigma}, \qquad S_3' := \tfrac{1}{3}\, \mathrm{tr}\big((\mathrm{dev}\, \boldsymbol{\Sigma})^3\big),$$

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

$$S_\varphi(\boldsymbol{\Sigma}) := \big(S_1(\boldsymbol{\Sigma}),\, S_2(\boldsymbol{\Sigma}),\, S_3(\boldsymbol{\Sigma})\big), \qquad \varphi(\boldsymbol{\Sigma}) := \omega(S_\varphi(\boldsymbol{\Sigma})),$$

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

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

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,

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

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

$$\boldsymbol{\Sigma} : \boldsymbol{D}_i = \big\langle S_\varphi,\, \partial_{S_\varphi} \omega \big\rangle \;\geq\; \omega(S_\varphi) - \omega(\mathbf{0}) = \omega(S_\varphi) \;\geq\; 0,$$

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

$$\boldsymbol{\Sigma} : \partial_{\boldsymbol{\Sigma}} \varphi = \frac{\partial \zeta}{\partial \omega} \big\langle S_\varphi,\, \partial_{S_\varphi} \omega \big\rangle \;\geq 0,$$

so thermodynamic consistency is still preserved.

Anisotropic materials. The push-forward $\tilde{\boldsymbol{M}} = \boldsymbol{F}_i \star \boldsymbol{M}$ represents either a covariant or contravariant mapping of the structural vector; see Eq. (24). Following [93], we may form the irreducible basis in terms of the mixed invariants of the stress and the structural vectors

$$\boldsymbol{\Sigma} : \tilde{\boldsymbol{M}}_\alpha = \mathrm{dev}\boldsymbol{\Sigma} : \tilde{\boldsymbol{M}}_\alpha + \frac{S_1}{3}, \quad \boldsymbol{\Sigma}^2 : \tilde{\boldsymbol{M}}_\alpha = (\mathrm{dev}\boldsymbol{\Sigma})^2 : \tilde{\boldsymbol{M}}_\alpha + 2\frac{S_1}{3}\, \mathrm{dev}\boldsymbol{\Sigma} : \tilde{\boldsymbol{M}}_\alpha + \frac{S_1^2}{9},$$

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

where $\tilde{\boldsymbol{M}}_\alpha := \tilde{\boldsymbol{n}}_\alpha \otimes \tilde{\boldsymbol{n}}_\alpha$ with $\alpha = 1, 2$. We introduce

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

with

$$A_{1\alpha} := \mathrm{dev}\boldsymbol{\Sigma} : \tilde{\boldsymbol{M}}_\alpha, \quad A_{2\alpha} := \sqrt{\tfrac{1}{2}(\mathrm{dev}\boldsymbol{\Sigma})^2 : \tilde{\boldsymbol{M}}_\alpha}, \quad A_3 := \boldsymbol{\Sigma} : \mathrm{sym}(\tilde{\boldsymbol{n}}_1 \otimes \tilde{\boldsymbol{n}}_2), \quad A_4 := \sqrt{\tfrac{1}{2}\, \boldsymbol{\Sigma}^2 : \mathrm{sym}(\tilde{\boldsymbol{n}}_1 \otimes \tilde{\boldsymbol{n}}_2)}.$$

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

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

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

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

ensuring thermodynamic consistency as in the isotropic case.

A.4. Helmholtz free energy

Mixed-variant mappings. In Eq. (25), for $k = r = 0$ and $s = t = 1$,

$$\tilde{\boldsymbol{M}}_1\,\tilde{\boldsymbol{M}}_2 = \frac{1}{\gamma}\,\boldsymbol{F}_i\,\boldsymbol{M}\,\boldsymbol{F}_i^{-1},$$

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

$$\gamma := (\boldsymbol{C}_i : \boldsymbol{M})(\boldsymbol{C}_i^{-1} : \boldsymbol{M}) = (\boldsymbol{n}\cdot\boldsymbol{C}_i\boldsymbol{n})(\boldsymbol{n}\cdot\boldsymbol{C}_i^{-1}\boldsymbol{n}) \;\geq\; 1$$

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

$$\left(\frac{\boldsymbol{C}_i\boldsymbol{M}}{\boldsymbol{C}_i : \boldsymbol{M}}\right)^{-1} := (\boldsymbol{C}_i : \boldsymbol{M})\,\boldsymbol{M}\boldsymbol{C}_i^{-1}, \qquad \left(\frac{\boldsymbol{M}\boldsymbol{C}_i^{-1}}{\boldsymbol{C}_i^{-1} : \boldsymbol{M}}\right)^{-1} := (\boldsymbol{C}_i^{-1} : \boldsymbol{M})\,\boldsymbol{C}_i\boldsymbol{M},$$

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

$$\tilde{\boldsymbol{M}}_1^{-1}\,\tilde{\boldsymbol{M}}_2^{-1} = \gamma\,\boldsymbol{F}_i^{-T}\,\boldsymbol{M}\,\boldsymbol{F}_i^{T},$$

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 $(\boldsymbol{F}, \operatorname{cof}\boldsymbol{F}, J)$. As an example, consider

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

Its second Gâteaux derivative in direction $\boldsymbol{H} \in \mathbb{M}^{3\times3}$ is

$$\mathfrak{D}^2_{\boldsymbol{F}}[f(\boldsymbol{F}, \boldsymbol{G})](\boldsymbol{H}, \boldsymbol{H}) = \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\operatorname{tr}\left((\boldsymbol{F}+\epsilon\boldsymbol{H})^T(\boldsymbol{F}+\epsilon\boldsymbol{H})\,\boldsymbol{G}\right)\Big|_{\epsilon=0} = 2\,\boldsymbol{G} : \boldsymbol{H}^T\boldsymbol{H}.$$

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

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

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

where $\times$ denotes the vector cross product. Let $\boldsymbol{B} \in Sym^{3\times3}_{+}$ and $\boldsymbol{G} = \boldsymbol{g}\otimes\boldsymbol{g}$. Then their outer product can be written as

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

which is positive semi-definite, with B_i and $\boldsymbol{p}_i$ denoting the eigenvalues and eigenvectors of $\boldsymbol{B}$, respectively. More generally, the outer tensor product of any two tensors $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{M}^{3\times3}$ is given by

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

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

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

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

$$\operatorname{tr}\big(\boldsymbol{C}_e\,(\boldsymbol{B}_i\#\tilde{\boldsymbol{M}})\big) = \operatorname{tr}\big(\boldsymbol{C}\,\boldsymbol{F}_i^{-1}(\boldsymbol{B}_i\#\tilde{\boldsymbol{M}})\boldsymbol{F}_i^{-T}\big)$$

is polyconvex, since $\boldsymbol{F}_i^{-1}(\boldsymbol{B}_i\#\tilde{\boldsymbol{M}})\boldsymbol{F}_i^{-T}$ is positive semi-definite. Finally, using Eq. (25), we obtain

$$\operatorname{tr}\big(\boldsymbol{C}\,\boldsymbol{F}_i^{-1}(\boldsymbol{B}_i\#\tilde{\boldsymbol{M}})\boldsymbol{F}_i^{-T}\big) = \operatorname{tr}\big(\boldsymbol{C}_i^{-1}\boldsymbol{C}\,(\boldsymbol{C}_i\#\boldsymbol{C}_i\boldsymbol{M})\big) = \operatorname{tr}\big(\boldsymbol{C}\boldsymbol{C}_i^{-1}\,(\boldsymbol{C}_i\#\boldsymbol{M}\boldsymbol{C}_i)\big).$$

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:

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

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

Here, a, b, c, α, and η denote material parameters, while $\boldsymbol{G}$ 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 $\boldsymbol{e}_1$, the reference metric reads

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

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

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

The corresponding rotation matrix is

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

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

$$\boldsymbol{G} = \boldsymbol{R}\,\boldsymbol{G}_0\,\boldsymbol{R}^T.$$

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

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

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

$$\psi_{iso}^{eq/neq}(\boldsymbol{A}) = a\,\mathrm{tr}\,\boldsymbol{A} + b\,\mathrm{tr}\,\mathrm{cof}\,\boldsymbol{A} + c\,\det\boldsymbol{A} - \tfrac{d_1}{2}\,\ln\det\boldsymbol{A}, \qquad d_1 = 2a+4b+2c.$$

Further, for the anisotropic Helmholtz free energy, we adopt

$$\psi_{ani}^{eq/neq}(\boldsymbol{A},\boldsymbol{B}) = \frac{\eta}{\alpha\,(\mathrm{tr}\,\boldsymbol{B})^\alpha}((\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B})^\alpha + (\mathrm{tr}\,\mathrm{cof}\,\boldsymbol{A}\,\boldsymbol{B})^\alpha) - \tfrac{d_2}{2}\,\ln\det\boldsymbol{A}, \qquad d_2 = 2\eta.$$

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

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

while the anisotropic contribution is given by

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

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,

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

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

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

with $n_x = 7$ (strain-like features $\boldsymbol{C}$ 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.

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

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

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

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

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

$$\mathcal{N}_\alpha : \mathbb{R}^{n_x+n_h} \to \mathbb{R}^{n_h}, \quad (\mathbf{x}_t, \mathbf{h}_{t-1}) \xrightarrow{GELU} 64 \xrightarrow{GELU} 64 \xrightarrow{GELU} 64 \xrightarrow{GELU} 64 \xrightarrow{ReLU} \boldsymbol{\alpha}.$$

Here, $n_x = 7$ (strain-like features $\boldsymbol{C}$ 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.

References

[1] L. Lu, R. Pestourie, W. Yao, Z. Wang, F. Verdugo, S.G. Johnson, Physics-informed neural networks with hard constraints for inverse design, SIAM J. Sci. Comput. 43 (6) (2021) B1105–B1132.

[2] A.A. Jadoon, K.A. Kalina, M.K. Rausch, R. Jones, J.N. Fuhg, Inverse design of anisotropic microstructures using physics-augmented neural networks, J. Mech. Phys. Solids 203 (2025) 106161. https://doi.org/https://doi.org/10.1016/j.jmps.2025.106161

[3] S. Ghouli, M. Flaschel, S. Kumar, L. De Lorenzis, A topology optimisation framework to design test specimens for one-shot identification or discovery of material models, J. Mech. Phys. Solids 203 (2025) 106210. https://doi.org/https://doi.org/10.1016/j.jmps.2025.106210

[4] J. Stollberg, T. Gangwar, O. Weeger, D. Schillinger, Multiscale topology optimization of functionally graded lattice structures based on physics-augmented neural network material models, Comput. Methods Appl. Mech. Eng. 438 (2025) 117808. https://doi.org/https://doi.org/10.1016/j.cma.2025.117808

[5] A. Danoun, E. Prulière, Y. Chemisky, FE-LSTM: a hybrid approach to accelerate multiscale simulations of architectured materials using recurrent neural networks and finite element analysis, Comput. Methods Appl. Mech. Eng. 429 (2024) 117192. https://doi.org/https://doi.org/10.1016/j.cma.2024.117192

[6] K.A. Kalina, L. Linden, J. Brummund, M. Kästner, FE^{ANN}: an efficient data-driven multiscale approach based on physics-constrained neural networks and automated data mining, Comput. Mech. 71 (5) (2023) 827-851. https://doi.org/10.1007/s00466-022-02260-0

[7] D.W. Abueidda, M.E. Mobasher, I-FENN for thermoelasticity based on physics-informed temporal convolutional network (PI-TCN), Comput. Mech. 74 (6) (2024) 1229-1259. https://doi.org/10.1007/s00466-024-02475-3

[8] J.N. Fuhg, A. Jadoon, O. Weeger, D.T. Seidl, R.E. Jones, Polyconvex neural network models of thermoelasticity, 2024, (????). https://arxiv.org/abs/2404.15562. 2404.15562

[9] J. Kehls, E. Kuhl, T. Brepols, K. Linka, H. Holthusen, Autoencoder-based non-intrusive model order reduction in continuum mechanics, 2025, (????). https://arxiv.org/abs/2509.02237. 2509.02237

[10] J.S. Hesthaven, S. Ubbiali, Non-intrusive reduced order modeling of nonlinear problems using neural networks, J. Comput. Phys. 363 (2018) 55–78.

[11] M. Flaschel, S. Kumar, L. De Lorenzis, Unsupervised discovery of interpretable hyperelastic constitutive laws, Comput. Methods Appl. Mech. Eng. 381 (2021) 113852. https://doi.org/https://doi.org/10.1016/j.cma.2021.113852

[12] K. Linka, M. Hillgärtner, K.P. Abdolazizi, R.C. Aydin, M. Itskov, C.J. Cyron, Constitutive artificial neural networks: a fast and general approach to predictive data-driven constitutive modeling by deep learning, J. Comput. Phys. 429 (2021) 110010. https://doi.org/https://doi.org/10.1016/j.jcp.2020.110010

[13] D.K. Klein, M. Fernández, R.J. Martin, P. Neff, O. Weeger, Polyconvex anisotropic hyperelasticity with neural networks, J. Mech. Phys. Solids 159 (2022) 104703. https://doi.org/https://doi.org/10.1016/j.jmps.2021.104703

[14] M. Chiachío, M. Megía, J. Chiachío, J. Fernandez, M.L. Jalón, Structural digital twin framework: formulation and technology integration, Autom. Constr. 140 (2022) 104333. https://doi.org/https://doi.org/10.1016/j.autcon.2022.104333

[15] E. Duede, W. Dolan, A. Bauer, I. Foster, K. Lakhani, Oil & water? Diffusion of AI within and across scientific fields, 2024, (????). https://arxiv.org/abs/2405.15828. 2405.15828

[16] S. Hajkowicz, C. Sanderson, S. Karimi, A. Bratanova, C. 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, Technol. Soc. 74 (2023). https://doi.org/10.1016/j.techsoc.2023.102260

[17] J. Ghaboussi, J.H. Garrett, X. Wu, Knowledge-based modeling of material behavior with neural networks, J. Eng. Mech. 117 (1) (1991) 132–153. https://doi.org/https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%290733-9399%281991%29117%3A1%28132%29 https://doi.org/10.1061/(ASCE)0733-9399(1991)117:1(132)

[18] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707. https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045

[19] F. Masi, I. Stefanou, P. Vannucci, V. Maffi-Berthier, Thermodynamics-based artificial neural networks for constitutive modeling, J. Mech. Phys. Solids 147 (2021) 104277. https://doi.org/https://doi.org/10.1016/j.jmps.2020.104277

[20] F. Masi, I. Stefanou, Multiscale modeling of inelastic materials with thermodynamics-based artificial neural networks (TANN), Comput. Methods Appl. Mech. Eng. 398 (2022) 115190. https://doi.org/https://doi.org/10.1016/j.cma.2022.115190

[21] K. Linka, E. Kuhl, A new family of constitutive artificial neural networks towards automated model discovery, Comput. Methods Appl. Mech. Eng. 403 (2023) 115731. https://doi.org/https://doi.org/10.1016/j.cma.2022.115731

[22] L. Linden, D.K. Klein, K.A. Kalina, J. Brummund, O. Weeger, M. Kästner, Neural networks meet hyperelasticity: a guide to enforcing physics, J. Mech. Phys. Solids 179 (2023) 105363. https://doi.org/10.1016/j.jmps.2023.105363

[23] M. Flaschel, S. Kumar, L. De Lorenzis, Discovering plasticity models without stress data, npj Comput. Mater. 8 (1) (2022). https://doi.org/10.1038/s41524-022-00752-4

[24] H. Xu, M. Flaschel, L. De Lorenzis, Discovering non-associated pressure-sensitive plasticity models with EUCLID, Adv. Model. Simul. Eng. Sci. 12 (1) (2025). https://doi.org/10.1186/s40323-024-00281-3

[25] E. Marino, M. Flaschel, S. Kumar, L. De Lorenzis, Automated identification of linear viscoelastic constitutive laws with EUCLID, Mech. Mater. 181 (2023) 104643. https://doi.org/https://doi.org/10.1016/j.mechmat.2023.104643

[26] M. Flaschel, S. Kumar, L. De Lorenzis, Automated discovery of generalized standard material models with EUCLID, Comput. Methods Appl. Mech. Eng. 405 (2023) 115867. https://doi.org/https://doi.org/10.1016/j.cma.2022.115867

[27] B. Halphen, Q.S. Nguyen, Sur les matériaux standard généralisés, J. Mec. 14 (1) (1975) 39–63.

[28] K. Hackl, Generalized standard media and variational principles in classical and finite strain elastoplasticity, J. Mech. Phys. Solids 45 (5) (1997) 667–688. https://doi.org/https://doi.org/10.1016/S0022-5096(96)00110-X

[29] M. Flaschel, P. Steinmann, L. De Lorenzis, E. Kuhl, Convex neural networks learn generalized standard material models, J. Mech. Phys. Solids 200 (2025) 106103. https://doi.org/https://doi.org/10.1016/j.jmps.2025.106103

[30] S. Huang, Z. He, B. Chem, C. Reina, Variational onsager neural networks (VONNs): a thermodynamics-based variational learning strategy for non-equilibrium PDEs, J. Mech. Phys. Solids 163 (2022) 104856. https://doi.org/https://doi.org/10.1016/j.jmps.2022.104856

[31] H. Holthusen, L. Lamm, T. Brepols, S. Reese, E. Kuhl, et al., Theory and implementation of inelastic constitutive artificial neural networks, Comput. Methods Appl. Mech. Eng. 428 (2024) 117063. https://doi.org/https://doi.org/10.1016/j.cma.2024.117063

[32] A. Benady, E. Baranger, L. Chamoin, Unsupervised learning of history-dependent constitutive material laws with thermodynamically-consistent neural networks in the modified constitutive relation error framework, Comput. Methods Appl. Mech. Eng. 425 (2024) 116967. https://doi.org/https://doi.org/10.1016/j.cma.2024.116967

[33] M. Rosenkranz, K.A. Kalina, J. Brummund, W. Sun, M. Kästner, Viscoelasticty with physics-augmented neural networks: model formulation and training methods without prescribed internal variables, Comput. Mech. 74 (6) (2024) 1279–1301. https://doi.org/10.1007/s00466-024-02477-1

[34] W.G. Dettmer, E.J. Muttio, R. Alhayki, D. Perić, A framework for neural network based constitutive modelling of inelastic materials, Comput. Methods Appl. Mech. Eng. 420 (2024) 116672. https://doi.org/https://doi.org/10.1016/j.cma.2023.116672

[35] F. As'ad, C. Farhat, A mechanics-informed deep learning framework for data-driven nonlinear viscoelasticity, Comput. Methods Appl. Mech. Eng. 417 (2023) 116463. https://doi.org/https://doi.org/10.1016/j.cma.2023.116463

[36] H. Holthusen, K. Linka, E. Kuhl, T. Brepols, A generalized dual potential for inelastic constitutive artificial neural networks: a JAX implementation at finite strains, J. Mech. Phys. Solids (2025) 106337. https://doi.org/https://doi.org/10.1016/j.jmps.2025.106337

[37] A.A. Jadoon, K.A. Meyer, J.N. Fuhg, Automated model discovery of finite strain elastoplasticity from uniaxial experiments, Comput. Methods Appl. Mech. Eng. 435 (2025) 117653. https://doi.org/https://doi.org/10.1016/j.cma.2024.117653

[38] B. Boes, J.-W. Simon, H. Holthusen, Accounting for plasticity: an extension of inelastic constitutive artificial neural networks, 2025, (????). https://arxiv.org/abs/2407.19326. 2407.19326

[39] T. van der Velden, B. Boes, T. Brepols, E. Kuhl, H. Holthusen, A note on constitutive artificial neural networks for finite strain plasticity (2025). https://doi.org/10.2139/ssrn.5588650

[40] F. Dammaß, K.A. Kalina, M. Kästner, Neural networks meet phase-field: a hybrid fracture model, Comput. Methods Appl. Mech. Eng. 440 (2025) 117937. https://doi.org/https://doi.org/10.1016/j.cma.2025.117937

[41] H. Holthusen, T. Brepols, K. Linka, E. Kuhl, Automated model discovery for tensional homeostasis: constitutive machine learning in growth and remodeling, Comput. Biol. Med. 186 (2025) 109691. https://doi.org/10.1016/j.compbiomed.2025.109691

[42] A.A. Jadoon, R.G. Patel, B.N. Granzow, R.E. Jones, D.T. Seidl, J.N. Fuhg, Thermodynamically consistent hybrid and permutation-invariant neural yield functions for anisotropic plasticity, 2025, (????). https://arxiv.org/abs/2508.15923. 2508.15923

[43] B. Amos, L. Xu, J.Z. Kolter, Input convex neural networks, 2017, (????). https://arxiv.org/abs/1609.07152. 1609.07152

[44] K.A. Kalina, J. Brummund, M. Kästner, A physics-augmented neural network framework for finite strain incompressible viscoelasticity, 2025, (????). https://arxiv.org/abs/2511.02959. 2511.02959

[45] S. Kumar, D.T. Seidl, B.N. Granzow, J. Yang, J.N. Fuhg, A comparative study of calibration techniques for finite strain elastoplasticity: numerically-exact sensitivities for FEMU and VFM, Comput. Methods Appl. Mech. Eng. 444 (2025) 118159. https://doi.org/https://doi.org/10.1016/j.cma.2025.118159

[46] R.E. Jones, A.L. Frankel, K.L. Johnson, A neural ordinary differential equation framework for modeling inelastic stress response via internal state variables, J. Mach. Learn. Model. Comput. 3 (3) (2022) 1-35. https://doi.org/10.1615/jmachlearnmodelcomput.2022042917

[47] R.E. Jones, J.N. Fuhg, An attention-based neural ordinary differential equation framework for modeling inelastic processes, 2025, (????). https://arxiv.org/abs/2502.10633. 2502.10633

[48] L. Borkowski, C. Sorini, A. Chattopadhyay, Recurrent neural network-based multiaxial plasticity model with regularization for physics-informed constraints, Comput. Struct. 258 (2022) 106678. https://doi.org/https://doi.org/10.1016/j.compstruc.2021.106678

[49] J.-Z. Li, Z.-P. Guan, J.-R. Chen, H.-C. Jin, A long short-term memory-based constitutive modeling framework for capturing strain path dependence in plastic deformation, Mech. Mater. 205 (2025) 105325. https://doi.org/https://doi.org/10.1016/j.mechmat.2025.105325

[50] B. Guo, Z. Lin, Q. He, History-aware neural operator: robust data-driven constitutive modeling of path-dependent materials, Comput. Methods Appl. Mech. Eng. 447 (2025) 118358. https://doi.org/https://doi.org/10.1016/j.cma.2025.118358

[51] M. Maurizi, C. Gao, F. Berto, Predicting stress, strain and deformation fields in materials and structures with graph neural networks, Sci. Rep. 12 (1) (2022). https://doi.org/10.1038/s41598-022-26424-3

[52] R.G. Patel, R.E. Jones, D.T. Seidl, B.N. Granzow, J.N. Fuhg, A general, automated method for building structural tensors of arbitrary order for anisotropic function representations, 2025, (????). https://arxiv.org/abs/2507.09088. 2507.09088

[53] V. Tac, F. Sahli Costabal, A.B. Tepole, Data-driven tissue mechanics with polyconvex neural ordinary differential equations, Comput. Methods Appl. Mech. Eng. 398 (2022) 115248. https://doi.org/https://doi.org/10.1016/j.cma.2022.115248

[54] J.N. Fuhg, N. Bouklas, et al., On physics-informed data-driven isotropic and anisotropic constitutive models through probabilistic machine learning and space-filling sampling, Comput. Methods Appl. Mech. Eng. 394 (2022) 114915. https://doi.org/https://doi.org/10.1016/j.cma.2022.114915

[55] K.A. Kalina, J. Brummund, W. Sun, M. Kästner, Neural networks meet anisotropic hyperelasticity: a framework based on generalized structure tensors and isotropic tensor functions, Comput. Methods Appl. Mech. Eng. 437 (2025) 117725. https://doi.org/https://doi.org/10.1016/j.cma.2024.117725

[56] J.A. McCulloch, E. Kuhl, Automated model discovery for textile structures: the unique mechanical signature of warp knitted fabrics, Acta Biomater. 189 (2024) 461–477. https://doi.org/https://doi.org/10.1016/j.actbio.2024.09.051
[57] V. Tac, V.D. Sree, M.K. Rausch, A.B. Tepole, Data-driven modeling of the mechanical behavior of anisotropic soft biological tissue, Eng. Comput. 38 (5) (2022) 4167-4182. https://doi.org/10.1007/s00366-022-01733-3
[58] O. Gültekin, A. Moeineddin, B. Cans1z, K. Sveric, A. Linke, M. Kaliske, A physics-informed neural network model for the anisotropic hyperelasticity of the human passive myocardium, Int. J. Numer. Methods Eng. 126 (14) (2025) e70067. https://doi.org/https://doi.org/10.1002/nme.70067
[59] D. Martonová, S. Leyendecker, G.A. Holzapfel, E. Kuhl, Discovering dispersion: how robust is automated model discovery for human myocardial tissue?, Biomech. Model. Mechanobiol. (2025). https://doi.org/10.1007/s10237-025-02005-x
[60] T. Vervenne, M. Peirlinck, N. Famaey, E. Kuhl, Constitutive neural networks for main pulmonary arteries: discovering the undiscovered, Biomech. Model. Mechanobiol. 24 (2) (2025) 615-634. https://doi.org/10.1007/s10237-025-01930-1
[61] J.N. Fuhg, G.A. Padmanabha, N. Bouklas, B. Bahmani, W. Sun, N.N. Vlassis, M. Flaschel, P. Carrara, L. De Lorenzis, A review on data-driven constitutive laws for solids, 2024, (????). https://arxiv.org/abs/2405.03658. 2405.03658
[62] J. Watson, C. Song, O. Weeger, T. Gruner, A.T. Le, K. Pompetzki, A. Hendawy, O. Arenz, W. Trojak, M. Cranmer, C. D'Eramo, F. Bülow, T. Goyal, J. Peters, M.W. Hoffman, Machine learning with physics knowledge for prediction: a survey, 2025, (????). https://arxiv.org/abs/2408.09840. 2408.09840
[63] P. Germain, Q.S. Nguyen, P. Suquet, Continuum thermodynamics, J. Appl. Mech. 50 (4b) (1983) 1010-1020. https://doi.org/10.1115/1.3167184
[64] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, Princeton, 1970. https://doi.org/10.1515/9781400873173. https://doi.org/doi:10.1515/9781400873173
[65] J. Sill, Monotonic networks, in: Neural Information Processing Systems, 1997. https://api.semanticscholar.org/CorpusID:10546146.
[66] H. Daniels, M. Velikova, Monotone and partially monotone neural networks, IEEE Trans. Neural Netw. 21 (6) (2010) 906-917. https://doi.org/10.1109/tnn.2010.2044803
[67] S. You, D. Ding, K. Canini, J. Pfeifer, M. Gupta, Deep lattice networks and partial monotonic functions, 2017, (????). https://arxiv.org/abs/1709.06680. 1709.06680
[68] F. As'ad, C. Farhat, A mechanics-informed neural network framework for data-driven nonlinear viscoelasticity, in: AIAA SCITECH 2023 Forum, American Institute of Aeronautics and Astronautics, 2023. https://doi.org/10.2514/6.2023-0949
[69] R. Hasani, M. Lechner, A. Amini, D. Rus, R. Grosu, Liquid time-constant networks, Proc. AAAI Conf. Artif. Intell. 35 (9) (2021) 7657–7666. https://doi.org/10.1609/aaai.v35i9.16936
[70] S.M. Cox, P.C. Matthews, Exponential time differencing for stiff systems, J. Comput. Phys. 176 (2) (2002) 430-455. https://doi.org/10.1006/jcph.2002.6995
[71] C. Eckart, The thermodynamics of irreversible processes. IV. The theory of elasticity and anelasticity, Phys. Rev. 73 (4) (1948) 373-382. https://doi.org/10.1103/physrev.73.373
[72] E. Kröner, Allgemeine kontinuumstheorie der versetzungen und eigenspannungen, Arch. Ration. Mech. Anal. 4 (1) (1959) 273-334. https://doi.org/10.1007/bf00281393
[73] F.c. Sidoroff, Un modèle viscoélastique non linéaire avec configuration intermédiaire, J. Mec. 13 (1974) 679–713.
[74] E.K. Rodriguez, A. Hoger, A.D. McCulloch, Stress-dependent finite growth in soft elastic tissues, J. Biomech. 27 (4) (1994) 455-467. https://doi.org/10.1016/0021-9290(94)90021-3
[75] D. Christ, S. Reese, A finite element model for shape memory alloys considering thermomechanical couplings at large strains, Int. J. Solids Struct. 46 (20) (2009) 3694-3709. https://doi.org/10.1016/j.ijsolstr.2009.06.017
[76] J. Casey, A convenient form of the multiplicative decomposition of the deformation gradient, Math. Mech. Solids 22 (3) (2016) 528-537. https://doi.org/10.1177/1081286515598662
[77] J.-P. Boehler, A simple derivation of representations for non-polynomial constitutive equations in some cases of anisotropy, ZAMM-J. Appl. Math. Mech./Z. Angew. Math. Mech. 59 (4) (1979) 157–167.
[78] A.J.M. Spencer, Constitutive Theory for Strongly Anisotropic Solids, Springer Vienna, 1984, p. 1-32. https://doi.org/10.1007/978-3-7091-4336-0_1
[79] J. Schröder, P. Neff, Invariant formulation of hyperelastic transverse isotropy based on polyconvex free energy functions, Int. J. Solids Struct. 40 (2) (2003) 401-445. https://doi.org/10.1016/s0020-7683(02)00458-4
[80] J.M. Ball, Convexity conditions and existence theorems in nonlinear elasticity, Arch. Ration. Mech. Anal. 63 (4) (1976) 337-403. https://doi.org/10.1007/bf00279992
[81] J.M. Ball, Constitutive inequalities and existence theorems in nonlinear elastostatics, in: Nonlinear Analysis and Mechanics: Heriot-Watt Symposium, 1, Pitman London, 1977, pp. 187–241.
[82] J.E. Marsden, T.J.R. Hughes, Mathematical Foundations of Elasticity, Courier Corporation, 1994.
[83] B. Svendsen, On the modelling of anisotropic elastic and inelastic material behaviour at large deformation, Int. J. Solids Struct. 38 (52) (2001) 9579-9599. https://doi.org/10.1016/s0020-7683(01)00140-8
[84] B.D. Coleman, W. Noll, Foundations of linear viscoelasticity, Rev. Mod. Phys. 33 (2) (1961) 239-249. https://doi.org/10.1103/revmodphys.33.239
[85] B.D. Coleman, W. Noll, The thermodynamics of elastic materials with heat conduction and viscosity, Arch. Ration. Mech. Anal. 13 (1) (1963) 167-178. https://doi.org/10.1007/bf01262690
[86] B.D. Coleman, M.E. Gurtin, Thermodynamics with internal state variables, J. Chem. Phys. 47 (2) (1967) 597-613. https://doi.org/10.1063/1.1711937
[87] W. Dettmer, S. Reese, On the theoretical and numerical modelling of Armstrong-Frederick kinematic hardening in the finite strain regime, Comput. Methods Appl. Mech. Eng. 193 (1-2) (2004) 87-116. https://doi.org/10.1016/j.cma.2003.09.005
[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) (1971) 433–455.
[89] P. Germain, Functional concepts in continuum mechanics, Meccanica 33 (5) (1998) 433-444. https://doi.org/10.1023/a:1004304224398
[90] Y.F. Dafalias, Issues on the constitutive formulation at large elastoplastic deformations, part 1: kinematics, Acta Mech. 69 (1-4) (1987) 119-138. https://doi.org/10.1007/bf01175717
[91] S. Reese, Meso-macro modelling of fibre-reinforced rubber-like composites exhibiting large elastoplastic deformation, Int. J. Solids Struct. 40 (4) (2003) 951-980. https://doi.org/10.1016/s0020-7683(02)00602-9
[92] B. Boes, J.-W. Simon, S. Reese, H. Holthusen, A novel continuum mechanical framework for decoupled material behavior in thickness and in-plane directions, Comput. Methods Appl. Mech. Eng. 415 (2023) 116192. https://doi.org/10.1016/j.cma.2023.116192
[93] Q.S. Zheng, Theory of representations for tensor functions–a unified invariant approach to constitutive equations, Appl. Mech. Rev. 47 (11) (1994) 545-587. https://doi.org/10.1115/1.3111066
[94] R. Glüge, S. Bucci, Does convexity of yield surfaces in plasticity have a physical significance?, Math. Mech. Solids 23 (9) (2017) 1364-1373. https://doi.org/10.1177/1081286517721599
[95] A. Janećka, M. Pavelka, Non-convex dissipation potentials in multiscale non-equilibrium thermodynamics, Continuum Mech. Thermodyn. 30 (4) (2018) 917-941. https://doi.org/10.1007/s00161-018-0667-1
[96] J.H. Argyris, G. Faust, J. Szimmat, E.P. Warnke, K.J. Willam, Recent developments in the finite element analysis of prestressed concrete reactor vessels, Nucl. Eng. Des. 28 (1) (1974) 42-75. https://doi.org/10.1016/0029-5493(74)90088-0
[97] H. Holthusen, C. Rothkranz, L. Lamm, T. Brepols, S. Reese, Inelastic material formulations based on a co-rotated intermediate configuration–application to bioengineered tissues, J. Mech. Phys. Solids 172 (2023) 105174. https://doi.org/10.1016/j.jmps.2022.105174
[98] Z. Liu, R. Ortigosa, A.J. Gil, J. Bonet, Large strain constitutive modelling of soft compressible and incompressible solids: generalised isotropic and anisotropic viscoelasticity, J. Mech. Phys. Solids 203 (2025) 106194. https://doi.org/https://doi.org/10.1016/j.jmps.2025.106194
[99] D.K. Klein, R. Ortigosa, J. Martínez-Frutos, O. Weeger, Finite electro-elasticity with physics-augmented neural networks, Comput. Methods Appl. Mech. Eng. 400 (2022) 115501. https://doi.org/https://doi.org/10.1016/j.cma.2022.115501

[100] I.N. Vladimirov, M.P. Pietryga, S. Reese, On the modelling of non-linear kinematic hardening at finite strains with application to springback–comparison of time integration algorithms, Int. J. Numer. Methods Eng. 75 (1) (2008) 1–28. https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.2234. https://doi.org/https://doi.org/10.1002/nme.2234
[101] D. Hendrycks, K. Gimpel, Gaussian error linear units (GELUs), 2023, (????). https://arxiv.org/abs/1606.08415. 1606.08415
[102] J. Bradbury, R. Frostig, P. Hawkins, M.J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, Q. Zhang, JAX: composable transformations of Python + NumPy programs, http://github.com/jax-ml/jax, (2018).
[103] C. Geuzaine, J.-F. Remacle, GMSH: a 3-D finite element mesh generator with built-in pre- and post-processing facilities, Int. J. Numer. Methods Eng. 79 (11) (2009) 1309–1331. https://doi.org/10.1002/nme.2579
[104] U. Ayachit, The ParaView Guide: A Parallel Visualization Application (2015).
[105] J. Schröder, P. Neff, V. Ebbing, Anisotropic polyconvex energies on the basis of crystallographic motivated structural tensors, J. Mech. Phys. Solids 56 (12) (2008) 3486–3506. https://doi.org/https://doi.org/10.1016/j.jmps.2008.08.008
[106] S. Reese, S. Govindjee, A theory of finite viscoelasticity and numerical aspects, Int. J. Solids Struct. 35 (26) (1998) 3455–3482. https://doi.org/https://doi.org/10.1016/S0020-7683(97)00217-5
[107] K.A. Kalina, L. Linden, J. Brummund, P. Metsch, M. Kästner, Automated constitutive modeling of isotropic hyperelasticity based on artificial neural networks, Comput. Mech. 69 (1) (2021) 213-232. https://doi.org/10.1007/s00466-021-02090-6
[108] J.L. Elman, Finding structure in time, Cogn. Sci. 14 (2) (1990) 179–211. https://doi.org/https://doi.org/10.1016/0364-0213(90)90002-E
[109] L.M. Wang, K. Linka, E. Kuhl, Automated model discovery for muscle using constitutive recurrent neural networks, J. Mech. Behav. Biomed. Mater. 145 (2023) 106021. https://doi.org/https://doi.org/10.1016/j.jmbbm.2023.106021
[110] R.D. Cook, D.S. Malkus, M.E. Plesha, Concepts and Applications of Finite Element Analysis, John Wiley & Sons, 2nd edition, 1981.
[111] J. Guan, X. Li, H. Yuan, J. Liu, Hyperelastic modeling based on generalized Landau invariants and multi-stage calibration, J. Mech. Phys. Solids 206 (2026) 106338. https://doi.org/https://doi.org/10.1016/j.jmps.2025.106338
[112] A. Anssari-Benam, A. Goriely, G. Saccomandi, Generalised invariants and pseudo-universal relationships for hyperelastic materials: a new approach to constitutive modelling, J. Mech. Phys. Solids 193 (2024) 105883. https://doi.org/https://doi.org/10.1016/j.jmps.2024.105883
[113] D. Martonová, A. Goriely, E. Kuhl, Generalized invariants meet constitutive neural networks: a novel framework for hyperelastic materials, J. Mech. Phys. Solids (2025) 106352. https://doi.org/https://doi.org/10.1016/j.jmps.2025.106352
[114] D. Wiedemann, M.A. Peter, Characterization of polyconvex isotropic functions, 2023, (????). https://arxiv.org/abs/2304.08385. 2304.08385
[115] G.-L. Geuken, P. Kurzeja, D. Wiedemann, J. Mosler, A novel neural network for isotropic polyconvex hyperelasticity satisfying the universal approximation theorem, J. Mech. Phys. Solids 203 (2025) 106209. https://doi.org/https://doi.org/10.1016/j.jmps.2025.106209
[116] F. Charton, J.S. Ellenberg, A.Z. Wagner, G. Williamson, PatternBoost: constructions in mathematics with a little help from AI, 2024, (????). https://arxiv.org/abs/2411.00566. 2411.00566
[117] E. Prume, C. Gierden, M. Ortiz, S. Reese, Direct data-driven algorithms for multiscale mechanics, Comput. Methods Appl. Mech. Eng. 433 (2025) 117525. https://doi.org/https://doi.org/10.1016/j.cma.2024.117525
[118] L. Stainier, A. Leygue, M. Ortiz, Model-free data-driven methods in mechanics: material data identification and solvers, Comput. Mech. 64 (2) (2019) 381-393. https://doi.org/10.1007/s00466-019-01731-1
[119] L. Linden, K.A. Kalina, J. Brummund, B. Riemer, M. Kästner, A dual-stage constitutive modeling framework based on finite strain data-driven identification and physics-augmented neural networks, 2025, (????). https://arxiv.org/abs/2504.15492. 2504.15492
[120] M. Flaschel, D. Martonová, C. Veil, E. Kuhl, Material fingerprinting: a shortcut to material model discovery without solving optimization problems, 2025, (????). https://arxiv.org/abs/2508.07831. 2508.07831
[121] A. Eghtesad, J. Tan, J.N. Fuhg, N. Bouklas, NN-EVP: a physics informed neural network-based elasto-viscoplastic framework for predictions of grain size-aware flow response, Int. J. Plast. 181 (2024) 104072. https://doi.org/https://doi.org/10.1016/j.ijplas.2024.104072
[122] M. Flaschel, H. Yu, N. Reiter, J. Hinrichsen, S. Budday, P. Steinmann, S. Kumar, L. De Lorenzis, Automated discovery of interpretable hyperelastic material models for human brain tissue with EUCLID, J. Mech. Phys. Solids 180 (2023) 105404. https://doi.org/https://doi.org/10.1016/j.jmps.2023.105404
[123] J.N. Fuhg, R.E. Jones, N. Bouklas, Extreme sparsification of physics-augmented neural networks for interpretable model discovery in mechanics, Comput. Methods Appl. Mech. Eng. 426 (2024) 116973. https://doi.org/https://doi.org/10.1016/j.cma.2024.116973
[124] M. Poyser, T.P. Breckon, Neural architecture search: a contemporary literature review for computer vision applications, Pattern Recognit. 147 (2024) 110052. https://doi.org/https://doi.org/10.1016/j.patcog.2023.110052
[125] A.A. Jadoon, D.T. Seidl, R.E. Jones, J.N. Fuhg, Input specific neural networks, 2025, (????). https://arxiv.org/abs/2503.00268. 2503.00268
[126] M. Itskov, On the application of the additive decomposition of generalized strain measures in large strain plasticity, Mech. Res. Commun. 31 (5) (2004) 507–517. https://doi.org/https://doi.org/10.1016/j.mechrescom.2004.02.006
[127] S. Reese, T. Brepols, M. Fassin, L. Poggenpohl, S. Wulfinghoff, Using structural tensors for inelastic material modeling in the finite strain regime - a novel approach to anisotropic damage, J. Mech. Phys. Solids 146 (2021) 104174. https://doi.org/10.1016/j.jmps.2020.104174