
NEURAL NETWORKS MEET ANISOTROPIC HYPERELASTICITY: A FRAMEWORK BASED ON GENERALIZED STRUCTURE TENSORS AND ISOTROPIC TENSOR FUNCTIONS

A PREPRINT

Karl A. Kalina

Chair of Computational and
Experimental Solid Mechanics
TU Dresden, 01062 Dresden, Germany

Jörg Brummund

Chair of Computational and
Experimental Solid Mechanics
TU Dresden, 01062 Dresden, Germany

WaiChing Sun

Department of Civil Engineering
and Engineering Mechanics
Columbia University, NY 10027, New York, United States

Markus Kästner*

Chair of Computational and
Experimental Solid Mechanics
TU Dresden, 01062 Dresden, Germany

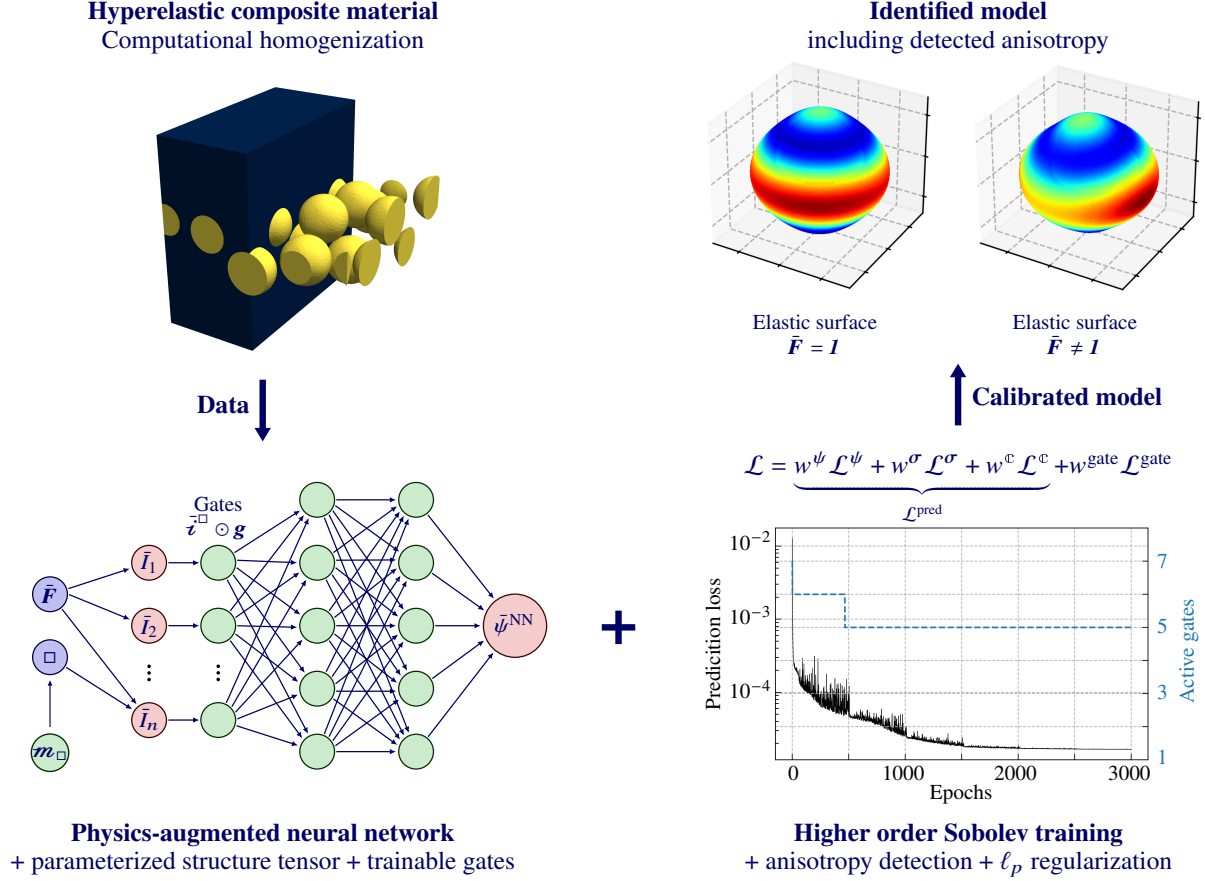
October 7, 2024

ABSTRACT

We present a data-driven framework for the multiscale modeling of anisotropic finite strain elasticity based on physics-augmented neural networks (PANNs). Our approach allows the efficient simulation of materials with complex underlying microstructures which reveal an overall anisotropic and nonlinear behavior on the macroscale. By using a set of invariants as input, an energy-type output and by adding several correction terms to the overall energy density functional, the model fulfills multiple physical principles by construction. The invariants are formed from the right Cauchy-Green deformation tensor and fully symmetric 2nd, 4th or 6th order structure tensors which enables to describe a wide range of symmetry groups. Besides the network parameters, the structure tensors are simultaneously calibrated during training so that the underlying anisotropy of the material is reproduced most accurately. In addition, sparsity of the model with respect to the number of invariants is enforced by adding a trainable gate layer and using ℓ_p regularization. Our approach works for data containing tuples of deformation, stress and material tangent, but also for data consisting only of tuples of deformation and stress, as is the case in real experiments. The developed approach is exemplarily applied to several representative examples, where necessary data for the training of the PANN surrogate model are collected via computational homogenization. We show that the proposed model achieves excellent interpolation and extrapolation behaviors. In addition, the approach is benchmarked against an NN model based on the components of the right Cauchy-Green deformation tensor.

*Corresponding author, email: markus.kaestner@tu-dresden.de.

Graphical abstract



Keywords anisotropic finite strain elasticity · computational homogenization · physics-augmented neural networks · generalized structure tensors · anisotropy detection

1 Introduction

One of the cornerstones of continuum solid mechanics are constitutive models which allow to mathematically describe the behavior of a variety of materials such as metals, elastomers or even active materials and even non-Newtonian fluids. During the last century, considerable efforts have been made to understand the mathematical and physical requirements that a constitutive model should fulfill [1, 2]. Based on this knowledge, numerous models, referred to as classical constitutive models in the following, have been formulated and parameterized using data from experiments or simulations at lower scales. However, especially for composite materials, which often show an extremely complex anisotropic and nonlinear behavior, these classical models are often not flexible and sufficiently accurate enough. For this reason, alternatives based on *machine learning* and in particular the use of *neural networks (NNs)* have recently become increasingly popular in constitutive modeling [3, 4]. Approaches of this kind indicate the potential of data-driven constitutive modeling, i.e., without having to decide on a specific model, it is possible to learn complex material behavior.

1.1 Application of neural networks in constitutive modeling

In the pioneering work of Ghaboussi et al. [5] from the early 1990s, NNs, in particular *feedforward neural networks (FNNs)*, were used for the first time to predict hysteresis in uniaxial and multiaxial stress states. To allow the FNN to learn history-dependent behavior, information from several prior time steps are used as input. After a brief period in the 1990s, the approach of using NNs for constitutive modeling has not been significantly pursued for the time being. With the increasing popularity of machine learning and the associated rapid advances in efficiency and accessibility, many

different data-driven methods² have emerged in mechanics in a very short time in recent years, see the review articles [3, 4, 15, 16, 17].

A rather important development in NN-based constitutive modeling and generally in scientific machine learning is to include fundamental physical knowledge, which is denoted as *physics-informed* [18, 19, 20, 21], *mechanics-informed* [22], *physics-augmented* [23, 24], *physics-constrained* [25], or *thermodynamics-based* [26]. This may be accomplished in two ways: either strongly, as in the case of adapted network architectures [27, 28], or weakly, as in the case of problem-specific loss terms for training, see [29, 30, 31]. As shown in [24, 26, 32, 33], these models enable the usage of sparse training data and a significant improvement of the model’s extrapolation capability. In the following, we will give a brief overview on NN-based constitutive modeling which is mainly limited on *elasticity*, i.e., perfectly path-independent behavior.

There are numerous works that model elasticity with NNs in this sense, e.g., in the initial works [34, 35] from the 2000s, the elastic potential of isotropic materials is approximated by using an FNN with three deformation-type invariants as input, which leads to the fulfillment of several requirements by construction, e.g., *thermodynamic consistency*, *objectivity*, or *material symmetry*. However, instead of calibrating the model with stress data, the strain energy density data were used directly to train these models. Meanwhile, approaches based on architectures that use the *hyperelastic potential* as output and *invariants* as input are very common, e.g. [24, 27, 28, 36, 37, 38, 39, 40, 41, 42]. Thereby, a special training method allows direct calibration of the NN using stress and strain tuples. In particular, the loss function involves the derivation of the energy with respect to the deformation, a technique known as *Sobolev training* [43, 44]. If also a loss term for the elasticity tensor is added, which requires to calculate the second derivative of the potential with respect to the deformation, it is called higher-order Sobolev training [45, 46]. Other models formulate NN-based potentials directly in terms of the components of the strain or deformation tensor [22, 44, 47], which offers more flexibility in cases of anisotropy. However, this has the disadvantage that the network design no longer enforces the material symmetry by construction. For finite symmetry groups, this problem can be overcome by group symmetrization, see [36, 48]. As shown in [49], it is also possible to use tensor feature equivariant neural networks to enforce material symmetry.

In addition, *polyconvex* NNs are used in several works [36, 39, 40, 50, 51, 52], which can be favorable in finite element (FE) simulations, improves the extrapolation capability [24, 46] and guarantees *rank-one convexity* and thus *ellipticity* [53, 54]. Various techniques are used for incorporating this condition, with the most widely spread being the application of *fully input convex neural networks (FICNNs)* introduced by Amos et al. [55]. Recently, Linden et al. [24] presented an approach based on FICNNs that fulfills all usual conditions of (an)isotropic compressible hyperelasticity by construction, i.e., thermodynamic consistency, symmetry of stress, objectivity, material symmetry, polyconvexity and thus ellipticity, volumetric growth condition, as well a stress- and energy-free undeformed state.

In the context of multiscale problems, NN-based approaches can be used as surrogate models which replace the computationally expensive simulation of *representative volume elements (RVEs)* and thus enable a significant speed up [25, 56]. Due to the high flexibility and simultaneously excellent prediction quality, NNs enriched with physical knowledge are excellently suited as surrogate models [16, 23, 25, 36, 45, 46, 57, 58, 59, 60, 61, 62]. In principle, the rule can be established that NNs with as much physics incorporated by design as possible are to be preferred, as these have a significantly better extrapolation behavior and thus less time-consuming RVE simulations have to be carried out [25].

1.2 Objectives and contributions of this work

As discussed in the literature overview given above, numerous very sophisticated approaches for constitutive modeling exist that combine modern machine learning methods with a reasonable physical basis. Restricting to elasticity, a description of the elastic potential by an NN with invariants as input is favorable if the underlying anisotropy can be described. However, most models formulated in this way are either limited to isotropy or assume that the symmetry group and the associated preferred directions are known in advance.

²It should be noted that in addition to NNs, there are several other machine learning/data-driven techniques for constitutive modeling. E.g., Gaussian process regression as applied in [6, 7, 8] also enables to model elasticity. As shown in [9], it is also useful to use splines to formulate the elastic energy density defined on a discretized invariant space. Other approaches based on sparse or symbolic regression allow an automated discovery of constitutive models [10, 11, 12, 13, 14], i.e., instead of only identifying the parameters of a predefined model, the algorithm selects a model from a large set of candidates. This offers the advantage of interpretability.

To the author's knowledge, there are only a few works that deal with the problem of *anisotropy detection* combined with NN-based constitutive modeling in the finite strain regime.³ In Linka et al. [28], a structure learning block based on a set of 2nd order *structure tensors* built as vector dyads $\mathbf{l}_i \otimes \mathbf{l}_i$ with \mathbf{l}_i being a preferred direction is presented. Besides the NN weights and biases, the preferred directions are determined during training to determine the material's anisotropy. In the same line, Thakolkaran et al. [37] detect the fiber orientation for transverse isotropy within the framework NN-EUCLID. Finally, Fuhg et al. [38] have proposed so called tensor-basis NNs which allow to discover both the type and orientation of the anisotropy. To do so, structure tensors and invariants for isotropy, transverse isotropy and orthotropy are used. The three papers mentioned above already allow the detection of anisotropy for several classes of materials without the need for prior knowledge. However, in all works a restriction to one 2nd order structure tensor or a combination of several 2nd order structure tensors has been made. For the generation of complete invariant sets, however, structure tensors of a higher order are necessary for numerous symmetry groups [53, 65, 66]. For example, a 4th order structure tensor is required for the cubic group and even a 6th order structure tensor for the hexagonal anisotropy class.

Thus, within this contribution, we present an invariant-based NN approach, where *generalized structure tensors* [67] up to 6th order are used for the generation of the invariant set. Simultaneous training of structure tensors and NN weights allows the identification of anisotropy. To do so, parameterized versions of generalized structure tensors fulfilling important properties are used. In addition, we introduce *trainable gates* in combination with a p -norm type penalty loss which allows us to remove unnecessary invariants from the model, which is also denoted as ℓ_p regularization [10, 68]. To allow for maximum accuracy in energy, stress and elasticity tensor, a higher-order Sobolev training is applied. Following the idea of *physics-augmented neural networks (PANNs)* [23, 24, 46, 69], the proposed model is formulated such that as many conditions as possible are fulfilled by construction. These are thermodynamic consistency, compatibility with the balance of angular momentum, objectivity, material symmetry, volumetric growth condition, as well as energy- and stress-free undeformed state. The performance of our approach is demonstrated for five different RVEs, where *interpolation* as well as *extrapolation* are considered. In addition, a comparison to a model based on the coordinates of the right Cauchy-Green deformation tensor is shown. The comprehensive RVE databases are generated by a computational homogenization approach.

The organization of the paper is as follows: In Sect. 2, the fundamentals of finite strain continuum mechanics, basic principles of hyperelasticity, a scale transition scheme and the theory of generalized structure tensors are summarized. After this, PANNs based on generalized structure tensors and a training which enables anisotropy detection are introduced in Sect. 4. The developed approach is exemplarily shown Sect. 5. After a discussion of the results, the paper is closed by concluding remarks and an outlook to necessary future work in Sect. 6.

Notation Within this work, tensors of rank one and two are given by boldface italic letters, i.e., $\mathbf{A}, \mathbf{B} \in \mathcal{L}_1$ or $\mathbf{C}, \mathbf{D} \in \mathcal{L}_2$, where \mathcal{L}_n denotes the space of tensors with rank $n \in \mathbb{N}$ with \mathbb{N} being the set of natural numbers without zero. Tensors with rank four and six are marked by blackboard symbols and bold upright sans serif letters, i.e., $\mathbb{A} \in \mathcal{L}_4$ and $\mathbb{A} \in \mathcal{L}_6$, respectively. Single and double contractions of two tensors are given by $\mathbf{C} \cdot \mathbf{D} = C_{kl} D_{li} \mathbf{e}_k \otimes \mathbf{e}_i$ and $\mathbf{C} : \mathbf{D} = C_{kl} D_{kl}$, respectively. Therein, $\mathbf{e}_k \in \mathcal{L}_1$ and \otimes denote a Cartesian basis vector and the dyadic product, where the Einstein summation convention is used. The cross product of two rank one tensors is given by $\mathbf{A} \times \mathbf{B} = e_{ijk} A_j B_k \mathbf{e}_i$, with e_{ijk} being the antisymmetric Levi-Civita symbol. Transpose and inverse of a second order tensor \mathbf{C} are given by \mathbf{C}^T and \mathbf{C}^{-1} , respectively. Additionally, $\text{tr } \mathbf{C}$, $\det \mathbf{C}$, $\text{cof } \mathbf{C} := \det(\mathbf{C}) \mathbf{C}^{-T}$ are used to indicate trace, determinant as well as cofactor, respectively. The Hadamard product of tensors, vectors or matrices, i.e., the element-wise product, is given by the symbol \odot . The space of unit vectors is given by $\mathcal{N} := \{\mathbf{n} \in \mathcal{L}_1 \mid \mathbf{n} \cdot \mathbf{n} = 1\}$. The sets $\mathcal{Sym} := \{\mathbf{A} \in \mathcal{L}_2 \mid \mathbf{A} = \mathbf{A}^T\}$, $\mathcal{Sym}_4 := \{\mathbb{A} \in \mathcal{L}_4 \mid A_{ijkl} = A_{jikl} = A_{ijlk} = A_{klij} = A_{kjil} = \dots\}$ and \mathcal{Sym}_6 equivalently defined, denote the spaces of fully symmetric 2nd, 4th and 6th order tensors, i.e., with full symmetry. The space of 4th order tensors which only have major and minor symmetry, e.g., the elasticity tensor, is given by the set $\overline{\mathcal{Sym}}_4 := \{\mathbb{A} \in \mathcal{L}_4 \mid A_{ijkl} = A_{jikl} = A_{ijlk} = A_{klij}\}$. Furthermore, the orthogonal group and special orthogonal group in the Euclidean vector space \mathbb{R}^3 are given by $\mathcal{O}(3) := \{\mathbf{A} \in \mathcal{L}_2 \mid \mathbf{A}^T \cdot \mathbf{A} = \mathbf{I}\}$ and $\mathcal{SO}(3) := \{\mathbf{A} \in \mathcal{L}_2 \mid \mathbf{A}^T \cdot \mathbf{A} = \mathbf{I}, \det \mathbf{A} = 1\}$, respectively, while $\mathcal{GL}^+(3) := \{\mathbf{A} \in \mathcal{L}_2 \mid \det \mathbf{A} > 0\}$ is the set of invertible second order tensors with positive determinant. Therein, $\mathbf{I} := \delta_{ij} \mathbf{e}_i \otimes \mathbf{e}_j \in \mathcal{L}_2$ is the second order identity tensor, where δ_{ij} denotes the Kronecker delta. Norms of rank one and two tensors or matrices are given by $|\mathbf{A}| := \sqrt{A_i A_i}$ and $\|\mathbf{C}\| := \sqrt{C_{ij} C_{ij}}$, respectively.

³For linear elasticity, the approaches developed by Cowin and Mehrabadi [63] or Moahker and Norris [64] can be used to identify the anisotropy class and orientation in the linear-elastic range. However, to apply these techniques, prior knowledge of the elasticity tensor is necessary. Furthermore, it is not possible to find out from the linear elastic regime alone to which anisotropy class the material belongs, since some classes coincide in the linear elastic limiting case, e.g. transverse isotropy and hexagonal anisotropy.

For reasons of readability, the arguments of functions are usually omitted within this work. However, energy functions are given with their arguments to show the dependencies, except when derivatives are written. Furthermore, in the following the symbol of a function is identical with the symbol of the function value itself.

2 Fundamentals

In this section, we introduce kinematics and stress measures common in finite strain continuum mechanics. Based on this, general principles of *anisotropic finite strain hyperelasticity* are summarized. Additionally, the concept of *generalized structure tensors* and *isotropic tensor functions* are introduced. Finally, to link the micro- and macroscale, we introduce an appropriate homogenization scheme.

2.1 Kinematics and stress measures

Let us consider the motion of a material body with reference configuration $\mathcal{B}_0 \subset \mathbb{R}^3$ at time $t_0 \in \mathbb{R}_{\geq 0}$ and current configuration $\mathcal{B} \subset \mathbb{R}^3$ at time $t \in \mathcal{T} := \{\tau \in \mathbb{R} \mid \tau \geq t_0\}$. To describe the body's motion, we introduce a smooth bijective mapping $\varphi : \mathcal{B}_0 \times \mathcal{T} \rightarrow \mathcal{B}$, mapping material points $\mathbf{X} \in \mathcal{B}_0$ to $\mathbf{x} = \varphi(\mathbf{X}, t) \in \mathcal{B}$. The displacement $\mathbf{u} \in \mathcal{L}_1$ of each material point is given by $\mathbf{u}(\mathbf{X}, t) := \varphi(\mathbf{X}, t) - \mathbf{X}$. As further kinematic quantities, the deformation gradient $\mathbf{F} := (\nabla_{\mathbf{X}} \varphi)^T \in \mathcal{GL}^+(3)$ and its determinant $J := \det \mathbf{F} \in \mathbb{R}_{>0}$ are defined. Finally, we introduce the right Cauchy-Green deformation tensor $\mathbf{C} := \mathbf{F}^T \cdot \mathbf{F} \in \mathcal{Sym}$ and the Green-Lagrange strain tensor $\mathbf{E} := \frac{1}{2}(\mathbf{C} - \mathbf{I}) \in \mathcal{Sym}$ as kinematic quantities which are invariant to rigid body motions.

Within finite strain continuum solid mechanics, various stress measures can be defined. In this work, we make use of the Cauchy stress tensor $\boldsymbol{\sigma} \in \mathcal{Sym}$, which is symmetric and is also known as true stress, as well as the 1st and 2nd Piola-Kirchhoff stress tensors $\mathbf{P} \in \mathcal{L}_2$ and $\mathbf{T} \in \mathcal{Sym}$. The latter two stress tensors are linked to the Cauchy stress by the pullback operations $\mathbf{P} := J \boldsymbol{\sigma} \cdot \mathbf{F}^{-T}$ and $\mathbf{T} := J \mathbf{F}^{-1} \cdot \boldsymbol{\sigma} \cdot \mathbf{F}^{-T}$, respectively.

For more details on basic principles in continuum solid mechanics the reader is referred to the textbooks of Haupt [1] or Holzapfel [2].

2.2 Physical conditions for anisotropic finite strain hyperelasticity

An elastic constitutive model relates deformation gradient to stress induced at a material point. In *hyperelasticity*, this mapping is not defined directly, but via an elastic potential, i.e.,

$$\psi : \mathcal{GL}^+(3) \rightarrow \mathbb{R}_{\geq 0}, \mathbf{F} \mapsto \psi(\mathbf{F}) \text{ and } \mathbf{P} = \frac{\partial \psi}{\partial \mathbf{F}}. \quad (1)$$

Relation (1)₂ implies energy conservation and path-independency. The model is thus a priori *thermodynamically consistent*, i.e., in accordance with the second law of thermodynamics [24, 27].

A mapping between the rates of 1st Piola-Kirchhoff stress $\dot{\mathbf{P}}$ and deformation gradient $\dot{\mathbf{F}}$ follows by the material time derivative $(\dot{\cdot})$ of Eq. (1)₂ and by introducing the material tangent $\mathbb{A} \in \mathcal{L}_4$, i.e.,

$$\dot{\mathbf{P}} = \mathbb{A} : \dot{\mathbf{F}} \text{ with } \mathbb{A} := \frac{\partial^2 \psi}{\partial \mathbf{F} \partial \mathbf{F}} \in \mathcal{L}_4. \quad (2)$$

For the other stress measures introduced in Sect. 2.1, the equivalent relationships $\dot{\boldsymbol{\sigma}} = \mathbb{c} : \mathbf{d}$ and $\dot{\mathbf{T}} = \mathbb{C} : \dot{\mathbf{E}}$ can be found. Thereby, $\mathbf{d} := \text{sym}(\mathbf{l}) \in \mathcal{Sym}$ is the rate of deformation tensor, which is the symmetric part of the velocity gradient $\mathbf{l} = \dot{\mathbf{F}} \cdot \mathbf{F}^{-1}$, and $\dot{\boldsymbol{\sigma}} = \dot{\boldsymbol{\sigma}} - \mathbf{l} \cdot \boldsymbol{\sigma} - \boldsymbol{\sigma} \cdot \mathbf{l}^T + \boldsymbol{\sigma} \text{tr} \mathbf{l}$ is the Truesdell rate of $\boldsymbol{\sigma}$. Relations of the introduced material tangents $\mathbb{c} \in \mathcal{Sym}_4$ and $\mathbb{C} \in \mathcal{Sym}_4$ to \mathbb{A} can be determined through some tensor calculations and follow to $c_{ijkl} = J^{-1} A_{iJKL} F_{jJ} F_{iL} - \delta_{ik} \sigma_{jl}$ as well as $C_{IJKL} = (A_{iJKL} - \delta_{ik} T_{jL}) F_{Ii}^{-1} F_{Kk}^{-1}$, respectively. Note that both \mathbb{C} and \mathbb{c} have a minor and a major symmetry, but are not completely symmetrical.

There are additional mathematical and physical requirements for the hyperelastic potential [1, 2]. The most common ones are briefly summarized below. For more details, please refer to the textbooks given above or the recent work Linden et al. [24]

In order to guarantee *symmetric Cauchy* and *2nd Piola-Kirchhoff stress tensors*, which follow from the balance of angular momentum, $\psi(\mathbf{F})$ has to be formulated such that the following condition holds:

$$\frac{\partial \psi}{\partial \mathbf{F}} \cdot \mathbf{F}^T = \mathbf{F} \cdot \frac{\partial \psi}{\partial \mathbf{F}^T}. \quad (3)$$

The principle of *material objectivity* states that the behavior of the material must not change with any rigid body motion. Furthermore, the potential should also account for the material's anisotropy, which is termed *material symmetry*. These two principles are given as

$$\psi(\mathbf{F}) = \psi(\mathbf{Q} \cdot \mathbf{F}) \quad \forall \mathbf{F} \in \mathcal{GL}^+(3), \mathbf{Q} \in \mathcal{SO}(3) \text{ and} \quad (4)$$

$$\psi(\mathbf{F}) = \psi(\mathbf{F} \cdot \mathbf{Q}^T) \quad \forall \mathbf{F} \in \mathcal{GL}^+(3), \mathbf{Q} \in \mathcal{G} \subseteq \mathcal{O}(3), \quad (5)$$

respectively. In Eq. (5), \mathcal{G} denotes the symmetry group of the material under consideration. There are 11 crystal symmetries following from 32 point groups with finite order. As the numbering of the groups is not uniform in the literature, we use the Schoenflies notation in addition to the group number. The symmetry groups are named *triclinic* $\mathcal{G}_1 (C_i)$, *monoclinic* $\mathcal{G}_2 (C_{2h})$, *rhombic* or *orthotropic* $\mathcal{G}_3 (D_{2h})$, *tetragonal* $\mathcal{G}_4 (C_{4h})$ and $\mathcal{G}_5 (D_{4h})$, *cubic* $\mathcal{G}_6 (T_h)$ and $\mathcal{G}_7 (O_h)$, *trigonal* $\mathcal{G}_8 (S_6)$ and $\mathcal{G}_9 (D_{3d})$, and *hexagonal* $\mathcal{G}_{10} (C_{6h})$ and $\mathcal{G}_{11} (D_{6h})$, see [53, 66] for more details. Three additional symmetry groups follow by also considering the continuous cylindrical and spherical groups, which are not finite. These groups are denoted as *transversely isotropic* $\mathcal{G}_{12} (C_{\infty h})$ and $\mathcal{G}_{13} (D_{\infty h})$ and *isotropic* $\mathcal{G}_{14} (\mathcal{K}_h = O(3))$.

Various coercivity conditions can also be taken into account for hyperelasticity, whereby the *volumetric growth condition* is the most widely used [2, 24]. It is given by $\psi(\mathbf{F}) \rightarrow \infty$ as $(J \rightarrow 0^+ \vee J \rightarrow \infty)$ and states that a material cannot be compressed to a volume of zero or extended to an infinite volume, requiring an increase in energy toward infinity.

Furthermore, the *undeformed* configuration of the material should be *energy- and stress-free*, i.e., $\psi(\mathbf{F} = \mathbf{I}) = 0$ and $\mathbf{P}(\mathbf{F} = \mathbf{I}) = \mathbf{0}$ should hold. It is also expected that the stored energy always increases, i.e., it is *non-negative*, if a deformation \mathbf{F} is applied, thus $\psi(\mathbf{F}) \geq 0 \quad \forall \mathbf{F} \in \mathcal{GL}^+(3)$.

Finally, we will also briefly discuss the concept of *ellipticity* which ensures traveling waves with real-valued and non-negative wave speeds [53, 54, 70]. In its local form, i.e., for a specific state $\mathbf{F} \in \mathcal{GL}^+(3)$, (strict) ellipticity requires

$$\forall \mathbf{a}, \mathbf{N} \in \mathcal{N} : (\mathbf{a} \otimes \mathbf{N}) : \mathbb{A}(\mathbf{F}) : (\mathbf{a} \otimes \mathbf{N})(>) \geq 0. \quad (6)$$

If the above condition (6) applies to all permissible deformation states, we speak of (strict) global ellipticity, i.e., $\forall \mathbf{F} \in \mathcal{GL}^+(3) : \forall \mathbf{a}, \mathbf{N} \in \mathcal{N} : (\mathbf{a} \otimes \mathbf{N}) : \mathbb{A}(\mathbf{F}) : (\mathbf{a} \otimes \mathbf{N})(>) \geq 0$. Note that global ellipticity of twice differentiable and smooth energy functions as typically considered in hyperelasticity is equivalent to *rank-one convexity*. Instead of enforcing global ellipticity directly, the mathematical concept of *polyconvexity*, which was introduced by Ball [71, 72], is often used in constitutive modeling[24, 36, 53, 54, 73].⁴ However, it should be noted that polyconvexity can be too restrictive, especially for multiscale modeling [46, 74, 75]

2.3 Concept of structure tensors and isotropic tensor functions

In order to describe anisotropic constitutive behavior, the concept of *structure tensors* can be used [1, 2, 53, 66]. Depending on the considered symmetry group, i.e., $\mathcal{G}_1 - \mathcal{G}_{14}$, these tensors are of orders up to six, where also odd tensor ranks or antisymmetric tensors are possible, see Xiao [65]. Within this work, we restrict ourselves to fully symmetric tensors of order two $\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{n_2} \in \mathit{Sym}$, four $\mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_{n_4} \in \mathit{Sym}_4$ and six $\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{n_6} \in \mathit{Sym}_6$. The structure tensors reflect the material's anisotropy and are thus invariant with respect to the symmetry transformations, i.e.,

$$\mathbf{G}_\alpha = \mathbf{Q} \cdot \mathbf{G}_\alpha \cdot \mathbf{Q}^T, \quad \mathbb{G}_\beta = \mathbf{Q} * \mathbb{G}_\beta, \quad \mathbf{G}_\gamma = \mathbf{Q} \star \mathbf{G}_\gamma \quad \forall \mathbf{Q} \in \mathcal{G}, \quad (7)$$

where $(\mathbf{Q} * \mathbb{G}_\beta)_{IJKL} = Q_{IM}Q_{JN}Q_{KP}Q_{LQ}G_{MNPQ}^\beta$ and $(\mathbf{Q} \star \mathbf{G}_\gamma)_{IJKLMN} = Q_{IO}Q_{JP}Q_{KQ}Q_{LR}Q_{MS}Q_{NT}G_{OPQRST}^\gamma$. If the structure tensors are appended to the list of arguments of $\psi(\mathbf{F})$, the energy is an *isotropic tensor function* [76] even if the material is anisotropic which means that

$$\psi(\mathbf{F}, \mathbf{S}_2, \mathbf{S}_4, \mathbf{S}_6) = \psi(\mathbf{F} \cdot \mathbf{Q}^T, \mathbf{Q} \cdot \mathbf{S}_2 \cdot \mathbf{Q}^T, \mathbf{Q} * \mathbf{S}_4, \mathbf{Q} \star \mathbf{S}_6) \quad \forall \mathbf{Q} \in \mathcal{O}(3), \quad (8)$$

whereby the sets $\mathbf{S}_2 := \{\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{n_2}\}$, $\mathbf{S}_4 := \{\mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_{n_4}\}$ and $\mathbf{S}_6 := \{\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{n_6}\}$ have been used to abbreviate the notation. In the following, for brevity, we summarize all structure tensors in the set \mathcal{S} .

The introduced concept of structure tensors can be used to fulfill numerous of the principles given in Sect. 2.2 by construction. To this end, the potential is formulated in terms of *invariants* from the right Cauchy-Green deformation

⁴A potential $\psi(\mathbf{F})$ is polyconvex if and only if there is a function such that $\psi(\mathbf{F}) = \mathcal{P}(\mathbf{F}, \text{cof } \mathbf{F}, \det \mathbf{F})$, with $\mathcal{P}(\mathbf{F}, \text{cof } \mathbf{F}, \det \mathbf{F})$ convex with respect to its arguments. Polyconvexity is sufficient for quasi-convexity, which ensures rank-one convexity and thus global ellipticity [53, 54]. Furthermore, sequential weak lower semi-continuity (s.w.l.s.) is guaranteed by polyconvexity. It is sufficient for the existence of minimizers if coercivity is also guaranteed [53, 54]. The above implications do not apply in reverse.

tensor \mathbf{C} and a set of structure tensors \mathcal{S} , i.e., $\psi(\mathbf{I})$ with $\mathbf{I} := (I_1, I_2, \dots, I_n) \in \mathbb{R}^n$. Thus, the constitutive relations according to Eqs. (1)₂ and (2)₂ follow then to

$$\mathbf{P} = \sum_{\alpha=1}^n \frac{\partial \psi}{\partial I_\alpha} \frac{\partial I_\alpha}{\partial \mathbf{F}} \quad \text{and} \quad \mathbb{A} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \frac{\partial^2 \psi}{\partial I_\alpha \partial I_\beta} \frac{\partial I_\alpha}{\partial \mathbf{F}} \otimes \frac{\partial I_\beta}{\partial \mathbf{F}} + \sum_{\alpha=1}^n \frac{\partial \psi}{\partial I_\alpha} \frac{\partial^2 I_\alpha}{\partial \mathbf{F} \partial \mathbf{F}}, \quad (9)$$

respectively. The partial derivatives of the invariants with respect to \mathbf{F} are also referred to as tensor generators [27]. With that, *thermodynamic consistency*, *symmetry of the Cauchy stress*, *material objectivity* and *material symmetry* are automatically fulfilled [24].

Remark 1. Note that it is not possible to build invariant sets for all introduced symmetry groups $\mathcal{G}_1 - \mathcal{G}_{14}$ with two fully symmetric 2nd order structure tensors or one fully symmetric 4th or 6th order structure tensor each, as done here, see Xiao [65]. However, a wide range of groups can be described with these structure tensors, cf. A.

2.4 Scale transition scheme

In the following, we distinguish between two different scales, the *microscale* and the *macroscale* with characteristic lengths $\ell \in \mathbb{R}_{\geq 0}$ and $\bar{\ell} \in \mathbb{R}_{\geq 0}$, respectively. The microscale is represented by a heterogeneous structure consisting of matrix and inhomogeneities, while the macroscale is considered to be homogeneous. For the characteristic lengths of both scales, the relationship $\ell \ll \bar{\ell}$ known as scale separation should hold [77]. To label macroscopic quantities, they are marked by an overline in the following, i.e., $(\bar{\cdot})$.

In order to connect microscopic and macroscopic quantities, a computational homogenization scheme is applied. Consequently, each macroscopic point $\bar{\mathbf{X}} \in \bar{\mathcal{B}}_0$ gets assigned properties resulting from the behavior of the microscale which is represented by an RVE with domain $\mathcal{B}_0^{\text{RVE}}$ in the vicinity of $\bar{\mathbf{X}}$ on the microscale. An effective macroscopic quantity can be determined by the volume average

$$\langle\langle \bar{\cdot} \rangle\rangle := \frac{1}{V^{\text{RVE}}} \int_{\mathcal{B}_0^{\text{RVE}}} (\bar{\cdot}) \, dV, \quad (10)$$

where $V^{\text{RVE}} \in \mathbb{R}_{\geq 0}$ is the RVE's volume. With that, macroscopic deformation gradient and 1st Piola-Kirchhoff stress tensor are defined by $\bar{\mathbf{F}} := \langle \mathbf{F} \rangle$ and $\bar{\mathbf{P}} := \langle \mathbf{P} \rangle$, respectively [25, 77]. Boundary conditions (BCs) for the microscopic BVP, which must be solved before volume averaging, can be derived from the Hill-Mandel condition. For the finite strain setting under consideration, it is given by the well known equation $\langle \mathbf{P} : \dot{\mathbf{F}} \rangle = \bar{\mathbf{P}} : \dot{\bar{\mathbf{F}}}$ [77]. Herein, we use periodic BCs to guarantee the fulfillment of this relation [25]. Within FE simulations, we realize the periodic BCs via the concept of master nodes, see Haasemann et al. [78]. This also allows the *algorithmically consistent tangent modulus of the RVE* to be calculated in a straightforward manner, which corresponds to the macroscopic material tangent, i.e., $\bar{\mathbb{A}} = \bar{\mathbb{A}}^{\text{algo}}$.

For the hyperelastic case considered here, the Hill-Mandel condition represents the equality of the rates of the averaged microscopic and macroscopic energies, allowing to calculate the effective potential by volume averaging, i.e. $\bar{\psi} = \langle \psi \rangle$. Thus, the application of the computational homogenization approach allows us to determine the mapping

$$\mathcal{H} : \mathcal{GL}^+(3) \rightarrow \mathbb{R} \times \mathcal{L}_2 \times \mathcal{L}_4, \quad \bar{\mathbf{F}} \mapsto (\bar{\psi}, \bar{\mathbf{P}}, \bar{\mathbb{A}}). \quad (11)$$

The Cauchy stress and the corresponding elasticity tensor can be calculated by applying push forward operations. This allows us to get $\mathcal{PF} : \mathcal{GL}^+(3) \times \mathcal{L}_2 \times \mathcal{L}_4 \rightarrow \text{Sym} \times \mathcal{L}_4$, $(\bar{\mathbf{F}}, \bar{\mathbf{P}}, \bar{\mathbb{A}}) \mapsto (\bar{\boldsymbol{\sigma}}, \bar{\mathbb{c}})$. All FE simulations in this work were carried out using an in-house code based on Matlab. To close the brief introduction on computational homogenization we want note that all of the definitions given in Sects. 2.1 and 2.2 are valid on the microscopic as well as the macroscopic scales.

3 Macroscale modeling with generalized structure tensors

As already mentioned in Sect. 2.3, it is favorable to formulate the elastic potential in terms of invariants from $\bar{\mathbf{C}}$ and a set of structure tensors \mathcal{S} . In the following, we introduce the concept of generalized structure tensors [67, 79] and give related sets of invariants.

For the *isotropic* case, i.e., the group \mathcal{G}_{14} , $\mathcal{S} = \{\mathbf{I}\}$, so that a complete and irreducible [80] set of invariants is given by the well known set

$$\bar{I}_1 := \text{tr} \bar{\mathbf{C}}, \quad \bar{I}_2 := \text{tr}(\text{cof} \bar{\mathbf{C}}) \quad \text{and} \quad \bar{I}_3 := \det \bar{\mathbf{C}}. \quad (12)$$

In order to model complex *anisotropic* behavior on the macroscale, generalized structure tensors can be used. Originally, this concept has been introduced to model materials which are characterized by fibers oriented in different directions

on the microscale, e.g., biological tissues or fiber reinforced composites [67]. As already mentioned, we restrict ourselves to fully symmetric structure tensors of 2nd, 4th and 6th order. These tensors and the related invariant basis are introduced in the following. Since these tensors correspond to the macroscale as introduced in Sect. 2.4, we mark them as macroscopic quantities in the following.

3.1 2nd order generalized structure tensor

A second order generalized structure tensor is defined by

$$\bar{\mathbf{G}} := \int_{\Omega} \rho(\theta, \varphi) \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \, dA \in \mathcal{S}ym \text{ with } \int_{\Omega} \rho(\theta, \varphi) \, d\Omega = 1, \quad (13)$$

where $\mathbf{N}(\theta, \varphi) \in \mathcal{N}$ is a fiber orientation vector and $\rho(\theta, \varphi) \in \mathbb{R}_{\geq 0}$ is a probability density function describing the orientation of the fibers on the microscale [67].⁵ The tensor $\bar{\mathbf{G}}$ is symmetric, positive semi-definite and has a trace of one, i.e., $\bar{\mathbf{G}}^T = \bar{\mathbf{G}}$, $\mathbf{A} \cdot \bar{\mathbf{G}} \cdot \mathbf{A} \geq 0 \, \forall \mathbf{A} \in \mathcal{L}_1$ and $\text{tr} \bar{\mathbf{G}} = 1$. Following Boehler [81], with $\mathcal{S} := \{\bar{\mathbf{G}}\}$, four additional mixed invariants follow:

$$\bar{K}_4 := \bar{\mathbf{C}} : \bar{\mathbf{G}}, \quad \bar{K}_5 := \bar{\mathbf{C}}^2 : \bar{\mathbf{G}}, \quad \bar{K}_6 := \bar{\mathbf{C}} : \bar{\mathbf{G}}^2, \quad \bar{K}_7 := \bar{\mathbf{C}}^2 : \bar{\mathbf{G}}^2. \quad (14)$$

The set of invariants is thus given by $\bar{\mathcal{I}}^{\bar{\mathbf{G}}} := (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{K}_4, \bar{K}_5, \bar{K}_6, \bar{K}_7) \in \mathbb{R}^7$. The structure tensor $\bar{\mathbf{G}}$ can also be given by its eigenvalues $\lambda_{\alpha} \in \mathbb{R}_{\geq 0}$ and the corresponding projection tensors $\mathbf{P}_{\alpha} \in \mathcal{S}ym$:

$$\bar{\mathbf{G}} = \sum_{\alpha=1}^{n_G} \lambda_{\alpha} \mathbf{P}_{\alpha} \text{ with } \mathbf{P}_{\alpha} \cdot \mathbf{P}_{\beta} = \delta_{\alpha\beta} \mathbf{P}_{(\beta)}, \quad \sum_{\alpha=1}^{n_G} \mathbf{P}_{\alpha} = \mathbf{I}, \quad (15)$$

where $n_G \in \{1, 2, 3\}$ is the number of non-equal eigenvalues. For $n_G = 1$ the tensor $\bar{\mathbf{G}}$ describes the *isotropic* group \mathcal{G}_{14} , whereas for $n_G = 2$ and $n_G = 3$ the material is *transversely isotropic* (\mathcal{G}_{13}) and *orthotropic* (\mathcal{G}_3), respectively [82]. Note that \bar{K}_6 and \bar{K}_7 would be omitted if $\bar{\mathbf{G}}$ has only two different eigenvalues, i.e., for transverse isotropy, see A.8.

Remark 2. Note that orthotropy is typically modeled with two 2nd order structure tensors given by $\bar{\mathbf{G}}_1 := \mathbf{a}_1 \otimes \mathbf{a}_1 \in \mathcal{S}ym$ and $\bar{\mathbf{G}}_2 := \mathbf{a}_2 \otimes \mathbf{a}_2 \in \mathcal{S}ym$ with $\mathbf{a}_{\alpha} \cdot \mathbf{a}_{\beta} = \delta_{\alpha\beta}$ [2]. However, as shown in A.4, the invariant set following from $\bar{\mathbf{C}}$, $\bar{\mathbf{G}}_1$ and $\bar{\mathbf{G}}_2$ can be expressed by $\bar{\mathcal{I}}^{\bar{\mathbf{G}}}$ if $\bar{\mathbf{G}}$ has 3 non-equal eigenvalues and two eigenvectors which are (anti)parallel with \mathbf{a}_1 and \mathbf{a}_2 , respectively.

3.2 4th order generalized structure tensor

In order to model more complex anisotropic behavior, we introduce an extension of Eq. (13) to a 4th order generalized structure tensor given by

$$\bar{\mathbb{G}} := \int_{\Omega} \rho(\theta, \varphi) \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \, dA \in \mathcal{S}ym_4, \quad (16)$$

see also [83, 84], where a similar quantity is introduced as fiber orientation tensor.⁶ The 4th order structure tensor given in Eq. (16) is fully symmetric, positive semi-definite with respect to vector dyads as well as symmetric 2nd order tensors and has a generalized trace of one:

$$\bar{G}_{KLMN} = \bar{G}_{LKMN} = \bar{G}_{KLN M} = \dots, \quad \mathbf{S} : \bar{\mathbb{G}} : \mathbf{S} \geq 0 \, \forall \mathbf{S} \in \mathcal{S}ym \text{ and } \bar{G}_{KKLL} = \bar{G}_{KLKL} = \bar{G}_{KLLK} = 1. \quad (17)$$

In order to identify a corresponding set of invariants for the case $\mathcal{S} := \{\bar{\mathbb{G}}\}$, we follow the procedure of Xiao [65]. Thus, we build a set of 2nd order tensors from $\bar{\mathbb{G}}$ and $\bar{\mathbf{C}}$ by polynomials. Due to the fact that $\bar{\mathbb{G}}$ is constant and due to the Cayley-Hamilton theorem applied for $\bar{\mathbf{C}}$, we get

$$\bar{\mathbf{H}}_1 := \bar{\mathbb{G}} : \bar{\mathbf{C}} \text{ and } \bar{\mathbf{H}}_2 := \bar{\mathbb{G}} : \bar{\mathbf{C}}^2. \quad (18)$$

According to Boehler [81] we can build a total of 21 invariants from $\bar{\mathbf{C}}$, $\bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$. This set is complete but not irreducible in general, i.e., for specific cases some invariants can be expressed by others or are even redundant. However,

⁵Note that the introduced parameterized forms of the generalized structure tensors as introduced in Sect. 4.1.1 also allow an application to materials with microstructures that are not characterized by a distribution of fibers. This will be demonstrated in the examples part 5 of this paper.

⁶Note that Fiber orientation tensors are defined in exactly the same way as generalized structure tensors. Fiber orientation tensors of order $n \in \mathbb{N}$ can be used for the characterization of fiber reinforced composites, generation of RVEs [85] or for mean-field homogenization [86].

in order to reduce the number of invariants for the generalized structure tensor of order four, only invariants up to an order less than or equal to 3 in $\bar{\mathbf{C}}$ are taken into account, i.e., $\bar{\mathbf{C}}^n$ with $n \leq 3$. This leads to a set comprising the invariants given in Eq. (12) and the additional invariants

$$\begin{aligned} \bar{L}_4 &:= \text{tr} \bar{\mathbf{H}}_1, \bar{L}_5 := \text{tr} \bar{\mathbf{H}}_1^2, \bar{L}_6 := \text{tr} \bar{\mathbf{H}}_1^3, \bar{L}_7 := \text{tr} \bar{\mathbf{H}}_2, \bar{L}_8 := \text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{C}}), \\ \bar{L}_9 &:= \text{tr}(\bar{\mathbf{H}}_2 \cdot \bar{\mathbf{C}}), \bar{L}_{10} := \text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{H}}_2), \bar{L}_{11} := \text{tr}(\bar{\mathbf{H}}_1^2 \cdot \bar{\mathbf{C}}). \end{aligned} \quad (19)$$

The invariant set is thus given by $\bar{\mathcal{I}}^{\bar{\mathbf{G}}} := (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{L}_4, \bar{L}_5, \dots, \bar{L}_{11}) \in \mathbb{R}^{11}$.

The formulation based on the generalized structure tensor $\bar{\mathbf{G}}$ includes the *tetragonal* group \mathcal{G}_5 and the *cubic* group \mathcal{G}_7 up to invariants of order 3 in $\bar{\mathbf{C}}$, cf. A.5 and A.6. Note that it is also possible to describe further anisotropy classes with $\bar{\mathbf{G}}$. However, this is not discussed here. The reader is referred to [87] for a discussion on fully symmetric and traceless 4th order structure tensors.

3.3 6th order generalized structure tensor

Furthermore, to enable the description of anisotropy classes that require structure tensors of orders higher than 4, we introduce a generalized 6th order structure tensor defined as

$$\bar{\mathbf{G}} := \int_{\Omega} \rho(\theta, \varphi) \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \otimes \mathbf{N}(\theta, \varphi) \, dA \in \mathcal{S}ym_6. \quad (20)$$

Again, similar to $\bar{\mathbf{G}} \in \mathcal{S}ym$ and $\bar{\mathbf{G}} \in \mathcal{S}ym_4$, this tensor is fully symmetric, positive semi-definite with respect to $\mathbf{S} \otimes \mathbf{A} \in \mathcal{L}_3$ and has a generalized trace of one:

$$\bar{G}_{IJKLMN} = \bar{G}_{JIKLMN} = \bar{G}_{IJLKMN} = \bar{G}_{IKJLMN} = \dots, \quad (21)$$

$$S_{IJA}K\bar{G}_{IJKLMN}S_{LM}A_N \geq 0 \quad \forall \mathbf{S} \in \mathcal{S}ym, \mathbf{A} \in \mathcal{L}_1 \text{ and} \quad (22)$$

$$\bar{G}_{KKLLMM} = 1. \quad (23)$$

To build an invariant basis, we adapt again the technique from Xiao [65], i.e., we build $\bar{\mathbf{H}}_1 := \mathbf{1} : \bar{\mathbf{G}} : \bar{\mathbf{C}}$, $\bar{\mathbf{H}}_2 := \bar{\mathbf{C}} : \bar{\mathbf{G}} : \bar{\mathbf{C}}$, $\bar{\mathbf{H}}_3 := \mathbf{1} : \bar{\mathbf{G}} : \bar{\mathbf{C}}^2$, $\bar{\mathbf{H}}_4 := \bar{\mathbf{C}} : \bar{\mathbf{G}} : \bar{\mathbf{C}}^2$, and $\bar{\mathbf{H}}_5 := \bar{\mathbf{C}}^2 : \bar{\mathbf{G}} : \bar{\mathbf{C}}^2$ and use the rules given in Boehler [81]. With that, one gets a total of 98 invariants, which can be reduced to 13 if only invariants with orders up to $\bar{\mathbf{C}}^3$ are used and redundant invariants are neglected. This leads to the set $\bar{\mathcal{I}}^{\bar{\mathbf{G}}} := (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{M}_4, \bar{M}_5, \dots, \bar{M}_{13}) \in \mathbb{R}^{13}$ comprising the invariants given in Eq. (12) and the additional invariants

$$\begin{aligned} \bar{M}_4 &:= \text{tr} \bar{\mathbf{H}}_1, \bar{M}_5 := \text{tr} \bar{\mathbf{H}}_1^2, \bar{M}_6 := \text{tr} \bar{\mathbf{H}}_1^3, \bar{M}_7 := \text{tr} \bar{\mathbf{H}}_2, \bar{M}_8 := \text{tr} \bar{\mathbf{H}}_3, \bar{M}_9 := \text{tr} \bar{\mathbf{H}}_4, \\ \bar{M}_{10} &:= \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_1^2), \bar{M}_{11} := \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_2), \bar{M}_{12} := \text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{H}}_2), \bar{M}_{13} := \text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{H}}_3). \end{aligned} \quad (24)$$

The formulation based on the generalized structure tensor $\bar{\mathbf{G}}$ includes the *hexagonal* group \mathcal{G}_{11} up to invariants of order 3 in $\bar{\mathbf{C}}$, see A.7.

3.4 Two 2nd order generalized structure tensors

Finally, several generalized structure tensors can also be used to extend the symmetry groups that can be described using the structure tensors introduced so far. By using a set given by two 2nd order structure tensors, i.e., $\mathcal{S} = \{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2\}$, we get 12 invariants according to Boehler [81]: 3 of $\bar{\mathbf{C}}$ as defined in Eq. (12), 4 each according to Eq. (14) for $\bar{\mathbf{C}}, \bar{\mathbf{G}}_1$ and $\bar{\mathbf{C}}, \bar{\mathbf{G}}_2$ and a further invariant defined by $\bar{N}_{12} := \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{G}}_1 \cdot \bar{\mathbf{G}}_2)$. Thus, we end up with the set $\bar{\mathcal{I}}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} := (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{N}_4, \dots, \bar{N}_{12}) \in \mathbb{R}^{12}$.

This formulation includes five different anisotropy classes. For the following discussion, we are taking up the argumentation from Olive et al. [87] in a slightly modified form. To this end, we use the spectral decomposition of the structure tensors according to Eq. (15):

$$\bar{\mathbf{G}}_1 = \sum_{\alpha=1}^{n_{G_1}} \lambda_{\alpha} \mathbf{P}_{\alpha} \text{ and } \bar{\mathbf{G}}_2 = \sum_{\alpha=1}^{n_{G_2}} \mu_{\alpha} \mathbf{M}_{\alpha}. \quad (25)$$

Note that we only consider positive semi-definite structure tensors. For the case that both structure tensors have only equal eigenvalues each, i.e., $n_{G_1} = n_{G_2} = 1$, the tensors describe the *isotropic* group \mathcal{G}_{14} . If $\bar{\mathbf{G}}_1$ has $n_{G_1} = 2$

non-equal eigenvalues and it holds $\bar{\mathbf{G}}_2 = c_1 \mathbf{I} + c_2 \bar{\mathbf{G}}_1$ with $c_1, c_2 \geq 0 \wedge c_1 + c_2 > 0$, the material is *transversely isotropic* (\mathcal{G}_{13}). The same holds if $\bar{\mathbf{G}}_1$ and $\bar{\mathbf{G}}_2$ change the roles. For the case that $\bar{\mathbf{G}}_1 \cdot \bar{\mathbf{G}}_2 - \bar{\mathbf{G}}_2 \cdot \bar{\mathbf{G}}_1 = \mathbf{0}$ holds and $\exists \bar{\mathbf{G}}_\alpha$ with $n_{\bar{\mathbf{G}}_\alpha} = 3$ for $\alpha \in \{1, 2\}$ or both structure tensors have $n_{\bar{\mathbf{G}}_1} = n_{\bar{\mathbf{G}}_2} = 2$ non-equal eigenvalues and it holds $\bar{\mathbf{G}}_1 \neq c_1 \mathbf{I} + c_2 \bar{\mathbf{G}}_2 \forall c_1, c_2 \in \mathbb{R}$, the material is *orthotropic* (\mathcal{G}_3). Both tensors describe the *monoclinic group* \mathcal{G}_2 if it holds $w_L = e_{LMN} G_{MP}^1 G_{PN}^2 \neq 0$ and $(\bar{\mathbf{G}}_\alpha \cdot \mathbf{w}) \times \mathbf{w} = \mathbf{0} \forall \alpha \in \{1, 2\}$. Finally, $\bar{\mathbf{G}}_1$ and $\bar{\mathbf{G}}_2$ describe the *triclinic group* \mathcal{G}_1 if and only if $w_L = e_{LMN} G_{MP}^1 G_{PN}^2 \neq 0$ and $\exists \bar{\mathbf{G}}_\alpha$ with $(\bar{\mathbf{G}}_\alpha \cdot \mathbf{w}) \times \mathbf{w} \neq \mathbf{0}, \alpha \in \{1, 2\}$.

4 Physics-augmented neural networks with anisotropy detection

To enable an efficient and accurate description of the complex behavior of composite materials at the macroscale, we use NNs as a replacement for the time-consuming RVE simulation in this work. In doing so, we combine the advantages of NNs with a rigorous physical formulation, which we refer to as *physics-augmented neural networks (PANNs)* [24]. In this section, we formulate a PANN that builds on the concept of generalized structure tensors as introduced in Sect. 3 and allows the description of a variety of symmetry groups. The training and model identification procedure tailored to the problem makes it possible to recognize the underlying anisotropy only from homogenized data. Note that an application of our approach to data from real experiments is also possible.

4.1 Model formulation

4.1.1 Parameterized generalized structure tensors

To find a set of generalized structure tensors accounting for the underlying anisotropy of the material, it is necessary to determine the probability density function $\rho(\theta, \varphi)$, or, in the case of two structure tensors, the probability density functions. Usually this is done by identifying the orientation distribution from 2D microscopy images or computer tomography (CT) scans of the microstructure, e.g., [88].

However, in the following we leave this idea behind us and do not directly use the integral formulations as given in the previous section. Instead, we consider *parameterized versions of the structure tensors*. This procedure has two main advantages: (i) It allows for the efficient determination of $\bar{\mathbf{G}} \in \text{Sym}$, $\bar{\mathbf{G}} \in \text{Sym}_4$, $\bar{\mathbf{G}} \in \text{Sym}_6$ or $\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2 \in \text{Sym}$ directly from available data during training, i.e., from homogenized quantities $(\bar{\mathbf{F}}, \bar{\psi}, \bar{\mathbf{P}}, \bar{\mathbf{A}})$ or also from real experimental data, and (ii) it enables to apply the modeling framework to arbitrary materials, e.g., crystals, composites or foams. The parameterized generalized structure tensors are constructed in such a way that the properties discussed in Sect. 3, i.e., symmetry, positive semi-definiteness and (generalized) trace of one, are guaranteed while maintaining the high flexibility of the original approach.

2nd order structure tensor(s) For the 2nd order structure tensor, the ansatz

$$\bar{\mathbf{G}} := \frac{1}{g_1 + g_2 + g_3} \mathbf{Q} \cdot \text{diag}(g_1, g_2, g_3) \cdot \mathbf{Q}^T \in \text{Sym}, \mathbf{Q} \in \mathcal{SO}(3), \quad (26)$$

$$[\mathbf{Q}] := \begin{bmatrix} \cos \varphi_1 \cos \varphi_2 & \cos \varphi_1 \sin \varphi_2 \sin \varphi_3 - \cos \varphi_3 \sin \varphi_1 & \sin \varphi_1 \sin \varphi_3 + \cos \varphi_1 \cos \varphi_3 \sin \varphi_2 \\ \cos \varphi_2 \sin \varphi_1 & \cos \varphi_1 \cos \varphi_3 + \sin \varphi_1 \sin \varphi_2 \sin \varphi_3 & \cos \varphi_3 \sin \varphi_1 \sin \varphi_2 - \cos \varphi_1 \sin \varphi_3 \\ -\sin \varphi_2 & \cos \varphi_2 \sin \varphi_3 & \cos \varphi_2 \cos \varphi_3 \end{bmatrix} \quad (27)$$

with $\mathbf{m}_{\bar{\mathbf{G}}} \in \mathcal{M}_{\bar{\mathbf{G}}} := \{\mathbf{m}_{\bar{\mathbf{G}}} \in \mathbb{R}^6 \mid g_1, g_2, g_3 \in [0, 1], g_i g_i \neq 0, \varphi_1 \in [0, \pi], \varphi_2 \in [-\pi/2, \pi/2], \varphi_3 \in [-\pi, \pi]\}$ is used. Note that it would be possible to reduce from six to five parameters due to the normalization condition [82], i.e., $\text{tr} \bar{\mathbf{G}} = 1$. From a numerical point of view, however, the chosen approach has proven to be advantageous, since the conditions on $\mathbf{m}_{\bar{\mathbf{G}}}$ in the optimization can be set quite easily. As already stated, for $g_1 = g_2 = g_3$, $\bar{\mathbf{G}}$ belongs to the *isotropic* group \mathcal{G}_{14} , for $g_1 = g_2 \neq g_3$ to the *transversely isotropic* group \mathcal{G}_{13} and for $g_1 \neq g_2 \neq g_3$ to the *orthotropic* group \mathcal{G}_3 , see A.8 and A.4. If two 2nd order tensors are used, the ansatz given above is made for each of the two tensors, respectively. Thus, one gets 12 trainable parameters in this case: $\mathbf{m}_{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} \in \mathbb{R}^{12}$.

4th and 6th order structure tensors For the 4th order structure tensor, we propose an ansatz motivated from the crystal symmetries, see also Ebbing [53]. It is given by a sum of dyadic products of three vectors, i.e.,

$$\bar{\mathbf{G}} := \frac{1}{n} \bar{\mathbb{S}} \text{ with } \bar{\mathbb{S}} := \sum_{\alpha=1}^3 \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha, n := \bar{\mathbb{S}}_{KKLL} \quad (28)$$

This special choice for $\bar{\mathbf{G}}$ allows to fulfill for symmetry, positive semi-definiteness and normalization. In contrast to the 2nd order structure tensor, only a subset of the structure tensors described by Eq. (16) can be described using the

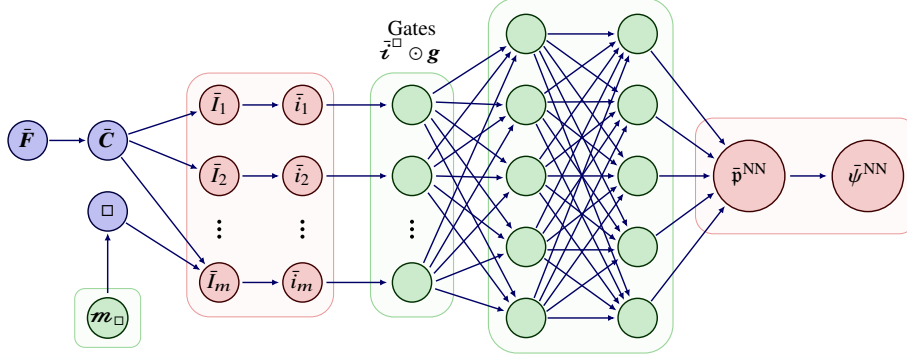


Figure 1: Illustration of the neural network $\bar{\psi}^{\text{NN}}(\bar{\mathbf{I}}^\square)$ for the representation of the elastic potential described by an invariant set $\bar{\mathbf{I}}^\square$ build from $\bar{\mathbf{C}} := \bar{\mathbf{F}}^T \cdot \bar{\mathbf{F}}$ and one or two structure tensor(s) $\square \in \{\bar{\mathbf{G}}, \bar{\mathbf{G}}, \bar{\mathbf{G}}, (\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)\}$, i.e., the mapping is $\bar{\psi}^{\text{NN}} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$, $\bar{\mathbf{I}}^\square \mapsto \bar{\psi}^{\text{NN}}(\bar{\mathbf{I}}^\square) := (n^{\text{out}} \circ g^{\text{NN}} \circ \ell^{\text{gate}} \circ n^{\text{in}})(\bar{\mathbf{I}}^\square)$. Therein, $n^{\text{in}}(\bar{\mathbf{I}}^\square)$ and $n^{\text{out}}(\bar{\mathbf{p}}^{\text{NN}})$ are non-trainable normalization layers that have to be fitted before training, $\ell^{\text{gate}}(\bar{\mathbf{I}}^\square)$ is a trainable gate layer and $g^{\text{NN}}(\bar{\mathbf{I}}^\square \circ \mathbf{g})$ is a standard PNN guaranteeing positive outputs. The vector \mathbf{m}_\square includes the parameters of the structure tensors and is also trainable.

selected approach. The vectors, given by

$$\mathbf{A}_\alpha := a_\alpha \mathbf{N}(\alpha) \in \mathcal{L}_1, [\mathbf{N}] := \begin{bmatrix} \sin \vartheta_\alpha \cos \varphi(\alpha) \\ \sin \vartheta_\alpha \sin \varphi(\alpha) \\ \cos \vartheta_\alpha \end{bmatrix}, \quad (29)$$

are described by the three parameters $a_\alpha \in [0, 1]$, $\vartheta_\alpha \in [0, \pi]$, $\varphi_\alpha \in [0, 2\pi]$ each. The parameters of the three vectors, which describe the 4th order structure tensor, are summarized in the set $\mathbf{m}_{\bar{\mathbf{G}}} \in \mathcal{M}_{\bar{\mathbf{G}}}$. For the special case $\mathbf{A}_\alpha \cdot \mathbf{A}_\beta = |\mathbf{A}_{(\alpha)}| |\mathbf{A}_{(\beta)}| \delta_{\alpha\beta}$, the resulting tensor $\bar{\mathbf{G}}$ describes the *cubic* anisotropy group \mathcal{G}_7 and for $\mathbf{A}_\alpha \cdot \mathbf{A}_\beta = 0$, $\alpha \neq \beta$, $|\mathbf{A}_1| = |\mathbf{A}_2| \neq |\mathbf{A}_3|$ it represents the *tetragonal* group \mathcal{G}_5 , see A.6 and A.5.

Note that the choice of parameterization of the unit vectors may affect the uniqueness of the minimizer, see [89]. However, we were able to achieve very good results with the parameterization used here.

Similarly, for the 6h order structure tensor, we use the ansatz

$$\bar{\mathbf{G}} := \frac{1}{n} \bar{\mathbf{S}} \text{ with } \bar{\mathbf{S}} := \sum_{\alpha=1}^3 \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha, n := \bar{S}_{KKLLMM}. \quad (30)$$

Again, this particular choice is made to ensure symmetry, a generalized trace of one and positive semi-definiteness for $\bar{\mathbf{G}}$. The parameters describing the 6th order structure tensor are summarized in the set $\mathbf{m}_{\bar{\mathbf{G}}} \in \mathcal{M}_{\bar{\mathbf{G}}}$. For the special case $\mathbf{A}_\alpha \cdot \mathbf{A}_\beta = \pm \frac{1}{2} |\mathbf{A}_\alpha| |\mathbf{A}_\beta|$, $\alpha \neq \beta$ and $|\mathbf{A}_1| = |\mathbf{A}_2| = |\mathbf{A}_3|$, the tensor describes the *hexagonal* group \mathcal{G}_{11} , see A.7.

4.1.2 Neural network-based model

In the following, we will restrict ourselves to models that only use a single 2nd, 4th, or 6th order structure tensor or two 2nd order structure tensors. As stated above, these models already cover a wide range of symmetry groups and have proven to be sufficient for the numerical examples shown later. However, the extension to models with other combinations of structure tensors is straightforward.

Neural network architecture We construct an elastic potential based on a feedforward neural network (FNN) [90] with the invariant set $\bar{\mathbf{I}}^\square$ as input, i.e., $\bar{\psi}^{\text{NN}}(\bar{\mathbf{I}}^\square)$, where $\square \in \{\bar{\mathbf{G}}, \bar{\mathbf{G}}, \bar{\mathbf{G}}, (\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)\}$. This network has a structure tailored to the problem. It consist of a non-trainable input normalization layer, a trainable gate layer, a trainable positive neural network (PNN), and a non-trainable output normalization layer, i.e., $\bar{\psi}^{\text{NN}} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$, $\bar{\mathbf{I}}^\square \mapsto \bar{\psi}^{\text{NN}}(\bar{\mathbf{I}}^\square) := (n^{\text{out}} \circ g^{\text{NN}} \circ \ell^{\text{gate}} \circ n^{\text{in}})(\bar{\mathbf{I}}^\square)$. The task of the trainable gate layer is to remove unneeded invariants from the model during training. The architecture is illustrated in Fig. 1.

To satisfy that the output of the PNN is always greater equal to zero, the activation function in the last hidden layer has to be greater equal to zero for all outputs of the former layer. Here, the softplus activation function is chosen. Furthermore, weights and bias of the output layer have to be non-negative. For brevity, the *weights* and *biases* of

the PNN interconnected between the normalization layers and the gate layer are summarized in the parameter set $\boldsymbol{w} \in \mathcal{PNN}$, where the set \mathcal{PNN} is introduced in B, Eq. (86).

As introduced in [46], *non-trainable normalization layers* for in- and output are used, which allows an efficient optimization without having to normalize the training data. The layers are given by

$$\boldsymbol{n}^{\text{in}} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \bar{\boldsymbol{I}}^\square \mapsto \bar{\boldsymbol{t}}^\square \text{ and } \boldsymbol{n}^{\text{out}} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}, \bar{\boldsymbol{z}}^{\text{NN}} \mapsto \bar{\boldsymbol{\psi}}^{\text{NN}}. \quad (31)$$

As already mentioned, the weights and biases of these normalization layers are non-trainable. Instead, they have to be fitted to the data before training.

Furthermore, to allow to automatically remove unneeded invariants from the model during training, we add a *trainable gate layer* between the input normalization layer and the PNN. The gate layer is defined by

$$\boldsymbol{g}^{\text{gate}} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \bar{\boldsymbol{t}}^\square \mapsto \bar{\boldsymbol{t}}^\square \odot \boldsymbol{g} \text{ with } g_\alpha := \min(1, \gamma \tanh(\epsilon q_\alpha)) \in [0, 1], \quad (32)$$

where $\gamma, \epsilon \in \mathbb{R}_{>0}$ are hyper parameters and $q_\alpha \in [0, 1]$ are trainable variables. Thus, we have the additional set $\boldsymbol{q} \in \mathcal{Gate} := \{\boldsymbol{q} \in \mathbb{R}^n \mid q_\alpha \in [0, 1]\}$.

The details on the chosen neural network architecture are given in B.

Additional terms for the construction of the elastic potential In addition, following Linden et al. [24], terms which guarantee zero energy and stress in the undeformed state as well as a term for the fulfillment of the volumetric growth condition are added. Thus, we end up with

$$\bar{\boldsymbol{\psi}}^\square(\bar{\boldsymbol{I}}^\square, \bar{\boldsymbol{J}}) := \bar{\boldsymbol{\psi}}^{\text{NN}}(\bar{\boldsymbol{I}}^\square) + \bar{\boldsymbol{\psi}}^{\text{en}} + \bar{\boldsymbol{\psi}}^{\text{str},\square}(\bar{\boldsymbol{I}}^\square, \bar{\boldsymbol{J}}) + \bar{\boldsymbol{\psi}}^{\text{gr}}(\bar{\boldsymbol{J}}). \quad (33)$$

Note that the additional terms are chosen in such a way that the material symmetry is not violated.

The term to guarantee zero energy in the undeformed state is independent of the symmetry group and is given by $\bar{\boldsymbol{\psi}}^{\text{en}} = -\bar{\boldsymbol{\psi}}^{\text{NN}}(\bar{\boldsymbol{I}}^\square)|_{\bar{\boldsymbol{F}}=\boldsymbol{I}}$. In contrast, the energy expression enforcing $\bar{\boldsymbol{P}}(\bar{\boldsymbol{F}} = \boldsymbol{I}) = \boldsymbol{0}$ depends on the chosen set of invariants [24], i.e., whether the set is constructed either with $\bar{\boldsymbol{G}}, \bar{\boldsymbol{G}}, \bar{\boldsymbol{G}}$ or $(\bar{\boldsymbol{G}}_1, \bar{\boldsymbol{G}}_2)$. The expressions $\bar{\boldsymbol{\psi}}^{\text{str},\square}(\bar{\boldsymbol{I}}^\square, \bar{\boldsymbol{J}})$ chosen in this work are summarized in C. For the growth term, which enforces the fulfillment of the volumetric growth condition introduced in Sect. 2.2, the expression $\bar{\boldsymbol{\psi}}^{\text{gr}}(\bar{\boldsymbol{J}}) := \lambda_{\text{gr}} (\bar{\boldsymbol{J}} + \bar{\boldsymbol{J}}^{-1} - 2)^3$ is chosen. The parameter λ_{gr} has to be chosen such that the energy grows fast enough during compression. Following [25], a value between 1×10^{-2} and 1×10^{-3} the material's initial stiffness has proven to be suitable.

The proposed models fulfill the following conditions by construction: *thermodynamic consistency, symmetric Cauchy and 2nd Piola-Kirchhoff stress, objectivity, material symmetry, energy- and stress-free undeformed state, as well as volumetric growth condition.*⁷

4.2 Training

To calibrate the NN-based models, the weights and biases of the PNN, collected into $\boldsymbol{w} \in \mathcal{PNN}$, the parameters of the generalized structure tensors $\boldsymbol{m}_\square \in \mathcal{M}_\square$, and the gate variables $\boldsymbol{q} \in \mathcal{Gate}$ have to be determined in a suitable training procedure. Thereby, we apply a strategy allowing to optimize all parameters simultaneously, which has the advantage that no further information from the microscopic geometry are necessary. Only data from RVE simulations are required. We collect these data into the set $\mathcal{D} := \{^1\mathcal{T}, ^2\mathcal{T}, \dots, ^k\mathcal{T}\}$. Thereby, the tuples are given by

$$^i\mathcal{T} := ({}^i\bar{\boldsymbol{F}}, {}^i\bar{\boldsymbol{\psi}}, {}^i\bar{\boldsymbol{\sigma}}, {}^i\bar{\boldsymbol{\alpha}})^{\text{RVE}} \in \mathcal{GL}^+(3) \times \mathbb{R}_{\geq 0} \times \text{Sym} \times \mathcal{L}_4. \quad (34)$$

Please note that a calibration based on $\bar{\boldsymbol{P}}$ and $\bar{\boldsymbol{A}}$ is equivalent, as the respective quantities can be calculated from each other by pullback or push-forward operations, see Sects. 2.1 and 2.2.

A model must be assessed against non-calibration data in order to look at its generalizability. Consequently, in accordance with standard machine learning procedure [24, 36, 45, 90], we split the whole dataset \mathcal{D} into *calibration* and *test* sets, respectively:

$$\mathcal{D} = \mathcal{D}_{\text{cal}} \cup \mathcal{D}_{\text{test}} \text{ and } \emptyset = \mathcal{D}_{\text{cal}} \cap \mathcal{D}_{\text{test}}. \quad (35)$$

⁷Note that the introduced energy expression is not polyconvex by construction. However, since a polyconvex energy turned out to be too restrictive for the fitting, this does not play a role here. For details on how an NN-based polyconvex energy can be constructed for anisotropic materials, which fulfills the material symmetry and at the same time guarantees a stress-free undeformed configuration, the reader is referred to Linden et al. [24].

The NN-based model's predictions for $\mathcal{D}_{\text{test}}$ can be verified once it has been calibrated with \mathcal{D}_{cal} . Thereby, the calibrated model should be able to generate reasonable predictions not only for the calibration but also for the test dataset, which will ensure good generalization of the model and allows to mimic the RVE's effective behavior for arbitrary states $\bar{\mathbf{F}} \in \mathcal{GL}^+(3)$.

To perform training based on the tuples given in Eq. (34), we define the following loss terms for the *mean squared error* (MSE) for energy, stress and elasticity tensor:

$$\mathcal{L}^\psi := \frac{1}{n^\psi} \left\| \bar{\psi} - \bar{\psi}(\bar{\mathbf{F}}, \mathbf{w}, \mathbf{m}_\square, \mathbf{q}) \right\|_{\text{MSE}} \quad \text{with } n^\psi := \max \left(\bar{\psi}^2, {}^2\bar{\psi}^2, \dots, |{}^{\mathcal{D}_{\text{cal}}}\bar{\psi}^2 \right), \quad (36)$$

$$\mathcal{L}^\sigma := \frac{1}{n^\sigma} \left\| \bar{\sigma} - \bar{\sigma}(\bar{\mathbf{F}}, \mathbf{w}, \mathbf{m}_\square, \mathbf{q}) \right\|_{\text{MSE}} \quad \text{with } n^\sigma := \frac{1}{3^2} \max \left(\|\bar{\sigma}\|^2, \|\bar{\sigma}\|^2, \dots, \|{}^{\mathcal{D}_{\text{cal}}}\bar{\sigma}\|^2 \right), \quad (37)$$

$$\mathcal{L}^c := \frac{1}{n^c} \left\| \bar{\mathbf{c}} - \bar{\mathbf{c}}(\bar{\mathbf{F}}, \mathbf{w}, \mathbf{m}_\square, \mathbf{q}) \right\|_{\text{MSE}} \quad \text{with } n^c := \frac{1}{6^2} \max \left(\|\bar{\mathbf{c}}\|^2, \|\bar{\mathbf{c}}\|^2, \dots, \|{}^{\mathcal{D}_{\text{cal}}}\bar{\mathbf{c}}\|^2 \right). \quad (38)$$

Therein, $\bar{\mathbf{c}} \in \mathbb{R}^{6 \times 6}$ denotes a matrix representation of the coordinates of the 4th order material tangent $\bar{\mathbf{c}}$ given in Voigt notation and $\|(\cdot) - (\cdot)\|_{\text{MSE}}$ is the standard MSE norm. The normalization factors n^ψ , n^σ and n^c ensure that the loss terms can be compared with each other. Thus, the prediction loss is given by $\mathcal{L}^{\text{pred}} = w^\psi \mathcal{L}^\psi + w^\sigma \mathcal{L}^\sigma + w^c \mathcal{L}^c$, where the non-trainable weights $w^\psi, w^\sigma, w^c \in \mathbb{R}_{\geq 0}$, with $w^\psi + w^\sigma + w^c = 1$, have to be chosen. In addition, to enforce sparsity of the model with respect to the number of invariants, a penalty term based on the p -norm [10, 68] of the gates

$$\mathcal{L}^{\text{gate}} := \frac{1}{n^{\text{gate}}} \left[\sum_{\alpha=1}^n (g_\alpha(q_\alpha) + \delta)^p \right]^{\frac{1}{p}} \quad \text{with } n^{\text{gate}} := [n(1 + \delta)^p]^{\frac{1}{p}} \quad (39)$$

is used, where $p \in \mathbb{R}_{\geq 0}$ and n is the number of invariants. The parameter $\delta \ll 1$ prevents division by zero when differentiating. By applying the introduced loss terms, the training is defined by the optimization

$$(\hat{\mathbf{w}}, \hat{\mathbf{m}}, \hat{\mathbf{q}}) = \arg \min_{\mathbf{w} \in \mathcal{PNN}, \mathbf{m}_\square \in \mathcal{M}_\square, \mathbf{q} \in \mathcal{Gate}} (w^\psi \mathcal{L}^\psi + w^\sigma \mathcal{L}^\sigma + w^c \mathcal{L}^c + w^{\text{gate}} \mathcal{L}^{\text{gate}}), \quad (40)$$

where the parameter w^{gate} has to be chosen such that the approximation quality of the model does not decrease. Since the second derivatives of the potential with respect to $\bar{\mathbf{F}}$ are necessary to compute $\bar{\mathbf{A}}$ and from this $\bar{\mathbf{c}}$, it is a *higher-order Sobolev training* [45, 46]. The proposed training method allows to calibrate all four models, $\bar{\psi}^{\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$, $\bar{\psi}^{\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$, $\bar{\psi}^{\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$, as well as $\bar{\psi}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}, \bar{\mathbf{J}})$. Note that the internal normalization layers of the PNN have to be fitted to the data before training.

Remark 3. To achieve optimal results after training, a two-step approach has shown to be favorable. Thereby, a pre-training with the *Adam optimizer* and a post-training with the *SLSQP optimizer* (Sequential Least Squares Programming), which is a quasi Newton optimizer, is carried out. As shown in [46], the optimization with SLSQP is clearly superior to Adam for comparatively small networks, since it converges faster and a lower loss is reached. However, it has shown to be difficult to find a suitable structure tensor only with SLSQP, which is due to the fact that a strong dependence on the initialization is typical for this optimizer. However, very good results can be achieved if good starting values are specified by applying the pre-training with Adam. The implementation of the models and the calibration workflow was realized using *Python*, *TensorFlow* and *SciPy*.

4.3 Data generation and model identification procedure

In order to generate the NN-based surrogate for RVE simulations, the following procedure is applied: (a) *sampling of the deformation space in a prescribed range*, (b) *generation of a database by computational homogenization*, and (c) *model identification procedure* which consists of a maximum of four substeps. The overall procedure is illustrated in Fig. 2.

The sampling technique, which is similar to [46], is described in D and the computational homogenization is done as described in Sect. 2.4. Finally, in step (c), the model which is suitable as a surrogate, i.e., $\bar{\psi}^{\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$, $\bar{\psi}^{\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$, $\bar{\psi}^{\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$ or $\bar{\psi}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}, \bar{\mathbf{J}})$, is chosen and calibrated. In order to obtain an RVE surrogate model that is as efficient as possible in macroscopic FE simulations [25], we want to find the model with the lowest possible order in the structure tensor that allows to describe the RVE's underlying symmetry group.⁸ This is done as follows: First, in (c.1), the model

⁸Note that models based on higher order structure tensors contain others, e.g., the model based on the 4th order structure tensor can also be used to describe transverse isotropy and orthotropy.

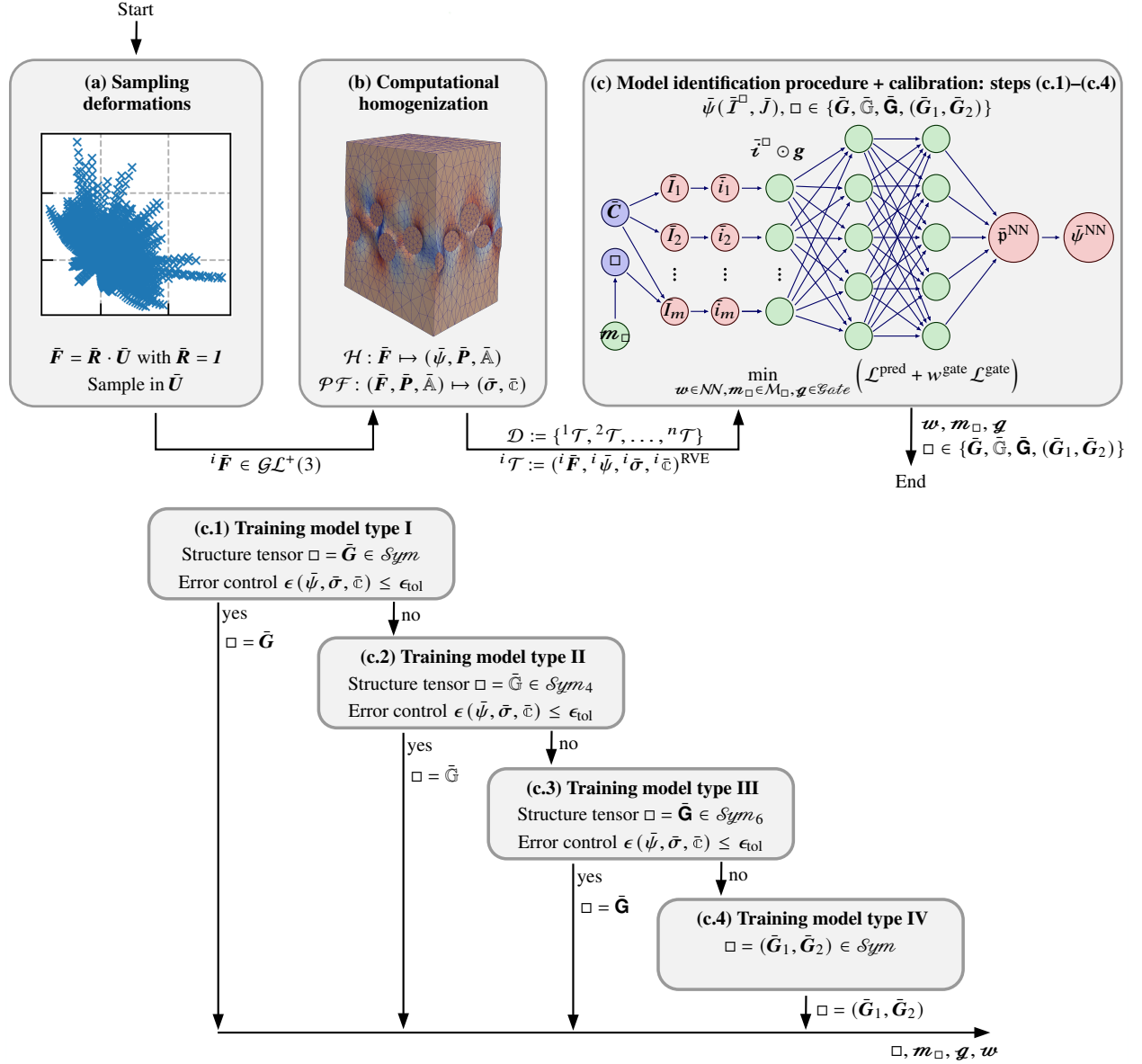


Figure 2: Data generation and model identification procedure: (a) the deformation space is sampled in a prescribed range by Latin Hypercube Sampling (LHS), (b) by prescribing the sampled deformation gradients ${}^i\bar{\mathbf{F}}$ in RVE simulations, corresponding energy ${}^i\bar{\psi}$, stress ${}^i\bar{\sigma}$ and elasticity tensor ${}^i\bar{\mathbb{C}}$ are calculated, (c) the dataset \mathcal{D} is used to identify and calibrate the NN-based model $\bar{\psi}(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$ via a higher-order Sobolev training with prediction loss $\mathcal{L}^{\text{pred}} = w^\psi \mathcal{L}^\psi + w^\sigma \mathcal{L}^\sigma + w^{\mathbb{C}} \mathcal{L}^{\mathbb{C}}$ and gate loss $\mathcal{L}^{\text{gate}}$ for sparsity. The step (c) involves the substeps (c.1)–(c.4): to identify which (set of) structure tensor(s) $\square \in \{\bar{\mathbf{G}}, \bar{\mathbb{G}}, \bar{\mathbf{G}}, (\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)\}$, i.e., 2nd, 4th, 6th, or two 2nd order structure tensor(s), is needed, the training is performed sequentially from low to high structure tensor order and an error control is performed after each training to decide whether the underlying anisotropy can be described with sufficient accuracy.

based on the 2nd order structure tensor is calibrated with the RVE data. After training, the relative error measures in energy, stress and material tangent

$$\epsilon^\psi = \frac{\sum_{i=1}^{|\mathcal{D}|} |\bar{\psi}^\square - i\bar{\psi}|}{\sum_{j=1}^{|\mathcal{D}|} |j\bar{\psi}|}, \quad \epsilon^\sigma = \frac{\sum_{i=1}^{|\mathcal{D}|} \|i\bar{\sigma}^\square - i\bar{\sigma}\|}{\sum_{j=1}^{|\mathcal{D}|} \|j\bar{\sigma}\|} \quad \text{and} \quad \epsilon^c = \frac{\sum_{i=1}^{|\mathcal{D}|} \|i\bar{c}^\square - i\bar{c}\|}{\sum_{j=1}^{|\mathcal{D}|} \|j\bar{c}\|} \quad (41)$$

are calculated. If all errors are below a given tolerance ϵ_{tol} , the model identification is complete. Here, we choose $\epsilon_{\text{tol}} = 1\%$. If the tolerance is exceeded, the model based on the 4th order structure tensor is trained and the errors are calculated again, substep (c.2). If this model is also insufficient, the 6th order structure tensor is used, substep (c.3). Finally, if it is not possible to reach errors below ϵ_{tol} with a model based on a single structure tensor, a training based on two 2nd order structure tensors is done (c.4). Note that this final step should lead to an error below the given tolerance, since the formulation then includes the triclinic group, i.e., fully anisotropic behavior. In each substep, the training as described in Sect. 4.2 is applied.

5 Numerical examples

In order to illustrate the ability of the developed NN-based elastic surrogate models to mimic the complex anisotropic behavior of RVEs with high precision, we will calibrate them with data from computational homogenizations in the following. With these data, the interpolation behavior of the models as well as the extrapolation behavior is investigated. Since we only use data generated by homogenization, we do not consider the case of noisy data in this paper.

5.1 Data generation via computational homogenization

In this work, two-phase soft composites consisting of matrix and spherical inclusions or matrix and fibers are considered. In order to provide data with varying overall anisotropic behavior, we consider five different RVEs depicted in Fig. 3: An RVE of a fiber reinforced material with stochastic fiber distribution (*stochastic fibers*), a unit cell with a hexagonal fiber arrangement (*hexagonal fibers*), a unit cell with one spherical inclusion (*cubic sphere*), an RVE with a plane-like arrangement of particles (*plane spheres*), and an RVE with an arrangement of particles in a chain-like structure (*chain spheres*). The geometries and the subsequent creation of the periodic mesh has been done with the Python tool *gmshModel*.⁹

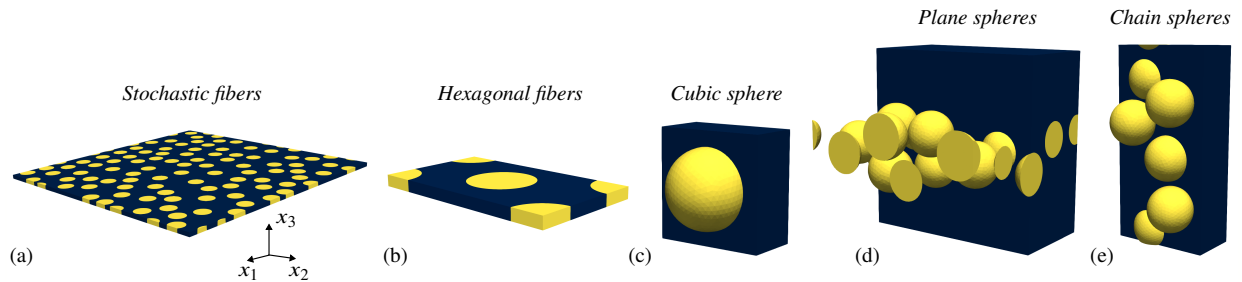


Figure 3: Considered RVEs for data generation: (a) fiber reinforced material (*stochastic fibers*), (b) unit cell with hexagonal fiber arrangement (*hexagonal fibers*), (c) unit cell with one spherical inclusion (*cubic sphere*), (d) particle reinforced plane-like microstructure (*plane spheres*), and (e) particle reinforced chain-like microstructure (*chain spheres*). The volume fractions of the fiber/particle phase are given by $\phi \in \{30, 30, 20, 12, 15\}\%$ from left to right.

All components, i.e., matrix, particles and fibers are assumed to be compressible and isotropic. For all, we choose the two-parametric neo-Hookean model according to Ciarlet [91] given by

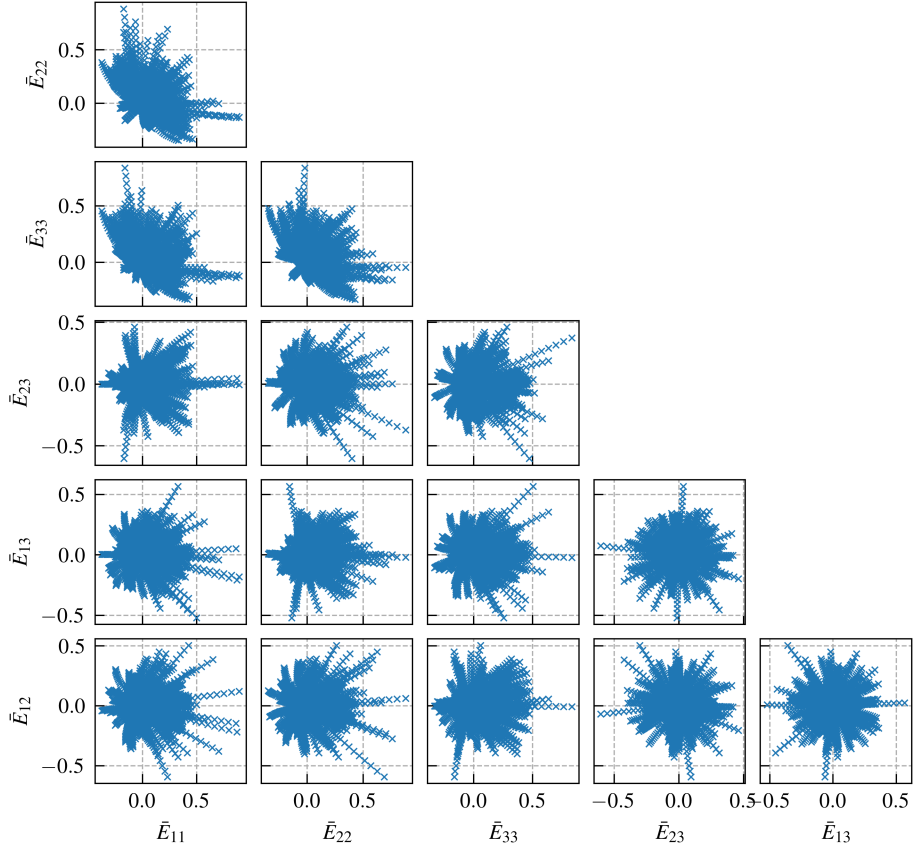
$$\psi(I_1, I_3) := \frac{1}{2} \left[\mu(I_1 - \ln I_3 - 3) + \frac{\lambda}{2}(I_3 - \ln I_3 - 1) \right] \quad \text{with} \quad \mu, \lambda > 0, \quad \mu = \frac{E}{2(1+\nu)}, \quad \lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}. \quad (42)$$

In the equation above, μ and λ denote Lamé parameters and E and ν are Young's modulus and Poisson's ratio, respectively. Note that the elastic potential (42) is also polyconvex [24] and thus guarantees ellipticity for all possible states $\mathbf{F} \in \mathcal{GL}^+(3)$. The chosen material parameters for the individual components are given in Tab. 1.

⁹The Python tool *gmshModel* is freely available under the link <https://gmshmodel.readthedocs.io/en/latest/>

Table 1: Chosen parameters for matrix material, spherical particles and fibers, described by the constitutive model (42).

Parameter	Symbol	Matrix material	Spherical particles	Fibers
Young's modulus	E	1 MPa	1000 MPa	10 MPa
Poisson's ratio	ν	0.4	0.3	0.44

Figure 4: Sampled deformation space comprising 141 loading paths with 20 increments each. Shown are sectional planes of the Green-Lagrange strain tensor $\bar{\mathbf{E}}$.

In order to investigate the RVEs' behavior within a predefined range of effective deformations $\bar{\mathbf{F}} \in \mathcal{GL}^+(3)$, a sampling technique similar to the approach used in [46] is applied. Details on this technique are given in D. With that, a total of 154 loading paths comprising 20 increments each are generated. The sampled states are shown within sectional planes of the Green-Lagrange strain tensor $\bar{\mathbf{E}}$ in Fig. 4. For all sampled states, the resulting effective material response is calculated by computational homogenization according to Sect. 2.4, i.e., $\mathcal{H} : \bar{\mathbf{F}} \mapsto (\bar{\psi}, \bar{\mathbf{P}}, \bar{\mathbb{A}})$ and $\mathcal{PF} : (\bar{\mathbf{F}}, \bar{\mathbf{P}}, \bar{\mathbb{A}}) \mapsto (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\epsilon}})$.¹⁰

Besides the determination of energy, stress and tangent corresponding to a deformation state, the *local ellipticity* condition (6) has been evaluated for all generated data. To this end, the positive semi-definiteness of the acoustic tensor $\bar{\boldsymbol{\gamma}}(\bar{\mathbf{F}}, \mathbf{N}) = \bar{A}_{iJKL}(\bar{\mathbf{F}}) N_J N_L \mathbf{e}_i \otimes \mathbf{e}_k$ [46, 53, 54] has been checked numerically by sampling $N(\vartheta, \varphi)$ parameterized by spherical coordinates $(\vartheta, \varphi) \in [0, \pi] \times [0, 2\pi]$ in $\pi/180$ steps for all states $\bar{\mathbf{F}}$ included in the dataset. Thereby, a loss of ellipticity was found in some states for all RVEs considered, except for the RVE plane spheres. Thus, the effective material response cannot be represented by a polyconvex model, which would guarantee ellipticity by construction, at least not for the complete data set.

The generated homogenized data will be made freely available in the final version of the article.

¹⁰Note that some simulations of the RVEs with spherical inclusions diverged before reaching the last increment, which is due to the fact that the large phase contrast can easily lead to extremely high matrix deformations. In that case, the corresponding states have been removed from the dataset of the RVE, respectively.

Remark 4. It is well known that polyconvex microscopic energies in multiscale problems do not necessarily imply macroscopic quasi-convexity and consequently rank-one convexity [74, 92]. For some regions of the effective deformation \bar{F} , even a loss of ellipticity is possible, cf. [92, 93].

5.2 Training and prediction of the neural networks

The obtained data \mathcal{D} for the different considered RVEs are now used for the calibration of the proposed NN-based macroscopic models introduced in Sect. 4, where the model identification procedure described in Sect. 4.3 is applied, i.e., it is automatically determined whether a 2nd, 4th or 6th order structure tensor or a set of two 2nd order structure tensors is required. To evaluate the performance of the developed invariant-based NN models, a typical approach based on the coordinates of the right Cauchy-Green deformation tensor \bar{C} will also be considered as a reference. Similar models are widespread, e.g., [22, 44]. The *coordinate-based reference model* is given by the energy expression

$$\bar{\psi}^{\text{coord}}(\bar{C}, \bar{J}) := \bar{\psi}^{\text{PNN}}(\bar{C}) - \bar{\psi}^{\text{PNN}}(\bar{C})|_{\bar{F}=\mathbf{I}} - \frac{\partial \bar{\psi}^{\text{PNN}}}{\partial \bar{C}} \Big|_{\bar{F}=\mathbf{I}} : (\bar{C} - \mathbf{I}) + \lambda_{\text{gr}} (\bar{J} + \bar{J}^{-1} - 2)^2, \quad (43)$$

where a PNN according to B, but without trainable gate, is chosen for the network. To enable a fair comparison, it is designed to fulfill the same conditions as the invariant-based models except for the material symmetry, i.e., it ensures thermodynamic consistency, symmetric Cauchy and 2nd Piola-Kirchhoff stress, objectivity, energy- and stress-free undeformed state, as well as volumetric growth condition.

The parameter λ_{gr} has to be chosen such that the energy grows fast enough during compression. According to [46], a value of around 1×10^{-2} or 1×10^{-3} the material's initial stiffness¹¹ has shown to be reasonable. Here we choose $\lambda_{\text{gr}} = 0.01$ MPa for all NN-based models. According to the hyperparameter study given in E.1, two hidden layers with 16 neurons each are chosen for the invariant-based NN models and three hidden layers with 16 neurons each for the coordinate-based model. For training of the invariant-based NN models, the parameter for the penalization of the gates was chosen to $w^{\text{gate}} = 5 \times 10^{-5}$, see E.2 for a study to determine this value. Following [10], we have chosen $p = \frac{1}{4}$ for the exponent in the p -norm. The parameters in the gate are chosen to $\gamma = 1.025$, $\epsilon = 2.5$ and $\delta = 1 \times 10^{-6}$, respectively.

To exclude random effects from initialization, 5 training runs with pre-training and post-training steps as described in Remark 3 were carried out each. The best model is then used for the discussion of the results, where this is decided by the measure $\mathcal{L}^{\text{pred}}/A$, where $A \in \mathbb{N}$ is the number of active gates. In the pre-training with Adam, an initial learning rate of 0.01 and an exponential learning rate decay, so that the learning rate is multiplied by 1/3 every 500 epochs, was selected. In addition, mini-batches were used, which introduces a regularization effect [90]. The mini-batches were recomposed in a random manner after each epoch. For the interpolation studies, the mini-batch size was set to 64, while it was set to 16 for the extrapolation. All training runs were carried out with 8 CPUs each, whereby a high performance cluster (HPC) equipped with Intel Xeon Platinum 8470 CPUs was used.

To mark that a structure tensor was calibrated by data from a specific RVE, we use the following superscripts: *stochastic fibers* \square^{stf} , *hexagonal fibers* \square^{hef} , *cubic sphere* \square^{cus} , *plane spheres* \square^{pls} , and *chain spheres* \square^{chs} .

5.2.1 Interpolation behavior: Training with stress and tangent

In this first study we mainly check the interpolation ability of the NN-based models. Thus, as described in Sect. 4.2, we divide the overall datasets into calibration and test sets with a ratio of 70/30, respectively, see Eq. (35). The prediction loss $\mathcal{L}^{\text{pred}} = 0.7\mathcal{L}^{\sigma} + 0.3\mathcal{L}^{\epsilon}$ is chosen, which is to weight the stresses somewhat more than the material tangent. It should be noted that experience has shown that it is not necessary to include the energy $\bar{\psi}$ itself in the loss, as this is adjusted very well with the Sobolev training used here [24, 29, 46].

Identified structure tensors For the RVEs *stochastic fibers* and *plane spheres*, a model based on a 2nd order generalized structure tensor \bar{G} was selected by the algorithm according to Fig. 2, respectively. The coordinates of these structure tensors are given by

$$[\bar{G}^{\text{stf}}] = \begin{bmatrix} 0.01 & 0.01 & 0. \\ 0.01 & 0.02 & 0. \\ 0. & 0. & 0.97 \end{bmatrix} \text{ and } [\bar{G}^{\text{pls}}] = \begin{bmatrix} 0.01 & 0. & -0.01 \\ 0. & 0. & 0.02 \\ -0.01 & 0.02 & 0.99 \end{bmatrix}. \quad (44)$$

The vectors of active/non-active gates for the two RVEs are given by

$$G(\mathbf{g}^{\text{stf}}) = (1, 0, 1, 0, 1, 1, 0) \text{ and } G(\mathbf{g}^{\text{pls}}) = (1, 1, 1, 1, 0, 0, 1), \quad (45)$$

¹¹In the anisotropic case the maximum initial stiffness has to be considered.

respectively, where $G(x) := [x > 0]$ is a function which is zero for all $x \leq 0$ and 1 else. The gates correspond to the invariants defined in Eqs. (12), (14), i.e., the additional invariants needed to describe orthotropy are removed from the model by the gates. The identified structure tensor for the RVE *stochastic fibers* corresponds to the expected result, i.e. transverse isotropy \mathcal{G}_{13} around \mathbf{e}_3 , as the fibers are aligned with the X_3 axis, which was identified as a preferred direction during training.¹² The result also makes sense for the RVE *plane spheres*, where the out-of-plane direction is a preferred direction. The slight deviation to ideal transverse isotropy with X_3 as preferred direction results from the stochastic arrangement of the particles within the plane of particles and from numerical inaccuracies during training.

In contrast, for the RVE *cubic sphere*, the selection procedure has identified that a 4th order generalized structure tensor $\bar{\mathbf{G}}$ is needed. This is not surprising, since it is well known that a 4th order structure tensor is needed for cubic anisotropy [53, 65, 66]. Interestingly, however, a 4th order structure tensor was identified during the optimization, which belongs to the tetragonal group \mathcal{G}_5 . This can be explained by the fact that the invariant set belonging to the cubic group \mathcal{G}_7 up to order $\bar{\mathbf{C}}^3$ can be represented by the invariant set of the tetragonal group, cf. A.6. The coordinates of the identified structure tensor in Voigt notation and the vector of active gates are given by

$$[\bar{\mathbf{G}}^{\text{cus}}] = \begin{bmatrix} 0. & 0. & 0. & -0. & -0. & 0. \\ 0. & 0.5 & 0. & -0. & -0. & 0. \\ 0. & 0. & 0.5 & -0. & -0. & 0. \\ -0. & -0. & -0. & 0. & 0. & -0. \\ -0. & -0. & -0. & 0. & 0. & -0. \\ 0. & 0. & 0. & -0. & -0. & 0. \end{bmatrix} \text{ and } G(\mathbf{g}^{\text{cus}}) = (1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0), \quad (46)$$

respectively. The gates correspond to the invariants defined in Eqs. (12), (19).

For the RVE *hexagonal fibers*, the algorithm has determined that even a 6th order structure tensor is required. This also corresponds to the expectation for a material that belongs to the hexagonal anisotropy group \mathcal{G}_{11} [65, 66]. The three vectors \mathbf{A}_1 , \mathbf{A}_2 and \mathbf{A}_3 specifying the 6th order tensor $\bar{\mathbf{G}} \in \mathcal{S}ym_6$ have been identified to

$$[\mathbf{A}_1^{\text{hef}}] = \begin{bmatrix} -0.66 \\ 0.38 \\ 0. \end{bmatrix}, \quad [\mathbf{A}_2^{\text{hef}}] = \begin{bmatrix} -0. \\ 0.76 \\ -0. \end{bmatrix} \text{ and } [\mathbf{A}_3^{\text{hef}}] = \begin{bmatrix} -0.66 \\ -0.38 \\ -0. \end{bmatrix}. \quad (47)$$

As one can see, the conditions $(|\mathbf{A}_\alpha| |\mathbf{A}_\beta|)^{-1} \mathbf{A}_{(\alpha)} \cdot \mathbf{A}_{(\beta)} = \pm \frac{1}{2}, \alpha \neq \beta, N \cdot \mathbf{A}_\alpha = 0$, with $N \in \mathcal{N}$ being the fiber direction, and $|\mathbf{A}_1| = |\mathbf{A}_2| = |\mathbf{A}_3|$ apply in good approximation to the three vectors that form $\bar{\mathbf{G}}$. The structure tensor determined thus corresponds to the expected structure tensor for the hexagonal group \mathcal{G}_{11} , cf. A.7. The vector of active/non-active gates is given by $G(\mathbf{g}^{\text{hef}}) = (1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1)$, where the gates correspond to the invariants defined in Eqs. (12), (24).

Finally, the identification algorithm has determined that two 2nd order structure tensors are needed for the RVE *chain spheres*. These two tensors are given by

$$[\bar{\mathbf{G}}_1^{\text{chs}}] = \begin{bmatrix} 0.08 & 0.02 & 0.04 \\ 0.02 & 0.53 & -0.01 \\ 0.04 & -0.01 & 0.39 \end{bmatrix} \text{ and } [\bar{\mathbf{G}}_2^{\text{chs}}] = \begin{bmatrix} 0.5 & -0.02 & 0.05 \\ -0.02 & 0.49 & -0. \\ 0.05 & -0. & 0.01 \end{bmatrix}. \quad (48)$$

To determine which symmetry group was determined, we apply the tests given in Sect. 3.4. Accordingly, since $w_L = e_{LMN} \bar{\mathbf{G}}_{MP}^1 \bar{\mathbf{G}}_{PN}^2 \neq 0$ and $\mathbf{v}_\alpha = (\bar{\mathbf{G}}_\alpha \cdot \mathbf{w}) \times \mathbf{w} \neq \mathbf{0}$ for $\alpha \in \{1, 2\}$, the algorithm has detected the *triclinic group* \mathcal{G}_1 . However, taking a certain numerical tolerance into account, the structural tensors can be assigned to the *monoclinic group* \mathcal{G}_2 to a good approximation. To show this, we consider \mathbf{w} , \mathbf{v}_1 and \mathbf{v}_2 and compute the norms given by $|\mathbf{w}| = 3.77 \times 10^{-2}$, $|\mathbf{v}_1| = 6.59 \times 10^{-5}$ and $|\mathbf{v}_2| = 1.90 \times 10^{-4}$. Thus, since the norms of the latter two vectors are approximately two orders of magnitude smaller compared to the norm of \mathbf{w} , one can say that the conditions which have to hold for the monoclinic group are fulfilled in good approximation, i.e., $\mathbf{w} \neq \mathbf{0} \wedge \mathbf{v}_1 = \mathbf{0} \wedge \mathbf{v}_2 = \mathbf{0}$. This means, $\bar{\mathbf{G}}_1$ and $\bar{\mathbf{G}}_2$ share one and only one (anti)parallel eigenvector in good approximation. The vector of active/non-active gates is given by $G(\mathbf{g}^{\text{chs}}) = (1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0)$. The gates correspond to the invariant set as introduced in Sect. 3.4.

Training process and model performance The training process is exemplary shown for the RVE *plane spheres* in Fig. 5, where the calibration loss is plotted for all 5 runs with thin lines and the best run is marked with a thick line, respectively. The deviation between the individual runs is relatively small, which demonstrates the robustness

¹² For the transverse isotropy with preferred direction X_3 there is an unlimited number of permissible structure tensors with $\text{tr } \bar{\mathbf{G}} = 1$, since all tensors given by $\bar{\mathbf{G}} = \alpha(\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2) + \beta \mathbf{e}_3 \otimes \mathbf{e}_3$ with $\alpha, \beta \in \mathbb{R}_{\geq 0}, \alpha \neq \beta$ and $2\alpha + \beta = 1$ are equivalent to each other, cf. A.8.

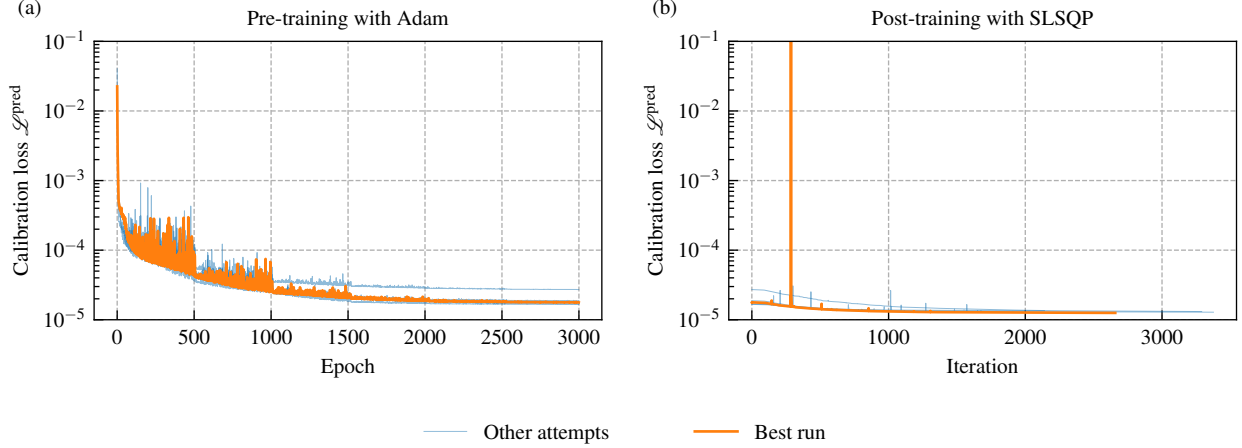


Figure 5: Training process of the invariant-based NN model $\bar{\psi}^{\bar{G}}(\bar{\mathcal{I}}^{\bar{G}}, \bar{\mathcal{J}})$ for the RVE plane spheres with the loss $\mathcal{L} = \mathcal{L}^{\text{pred}} + 5 \times 10^{-5} \mathcal{L}^{\text{gate}}$, $\mathcal{L}^{\text{pred}} = 0.7\mathcal{L}^{\sigma} + 0.3\mathcal{L}^{\epsilon}$: (a) pre-training with Adam optimizer and (b) post-training with SLSQP optimizer. Shown is the prediction loss for five training runs.

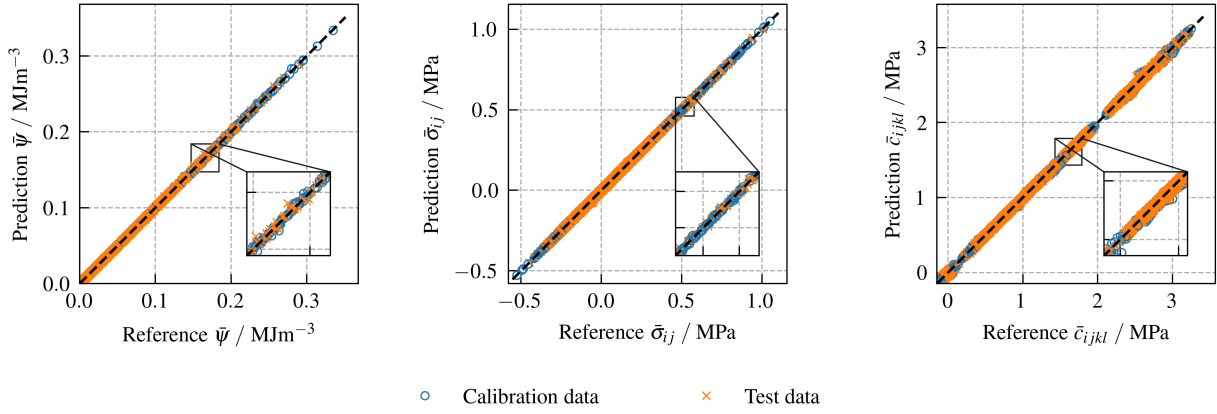


Figure 6: Predictions of the invariant-based NN model $\bar{\psi}^{\bar{G}}(\bar{\mathcal{I}}^{\bar{G}}, \bar{\mathcal{J}})$ for the RVE plane spheres compared to reference values: Shown are the energy $\bar{\psi} \in \mathbb{R}_{\geq 0}$ as well as the coordinates of $\bar{\sigma} \in \text{Sym}$ and $\bar{c} \in \text{Sym}_{\mathcal{A}}$. The ratio of calibration and test data is 70/30 and the training has been done with the loss $\mathcal{L} = \mathcal{L}^{\text{pred}} + 5 \times 10^{-5} \mathcal{L}^{\text{gate}}$, where $\mathcal{L}^{\text{pred}} = 0.7\mathcal{L}^{\sigma} + 0.3\mathcal{L}^{\epsilon}$.

of the method. As one can be seen in the plots, a very good result is already achieved during pre-training with Adam. Post-training with SLSQP, however, can noticeably reduce the loss again for other RVEs.

The performance of the calibrated model is depicted in Fig. 6 for energy, stress and material tangent, where the RVE *plane spheres* is exemplary considered here. As one can see, the accuracy of the invariant-based NN model is very good for all three quantities. Thereby, the accurate prediction quality for the material tangent \bar{c} is to be particularly emphasized. As shown in the zoom plot, even this rather difficult to model quantity is predicted with good precision. Note that there are only a few studies in the literature in which the NN's predictive quality of this quantity is considered at all, e.g., [45, 46, 62]. Works in which a loss is used for the material tangent and at the same time an energy-based NN model is used are even rarer [45], which is due to the fact that this requires expensive higher-order Sobolev training.

To avoid redundancy, the training process and the model predictions are not shown for the other four RVEs here. Instead, the final loss terms and the error measures as defined in Eq. (41) for all RVEs are given in Tab. 2. There, also the results for the coordinate-based model (43) are given for comparison. Although the coordinate-based model achieves slightly higher accuracies for most RVEs, the differences do not play a major role due to the very good precision also achieved with the invariant-based approach. It should also be noted that the coordinate-based model requires 3 instead of two hidden layers in order to achieve acceptable results, cf. the study given in E.1.

Table 2: Interpolation study on the performance of the developed invariant-based NN models $\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$ and the coordinate-based reference model $\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$ for the five considered RVEs and the prediction loss $\mathcal{L}^{\text{pred}} = 0.7\mathcal{L}^\sigma + 0.3\mathcal{L}^\epsilon$. The loss term for training the invariant-based models was $\mathcal{L} = \mathcal{L}^{\text{pred}} + 5 \times 10^{-5} \mathcal{L}^{\text{gate}}$ and an architecture with two hidden layers with 16 neurons each was chosen. The loss term for the coordinate-based model was $\mathcal{L} = \mathcal{L}^{\text{pred}}$ and an architecture with three hidden layers with 16 neurons each was chosen. The overall datasets were divided into calibration and test sets with a ratio of 70/30, respectively. All models were trained 5 times, where the best training run was selected. Given are the loss values after training and the error measures for $\bar{\psi}$, $\bar{\sigma}$ and $\bar{\epsilon}$, cf. Eqs. (37), (38) and (41).

RVE	Model	\square	Active gates	$\mathcal{L}_{\text{cal}}^{\text{pred}}$	$\mathcal{L}_{\text{test}}^{\text{pred}}$	$\epsilon^\psi / \%$	$\epsilon^\sigma / \%$	$\epsilon^\epsilon / \%$
Stochastic fibers	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	4	2.520×10^{-6}	3.031×10^{-6}	0.19	0.39	0.19
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	4.145×10^{-6}	6.305×10^{-6}	0.11	0.31	0.29
Hexagonal fibers	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	9	2.059×10^{-5}	2.352×10^{-5}	0.39	0.91	0.47
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	8.319×10^{-6}	1.063×10^{-5}	0.13	0.42	0.39
Cubic sphere	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	8	2.265×10^{-5}	2.341×10^{-5}	0.27	0.8	0.62
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	2.488×10^{-5}	3.363×10^{-5}	0.22	0.8	0.85
Plane spheres	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	5	1.349×10^{-5}	1.614×10^{-5}	0.41	0.92	0.44
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	4.563×10^{-6}	5.448×10^{-6}	0.1	0.31	0.29
Chain spheres	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$(\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)$	8	2.037×10^{-5}	2.242×10^{-5}	0.37	0.92	0.54
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	9.665×10^{-6}	1.297×10^{-5}	0.13	0.42	0.42

Elastic surface plots Finally, after the discussion of the achieved loss terms and error measures, we want to focus on the material tangents corresponding to the different RVEs. To visualize the effective anisotropic behavior of the four considered RVEs and to illustrate the high prediction quality of the calibrated NN models for the effective 4th order tensors $\bar{\mathbb{c}}$, the technique described in Nordmann et al. [94] to visualize the material’s effective Young’s modulus in all spatial directions is applied. We call such a plot *elastic surface* in the following. Since $\bar{\mathbb{c}}(\bar{\mathbf{F}})$ depends on the deformation within the considered nonlinear setting, it is not constant. Thus, the material tangent is not only visualized for the undeformed state, i.e., $\bar{\mathbb{c}}(\bar{\mathbf{F}} = \mathbf{I})$ which is the elasticity tensor, but also for a state with $\bar{\mathbf{F}} \neq \mathbf{I}$. Exemplary, the state

$$[\bar{\mathbf{F}}] = \begin{bmatrix} 0.89 & 0.09 & 0.02 \\ 0.09 & 1.4 & 0.18 \\ 0.02 & 0.18 & 0.94 \end{bmatrix} \quad (49)$$

is chosen here. The 3D plots of the Young’s modulus are given in Figs. 7–11 for the five different RVEs, see [94] for details on this visualization technique. As can be clearly seen, each RVE is characterized by a special anisotropy. All RVEs show a significant change in the elastic surface as a result of the imposed deformation. Both the shape of the surface shown and the amount of the maximum modulus \bar{E} change significantly. This effect is partly due to the already nonlinear behavior of the individual phases given by Eq. (42), but mainly due to the change in the microstructure as a result of the applied deformation. For both, undeformed as well as deformed state, the plots underpin the excellent prediction quality of the invariant-based NN models.

Once again it should be noted that a smaller NN architecture compared to the coordinate-based model is needed to achieve a similar accuracy in the predictions with the invariant-based approach. In addition, the trained model can be interpreted in a certain way, since the anisotropy group can also be inferred beyond the linear elastic case by analyzing the structure tensors. Thus, the advantage of our approach becomes clear even from this first study.

5.2.2 Interpolation behavior: Training with stress

In order to show that the underlying anisotropy and thus a meaningful structural tensor can be identified with the proposed approach by using deformation-stress tuples alone, only $\mathcal{L}^{\text{pred}} = \mathcal{L}^\sigma$ is now used as loss. The overall datasets are again divided into calibration and test sets with a ratio of 70/30, respectively, see Eq. (35). In order to decide when the selection procedure shown in Fig. 2 terminates, the only the error measures ϵ^ψ and ϵ^σ , so not ϵ^ϵ , are chosen since the material tangent is not directly trained within this study, cf. Eq (41) for the definition of the errors.

We will discuss the results for the RVE *plane spheres* as an example, cf. Fig. 3(d). For this RVE, the following structure tensor was identified after calibration with $\mathcal{L}^{\text{pred}} = \mathcal{L}^\sigma$:

$$[\bar{\mathbf{G}}^{\text{pls}}] = \begin{bmatrix} 0.08 & 0. & -0. \\ 0. & 0.07 & 0.02 \\ -0. & 0.02 & 0.85 \end{bmatrix}. \quad (50)$$

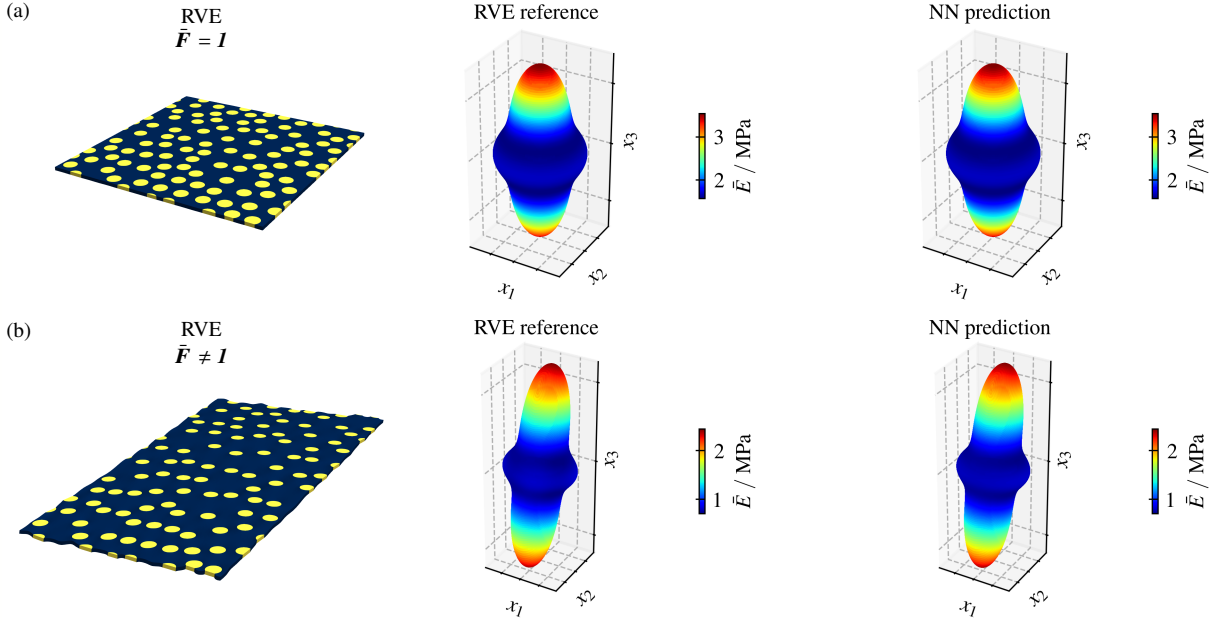


Figure 7: RVE stochastic fibers and corresponding elastic surfaces from the homogenized tangent tensor $\bar{\mathbf{c}}$ as well as the NN prediction from the model $\bar{\psi}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$: (a) undeformed state with $\bar{\mathbf{F}} = \mathbf{I}$ and (b) deformed state with $\bar{\mathbf{F}} \neq \mathbf{I}$ as given in Eq. (49).

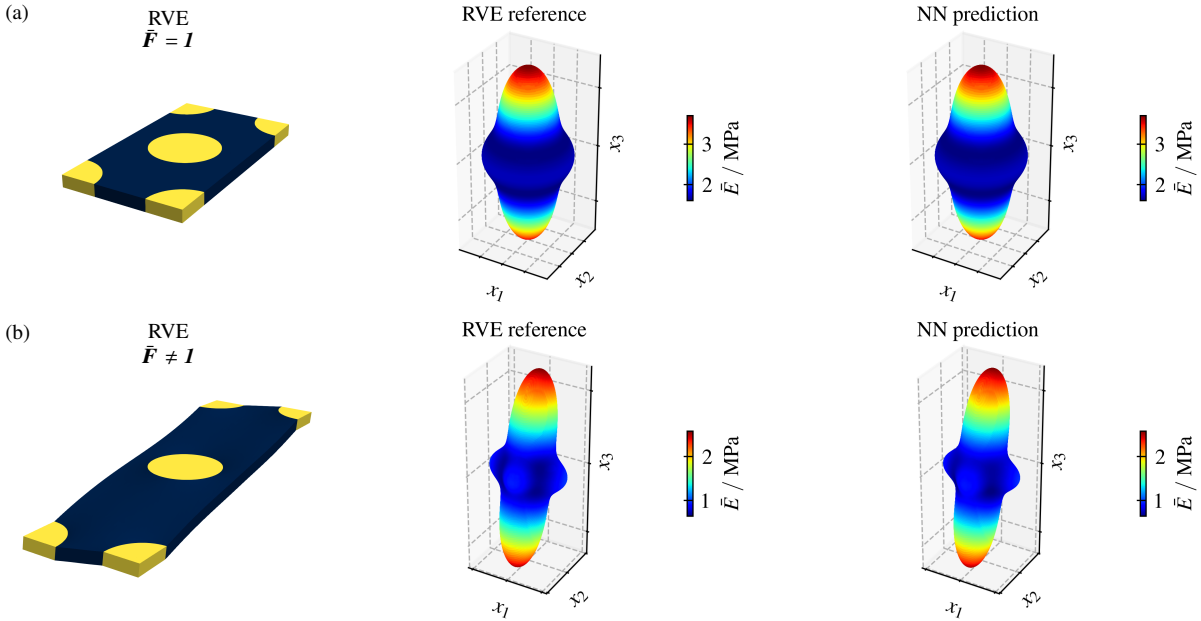


Figure 8: RVE hexagonal fibers and corresponding elastic surfaces from the homogenized tangent tensor $\bar{\mathbf{c}}$ as well as the NN prediction from the model $\bar{\psi}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$: (a) undeformed state with $\bar{\mathbf{F}} = \mathbf{I}$ and (b) deformed state with $\bar{\mathbf{F}} \neq \mathbf{I}$ as given in Eq. (49).

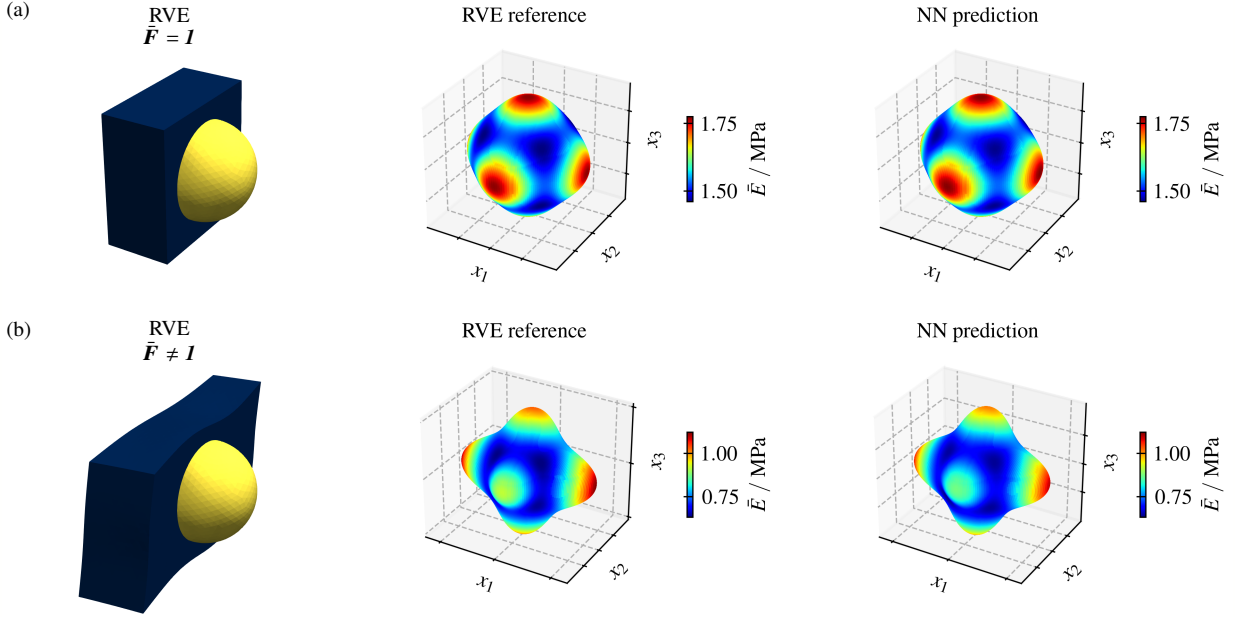


Figure 9: RVE cubic sphere and corresponding elastic surfaces from the homogenized tangent tensor $\bar{\mathbf{c}}$ as well as the NN prediction from the model $\bar{\psi}^{\bar{\mathbf{c}}}(\bar{\mathbf{I}}^{\bar{\mathbf{c}}}, \bar{\mathbf{J}})$: (a) undeformed state with $\bar{\mathbf{F}} = \mathbf{I}$ and (b) deformed state with $\bar{\mathbf{F}} \neq \mathbf{I}$ as given in Eq. (49).

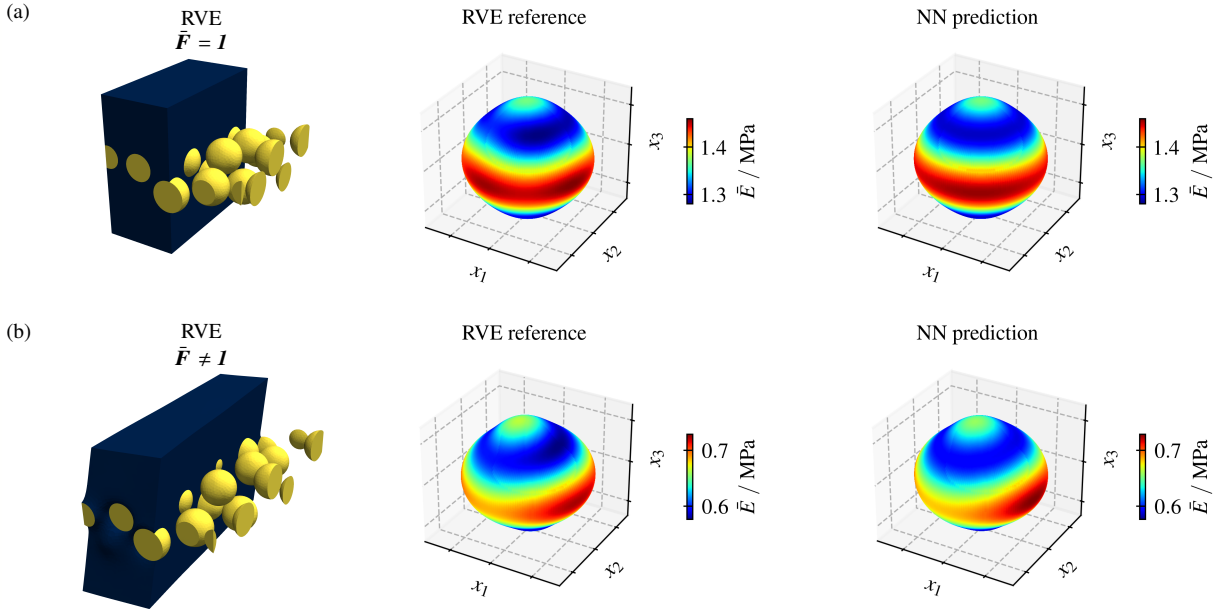


Figure 10: RVE plane spheres and corresponding elastic surfaces from the homogenized tangent tensor $\bar{\mathbf{c}}$ as well as the NN prediction from the model $\bar{\psi}^{\bar{\mathbf{c}}}(\bar{\mathbf{I}}^{\bar{\mathbf{c}}}, \bar{\mathbf{J}})$: (a) undeformed state with $\bar{\mathbf{F}} = \mathbf{I}$ and (b) deformed state with $\bar{\mathbf{F}} \neq \mathbf{I}$ as given in Eq. (49).

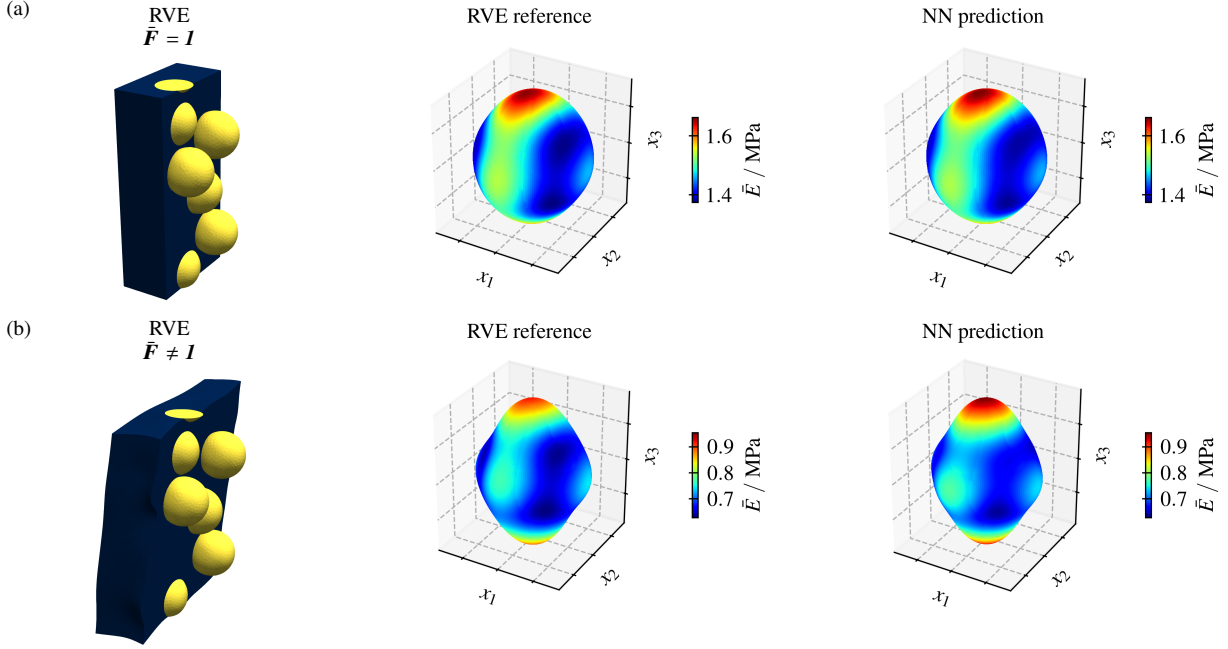


Figure 11: RVE chain spheres and corresponding elastic surfaces from the homogenized tangent tensor $\bar{\mathbb{c}}$ as well as the NN prediction from the model $\bar{\psi}_{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{\mathbf{J}}^{\bar{\mathbf{G}}_1}, \bar{\mathbf{G}}_2, \bar{\mathbf{J}})$: (a) undeformed state with $\bar{\mathbf{F}} = \mathbf{I}$ and (b) deformed state with $\bar{\mathbf{F}} \neq \mathbf{I}$ as given in Eq. (49).

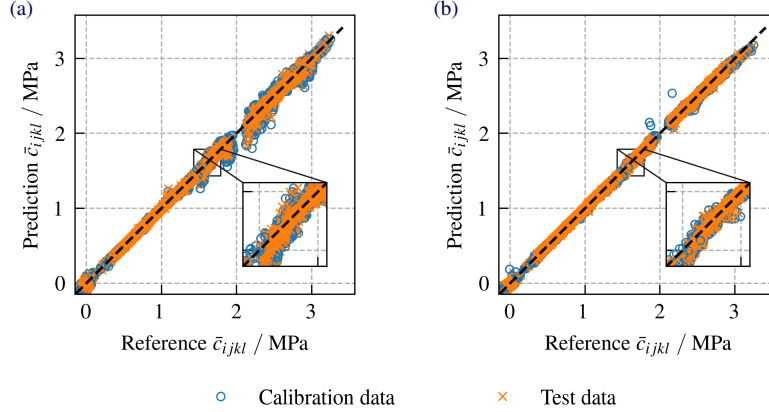


Figure 12: Model predictions of the material tangent $\bar{\mathbb{c}} \in \overline{\mathcal{S}ym}_4$ compared to reference values from the RVE plane spheres for NN models trained with the prediction loss $\mathcal{L}^{\text{pred}} = \mathcal{L}^\sigma$: (a) coordinate-based NN model $\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$ and (b) invariant-based NN model $\bar{\psi}^{\bar{\mathbf{G}}}(\bar{\mathbf{J}}^{\bar{\mathbf{G}}}, \bar{\mathbf{J}})$. The ratio of calibration and test data is 70/30.

Again, as in the former study, the invariant-based NN model was able to identify the X_3 axis as the preferred direction and to find a valid structure tensor. Note that the difference to the identified structure tensor given in Eq. (44) is not problematic, since both structure tensors are equivalent, see Footnote 12. The structure tensor describes *transverse isotropy* \mathcal{G}_{13} since it is characterized by two approximately equal eigenvalues: $(\lambda_1, \lambda_2, \lambda_3)^{\text{pls}} = (0.08, 0.07, 0.85)$. In addition, the vector of active/non-active gates is given by $G(\mathbf{g}^{\text{pls}}) = (1, 1, 1, 1, 0, 0, 1)$, i.e., the additional invariants needed to describe orthotropy are removed from the model by the gates.

Comparing the predictions of the invariant-based model with the coordinate-based model, it can first be observed that similarly good results for energy and stress are achieved with both models, which is not surprising given the selected loss, see Tab. 3 third line. However, as shown in Fig. 12, the accuracy in the material tangent, which was not directly used for training, is noticeably better if the invariant-based NN model is used. This is due to the fact that, if the structure tensor is identified correctly, the model structure ensures the material symmetry condition (5) to be fulfilled and is

Table 3: Interpolation study on the performance of the developed invariant-based NN models $\tilde{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$ and the coordinate-based reference model $\tilde{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$ for the five considered RVEs and the prediction loss $\mathcal{L}^{\text{pred}} = \mathcal{L}^\sigma$. The loss term for training the invariant-based models was $\mathcal{L} = \mathcal{L}^{\text{pred}} + 5 \times 10^{-5} \mathcal{L}^{\text{gate}}$ and an architecture with two hidden layers with 16 neurons each was chosen. The loss term for the coordinate-based model was $\mathcal{L} = \mathcal{L}^{\text{pred}}$ and an architecture with three hidden layers with 16 neurons each was chosen. The overall datasets were divided into calibration and test sets with a ratio of 70/30, respectively. All models were trained 5 times, where the best training run was selected. Given are the loss values after training and the error measures for $\tilde{\psi}$, $\tilde{\sigma}$ and $\tilde{\epsilon}$, cf. Eqs. (37) and (41).

RVE	Model	\square	Active gates	$\mathcal{L}_{\text{cal}}^{\text{pred}}$	$\mathcal{L}_{\text{test}}^{\text{pred}}$	$\epsilon^\psi / \%$	$\epsilon^\sigma / \%$	$\epsilon^\epsilon / \%$
Stochastic fibers	$\tilde{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	4	1.771×10^{-6}	2.593×10^{-6}	0.19	0.41	0.28
	$\tilde{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	6.746×10^{-7}	1.650×10^{-6}	0.08	0.29	1.1
Hexagonal fibers	$\tilde{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	7	1.194×10^{-5}	1.679×10^{-5}	0.38	0.93	0.96
	$\tilde{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	1.728×10^{-6}	3.592×10^{-6}	0.1	0.43	1.39
Cubic sphere	$\tilde{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	8	1.452×10^{-5}	2.427×10^{-5}	0.22	0.76	1.24
	$\tilde{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	6.994×10^{-6}	1.187×10^{-5}	0.16	0.64	2.03
Plane spheres	$\tilde{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	5	1.000×10^{-5}	1.314×10^{-5}	0.41	0.92	0.61
	$\tilde{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	1.002×10^{-6}	2.238×10^{-6}	0.06	0.29	0.97
Chain spheres	$\tilde{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$(\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)$	7	1.274×10^{-5}	1.462×10^{-5}	0.31	0.89	1.25
	$\tilde{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	1.820×10^{-6}	3.056×10^{-6}	0.08	0.35	1.06

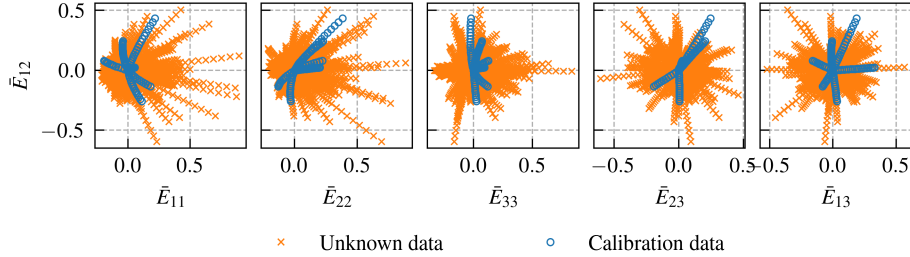


Figure 13: Dataset \mathcal{D} for the extrapolation study with: sampled deformation space visualized in five sectional planes of the Green-Lagrange strain tensor. Only 6 load cases out of \mathcal{D} comprising a total number of 120 tuples are used for calibration.

therefore much more suitable for the specific problem of describing the RVE's effective constitutive behavior. To be more precise, referring to Eq. (9)₁, the stress results from the derivative of the energy with respect to the invariants multiplied with derivatives of the invariants with respect to the deformations, the tensor generators [27]. This is similar for the material tangent, which is obtained by further derivation, cf. Eq. (9)₂.

For the other RVEs considered within this work, similar results are achieved. The final prediction loss values and the error measures defined in Eq. (41) are given in Tab. 3. As can be seen there, the invariant-based NN approach is superior regarding the error ϵ^ϵ in the material tangent over the coordinate-based model for all RVEs except for *chain spheres*.

5.2.3 Extrapolation behavior

After the detailed analysis of the model's ability to learn the effective constitutive behavior of different anisotropic RVEs from a large dataset, the *extrapolation behavior* is now studied. To mimic the situation of sparse data for training, only six load cases with 20 increments each are used. For visualization, the chosen load paths are shown in Fig. 13 within five sectional planes of the coordinates of the Green-Lagrange strain tensor $\bar{\mathbf{E}}$. Accordingly, large parts of the strain space are not covered by the sparse calibration data for the extrapolation study. In order to decide when the selection procedure shown in Fig. 2 terminates, the error measures ϵ^ψ , ϵ^σ and ϵ^ϵ are now calculated using only the calibration data \mathcal{D}_{cal} , cf. Eq. (41).

To compare the predictive capability of the invariant-based NN approach and the coordinate-based reference NN model, we again consider the RVE *plane spheres* exemplarily. Even for the sparse data set used for calibration, a reasonable structure tensor was identified during training:

$$[\bar{\mathbf{G}}^{\text{pls}}] = \begin{bmatrix} 0.5 & -0.01 & -0.01 \\ -0.01 & 0.5 & -0.01 \\ -0.01 & -0.01 & 0. \end{bmatrix}. \quad (51)$$

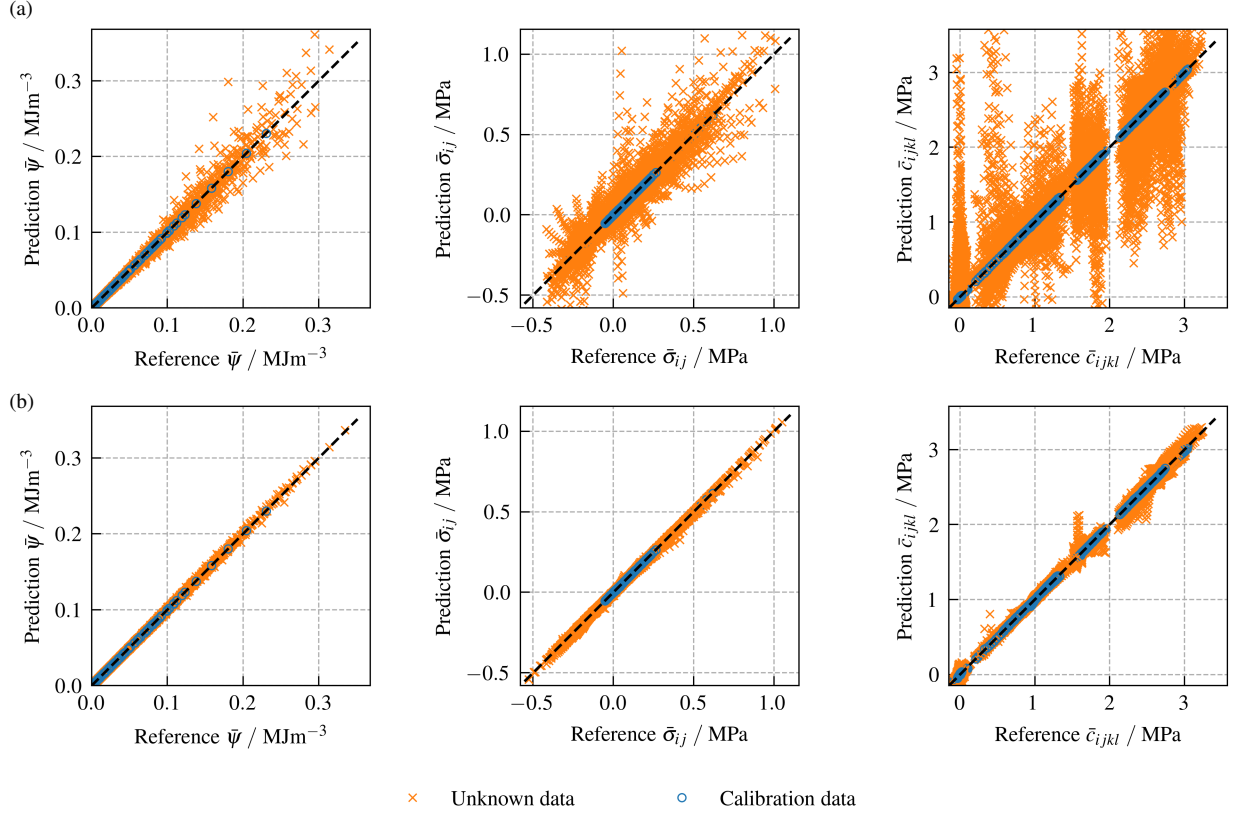


Figure 14: Predictions for the extrapolation behavior of NN-based models for the RVE plane spheres: (a) coordinate-based model $\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$ with three hidden layers including 16 neurons each and (b) invariant-based model $\bar{\psi}^{\mathbf{G}}(\bar{\mathbf{I}}^{\mathbf{G}}, \bar{\mathbf{J}})$ with two hidden layers including 16 neurons each. The prediction loss was $\mathcal{L}^{\text{pred}} = 0.7\mathcal{L}^{\sigma} + 0.3\mathcal{L}^{\text{c}}$ and the loss term for training the invariant-based models was $\mathcal{L} = \mathcal{L}^{\text{pred}} + 5 \times 10^{-5} \mathcal{L}^{\text{gate}}$. Only six load cases were used for calibration. All models were trained 5 times, where the best training run was selected.

Thus, the underlying *transverse isotropy* \mathcal{G}_{13} with the X_3 -axis as preferred direction has been detected. Also the vector of active/non-active gates given by $G(\mathbf{g}^{\text{pls}}) = (1, 1, 1, 0, 0, 1, 1)$ corresponds to this since two reducible invariants have been removed from the model by the gates.

The predictions for energy, stress and material tangent of the coordinate-based model $\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$ and the invariant-based models $\bar{\psi}^{\square}(\bar{\mathbf{I}}^{\square}, \bar{\mathbf{J}})$ trained with the reduced data and the prediction loss $\mathcal{L}^{\text{pred}} := \mathcal{L}^{\sigma} + \mathcal{L}^{\text{c}}$ are given in Fig. 14(a),(b). As shown there, the calibration data have been learned quite well by both models. However, the predictions for the unknown data, which require the model to extrapolate from the calibration domain, are very poor for the coordinate-based model, cf. Fig. 14(a). In contrast, as shown in Fig. 14(b), the extrapolation behavior of the invariant-based approach is significantly improved.

As shown by the loss terms for the test data given in Tab. 4, the invariant-based approach shows a clearly better extrapolation behavior compared to the coordinated-based approach for all five considered RVEs. The test loss is consistently one or even two orders of magnitude smaller. This advantageous behavior can again be attributed to the fact that the material symmetry is built into the model by construction through the use of structure tensors and invariants. At least if a meaningful structure tensor is detected during training with the sparse data set. It should be noted that for the RVEs *hexagonal fibers* and *cubic sphere*, the algorithm given in Fig. 2 did not select 6th and 4th order structure tensors as in the interpolation study, cf. 5.2.1. This is due to the fact that the tolerance in the error measures ϵ^{ψ} , ϵ^{σ} and ϵ^{c} for the calibration data was not undershot. Nevertheless, although the expected structural tensor set could not be detected for these two cases and a set comprising two 2nd order structure tensors was chosen instead, the extrapolation behavior is still noticeably better than with the coordinate-based approach.

Table 4: Extrapolation study on the performance of the developed invariant-based NN models $\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$ and the coordinate-based reference model $\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$ for the five considered RVEs and the prediction loss $\mathcal{L}^{\text{pred}} = 0.7\mathcal{L}^\sigma + 0.3\mathcal{L}^\epsilon$. The loss term for training the invariant-based models was $\mathcal{L} = \mathcal{L}^{\text{pred}} + 5 \times 10^{-5} \mathcal{L}^{\text{gate}}$ and an architecture with two hidden layers with 16 neurons each was chosen. The loss term for the coordinate-based model was $\mathcal{L} = \mathcal{L}^{\text{pred}}$ and an architecture with three hidden layers with 16 neurons each was chosen. Only six load cases were used for calibration. All models were trained 5 times, where the best training run was selected. Given are the prediction loss values after training, cf. Eqs. (37) and (38).

RVE	Model	\square	Active gates	$\mathcal{L}_{\text{cal}}^{\text{pred}}$	$\mathcal{L}_{\text{test}}^{\text{pred}}$
Stochastic fibers	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	5	7.759×10^{-6}	1.141×10^{-4}
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	2.019×10^{-7}	1.117×10^{-2}
Hexagonal fibers	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$(\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)$	8	2.115×10^{-6}	1.003×10^{-3}
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	3.740×10^{-7}	1.522×10^{-2}
Cubic sphere	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$(\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)$	6	2.214×10^{-5}	3.329×10^{-3}
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	5.511×10^{-7}	1.063×10^{-2}
Plane spheres	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$\bar{\mathbf{G}}$	5	1.195×10^{-5}	1.248×10^{-4}
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	2.904×10^{-7}	1.590×10^{-2}
Chain spheres	$\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{\mathbf{J}})$	$(\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)$	7	9.537×10^{-6}	4.285×10^{-4}
	$\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{\mathbf{J}})$	–	–	3.246×10^{-7}	1.260×10^{-2}

6 Conclusions

In the present work, an NN-based approach for the automated modeling of anisotropic finite strain elasticity including detection of anisotropy type and orientation is proposed. To this end, an invariant-based approach is formulated in such a way that important physical conditions are fulfilled a priori, i.e., by construction, whereby these models are denoted as PANNs [23, 24, 46]. The invariants are built from the deformation tensor and generalized structure tensor(s), where 2nd, 4th and 6th order structure tensors are used to cover a wide range of anisotropies. Thus, special attention is paid to the principle of material symmetry. The ability of our approach is investigated and compared to an NN model based on the coordinates of the deformation tensor by a calibration to data generated via computational homogenization of five different RVEs.

The article begins with a brief review of the kinematics of finite strains and stress measures in continuum solid mechanics, a summary of common principles in hyperelasticity, the concept of structure tensors and a scale transition scheme. In the following section, the generalized structure tensor approach [67] is introduced, where 2nd, 4th and 6th order tensors are given. Based on this theoretical foundation, PANNs are formulated that are based on parameterized versions of the generalized structure tensors, using a network with an additional gate layer and a penalty loss of p -norm type [10] to enforce sparsity with respect to the number of invariants involved in the model. Furthermore, a strategy is presented to decide which set of structure tensor(s) is required. Finally, the invariant-based NN models are used as macroscopic surrogates for computational homogenizations in an examples section, where a comparison with a model based on the coordinates of the right Cauchy-Green deformation tensor is included. A detailed analysis of the interpolation and extrapolation capability of the models was carried out with the generated data sets. It turned out that both approaches, the invariant-based one and the coordinate-based one, deliver very good results for the interpolation case. However, the presented invariant-based approach requires fewer hidden layers for an equivalent prediction quality. This applies to the energy, the stress and the material tangent tensor. In terms of extrapolation behavior, the presented invariant-based approach clearly beats the coordinate-based model. Even with a sparse data set, the prediction quality for states not included in the training remains acceptable, whereas the coordinate-based model exhibits huge errors in the extrapolation regime.

In summary, the presented PANN based on invariants formulated with generalized structural tensors is a very accurate surrogate model for computationally expensive RVE simulations in finite strain elasticity. The a priori inclusion of principles from constitutive modeling in the NN model, in particular the *principle of material symmetry*, ensures that the underlying physics is not violated even during extrapolation, thus guaranteeing good generalization. This also enables comparatively small network architectures. The use of ℓ_p regularization enables the elimination of unneeded invariants from the model.

Various applications and extensions of our approach are planned for the future. For example, an additional sparsification of the network as done in [68, 95] is possible. Furthermore, in order to exploit the advantages of the developed invariant-based PANN approach, an integration into multiscale schemes as FE^{ANN} [25] and the application to real experimental data are planned. In addition, an extension of our approach to further anisotropy classes is possible,

which requires the use of sets of structure tensors. Finally, the application of the developed concepts in the NN-based modeling of inelastic material behavior [39, 45, 58, 69, 96, 97, 98, 99], damage [100, 101] and coupled problems [46, 75, 102, 103] is possible.

Acknowledgment

All presented computations were performed on a PC-Cluster at the Center for Information Services and High Performance Computing (ZIH) at TU Dresden. The authors thus thank the ZIH for generous allocations of computer time. This work was supported by a postdoc fellowship of the German Academic Exchange Service (DAAD) to Karl A. Kalina. This support is gratefully acknowledged. Finally, the authors thank the German Research Foundation (DFG) for the support within the Research Training Group GRK 2868 D³–Project Number 493401063.

CRedit authorship contribution statement

Karl A. Kalina: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition. **Jörg Brummund:** Formal analysis, Writing – review & editing. **WaiChing Sun:** Conceptualization, Resources, Writing – review & editing. **Markus Kästner:** Resources, Writing – review & editing, Funding acquisition.

A Structure tensors and invariant sets for specific symmetry groups

Within this appended section, we summarize and discuss the structure tensors and invariants sets for specific symmetry groups. As the numbering of the groups is not uniform in the literature, we use the Schoenflies notation in addition to the group number. In the following, structure tensors and invariants are given for: *isotropy* \mathcal{G}_{14} ($\mathcal{K}_h = O(3)$), *triclinic anisotropy* \mathcal{G}_1 (C_i), *monoclinic anisotropy* \mathcal{G}_2 (C_{2h}), *orthotropy* \mathcal{G}_3 (\mathcal{D}_{2h}), *tetragonal anisotropy* \mathcal{G}_5 (\mathcal{D}_{4h}), *cubic anisotropy* \mathcal{G}_7 (O_h), *hexagonal anisotropy* \mathcal{G}_{11} (\mathcal{D}_{6h}), as well as *transverse isotropy* \mathcal{G}_{13} ($\mathcal{D}_{\infty h}$).

For the following discussions, we make use of the spectral decomposition of a symmetric 2nd order tensor

$$\mathbf{S} = \sum_{\alpha=1}^n \lambda_{\alpha} \mathbf{P}_{\alpha} \text{ with } \mathbf{P}_{\alpha} \cdot \mathbf{P}_{\beta} = \delta_{\alpha(\beta)} \mathbf{P}_{\beta} \text{ and } \sum_{\alpha=1}^n \mathbf{P}_{\alpha} = \mathbf{I} \quad (52)$$

with $n \in \{1, 2, 3\}$ non-equal eigenvalues. The projection tensors $\mathbf{P}_{\alpha} \in \mathcal{S}ym$ related to \mathbf{S} can be represented by Sylvester's formula [76]

$$\mathbf{P}_{\alpha} = \delta_{1n} \mathbf{I} + \prod_{\substack{\beta=1 \\ \beta \neq \alpha}}^n \frac{\mathbf{S} - \lambda_{\beta} \mathbf{I}}{\lambda_{\alpha} - \lambda_{\beta}}. \quad (53)$$

Furthermore, we use the Cayley-Hamilton theorem [76], which states that a 2nd order tensor fulfills its own eigenvalue equation, i.e.,

$$\mathbf{S}^3 - I_1 \mathbf{S}^2 + I_2 \mathbf{S} - I_3 \mathbf{I} = \mathbf{0} \text{ with } I_1 = \text{tr } \mathbf{S}, \quad I_2 = \text{tr}(\text{cof } \mathbf{S}), \quad I_3 = \det \mathbf{S}, \quad (54)$$

where I_1, I_2, I_3 are the principal invariants.

A.1 Isotropy \mathcal{G}_{14}

By applying the Cayley-Hamilton theorem (54), one can find the well known relation that the invariant set $\bar{J}_1 := \text{tr } \bar{\mathbf{C}}$, $\bar{J}_2 := \frac{1}{2} \text{tr } \bar{\mathbf{C}}^2$, $\bar{J}_3 := \frac{1}{3} \text{tr } \bar{\mathbf{C}}^3$ can be expressed by $\bar{I}_1 := \text{tr } \bar{\mathbf{C}}$, $\bar{I}_2 := \text{tr}(\text{cof } \bar{\mathbf{C}})$, $\bar{I}_3 := \det \bar{\mathbf{C}}$ as

$$\bar{J}_1 = \bar{I}_1, \quad \bar{J}_2 = \frac{1}{2} \bar{I}_1^2 - \bar{I}_2, \quad \bar{J}_3 = \frac{1}{3} \bar{I}_1^3 - \bar{I}_1 \bar{I}_2 + \bar{I}_3. \quad (55)$$

It is thus equivalent to use the sets $\bar{\mathbf{I}}^{\text{iso}} = (\bar{I}_1, \bar{I}_2, \bar{I}_3) \in \mathbb{R}^3$ and $\bar{\mathbf{I}}^{\text{iso},*} = (\bar{J}_1, \bar{J}_2, \bar{J}_3) \in \mathbb{R}^3$ for *isotropy* \mathcal{G}_{14} .

A.2 Triclinic anisotropy \mathcal{G}_1

The *triclinic anisotropy* \mathcal{G}_1 can be modeled with two symmetric 2nd order structure tensors $\bar{\mathbf{A}}_{\text{tri}}, \bar{\mathbf{B}}_{\text{tri}} \in \mathcal{Sym}$, cf. Olive et al. [87]. For these tensors, the following conditions must hold:

$$w_L = e_{LMN} A_{MP}^{\text{tri}} B_{PN}^{\text{tri}} \neq 0 \wedge [(\bar{\mathbf{A}}_{\text{tri}} \cdot \mathbf{w}) \times \mathbf{w} \neq \mathbf{0} \vee (\bar{\mathbf{B}}_{\text{tri}} \cdot \mathbf{w}) \times \mathbf{w} \neq \mathbf{0}] . \quad (56)$$

The statement of the above condition is that the eigensystems of the tensors $\bar{\mathbf{A}}_{\text{tri}}, \bar{\mathbf{B}}_{\text{tri}}$ are completely rotated to each other, i.e., there is no eigenvector of $\bar{\mathbf{A}}_{\text{tri}}$ which is (anti)parallel to an eigenvector of $\bar{\mathbf{B}}_{\text{tri}}$. Note that it is only possible to fulfill condition (56) if both structure tensors have three non-equal eigenvalues each.

According to Boehler [81], we get 9 additional invariants given by

$$\bar{P}_4 = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{A}}_{\text{tri}}), \bar{P}_5 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{A}}_{\text{tri}}), \bar{P}_6 = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{A}}_{\text{tri}}^2), \bar{P}_7 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{A}}_{\text{tri}}^2), \bar{P}_8 = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{B}}_{\text{tri}}), \quad (57)$$

$$\bar{P}_9 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{B}}_{\text{tri}}), \bar{P}_{10} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{B}}_{\text{tri}}^2), \bar{P}_{11} = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{B}}_{\text{tri}}^2), \bar{P}_{12} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{A}}_{\text{tri}} \cdot \bar{\mathbf{B}}_{\text{tri}}). \quad (58)$$

Thus, we end up with the following set $\bar{\mathbf{I}}^{\text{tri}} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{P}_4, \bar{P}_5, \dots, \bar{P}_{12}) \in \mathbb{R}^{12}$. Note that the given set is complete but might be irreducible. It is also possible to model *triclinic* anisotropy with two skew symmetric 2nd order tensors [65, 66].

A.3 Monoclinic anisotropy \mathcal{G}_2

The *monoclinic anisotropy* \mathcal{G}_2 can also be modeled with a set of two symmetric 2nd order structure tensors $\bar{\mathbf{A}}_{\text{mon}}, \bar{\mathbf{B}}_{\text{mon}} \in \mathcal{Sym}$ [87]. For these tensors, the following conditions must hold:

$$w_L = e_{LMN} A_{MP}^{\text{mon}} B_{PN}^{\text{mon}} \neq 0 \wedge [(\bar{\mathbf{A}}_{\text{mon}} \cdot \mathbf{w}) \times \mathbf{w} = \mathbf{0} \wedge (\bar{\mathbf{B}}_{\text{mon}} \cdot \mathbf{w}) \times \mathbf{w} = \mathbf{0}] . \quad (59)$$

The statement of the above condition is that the tensors $\bar{\mathbf{A}}_{\text{mon}}, \bar{\mathbf{B}}_{\text{mon}}$ share one and only one eigenvector, i.e., it is parallel or antiparallel.

According to Boehler [81], we get 9 additional invariants given by

$$\bar{Q}_4 = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{A}}_{\text{mon}}), \bar{Q}_5 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{A}}_{\text{mon}}), \bar{Q}_6 = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{A}}_{\text{mon}}^2), \bar{Q}_7 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{A}}_{\text{mon}}^2), \bar{Q}_8 = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{B}}_{\text{mon}}), \quad (60)$$

$$\bar{Q}_9 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{B}}_{\text{mon}}), \bar{Q}_{10} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{B}}_{\text{mon}}^2), \bar{Q}_{11} = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{B}}_{\text{mon}}^2), \bar{Q}_{12} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{A}}_{\text{mon}} \cdot \bar{\mathbf{B}}_{\text{mon}}). \quad (61)$$

Thus, we end up with the following set $\bar{\mathbf{I}}^{\text{mon}} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{Q}_4, \bar{Q}_5, \dots, \bar{Q}_{12}) \in \mathbb{R}^{12}$. Note that the given set is complete but might be irreducible. It is also possible to model *monoclinic* anisotropy with one symmetric and one skew symmetric 2nd order tensor [65, 66].

A.4 orthotropy \mathcal{G}_3

Typically, *orthotropy* \mathcal{G}_3 is modeled with two structure tensors $\bar{\mathbf{G}}_1 := \mathbf{a}_1 \otimes \mathbf{a}_1 \in \mathcal{Sym}$ and $\bar{\mathbf{G}}_2 := \mathbf{a}_2 \otimes \mathbf{a}_2 \in \mathcal{Sym}$ with $\mathbf{a}_\alpha \cdot \mathbf{a}_\beta = \delta_{\alpha\beta}$. Applying the rules from Boehler [81] and accounting for the orthogonality of \mathbf{a}_1 and \mathbf{a}_2 as well as the fact that both tensors have two non-equal eigenvalues, respectively, we can build four additional invariants from $\bar{\mathbf{C}}, \bar{\mathbf{G}}_1$ and $\bar{\mathbf{G}}_2$:

$$\bar{S}_4 = \text{tr}(\bar{\mathbf{G}}_1 \cdot \bar{\mathbf{C}}), \bar{S}_5 = \text{tr}(\bar{\mathbf{G}}_1 \cdot \bar{\mathbf{C}}^2), \bar{S}_6 = \text{tr}(\bar{\mathbf{G}}_2 \cdot \bar{\mathbf{C}}), \bar{S}_7 = \text{tr}(\bar{\mathbf{G}}_1 \cdot \bar{\mathbf{C}}^2). \quad (62)$$

As we will show in the following, it is also possible to build an equivalent invariant set by only using a single 2nd order structure tensor $\bar{\mathbf{G}}_{\text{orth}} \in \mathcal{Sym}$ with three different eigenvalues $\lambda_1 \neq \lambda_2 \neq \lambda_3$. According to Boehler [81], we can build the following four additional invariants from $\bar{\mathbf{C}}$ and $\bar{\mathbf{G}}_{\text{orth}}$:

$$\bar{T}_4 = \text{tr}(\bar{\mathbf{G}}_{\text{orth}} \cdot \bar{\mathbf{C}}), \bar{T}_5 = \text{tr}(\bar{\mathbf{G}}_{\text{orth}} \cdot \bar{\mathbf{C}}^2), \bar{T}_6 = \text{tr}(\bar{\mathbf{G}}_{\text{orth}}^2 \cdot \bar{\mathbf{C}}), \bar{T}_7 = \text{tr}(\bar{\mathbf{G}}_{\text{orth}}^2 \cdot \bar{\mathbf{C}}^2). \quad (63)$$

By applying Sylvester's formula (53), we get

$$\mathbf{P}_1 = \mathbf{n}_1 \otimes \mathbf{n}_1 = \frac{1}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)} \left(\bar{\mathbf{G}}_{\text{orth}}^2 - (\lambda_2 + \lambda_3) \bar{\mathbf{G}}_{\text{orth}} + \lambda_2 \lambda_3 \mathbf{I} \right) \text{ and} \quad (64)$$

$$\mathbf{P}_2 = \mathbf{n}_2 \otimes \mathbf{n}_2 = \frac{1}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3)} \left(\bar{\mathbf{G}}_{\text{orth}}^2 - (\lambda_1 + \lambda_3) \bar{\mathbf{G}}_{\text{orth}} + \lambda_1 \lambda_3 \mathbf{I} \right). \quad (65)$$

If $\bar{\mathbf{G}}_{\text{orth}}$ is chosen in such a way that $\mathbf{n}_1 = \mathbf{a}_1$ and $\mathbf{n}_2 = \mathbf{a}_2$, we find

$$\bar{S}_4 = \frac{1}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)} (\bar{T}_6 - (\lambda_2 + \lambda_3)\bar{T}_4 + \lambda_2\lambda_3\bar{T}_1), \quad (66)$$

$$\bar{S}_5 = \frac{1}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)} (\bar{T}_7 - (\lambda_2 + \lambda_3)\bar{T}_5 + \lambda_2\lambda_3(\bar{T}_1^2 - 2\bar{T}_2)), \quad (67)$$

$$\bar{S}_6 = \frac{1}{(\lambda_2 - \lambda_2)(\lambda_2 - \lambda_3)} (\bar{T}_6 - (\lambda_1 + \lambda_3)\bar{T}_4 + \lambda_1\lambda_3\bar{T}_1), \quad (68)$$

$$\bar{S}_7 = \frac{1}{(\lambda_2 - \lambda_2)(\lambda_2 - \lambda_3)} (\bar{T}_7 - (\lambda_1 + \lambda_3)\bar{T}_5 + \lambda_1\lambda_3(\bar{T}_1^2 - 2\bar{T}_2)). \quad (69)$$

Thus, it is equivalent to use the set $\bar{\mathbf{I}}^{\text{orth}} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{S}_4, \bar{S}_5, \bar{S}_6, \bar{S}_7) \in \mathbb{R}^7$ and $\bar{\mathbf{I}}^{\text{orth},*} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{T}_4, \bar{T}_5, \bar{T}_6, \bar{T}_7) \in \mathbb{R}^7$.

A.5 tetragonal anisotropy \mathcal{G}_5

Following [66], *tetragonal anisotropy* \mathcal{G}_5 can be modeled with the structure tensors

$$\bar{\mathbb{D}}_{\text{tet}} := \sum_{\alpha=1}^2 \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \text{ with } \mathbf{A}_\alpha \cdot \mathbf{A}_\beta = \delta_{\alpha\beta} \text{ and } \bar{\mathbf{M}}_{\text{tet}} := \mathbf{N} \otimes \mathbf{N}, \mathbf{N} \in \mathcal{N}, \mathbf{N} \cdot \mathbf{A}_\alpha = 0. \quad (70)$$

A complete invariant set can be formed from $\bar{\mathbf{C}}, \bar{\mathbf{H}}_1 := \bar{\mathbb{D}}_{\text{tet}} : \bar{\mathbf{C}}, \bar{\mathbf{H}}_2 := \bar{\mathbb{D}}_{\text{tet}} : \bar{\mathbf{C}}^2$ and $\bar{\mathbf{M}}_{\text{tet}}$. According to Boehler [81], one can build a set comprising 30 invariants from these four 2nd order tensors. By exploiting the orthogonality of the three vectors $\mathbf{A}_1, \mathbf{A}_2$ and \mathbf{N} , we find $\bar{\mathbb{D}}_{\text{tet}} : \bar{\mathbb{D}}_{\text{tet}} = \bar{\mathbb{D}}_{\text{tet}}$. Furthermore, by using $\mathbf{A}_\alpha = \mathbf{Q} \cdot \mathbf{e}_\alpha$ and $\mathbf{N} = \mathbf{Q} \cdot \mathbf{e}_3$, with $\mathbf{Q} \in \mathcal{SO}(3)$ and \mathbf{e}_α denoting the Cartesian basis vectors, we find that $Q_{I\alpha}Q_{J\beta}\bar{C}_{IJ} = \bar{C}_{\alpha\beta}^*$, where $\bar{C}_{\alpha\beta}^*$ are the in-plane coordinates of $\bar{\mathbf{C}}$ with respect to $\mathbf{A}_1, \mathbf{A}_2$ and $Q_{I3}Q_{J3}\bar{C}_{IJ} = \bar{C}_{33}^*$ is the out-of-plane coordinate in the direction of \mathbf{N} . From these relations we can find that $\text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{M}}_{\text{tet}}) = \text{tr}(\bar{\mathbf{H}}_1^2 \cdot \bar{\mathbf{M}}_{\text{tet}}) = \text{tr}(\bar{\mathbf{H}}_2 \cdot \bar{\mathbf{M}}_{\text{tet}}) = \text{tr}(\bar{\mathbf{H}}_2^2 \cdot \bar{\mathbf{M}}_{\text{tet}}) = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_1 \cdot \bar{\mathbf{M}}_{\text{tet}}) = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_2 \cdot \bar{\mathbf{M}}_{\text{tet}}) = \text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{H}}_2 \cdot \bar{\mathbf{M}}_{\text{tet}}) = 0$ and $\text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{M}}_{\text{tet}}) = \text{tr} \bar{\mathbf{C}} - \text{tr} \bar{\mathbf{H}}_1$ and $\text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{M}}_{\text{tet}}) = \text{tr} \bar{\mathbf{C}}^2 - \text{tr} \bar{\mathbf{H}}_2$ can be expressed by $\bar{\mathbf{C}}, \bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$. Thus, $\bar{\mathbf{M}}_{\text{tet}}$ can be removed from the list of structure tensors for the tetragonal symmetry group \mathcal{G}_5 . By accounting for the relations given above and by neglecting redundant expressions, e.g., $\text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_1) = \text{tr} \bar{\mathbf{H}}_1^2$, we end up with the following set $\bar{\mathbf{I}}^{\text{tet}} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{U}_4, \bar{U}_5, \dots, \bar{U}_{13}) \in \mathbb{R}^{13}$, where

$$\begin{aligned} \bar{U}_4 &:= \text{tr} \bar{\mathbf{H}}_1, \bar{U}_5 := \text{tr} \bar{\mathbf{H}}_1^2, \bar{U}_6 := \text{tr} \bar{\mathbf{H}}_1^3, \bar{U}_7 := \text{tr} \bar{\mathbf{H}}_2, \bar{U}_8 := \text{tr} \bar{\mathbf{H}}_2^2, \bar{U}_9 := \text{tr} \bar{\mathbf{H}}_2^3, \\ \bar{U}_{10} &:= \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{H}}_1), \bar{U}_{11} := \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{H}}_1^2), \bar{U}_{12} := \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_2^2), \bar{U}_{13} := \text{tr}(\bar{\mathbf{H}}_1^2 \cdot \bar{\mathbf{H}}_2^2). \end{aligned} \quad (71)$$

Note that the given invariant set might be reducible. We do not proof this here. Also note that the 4th order structure tensor

$$\bar{\mathbb{D}}_{\text{tet}}^* := \sum_{\alpha=1}^3 \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \text{ with } \mathbf{A}_\alpha \cdot \mathbf{A}_\beta = 0, \alpha \neq \beta, |\mathbf{A}_1| = |\mathbf{A}_2| \neq |\mathbf{A}_3| \quad (72)$$

build from three orthogonal vectors \mathbf{A}_α is equivalent to (70).

A.6 cubic anisotropy \mathcal{G}_7

Following [53, 65, 66], *cubic anisotropy* \mathcal{G}_7 can be modeled with the structure tensor

$$\bar{\mathbb{G}}_{\text{cub}} := \sum_{\alpha=1}^3 \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \text{ with } \mathbf{A}_\alpha \cdot \mathbf{A}_\beta = \delta_{\alpha\beta}. \quad (73)$$

A complete invariant set can be formed from $\bar{\mathbf{C}}, \bar{\mathbf{H}}_1 := \bar{\mathbb{G}}_{\text{cub}} : \bar{\mathbf{C}}$ and $\bar{\mathbf{H}}_2 := \bar{\mathbb{G}}_{\text{cub}} : \bar{\mathbf{C}}^2$, cf Xiao [65]. According to Boehler [81], one can build a set comprising 21 invariants from the three 2nd order tensors $\bar{\mathbf{C}}, \bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$. We can reduce this set by exploiting the orthogonality of the vectors \mathbf{A}_α , i.e., it holds $\bar{\mathbb{G}}_{\text{cub}} : \bar{\mathbb{G}}_{\text{cub}} = \bar{\mathbb{G}}_{\text{cub}}$. Furthermore, by using $\mathbf{A}_\alpha = \mathbf{Q} \cdot \mathbf{e}_\alpha$, with $\mathbf{Q} \in \mathcal{SO}(3)$ and \mathbf{e}_α denoting the Cartesian basis vectors, we find that $Q_{I\alpha}Q_{J\beta}\bar{C}_{IJ} = \bar{C}_{\alpha\beta}^*$, where $\bar{C}_{\alpha\beta}^*$ are the coordinates of the right Cauchy-Green deformation tensor with respect to the coordinate system

formed by \mathbf{A}_α , $\alpha \in \{1, 2, 3\}$. By applying these relations and neglecting redundant expressions, we find the set $\bar{\mathcal{I}}^{\text{cub}} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{V}_4, \bar{V}_5, \dots, \bar{V}_{11}) \in \mathbb{R}^{11}$ with

$$\begin{aligned} \bar{V}_4 &= \text{tr} \bar{\mathbf{H}}_1^2, \bar{V}_5 = \text{tr} \bar{\mathbf{H}}_1^3, \bar{V}_6 = \text{tr} \bar{\mathbf{H}}_2^2, \bar{V}_7 = \text{tr} \bar{\mathbf{H}}_2^3, \bar{V}_8 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{H}}_1), \\ \bar{V}_9 &= \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{H}}_1^2), \bar{V}_{10} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_2^2), \bar{V}_{11} = \text{tr}(\bar{\mathbf{H}}_1^2 \cdot \bar{\mathbf{H}}_2^2). \end{aligned} \quad (74)$$

Note that the given invariant set might be reducible. We do not proof this at this point.

Remark 5. Note that the invariant set belonging to the cubic group can be expressed by invariants from the tetragonal group \mathcal{G}_5 up to order $\bar{\mathbf{C}}^3$:

$$\bar{V}_4 = \bar{U}_5 + (\bar{I}_1 - \bar{U}_4)^2, \bar{V}_5 = \bar{U}_6 + (\bar{I}_1 - \bar{U}_4)^3, \bar{V}_8 = \bar{U}_{10} + (\bar{I}_1 - \bar{U}_4)(1/2\bar{I}_1^2 - \bar{I}_2 - \bar{U}_7). \quad (75)$$

A.7 hexagonal anisotropy \mathcal{G}_{11}

According to [65, 66], the *hexagonal anisotropy* group \mathcal{G}_{11} can be modeled with the structure tensors

$$\bar{\mathbf{G}}_{\text{hex}} := \sum_{\alpha=1}^3 \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \otimes \mathbf{A}_\alpha \text{ and } \bar{\mathbf{M}}_{\text{hex}} := \mathbf{N} \otimes \mathbf{N}, \quad (76)$$

where $\mathbf{A}_\alpha \cdot \mathbf{A}_\beta = \pm \frac{1}{2}, \alpha \neq \beta$, $\mathbf{N} \cdot \mathbf{A}_\alpha = 0$ and $|\mathbf{A}_1| = |\mathbf{A}_2| = |\mathbf{A}_3| = |\mathbf{N}| = 1$. By building the 2nd order tensors $\bar{\mathbf{H}}_1 := \bar{\mathbf{C}} : \bar{\mathbf{G}} : \bar{\mathbf{C}}$ and $\bar{\mathbf{H}}_2 := \bar{\mathbf{C}}^2 : \bar{\mathbf{G}} : \bar{\mathbf{C}}^2$, cf. [65], one can build a set comprising 30 invariants. However, similar to the tetragonal symmetry discussed in A.5, the following invariants are equal to zero: $\text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{M}}_{\text{hex}}) = \text{tr}(\bar{\mathbf{H}}_1^2 \cdot \bar{\mathbf{M}}_{\text{hex}}) = \text{tr}(\bar{\mathbf{H}}_2 \cdot \bar{\mathbf{M}}_{\text{hex}}) = \text{tr}(\bar{\mathbf{H}}_2^2 \cdot \bar{\mathbf{M}}_{\text{hex}}) = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_1 \cdot \bar{\mathbf{M}}_{\text{hex}}) = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_2 \cdot \bar{\mathbf{M}}_{\text{hex}}) = \text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{H}}_2 \cdot \bar{\mathbf{M}}_{\text{hex}}) = 0$. By neglecting redundant expressions, we find the set $\bar{\mathcal{I}}^{\text{hex}} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{W}_4, \bar{W}_5, \dots, \bar{W}_{24}) \in \mathbb{R}^{24}$ with

$$\begin{aligned} \bar{W}_4 &= \text{tr} \bar{\mathbf{H}}_1, \bar{W}_5 = \text{tr} \bar{\mathbf{H}}_1^2, \bar{W}_6 = \text{tr} \bar{\mathbf{H}}_1^3, \bar{W}_7 = \text{tr} \bar{\mathbf{H}}_2, \bar{W}_8 = \text{tr} \bar{\mathbf{H}}_2^2, \bar{W}_9 = \text{tr} \bar{\mathbf{H}}_2^3, \bar{W}_{10} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{M}}_{\text{hex}}), \\ \bar{W}_{11} &= \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{M}}_{\text{hex}}), \bar{W}_{12} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_1), \bar{W}_{13} = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{H}}_1), \bar{W}_{14} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_1^2), \bar{W}_{15} = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{H}}_1^2), \\ \bar{W}_{16} &= \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_2), \bar{W}_{17} = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{H}}_2), \bar{W}_{18} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_2^2), \bar{W}_{19} = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{H}}_2^2), \bar{W}_{20} = \text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{H}}_2), \\ \bar{W}_{21} &= \text{tr}(\bar{\mathbf{H}}_1^2 \cdot \bar{\mathbf{H}}_2), \bar{W}_{22} = \text{tr}(\bar{\mathbf{H}}_1 \cdot \bar{\mathbf{H}}_2^2), \bar{W}_{23} = \text{tr}(\bar{\mathbf{H}}_1^2 \cdot \bar{\mathbf{H}}_2^2), \bar{W}_{24} = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{H}}_1 \cdot \bar{\mathbf{H}}_2). \end{aligned} \quad (77)$$

Note again that the given invariant set for the hexagonal anisotropy group \mathcal{G}_{11} might be reducible. We do not proof this at this point.

Alternatively, if in addition to $\bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$, the tensors $\bar{\mathbf{H}}_3 := \mathbf{I} : \bar{\mathbf{G}} : \bar{\mathbf{C}}$ and $\bar{\mathbf{H}}_4 := \mathbf{I} : \bar{\mathbf{G}} : \bar{\mathbf{C}}^2$ are used to build an invariant set for the hexagonal anisotropy group \mathcal{G}_{11} , we can find that the invariants $\text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{M}}_{\text{hex}}) = \text{tr} \bar{\mathbf{C}} - \frac{2}{3} \text{tr} \bar{\mathbf{H}}_3$ and $\text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{M}}_{\text{hex}}) = \text{tr} \bar{\mathbf{C}}^2 - \frac{2}{3} \text{tr} \bar{\mathbf{H}}_4$ can be expressed by $\bar{\mathbf{C}}$ and $\bar{\mathbf{G}}_{\text{hex}}$. Thus, $\bar{\mathbf{M}}_{\text{hex}}$ is not needed anymore in this case.

A.8 Transverse isotropy \mathcal{G}_{13}

In the case of *transverse isotropy* \mathcal{G}_{13} , the structure tensor is given by $\bar{\mathbf{G}}_{\text{ti}} := a_1 \mathbf{a}_1 \otimes \mathbf{a}_1 + a_2 (\mathbf{I} - \mathbf{a}_1 \otimes \mathbf{a}_1) \in \mathcal{S}ym$ with $|\mathbf{a}_1| = 1$, $a_1, a_2 \in \mathbb{R}_{\geq 0}$ for positive semi-definiteness, $a_1 \neq a_2$, where a_2 is chosen to zero typically. Thus, according to Boehler [81], we get four additional invariants given by

$$\bar{R}_4 = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{G}}_{\text{ti}}), \bar{R}_5 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{G}}_{\text{ti}}), \bar{R}_6 = \text{tr}(\bar{\mathbf{C}} \cdot \bar{\mathbf{G}}_{\text{ti}}^2), \bar{R}_7 = \text{tr}(\bar{\mathbf{C}}^2 \cdot \bar{\mathbf{G}}_{\text{ti}}^2) \quad (78)$$

However, since $\bar{\mathbf{G}}_{\text{ti}}$ has only two non-equal eigenvalues $\lambda_1 \neq \lambda_2$, one can find that $\bar{\mathbf{G}}_{\text{ti}}^2$ can be expressed as

$$\bar{\mathbf{G}}_{\text{ti}}^2 = (\lambda_1 + \lambda_2) \bar{\mathbf{G}}_{\text{ti}} + \lambda_1 \lambda_2 \mathbf{I} \quad (79)$$

by applying Eqs. (52) and (53). Thus, the invariants \bar{R}_6, \bar{R}_7 given in Eq. (78) are redundant since they can be expressed as $\bar{R}_6 = (\lambda_1 + \lambda_2) \bar{R}_4 + \lambda_1 \lambda_2 \bar{I}_1$ and $\bar{R}_7 = (\lambda_1 + \lambda_2) \bar{R}_5 + \lambda_1 \lambda_2 (\bar{I}_1^2 - 2\bar{I}_2)$. A complete set for transverse isotropy is thus given by $\bar{\mathcal{I}}^{\text{ti}} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{R}_4, \bar{R}_5) \in \mathbb{R}^5$. Note that it is equivalent to choose the set $\bar{\mathcal{I}}^{\text{ti},*} = (\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{R}_6, \bar{R}_7) \in \mathbb{R}^5$.

B Neural network architecture

In this work, NNs combined with internal in- and output normalization layers are employed to represent the elastic potential [46]. This avoids prior normalization of the training data and still limits the weights and biases to a range that is appropriate for an efficient optimization. It also makes the process of integrating calibrated models into FE codes simpler. More specifically, after training, the internal normalization layers are included into the architecture by just multiplying the normalization weights. In addition, an architecture with an additional gate layer is introduced.

B.1 Neural network with internal normalization layers

A model which have to be trained by the data \mathcal{D} consisting of tuples ${}^i\mathcal{T} := ({}^i\mathcal{X}, {}^iY) \in \mathbb{R}^n \times \mathbb{R}$, with the generalized vector ${}^i\mathcal{X} := ({}^iX_1, {}^iX_2, \dots, {}^iX_n)$, is given by

$$f^{\text{NN}} : \mathbb{R}^n \rightarrow \mathbb{R}, \mathcal{X} \mapsto f^{\text{NN}}(\mathcal{X}) := (n^{\text{out}} \circ g^{\text{NN}} \circ \boldsymbol{n}^{\text{in}})(\mathcal{X}). \quad (80)$$

Therein, the trainable network is represented by the function $g^{\text{NN}}(\boldsymbol{x})$, where a PNN, i.e., an NN enforcing positive outputs for all possible normalized input vectors $\boldsymbol{x} \in \mathbb{R}^n$, is used. The internal normalization layers are $\boldsymbol{n}^{\text{in}}(\mathcal{X})$ and $\boldsymbol{n}^{\text{out}}(y)$. In order to not disturb the required positivity, we introduce these normalization layers as

$$n_{\alpha}^{\text{in}} : \mathbb{R} \rightarrow [x_{\alpha}^{\text{min}}, x_{\alpha}^{\text{max}}] \subset \mathbb{R}, X_{\alpha} \mapsto n_{\alpha}^{\text{in}}(X_{\alpha}) := X_{\alpha} \frac{x_{\alpha}^{\text{max}} - x_{\alpha}^{\text{min}}}{X_{\alpha}^{\text{max}} - X_{\alpha}^{\text{min}}} + \frac{x_{\alpha}^{\text{max}} X_{\alpha}^{\text{min}} - x_{\alpha}^{\text{min}} X_{\alpha}^{\text{max}}}{X_{\alpha}^{\text{min}} - X_{\alpha}^{\text{max}}} \quad \text{and} \quad (81)$$

$$n^{\text{out}} : [y^{\text{min}}, y^{\text{max}}] \rightarrow \mathbb{R}, y \mapsto n^{\text{out}}(y) := \frac{y^{\text{max}} - y^{\text{min}}}{y^{\text{max}} - y^{\text{min}}} y, \quad (82)$$

whereby there is no summation over the index $(\cdot)_{\alpha}$. The values $X_{\alpha}^{\text{min}}, X_{\alpha}^{\text{max}}, Y^{\text{min}}$ and Y^{max} have to be determined from the data before training, whereas $x_{\alpha}^{\text{min}}, x_{\alpha}^{\text{max}}, y^{\text{min}}$ and y^{max} have to be prescribed.

The PNN $g^{\text{NN}}(\boldsymbol{x})$ with H hidden layers is given by

$$o_{\alpha}^{[1]} = \mathcal{A} \left(\sum_{\beta=1}^n w_{\alpha\beta}^{[1]} x_{\beta} + b_{\alpha}^{[1]} \right), \quad \alpha \in \{1, 2, \dots, N^{\text{nn},1}\}, \quad (83)$$

$$o_{\alpha}^{[h]} = \mathcal{A} \left(\sum_{\beta=1}^{N^{\text{nn},h-1}} w_{\alpha\beta}^{[h]} o_{\beta}^{[h-1]} + b_{\alpha}^{[h]} \right), \quad \alpha \in \{1, 2, \dots, N^{\text{nn},h}\}, h \in \{2, \dots, H\}, \quad (84)$$

$$g^{\text{NN}}(\boldsymbol{x}) = \sum_{\alpha=1}^{N^{\text{nn},H}} W_{\alpha} o_{\alpha}^{[H]} + B \in \mathbb{R}, \quad (85)$$

whereby there are no restrictions for the weights expect for the output layer, i.e., $w_{\alpha\beta}^{[1]}, b_{\alpha}^{[1]}, w_{\alpha\beta}^{[h]}, b_{\alpha}^{[h]} \in \mathbb{R}$. The activation function in the final hidden layer must be greater equal to zero for all possible output values of the previous layer in order to meet the condition that $g^{\text{NN}}(\boldsymbol{x}) \geq 0 \forall \boldsymbol{x} \in \mathbb{R}^n$. Thus, the softplus activation function $\text{SP}(z) := \ln(1 + \exp z)$ is chosen and $W_{\alpha}, B \in \mathbb{R}_{\geq 0}$, the output layer's weights and bias, are restricted to be non-negative. All weights and bias values of the PNN are summarized in the k -dimensional vector

$$\boldsymbol{w} \in \mathcal{PNN} := \left\{ w_{\alpha\beta}^{[h]}, b_{\alpha}^{[h]} \in \mathbb{R}; W_{\alpha}, B \in \mathbb{R}_{\geq 0} \mid h \in \{1, \dots, H\} \right\}. \quad (86)$$

For the sake of clarity, a somewhat simplified mathematical notation is used here for the quantity \mathcal{PNN} .

B.2 Neural network with internal normalization layers and gate layer

If, to enable sparsity of the model with respect to the number of inputs, a trainable gate layer is included into the network in addition, the model given by Eq. (83) is modified according to

$$f^{\text{NN}} : \mathbb{R}^n \rightarrow \mathbb{R}, \mathcal{X} \mapsto f^{\text{NN}}(\mathcal{X}) := (n^{\text{out}} \circ g^{\text{NN}} \circ \ell^{\text{gate}} \circ \boldsymbol{n}^{\text{in}})(\mathcal{X}). \quad (87)$$

Thereby, the gate layer is defined by

$$\ell^{\text{gate}} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \boldsymbol{x} \mapsto \boldsymbol{x} \odot \boldsymbol{g} \quad \text{with} \quad g_{\alpha} := \min(1, \gamma \tanh(\epsilon q_{\alpha})) \in [0, 1] \quad (88)$$

where $\gamma, \epsilon \in \mathbb{R}$ are hyper parameters and $q_{\alpha} \in [0, 1]$ are trainable variables. Thus, we have the additional set $\boldsymbol{q} \in \mathcal{Gate} := \{\boldsymbol{q} \in \mathbb{R}^n \mid q_{\alpha} \in [0, 1]\}$. To enforce sparsity, ℓ_p regularization is used within this work.

C Terms to guarantee zero stress in the undeformed state

As discussed in Linden et al. [24], the energy expression $\bar{\psi}^{\text{str}, \square}$ enforcing $\bar{\mathbf{P}}(\bar{\mathbf{F}} = \mathbf{I}) = \mathbf{0}$ depends on the chosen set of invariants, i.e., in our case whether the set is constructed either with $\bar{\mathbf{G}}, \bar{\mathbb{G}}, \bar{\mathbf{G}},$ or $(\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2)$.

2nd order structure tensor For the invariant set build with $\bar{\mathbf{C}}$ and the 2nd order structure tensor $\bar{\mathbf{G}}$, i.e., $\bar{\mathbf{I}}^{\bar{\mathbf{G}}} \in \mathbb{R}^7$, we use the following expression to enforce a stress-free undeformed state:

$$\bar{\psi}^{\text{str},\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{J}) = -\mathbf{m}^{\bar{\mathbf{G}}}(\bar{J} - 1) - \mathbf{n}^{\bar{\mathbf{G}}}(\bar{K}_4 - 1) - \mathbf{o}^{\bar{\mathbf{G}}}(\bar{K}_6 - \bar{K}_6(\mathbf{I})) \quad \text{with} \quad (89)$$

$$\mathbf{m}^{\bar{\mathbf{G}}} = 2 \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_1} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_2} + \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_3} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (90)$$

$$\mathbf{n}^{\bar{\mathbf{G}}} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{K}_4} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{K}_5} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}} \quad \text{and} \quad \mathbf{o}^{\bar{\mathbf{G}}} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{K}_6} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{K}_7} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}. \quad (91)$$

4th order structure tensor Similarly, for the invariant set $\bar{\mathbf{I}}^{\bar{\mathbf{G}}} \in \mathbb{R}^{11}$ build with $\bar{\mathbf{C}}$ and the 4th order structure tensor $\bar{\mathbf{G}}$, the expression

$$\bar{\psi}^{\text{str},\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{J}) = -\mathbf{m}^{\bar{\mathbf{G}}}[\bar{J} - 1] - \mathbf{n}^{\bar{\mathbf{G}}}[\bar{L}_4 - \bar{L}_4(\mathbf{I})] - \mathbf{o}^{\bar{\mathbf{G}}}[\bar{L}_5 - \bar{L}_5(\mathbf{I})] - \mathbf{p}^{\bar{\mathbf{G}}}[\bar{L}_6 - \bar{L}_6(\mathbf{I})] - \mathbf{q}^{\bar{\mathbf{G}}}[\bar{L}_{11} - \bar{L}_{11}(\mathbf{I})] \quad \text{with} \quad (92)$$

$$\mathbf{m}^{\bar{\mathbf{G}}} = 2 \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_1} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_2} + \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_3} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (93)$$

$$\mathbf{n}^{\bar{\mathbf{G}}} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{L}_4} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{L}_7} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{L}_8} + 3 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{L}_9} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (94)$$

$$\mathbf{o}^{\bar{\mathbf{G}}} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{L}_5} + \frac{3}{2} \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{L}_{10}} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (95)$$

$$\mathbf{p}^{\bar{\mathbf{G}}} = \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{L}_6} \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}} \quad \text{and} \quad \mathbf{q}^{\bar{\mathbf{G}}} = \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{L}_{11}} \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}} \quad (96)$$

is used. Thereby, only invariants up to the order $\bar{\mathbf{C}}^3$ are used into the set.

6th order structure tensor Finally, for the model based on the 6th order structure tensor $\bar{\mathbf{G}}$, we can use the expression

$$\bar{\psi}^{\text{str},\bar{\mathbf{G}}}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}}, \bar{J}) = -\mathbf{m}^{\bar{\mathbf{G}}}[\bar{J} - 1] - \mathbf{n}^{\bar{\mathbf{G}}}[\bar{M}_4 - \bar{M}_4(\mathbf{I})] - \mathbf{o}^{\bar{\mathbf{G}}}[\bar{M}_5 - \bar{M}_5(\mathbf{I})] - \mathbf{p}^{\bar{\mathbf{G}}}[\bar{M}_6 - \bar{M}_6(\mathbf{I})] - \mathbf{q}^{\bar{\mathbf{G}}}[\bar{M}_{10} - \bar{M}_{10}(\mathbf{I})], \quad (97)$$

where the introduced terms are given by

$$\mathbf{m}^{\bar{\mathbf{G}}} = 2 \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_1} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_2} + \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_3} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (98)$$

$$\mathbf{n}^{\bar{\mathbf{G}}} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_4} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_7} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_8} + 3 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_9} + 3 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_{11}} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (99)$$

$$\mathbf{o}^{\bar{\mathbf{G}}} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_5} + \frac{3}{2} \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_{12}} + \frac{3}{2} \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_{13}} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (100)$$

$$\mathbf{p}^{\bar{\mathbf{G}}} = \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_6} \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}} \quad \text{and} \quad \mathbf{q}^{\bar{\mathbf{G}}} = \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{M}_{10}} \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}} \quad (101)$$

to enforce zero stress in the undeformed state.

Two 2nd order structure tensor For the invariant set build with $\bar{\mathbf{C}}$ and the two 2nd order structure tensors $\bar{\mathbf{G}}_1$ and $\bar{\mathbf{G}}_2$, i.e., $\bar{\mathbf{I}}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} \in \mathbb{R}^{12}$, we use the following expression to enforce a stress-free undeformed state:

$$\begin{aligned} \bar{\psi}^{\text{str}, \bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{\mathbf{I}}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}, \bar{\mathbf{J}}) = & -\mathbf{m}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{\mathbf{J}} - 1) - \mathbf{n}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{N}_4 - 1) - \mathbf{o}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{N}_6 - \bar{N}_6(\mathbf{I})) \dots \\ & - \mathbf{p}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{N}_8 - 1) - \mathbf{q}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{N}_{10} - \bar{N}_{10}(\mathbf{I})) - \mathbf{r}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2}(\bar{N}_{12} - \bar{N}_{12}(\mathbf{I})) \text{ with} \end{aligned} \quad (102)$$

$$\mathbf{m}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} = 2 \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_1} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_2} + \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{I}_3} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (103)$$

$$\mathbf{n}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{N}_4} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{K}_5} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad \mathbf{o}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{N}_6} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{K}_7} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad (104)$$

$$\mathbf{p}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{N}_8} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{N}_9} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}, \quad \mathbf{q}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{N}_{10}} + 2 \frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{N}_{11}} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}} \text{ and} \quad (105)$$

$$\mathbf{r}^{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2} = \left(\frac{\partial \bar{\psi}^{\text{NN}}}{\partial \bar{N}_{12}} \right) \Bigg|_{\bar{\mathbf{F}}=\mathbf{I}}. \quad (106)$$

Note that the proposed stress normalization terms are not polyconvex by construction. However, as the selected NN architecture is also not polyconvex, this is not a disadvantage.

D Sampling technique for data generation

In this appended section, we describe the details on the sampling technique that has been applied for the data generation. The used technique is similar to the one presented in Kalina et al. [46] but with the difference that the invariant space of the underlying anisotropy is not exploited here.

From the principle of material frame invariance we find that $\bar{\psi}(\bar{\mathbf{F}})$ only depends on the right stretch tensor $\bar{\mathbf{U}}$, where $\bar{\mathbf{F}} = \bar{\mathbf{R}} \cdot \bar{\mathbf{U}}$, $\bar{\mathbf{R}} \in \mathcal{SO}(3)$. Thus, for sampling we pragmatically choose $\bar{\mathbf{R}}^{\text{samp}} := \mathbf{I}$. Since $\bar{\mathbf{U}}$ has to be positive definite, i.e., $s \cdot \bar{\mathbf{U}} \cdot s > 0 \forall s \in \mathcal{L}_1, s \neq \mathbf{0}$, it is not easily possible to directly sample the coordinates \bar{U}_{KL} . Instead, we make use of the relation

$$\bar{\mathbf{U}}^{\text{samp}} = \bar{\mathbf{Q}}^T(\theta_1, \theta_2, \theta_3) \cdot \text{diag}(\bar{\lambda}_1, \bar{\lambda}_2, \bar{J} \bar{\lambda}_1^{-1} \bar{\lambda}_2^{-1}) \cdot \bar{\mathbf{Q}}(\theta_1, \theta_2, \theta_3) \in \mathcal{SO}(3) \quad (107)$$

and sample in the six-dimensional space defined by $\mathcal{S} := (\bar{\lambda}_1, \bar{\lambda}_2, \bar{J}, \theta_1, \theta_2, \theta_3) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ by Latin Hypercube Sampling (LHS). This guarantees the desired property of all samples $s \bar{\mathbf{U}}^{\text{samp}}$, $s \in \{1, 2, \dots, n_{\text{samp}}\}$. For each state $s \bar{\mathbf{F}}^{\text{samp}} = \bar{\mathbf{R}} \cdot s \bar{\mathbf{U}}^{\text{samp}} = s \bar{\mathbf{U}}^{\text{samp}}$, a loading path from $\bar{\mathbf{F}} = \mathbf{I}$ to the final deformation is generated by linearly interpolating in $\bar{\lambda}_1, \bar{\lambda}_2, \bar{J}$ within n_{inc} increments and keeping $\theta_1, \theta_2, \theta_3$ fixed. Thus, we have a set containing $k = n_{\text{sampe}} \cdot n_{\text{inc}}$ states: $\mathcal{A} := \{^1 \bar{\mathbf{U}}, ^2 \bar{\mathbf{U}}, \dots, ^k \bar{\mathbf{U}}\}$.

Finally, to avoid duplicate states in the data set and to reduce the number of states, filtering is performed using the Green-Lagrange strain $\bar{\mathbf{E}}$ as a measure for comparing the states. To start the process, all states of the first loading path are added to a set \mathcal{U} which only contains unique states, i.e., $\mathcal{U} := \{^1 \bar{\mathbf{E}}^{\text{un}}, ^2 \bar{\mathbf{E}}^{\text{un}}, \dots, ^{n_{\text{inc}}} \bar{\mathbf{E}}^{\text{un}}\}$. Then, to decide whether another state $^a \bar{\mathbf{E}}$ with $a \in \{n_{\text{inc}} + 1, \dots, k\}$ is inimitably, the relative distance

$$^a b d := \frac{\| ^a \bar{\mathbf{E}} - ^b \bar{\mathbf{E}} \|^2}{\text{ReLU}(\| ^b \bar{\mathbf{E}} \|^2 - \delta) + \delta} \text{ with } \delta := \frac{1}{3} \max \left(\| ^1 \bar{\mathbf{E}} \|^2, \| ^2 \bar{\mathbf{E}} \|^2, \dots, \| ^k \bar{\mathbf{E}} \|^2 \right) \text{ and } ^b \bar{\mathbf{E}} \in \mathcal{U} \quad (108)$$

of the state to all states currently included in \mathcal{U} is calculated. Thereby, δ is a heuristic parameter to prevent small values to be overrepresented [46]. If $^a b d \geq d_{\text{tol}}$ only applies to one $b \in \{1, 2, \dots, |\mathcal{U}|\}$, the state is unique and must be added to \mathcal{U} . To facilitate the subsequent application of a state in the computational homogenization, the complete loading path belonging to a new unique state is added to the set \mathcal{U} . After filtering, the deformation gradients belonging to the identified set of unique states are saved.

To generate the deformation tuples by using the algorithm described above, the parameters given in Tab. 5 have been chosen. With that, a total of 154 loading paths have been identified. Note that there exist several alternative sampling technique in the literature, e.g., Kunc and Fritzen [104] or Fuhs and Bouklas [7].

Table 5: Chosen parameters for the sampling algorithm. With that, a set containing a total of 154 loading paths have been generated.

range λ_1	range $\bar{\lambda}_2$	range \bar{J}	range θ_1	range θ_2	range θ_3	n_{samp}	n_{inc}	d_{tol}
[0.8, 1.4]	[0.8, 1.4]	[0.9, 1.2]	[0, π]	$[-\pi/2, \pi/2]$	$[-\pi, \pi]$	5×10^3	20	0.15

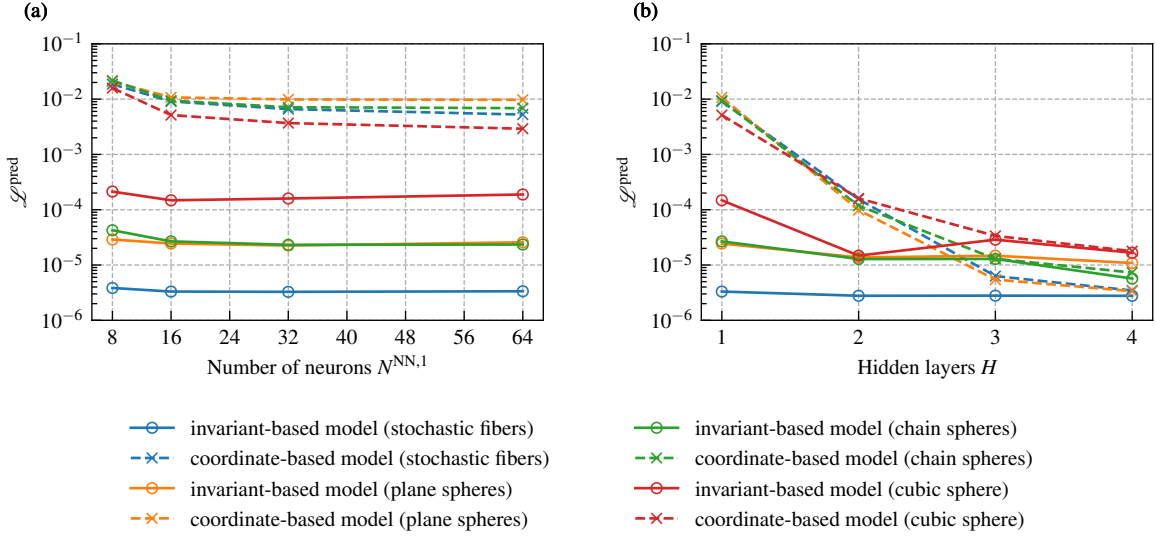


Figure 15: Hyperparameter study for the invariant-based model $\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{J})$ and the coordinate-based model $\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{J})$ with the loss $\mathcal{L} = 0.7\mathcal{L}^\sigma + 0.3\mathcal{L}^c$: (a) NNs with one hidden layer and varied number of neurons $N^{\text{NN},1}$ and (b) NNs consisting of H hidden layers with 16 neurons each. The trainable gate layer was deactivated for the invariant-based models. Shown are the results of the best run out of 5 training runs.

E Hyperparameter study

Within this appended section, we provide a hyperparameter study for the considered network architectures. In addition, a study on the optimal choice of the weight for ℓ_p regularization is shown. To exclude random effects from initialization, 5 training runs with pre-training and post-training step as described in Remark 3 have carried out each. The best training run is used. The overall datasets are divided into calibration and test sets with a ratio of 70/30, respectively, see Eq. (35). To reduce the number of trainings, we set the knowledge of the required order of structure tensor for the respective RVEs here: *stochastic fibers* ($\bar{\mathbf{G}} \in \text{Sym}$), *hexagonal fibers* ($\bar{\mathbf{G}} \in \text{Sym}_6$), *cubic sphere* ($\bar{\mathbf{G}} \in \text{Sym}_4$), *plane spheres* ($\bar{\mathbf{G}} \in \text{Sym}$), and *chain spheres* ($\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2 \in \text{Sym}$), cf. Tab. 2.

E.1 Number of hidden layers and neurons

Here, the hyperparameters of the networks used in the invariant-based model $\bar{\psi}^\square(\bar{\mathbf{I}}^\square, \bar{J})$ according to Eq. (33) and the coordinate-based model $\bar{\psi}^{\text{coord}}(\bar{\mathbf{C}}, \bar{J})$ given in Eq. (43) are varied. For the loss $\mathcal{L} = 0.7\mathcal{L}^\sigma + 0.3\mathcal{L}^c$. Within this hyperparameter study, the trainable gate layer is deactivated for the invariant-based models.

Variation number of neurons in one hidden layer In a first study, we consider NNs with one hidden layer and vary the number of neurons $N^{\text{NN},1} \in \{8, 16, 32, 64\}$. The results are shown in Fig. 15(a). As can be seen, the loss achieved with the coordinate-based model does not fall below 5×10^{-3} for any of the RVEs, and there is no noticeable improvement when the number of neurons is increased. In contrast, the invariant-based approach achieves very good results with just one hidden layer for four of the five RVEs. Only for the RVE cubic sphere does the loss remain above 1×10^{-4} even with 64 neurons in the hidden layer.

Variation number of hidden layers Secondly, we consider NNs with $H \in \{1, 2, 3, 4\}$ hidden layers with $N^{\text{NN},h} = 16$, $h \in \{1, \dots, H\}$. The results of this study are shown in Fig. 15(b). As can be seen, the loss achieved with the coordinate-based model now drops significantly for all RVEs considered when the number of hidden layers is increased. With two hidden layers, the loss values are around 1×10^{-4} and with three hidden layers, values clearly below 1×10^{-4} and in some cases below 1×10^{-4} are achieved for all RVEs. With four hidden layers, there is then no further significant

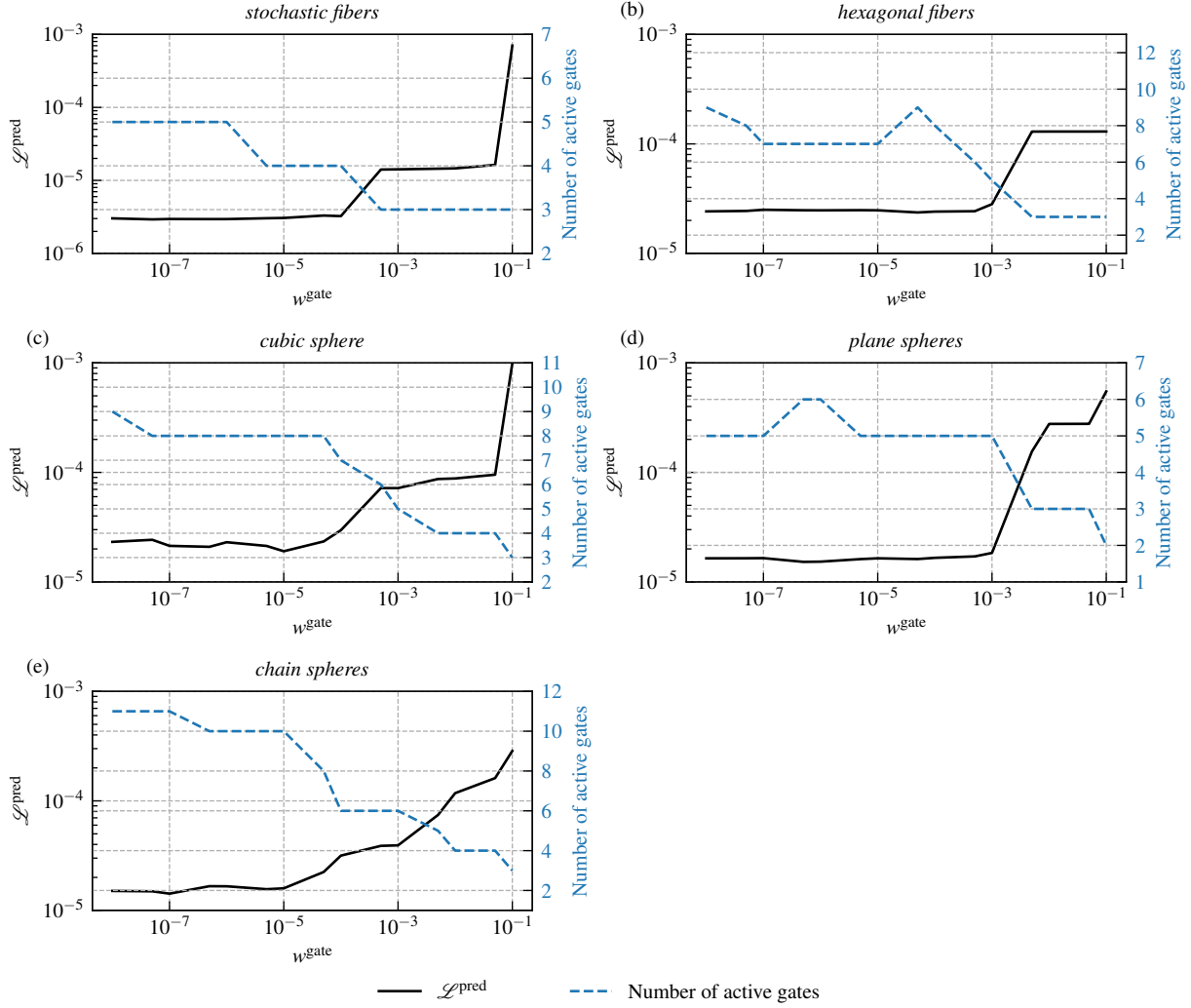


Figure 16: Variation of the weight w^{gate} for the gate loss term $\mathcal{L}^{\text{gate}}$: (a)–(e) RVEs stochastic fibers, hexagonal fibers, cubic fibers, plane spheres, and chain spheres, respectively. The prediction loss was chosen to $\mathcal{L}^{\text{pred}} = 0.7\mathcal{L}^{\sigma} + 0.3\mathcal{L}^{\text{c}}$. Shown are the results of the best run out of 5 training runs.

improvement. In contrast, the invariant-based approach results in almost no dependence on the number of hidden layers for four of the five RVEs. Even with one hidden layer of 16 neurons, the results here are very good. Only for the RVE cubic sphere there is a noticeable improvement above a number of two hidden layers, so that a loss below 1×10^{-4} can be achieved from $H \geq 2$. It should be noted that the finding that relatively small networks are required for invariant-based NN approaches is consistent with the literature, cf. [24, 28, 36, 46, 69, 75].

The study thus results in the following choice for the NNs used in this work: For the invariant-based approach, architectures with 2 hidden layers of 16 neurons each are used. For the coordinate-based model, on the other hand, networks with 3 hidden layers of 16 neurons each are used.

E.2 Weighting of the gate loss

Here, the weight w^{gate} for the loss term $\mathcal{L}^{\text{gate}}$ defined in Eq. (39) is varied systematically. Thus, only the invariant-based model $\bar{\psi}^{\square}(\bar{\mathbf{I}}^{\square}, \bar{\mathbf{J}})$ is considered. According to the study given in E.1, architectures with 2 hidden layers of 16 neurons each are used. The parameters for the gates and the exponent in the p -norm are chosen to $\gamma = 1.025$, $\epsilon = 2.5$, $\delta = 1 \times 10^{-6}$, and $p = \frac{1}{4}$, respectively. The loss term for the training is given by $\mathcal{L} = \mathcal{L}^{\text{pred}} + w^{\text{gate}}\mathcal{L}^{\text{gate}}$, where the prediction loss is chosen as $\mathcal{L}^{\text{pred}} = 0.7\mathcal{L}^{\sigma} + 0.3\mathcal{L}^{\text{c}}$. The weight is varied as follows: $w^{\text{gate}} \in \{1 \times 10^{-8}, 5 \times 10^{-8}, \dots, 1 \times 10^{-1}\}$.

The results of this study are given in Fig. 16 for the five considered RVEs. On the left vertical axis of each subplot, the prediction loss is plotted and on the right vertical axis (blue) the number of active gates, i.e., gates for which the condition $g_\alpha > 0$ holds. As one can see, the number of active gates after training decreases with an increasing w^{gate} . However, if the weight is set too high, this leads to an excessive weighting of the penalty term based on the p -norm. This initially leads to the elimination of invariants required to describe the anisotropy from the model and, if the value is increased further, to a disproportionate deterioration in the prediction capability. Thus, the aim is now to find a value for the weight that leads to a model with as few invariants as possible, but at the same time does not negatively affect the prediction quality. Accordingly, a value of 5×10^{-5} has proven to be suitable for all RVEs.

References

- [1] Peter Haupt. *Continuum Mechanics and Theory of Materials*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000. ISBN 978-3-662-04109-3.
- [2] Gerhard A. Holzapfel. *Nonlinear Solid Mechanics - A Continuum Approach for Engineering*. John Wiley & Sons, Chichester, 2000. ISBN 978-0-471-82319-3.
- [3] Johannes Dornheim, Lukas Morand, Hemanth Janarthanam Nallani, and Dirk Helm. Neural Networks for Constitutive Modeling: From Universal Function Approximators to Advanced Models and the Integration of Physics. *Archives of Computational Methods in Engineering*, October 2023. ISSN 1886-1784. doi:10.1007/s11831-023-10009-y.
- [4] Jan Niklas Fuhg, Govinda Anantha Padmanabha, Nikolaos Bouklas, Bahador Bahmani, WaiChing Sun, Nikolaos N. Vlassis, Moritz Flaschel, Pietro Carrara, and Laura De Lorenzis. A review on data-driven constitutive laws for solids, May 2024.
- [5] J. Ghaboussi, J. H. Garrett, and X. Wu. Knowledge-Based Modeling of Material Behavior with Neural Networks. *Journal of Engineering Mechanics*, 117(1):132–153, 1991. ISSN 0733-9399, 1943-7889. doi:10.1061/(ASCE)0733-9399(1991)117:1(132).
- [6] Ari L. Frankel, Reese E. Jones, and Laura P. Swiler. Tensor Basis Gaussian Process Models of Hyperelastic Materials. *Journal of Machine Learning for Modeling and Computing*, 1(1), 2020. ISSN 2689-3967, 2689-3975. doi:10.1615/2020033325.
- [7] Jan N. Fuhg and Nikolaos Bouklas. On physics-informed data-driven isotropic and anisotropic constitutive models through probabilistic machine learning and space-filling sampling. *Computer Methods in Applied Mechanics and Engineering*, 394:114915, May 2022. ISSN 0045-7825. doi:10.1016/j.cma.2022.114915.
- [8] Nathan Ellmer, Rogelio Ortigosa, Jesús Martínez-Frutos, and Antonio J. Gil. Gradient enhanced gaussian process regression for constitutive modelling in finite strain hyperelasticity. *Computer Methods in Applied Mechanics and Engineering*, 418:116547, January 2024. ISSN 0045-7825. doi:10.1016/j.cma.2023.116547.
- [9] Simon Wiesheier, Miguel Angel Moreno-Mateos, and Paul Steinmann. Versatile data-adaptive hyperelastic energy functions for soft materials. *Computer Methods in Applied Mechanics and Engineering*, 430:117208, October 2024. ISSN 0045-7825. doi:10.1016/j.cma.2024.117208.
- [10] Moritz Flaschel, Siddhant Kumar, and Laura De Lorenzis. Unsupervised discovery of interpretable hyperelastic constitutive laws. *Computer Methods in Applied Mechanics and Engineering*, 381:113852, August 2021. ISSN 00457825. doi:10.1016/j.cma.2021.113852.
- [11] Moritz Flaschel, Siddhant Kumar, and Laura De Lorenzis. Automated discovery of generalized standard material models with EUCLID. *Computer Methods in Applied Mechanics and Engineering*, 405:115867, February 2023. ISSN 0045-7825. doi:10.1016/j.cma.2022.115867.
- [12] Knut Andreas Meyer and Fredrik Ekre. Thermodynamically consistent neural network plasticity modeling and discovery of evolution laws. *Journal of the Mechanics and Physics of Solids*, 180:105416, November 2023. ISSN 0022-5096. doi:10.1016/j.jmps.2023.105416.
- [13] Rasul Abdusalamov, Markus Hillgärtner, and Mikhail Itskov. Automatic generation of interpretable hyperelastic material models by symbolic regression. *International Journal for Numerical Methods in Engineering*, 124(9): 2093–2104, 2023. ISSN 1097-0207. doi:10.1002/nme.7203.
- [14] Georgios Kissas, Siddhartha Mishra, Eleni Chatzi, and Laura De Lorenzis. The language of hyperelastic materials. *Computer Methods in Applied Mechanics and Engineering*, 428:117053, August 2024. ISSN 0045-7825. doi:10.1016/j.cma.2024.117053.
- [15] Frederic E. Bock, Roland C. Aydin, Christian J. Cyron, Norbert Huber, Surya R. Kalidindi, and Benjamin Klusemann. A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics. *Frontiers in Materials*, 6:110, May 2019. ISSN 2296-8016. doi:10.3389/fmats.2019.00110.

- [16] Xin Liu, Su Tian, Fei Tao, and Wenbin Yu. A review of artificial neural networks in the constitutive modeling of composite materials. *Composites Part B: Engineering*, 224:109152, November 2021. ISSN 13598368. doi:10.1016/j.compositesb.2021.109152.
- [17] Xiaoyang Zheng, Xubo Zhang, Ta-Te Chen, and Ikumu Watanabe. Deep Learning in Mechanical Metamaterials: From Prediction and Generation to Inverse Design. *Advanced Materials*, 35(45):2302530, 2023. ISSN 1521-4095. doi:10.1002/adma.202302530.
- [18] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 00219991. doi:10.1016/j.jcp.2018.10.045.
- [19] Alexander Henkes, Henning Wessels, and Rolf Mahnken. Physics informed neural networks for continuum micromechanics. *Computer Methods in Applied Mechanics and Engineering*, 393:114790, 2022. ISSN 0045-7825. doi:10.1016/j.cma.2022.114790.
- [20] Jan-Hendrik Bastek and Dennis M. Kochmann. Physics-Informed Neural Networks for shell structures. *European Journal of Mechanics - A/Solids*, 97:104849, January 2023. ISSN 0997-7538. doi:10.1016/j.euromechsol.2022.104849.
- [21] Ali Harandi, Ahmad Moeineddin, Michael Kaliske, Stefanie Reese, and Shahed Rezaei. Mixed formulation of physics-informed neural networks for thermo-mechanically coupled systems and heterogeneous domains. *International Journal for Numerical Methods in Engineering*, 125(4):e7388, 2024. ISSN 1097-0207. doi:10.1002/nme.7388.
- [22] Faisal As’ad, Philip Avery, and Charbel Farhat. A mechanics-informed artificial neural network approach in data-driven constitutive modeling. *International Journal for Numerical Methods in Engineering*, 123(12):2738–2759, 2022. ISSN 1097-0207. doi:10.1002/nme.6957.
- [23] Dominik K. Klein, Rogelio Ortigosa, Jesús Martínez-Frutos, and Oliver Weeger. Finite electro-elasticity with physics-augmented neural networks. *Computer Methods in Applied Mechanics and Engineering*, 400:115501, 2022. ISSN 0045-7825. doi:10.1016/j.cma.2022.115501.
- [24] Lennart Linden, Dominik K. Klein, Karl A. Kalina, Jörg Brummund, Oliver Weeger, and Markus Kästner. Neural networks meet hyperelasticity: A guide to enforcing physics. *Journal of the Mechanics and Physics of Solids*, 179:105363, 2023. ISSN 0022-5096. doi:10.1016/j.jmps.2023.105363.
- [25] Karl A. Kalina, Lennart Linden, Jörg Brummund, and Markus Kästner. FE^{ANN}: An efficient data-driven multiscale approach based on physics-constrained neural networks and automated data mining. *Computational Mechanics*, 71:827, 2023. ISSN 1432-0924. doi:10.1007/s00466-022-02260-0.
- [26] Filippo Masi, Ioannis Stefanou, Paolo Vannucci, and Victor Maffi-Berthier. Thermodynamics-based Artificial Neural Networks for constitutive modeling. *Journal of the Mechanics and Physics of Solids*, 147:104277, 2021. ISSN 0022-5096. doi:10.1016/j.jmps.2020.104277.
- [27] Karl A. Kalina, Lennart Linden, Jörg Brummund, Philipp Metsch, and Markus Kästner. Automated constitutive modeling of isotropic hyperelasticity based on artificial neural networks. *Computational Mechanics*, 69(1):213–232, 2022. ISSN 1432-0924. doi:10.1007/s00466-021-02090-6.
- [28] Kevin Linka, Markus Hillgärtner, Kian P. Abdolazizi, Roland C. Aydin, Mikhail Itskov, and Christian J. Cyron. Constitutive artificial neural networks: A fast and general approach to predictive data-driven constitutive modeling by deep learning. *Journal of Computational Physics*, 429:110010, 2021. ISSN 00219991. doi:10.1016/j.jcp.2020.110010.
- [29] Max Rosenkranz, Karl A. Kalina, Jörg Brummund, and Markus Kästner. A comparative study on different neural network architectures to model inelasticity. *International Journal for Numerical Methods in Engineering*, page nme.7319, 2023. ISSN 0029-5981, 1097-0207. doi:10.1002/nme.7319.
- [30] Patrick Weber, Jeremy Geiger, and Werner Wagner. Constrained neural network training and its application to hyperelastic material modeling. *Computational Mechanics*, 68(5):1179–1204, 2021. ISSN 0178-7675, 1432-0924. doi:10.1007/s00466-021-02064-8.
- [31] Patrick Weber, Werner Wagner, and Steffen Freitag. Physically enhanced training for modeling rate-independent plasticity with feedforward neural networks. *Computational Mechanics*, April 2023. ISSN 1432-0924. doi:10.1007/s00466-023-02316-9.
- [32] Jan Niklas Fuhg, Craig M. Hamel, Kyle Johnson, Reese Jones, and Nikolaos Bouklas. Modular machine learning-based elastoplasticity: Generalization in the context of limited data. *Computer Methods in Applied Mechanics and Engineering*, 407:115930, 2023. ISSN 0045-7825. doi:10.1016/j.cma.2023.115930.

- [33] Filippo Masi and Itai Einav. Neural integration for constitutive equations using small data. *Computer Methods in Applied Mechanics and Engineering*, 420:116698, February 2024. ISSN 0045-7825. doi:10.1016/j.cma.2023.116698.
- [34] Y. Shen, K. Chandrashekhara, W. F. Breig, and L. R. Oliver. Neural Network Based Constitutive Model for Rubber Material. *Rubber Chemistry and Technology*, 77(2):257–277, 2004. ISSN 1943-4804, 0035-9475. doi:10.5254/1.3547822.
- [35] G. Liang and K. Chandrashekhara. Neural network based constitutive model for elastomeric foams. *Engineering Structures*, 30(7):2002–2011, 2008. ISSN 01410296. doi:10.1016/j.engstruct.2007.12.021.
- [36] Dominik K. Klein, Mauricio Fernández, Robert J. Martin, Patrizio Neff, and Oliver Weeger. Polyconvex anisotropic hyperelasticity with neural networks. *Journal of the Mechanics and Physics of Solids*, page 104703, 2021. ISSN 00225096. doi:10.1016/j.jmps.2021.104703.
- [37] Prakash Thakolkaran, Akshay Joshi, Yiwen Zheng, Moritz Flaschel, Laura De Lorenzis, and Siddhant Kumar. NN-EUCLID: Deep-learning hyperelasticity without stress data. *Journal of the Mechanics and Physics of Solids*, 169:105076, 2022. ISSN 0022-5096. doi:10.1016/j.jmps.2022.105076.
- [38] Jan N. Fuhg, Nikolaos Bouklas, and Reese E. Jones. Learning hyperelastic anisotropy from data via a tensor basis neural network. *Journal of the Mechanics and Physics of Solids*, 168:105022, 2022. ISSN 00225096. doi:10.1016/j.jmps.2022.105022.
- [39] Vahidullah Taç, Kevin Linka, Francisco Sahli-Costabal, Ellen Kuhl, and Adrian Buganza Tepole. Benchmarking physics-informed frameworks for data-driven hyperelasticity. *Computational Mechanics*, 73(1):49–65, January 2024. ISSN 1432-0924. doi:10.1007/s00466-023-02355-2.
- [40] Bahador Bahmani and WaiChing Sun. Physics-constrained symbolic model discovery for polyconvex incompressible hyperelastic materials. *International Journal for Numerical Methods in Engineering*, n/a(n/a):e7473, 2024. ISSN 1097-0207. doi:10.1002/nme.7473.
- [41] Antoine Benady, Emmanuel Baranger, and Ludovic Chamoin. NN-mCRE: A modified constitutive relation error framework for unsupervised learning of nonlinear state laws with physics-augmented neural networks. *International Journal for Numerical Methods in Engineering*, 125(8):e7439, 2024. ISSN 1097-0207. doi:10.1002/nme.7439.
- [42] Mathias Peirlinck, Kevin Linka, Juan A. Hurtado, and Ellen Kuhl. On automated model discovery and a universal material subroutine for hyperelastic materials. *Computer Methods in Applied Mechanics and Engineering*, 418:116534, January 2024. ISSN 0045-7825. doi:10.1016/j.cma.2023.116534.
- [43] Wojciech M Czarnecki, Simon Osindero, Max Jaderberg, Grzegorz Swirszcz, and Razvan Pascanu. Sobolev Training for Neural Networks. In *Advances in Neural Information Processing Systems*, pages 4278–4287, 2017.
- [44] Nikolaos N. Vlassis, Ran Ma, and WaiChing Sun. Geometric deep learning for computational mechanics Part I: Anisotropic hyperelasticity. *Computer Methods in Applied Mechanics and Engineering*, 371:113299, 2020. ISSN 0045-7825. doi:10.1016/j.cma.2020.113299.
- [45] Nikolaos N. Vlassis and WaiChing Sun. Sobolev training of thermodynamic-informed neural networks for interpretable elasto-plasticity models with level set hardening. *Computer Methods in Applied Mechanics and Engineering*, 377:113695, 2021. ISSN 00457825. doi:10.1016/j.cma.2021.113695.
- [46] Karl A. Kalina, Philipp Gebhart, Jörg Brummund, Lennart Linden, WaiChing Sun, and Markus Kästner. Neural network-based multiscale modeling of finite strain magneto-elasticity with relaxed convexity criteria. *Computer Methods in Applied Mechanics and Engineering*, 421:116739, March 2024. ISSN 00457825. doi:10.1016/j.cma.2023.116739.
- [47] Nikolaos N. Vlassis, Puhan Zhao, Ran Ma, Tommy Sewell, and WaiChing Sun. Molecular dynamics inferred transfer learning models for finite-strain hyperelasticity of monoclinic crystals: Sobolev training and validations against physical constraints. *International Journal for Numerical Methods in Engineering*, 123(17):3922–3949, 2022. ISSN 1097-0207. doi:10.1002/nme.6992.
- [48] Mauricio Fernández, Mostafa Jamshidian, Thomas Böhlke, Kristian Kersting, and Oliver Weeger. Anisotropic hyperelastic constitutive models for finite deformations combining material theory and data-driven approaches with application to cubic lattice metamaterials. *Computational Mechanics*, 2020. ISSN 0178-7675, 1432-0924. doi:10.1007/s00466-020-01954-7.
- [49] Kévin Garanger, Julie Kraus, and Julian J. Rimoli. Symmetry-enforcing neural networks with applications to constitutive modeling. *Extreme Mechanics Letters*, 71:102188, September 2024. ISSN 2352-4316. doi:10.1016/j.eml.2024.102188.

- [50] Vahidullah Tac, Francisco Sahli Costabal, and Adrian B. Tepole. Data-driven tissue mechanics with polyconvex neural ordinary differential equations. *Computer Methods in Applied Mechanics and Engineering*, 398:115248, 2022. ISSN 0045-7825. doi:10.1016/j.cma.2022.115248.
- [51] Peiyi Chen and Johann Guilleminot. Polyconvex neural networks for hyperelastic constitutive models: A rectification approach. *Mechanics Research Communications*, 125:103993, 2022. ISSN 00936413. doi:10.1016/j.mechrescom.2022.103993.
- [52] Kevin Linka and Ellen Kuhl. A new family of Constitutive Artificial Neural Networks towards automated model discovery. *Computer Methods in Applied Mechanics and Engineering*, 403:115731, 2023. ISSN 00457825. doi:10.1016/j.cma.2022.115731.
- [53] Vera Ebbing. *Design of polyconvex energy functions for all anisotropy classes*. PhD thesis, Inst. für Mechanik, Abt. Bauwissenschaften, Essen, 2010.
- [54] Jörg Schröder, Patrizio Neff, and International Centre for Mechanical Sciences, editors. *Poly-, Quasi- and Rank-One Convexity in Applied Mechanics: CISM Course on Poly-, Quasi- and Rank-One Convexity in Applied Mechanics, Held in Udine from September 24 to September 28, 2007*. Number 516 in CISM Courses and Lectures. Springer, Berlin, 1. ed edition, 2010. ISBN 978-3-7091-0173-5.
- [55] Brandon Amos, Lei Xu, and J. Zico Kolter. Input Convex Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 146–155. PMLR, 2017.
- [56] Felix Fritzen, Mauricio Fernández, and Fredrik Larsson. On-the-Fly Adaptivity for Nonlinear Twoscale Simulations Using Artificial Neural Networks and Reduced Order Modeling. *Frontiers in Materials*, 6:75, May 2019. ISSN 2296-8016. doi:10.3389/fmats.2019.00075.
- [57] B. A. Le, J. Yvonnet, and Q.-C. He. Computational homogenization of nonlinear elastic materials using neural networks: NEURAL NETWORKS-BASED COMPUTATIONAL HOMOGENIZATION. *International Journal for Numerical Methods in Engineering*, 104(12):1061–1084, December 2015. ISSN 00295981. doi:10.1002/nme.4953.
- [58] Alexander Malik, Martin Abendroth, Geraf Hütter, and Bjoern Kiefer. A Hybrid Approach Employing Neural Networks to Simulate the Elasto-Plastic Deformation Behavior of 3D-Foam Structures. *Advanced Engineering Materials*, n/a(n/a):2100641, 2021. ISSN 1527-2648. doi:10.1002/adem.202100641.
- [59] M. A. Maia, I. B. C. M. Rocha, P. Kerfriden, and F. P. van der Meer. Physically recurrent neural networks for path-dependent heterogeneous materials: Embedding constitutive models in a data-driven surrogate. *Computer Methods in Applied Mechanics and Engineering*, 407:115934, March 2023. ISSN 0045-7825. doi:10.1016/j.cma.2023.115934.
- [60] Jindong Jiang, Jiajun Wu, Qiang Chen, George Chatzigeorgiou, and Fodil Meraghni. Physically informed deep homogenization neural network for unidirectional multiphase/multi-inclusion thermoconductive composites. *Computer Methods in Applied Mechanics and Engineering*, 409:115972, May 2023. ISSN 0045-7825. doi:10.1016/j.cma.2023.115972.
- [61] Filippo Masi and Ioannis Stefanou. Multiscale modeling of inelastic materials with Thermodynamics-based Artificial Neural Networks (TANN). *Computer Methods in Applied Mechanics and Engineering*, 398:115190, August 2022. ISSN 0045-7825. doi:10.1016/j.cma.2022.115190.
- [62] Hamidreza Eivazi, Jendrik-Alexander Tröger, Stefan Wittek, Stefan Hartmann, and Andreas Rausch. FE2 Computations with Deep Neural Networks: Algorithmic Structure, Data Generation, and Implementation. *Mathematical and Computational Applications*, 28(4):91, August 2023. ISSN 2297-8747. doi:10.3390/mca28040091.
- [63] S. C. Cowin and M. M. Mehrabadi. On the identification of material symmetry for anisotropic elastic materials. *The Quarterly Journal of Mechanics and Applied Mathematics*, 40(4):451–476, November 1987. ISSN 0033-5614. doi:10.1093/qjmam/40.4.451.
- [64] Maher Moakher and Andrew N. Norris. The Closest Elastic Tensor of Arbitrary Symmetry to an Elasticity Tensor of Lower Symmetry. *Journal of Elasticity*, 85(3):215–263, December 2006. ISSN 1573-2681. doi:10.1007/s10659-006-9082-0.
- [65] H. Xiao. On Isotropic Extension of Anisotropic Tensor Functions. *ZAMM - Journal of Applied Mathematics and Mechanics*, 76, 1996.
- [66] Nikolas Apel. *Approaches to the Description of Anisotropic Material Behaviour at Finite Elastic and Plastic Deformations, Theory and Numerics*. PhD thesis, Stuttgart University, 2004.
- [67] T. Christian Gasser, Ray W Ogden, and Gerhard A Holzapfel. Hyperelastic modelling of arterial layers with distributed collagen fibre orientations. *Journal of The Royal Society Interface*, 3(6):15–35, February 2006. doi:10.1098/rsif.2005.0073.

- [68] Jeremy A. McCulloch, Skyler R. St. Pierre, Kevin Linka, and Ellen Kuhl. On sparse regression, Lp-regularization, and automated model discovery. *International Journal for Numerical Methods in Engineering*, 125(14):e7481, 2024. ISSN 1097-0207. doi:10.1002/nme.7481.
- [69] Max Rosenkranz, Karl A. Kalina, Jörg Brummund, WaiChing Sun, and Markus Kästner. Viscoelasticity with physics-augmented neural networks: Model formulation and training methods without prescribed internal variables. *Computational Mechanics*, May 2024. ISSN 1432-0924. doi:10.1007/s00466-024-02477-1.
- [70] Jerrold E. Marsden, Thomas J. R. Hughes, and D. E. Carlson. *Mathematical Foundations of Elasticity*, volume 51. Prentice-Hall, 1984.
- [71] John M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Archive for Rational Mechanics and Analysis*, 63(4):337–403, 1976. ISSN 1432-0673. doi:10.1007/BF00279992.
- [72] J. M. Ball. Constitutive inequalities and existence theorems in nonlinear elasto-statics. In R.J. Knops, editor, *Herriot Watt Symposium: Nonlinear Analysis and Mechanics*, volume 1, pages 187–241, London, 1977. Pitman. ISBN 978-0-273-01128-6.
- [73] Jörg Schröder and Patrizio Neff. Invariant formulation of hyperelastic transverse isotropy based on polyconvex free energy functions. *International Journal of Solids and Structures*, 40(2):401–445, 2003. ISSN 0020-7683. doi:10.1016/S0020-7683(02)00458-4.
- [74] R. Abeyaratne and N. Triantafyllidis. An Investigation of Localization in a Porous Elastic Material Using Homogenization Theory. *Journal of Applied Mechanics*, 51(3):481–486, 1984. ISSN 0021-8936. doi:10.1115/1.3167661.
- [75] Dominik K. Klein, Rogelio Ortigosa, Jesús Martínez-Frutos, and Oliver Weeger. Nonlinear electro-elastic finite element analysis with neural network constitutive models. *Computer Methods in Applied Mechanics and Engineering*, 425:116910, May 2024. ISSN 0045-7825. doi:10.1016/j.cma.2024.116910.
- [76] Mikhail Itskov. *Tensor Algebra and Tensor Analysis for Engineers: With Applications to Continuum Mechanics*. Mathematical Engineering. Springer International Publishing, Cham, 2015. ISBN 978-3-319-16341-3 978-3-319-16342-0. doi:10.1007/978-3-319-16342-0.
- [77] Jörg Schröder and Klaus Hackl, editors. *Plasticity and beyond: Microstructures, Crystal-Plasticity and Phase Transitions*. Number 550 in Courses and Lectures / International Centre for Mechanical Sciences. Springer, Wien, 2014. ISBN 978-3-7091-1624-1 978-3-7091-1625-8.
- [78] G. Haasemann, M. Kästner, and V. Ulbricht. Multi-Scale Modelling and Simulation of Textile Reinforced Materials. In C. A. Motosoares, J. A. C. Martins, H. C. Rodrigues, Jorge A. C. Ambrósio, C. A. B. Pina, C. M. Motosoares, E. B. R. Pereira, and J. Folgado, editors, *III European Conference on Computational Mechanics*, pages 510–510. Springer Netherlands, Dordrecht, 2006. ISBN 978-1-4020-4994-1 978-1-4020-5370-2. doi:10.1007/1-4020-5370-3_510.
- [79] Alexander E. Ehret and Mikhail Itskov. A polyconvex hyperelastic model for fiber-reinforced materials in application to soft tissues. *Journal of Materials Science*, 42(21):8853–8863, August 2007. ISSN 0022-2461, 1573-4803. doi:10.1007/s10853-007-1812-6.
- [80] S. Pennisi and M. Trovato. On the irreducibility of professor G.F. Smith’s representations for isotropic functions. *International Journal of Engineering Science*, 25(8):1059–1065, January 1987. ISSN 00207225. doi:10.1016/0020-7225(87)90097-8.
- [81] J. P. Boehler. On Irreducible Representations for Isotropic Scalar Functions. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 57(6):323–327, 1977. ISSN 1521-4001. doi:10.1002/zamm.19770570608.
- [82] Maximilian P. Wollner, Michele Terzano, Malte Rolf-Pissarczyk, and Gerhard A. Holzapfel. A general model for anisotropic pseudo-elasticity and viscoelasticity at finite strains. *Journal of the Mechanics and Physics of Solids*, 180:105403, November 2023. ISSN 0022-5096. doi:10.1016/j.jmps.2023.105403.
- [83] Suresh G. Advani and Charles L. Tucker, III. The Use of Tensors to Describe and Predict Fiber Orientation in Short Fiber Composites. *Journal of Rheology*, 31(8):751–784, November 1987. ISSN 0148-6055. doi:10.1122/1.549945.
- [84] Julian Karl Bauer, Matti Schneider, and Thomas Böhlke. On the Phase Space of Fourth-Order Fiber-Orientation Tensors. *Journal of Elasticity*, 153(2):161–184, March 2023. ISSN 1573-2681. doi:10.1007/s10659-022-09977-2.
- [85] Alok Mehta and Matti Schneider. A sequential addition and migration method for generating microstructures of short fibers with prescribed length distribution. *Computational Mechanics*, 70(4):829–851, October 2022. ISSN 1432-0924. doi:10.1007/s00466-022-02201-x.

- [86] Loredana Kehler, Jeffrey T Wood, and Thomas Böhlke. Mean-field homogenization of thermoelastic material properties of a long fiber-reinforced thermoset and experimental investigation. *Journal of Composite Materials*, 54(25):3777–3799, October 2020. ISSN 0021-9983. doi:10.1177/0021998320920695.
- [87] Marc Olive, Boris Kolev, Rodrigue Desmorat, and Boris Desmorat. Characterization of the symmetry class of an elasticity tensor using polynomial covariants. *Mathematics and Mechanics of Solids*, 27(1):144–190, January 2022. ISSN 1081-2865, 1741-3028. doi:10.1177/10812865211010885.
- [88] Malte Rolf-Pissarczyk, Maximilian P. Wollner, Douglas R. Q. Pacheco, and Gerhard A. Holzapfel. Efficient computational modelling of smooth muscle orientation and function in the aorta. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 477(2255):20210592, November 2021. doi:10.1098/rspa.2021.0592.
- [89] A. Mota, Q. Chen, J.w. Foulk III, J.t. Ostien, and Z. Lai. A Cartesian parametrization for the numerical analysis of material instability. *International Journal for Numerical Methods in Engineering*, 108(2):156–180, 2016. ISSN 1097-0207. doi:10.1002/nme.5228.
- [90] Stefan Kollmannsberger, Davide D’Angella, Moritz Jokeit, and Leon Herrmann. *Deep Learning in Computational Mechanics: An Introductory Course*, volume 977 of *Studies in Computational Intelligence*. Springer International Publishing, Cham, 2021. ISBN 978-3-030-76586-6 978-3-030-76587-3. doi:10.1007/978-3-030-76587-3.
- [91] P.G. Ciarlet. *Mathematical Elasticity Volume I: Three-Dimensional Elasticity*. Elsevier, North-Holland, 1988. ISBN 978-0-08-087541-5.
- [92] Elten Polukhov and Marc-André Keip. Multiscale stability analysis of periodic magnetorheological elastomers. *Mechanics of Materials*, page 103699, 2020. ISSN 01676636. doi:10.1016/j.mechmat.2020.103699.
- [93] Stephan Rudykh and Katia Bertoldi. Stability of anisotropic magnetorheological elastomers in finite deformations: A micromechanical approach. *Journal of the Mechanics and Physics of Solids*, 61(4):949–967, 2013. ISSN 00225096. doi:10.1016/j.jmps.2012.12.008.
- [94] J. Nordmann, M. Aßmus, and H. Altenbach. Visualising elastic anisotropy: Theoretical background and computational implementation. *Continuum Mechanics and Thermodynamics*, 30(4):689–708, July 2018. ISSN 0935-1175, 1432-0959. doi:10.1007/s00161-018-0635-9.
- [95] Jan Niklas Fuhg, Reese Edward Jones, and Nikolaos Bouklas. Extreme sparsification of physics-augmented neural networks for interpretable model discovery in mechanics. *Computer Methods in Applied Mechanics and Engineering*, 426:116973, June 2024. ISSN 0045-7825. doi:10.1016/j.cma.2024.116973.
- [96] Kian P. Abdolazizi, Kevin Linka, and Christian J. Cyron. Viscoelastic Constitutive Artificial Neural Networks (vCANNs) – a framework for data-driven anisotropic nonlinear finite viscoelasticity. *Journal of Computational Physics*, page 112704, December 2023. ISSN 0021-9991. doi:10.1016/j.jcp.2023.112704.
- [97] Hagen Holthausen, Lukas Lamm, Tim Brepols, Stefanie Reese, and Ellen Kuhl. Theory and implementation of inelastic Constitutive Artificial Neural Networks. *Computer Methods in Applied Mechanics and Engineering*, 428:117063, August 2024. ISSN 0045-7825. doi:10.1016/j.cma.2024.117063.
- [98] Betim Bahtiri, Behrouz Arash, Sven Scheffler, Maximilian Jux, and Raimund Rolfes. A thermodynamically consistent physics-informed deep learning material model for short fiber/polymer nanocomposites. *Computer Methods in Applied Mechanics and Engineering*, 427:117038, July 2024. ISSN 0045-7825. doi:10.1016/j.cma.2024.117038.
- [99] Birte Boes, Jaan-Willem Simon, and Hagen Holthausen. Accounting for plasticity: An extension of inelastic Constitutive Artificial Neural Networks, July 2024.
- [100] Vahidullah Tac, Ellen Kuhl, and Adrian Buganza Tepole. Data-driven continuum damage mechanics with built-in physics. *Extreme Mechanics Letters*, 71:102220, September 2024. ISSN 2352-4316. doi:10.1016/j.eml.2024.102220.
- [101] Martin Zlatić and Marko Čanadija. Recovering Mullins damage hyperelastic behaviour with physics augmented neural networks. *Journal of the Mechanics and Physics of Solids*, 193:105839, December 2024. ISSN 0022-5096. doi:10.1016/j.jmps.2024.105839.
- [102] Martin Zlatić and Marko Čanadija. Incompressible rubber thermoelasticity: A neural network approach. *Computational Mechanics*, 71(5):895–916, 2023. ISSN 1432-0924. doi:10.1007/s00466-023-02278-y.
- [103] Jan N. Fuhg, Asghar Jadoon, Oliver Weeger, D. Thomas Seidl, and Reese E. Jones. Polyconvex neural network models of thermoelasticity. *Journal of the Mechanics and Physics of Solids*, 192:105837, November 2024. ISSN 0022-5096. doi:10.1016/j.jmps.2024.105837.

- [104] Oliver Kunc and Felix Fritzen. Finite Strain Homogenization Using a Reduced Basis and Efficient Sampling. *Mathematical and Computational Applications*, 24(2):56, June 2019. ISSN 2297-8747. doi:10.3390/mca24020056.