Mechanics Research Communications 138 (2024) 104281

Contents lists available at ScienceDirect

Mechanics Research Communications

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

# Operator learning for homogenizing hyperelastic materials, without PDE data

Hao Zhang, Johann Guilleminot *

*Department of Civil and Environmental Engineering, Duke University, Durham, NC 27708, USA*

ARTICLE INFO

*Keywords:*
Computational homogenization
Hyperelastic materials
Multiscale analysis
Operator learning

ABSTRACT

In this work, we address operator learning for stochastic homogenization in nonlinear elasticity. A Fourier neural operator is employed to learn the map between the input field describing the material at fine scale and the deformation map. We propose a variationally-consistent loss function that does not involve solution field data. The methodology is tested on materials described either by piecewise constant fields at microscale, or by random fields at mesoscale. High prediction accuracy is obtained for both the solution field and the homogenized response. We show, in particular, that the accuracy achieved with the proposed strategy is comparable to that obtained with the conventional data-driven training method.

## 1. Introduction

Multiscale analysis aims to exchange information across scales, generally by predicting some properties of interest at a coarse scale based on fine-scale material description. While homogenization can be performed analytically for a few particular cases involving "simple" microstructures, computational homogenization is generally required to tackle complex systems. The cost associated with this task is exacerbated when the so-called separation of scales cannot be stated: in this case, the homogenization operator does not completely filter subscale fluctuations, and the apparent physical quantities thus estimated exhibit statistical fluctuations. This setting requires concurrent multiscale computations where mechanistic information is exchanged back-and-forth between the material and structural scales (see, e.g., [1] for an example).

Regardless of the setting (periodic or non-periodic, with or without scale separation), homogenization comes with substantial computational burden, and many strategies have been proposed in the literature to circumvent this issue. Most frameworks involve reduced-order modeling for the microscopic problem (see, e.g., [2]), or the construction of surrogate models based on either approximation theory (see, e.g., [3, 4]) or, more recently, on the use of machine learning techniques (see, e.g., [5] and the references therein). We note that similar developments have occurred in many disciplines, and that providing an exhaustive review is beyond the scope of this paper. Examples include [6], where reduced-order modeling is combined with deep learning techniques to homogenize nonlinear hyperelastic materials, and [7], in which convolutional neural networks (CNNs) are employed to study the homogenization of two-phase heterogeneous materials. The case of history-dependent behaviors was considered in [8], using long short-term memory (LSTM) networks. In [9], the analysis of unit cell problems was realized using physics-informed neural networks (PINNs). Other relevant data-driven approaches can be found in [10, 11]; see [12] for a review.

More recently, a few operator learning techniques have emerged as a means to approximate mappings between function spaces (and to realize, in particular, zero-shot predictions at different resolutions). Such methods can, at least in principle, enable the approximation of the solution map relating the material input field to the deformation map (taken as the output field) — the homogenized response being then obtained by post-processing. Popular neural operator frameworks include (1) DeepONet [13] and variations thereof; (2) PCA-Net [14]; and (3) Fourier neural operators (FNO) [15] and its physics-informed variation (PINO) [16] (see [17] for an alternative kernel-based framework). A few attempts have been reported regarding the use of neural operators in multiscale analysis; see [18] for an application example, as well as [19] for approximation results in the case of elliptic operators. While these approaches are data-driven by construction, their practical use in such a context has yet to be fully demonstrated — especially in the context of anisotropic, nonlinear (and potentially time-dependent) responses. Their deployment as surrogates when the underlying governing equations (specified as partial differential equations (PDEs)) are known appears as a logical step to gain insights in the learning process, and to develop ad hoc theories for approximation capabilities (see [19] for thorough discussion). In this context, one common criticism about

* Corresponding author.
*E-mail address:* johann.guilleminot@duke.edu (J. Guilleminot).

https://doi.org/10.1016/j.mechrescom.2024.104281
Received 1 December 2023; Received in revised form 29 March 2024; Accepted 22 April 2024
Available online 6 May 2024

such surrogates is the necessity to proceed with offline computations, used to build training and validation datasets. This task can be computationally intensive, especially in the case of nonlinear materials. In this work, we seek to accelerate computational homogenization for hyperelastic random materials by combining operator learning and a training loss based on total energy minimization (as proposed in, e.g., [20,21] in the context of incremental homogenization). Leveraging the results presented in [16], FNO is employed as the approximator. The main contribution lies in the definition of a loss function that only involves the strain energy functional, which effectively enables training without PDE data. We further illustrate the relevance of this strategy for two material models presenting different types of continuity, and compare the performance of the proposed training method with the standard data-driven approach. The derivation of approximation results in nonlinear elasticity is left for future work.

## 2. Operator learning for homogenization

### 2.1. Modeling at microscale

Consider a heterogeneous hyperelastic solid occupying an open bounded domain $\Omega \subset \mathbb{R}^2$, with smooth boundary $\partial\Omega$, at rest (extension to the three-dimensional case is straightforward). Following standard convention, $\Omega$ is referred to as a statistical volume element (SVE) for random media. Let $\boldsymbol{\varphi} : \Omega \to \Omega^{\boldsymbol{\varphi}}$ denote the deformation map relating the position of a material point in the reference configuration (i.e., in $\Omega$) to its position in the current (deformed) configuration, denoted by $\Omega^{\boldsymbol{\varphi}}$: $\boldsymbol{x}^{\boldsymbol{\varphi}} = \boldsymbol{\varphi}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})$, where $\boldsymbol{x}^{\boldsymbol{\varphi}} \in \Omega^{\boldsymbol{\varphi}}$ and $\boldsymbol{x} \in \Omega$ are the position vectors of the material point in the current and reference configurations, respectively, and $\boldsymbol{u} : \Omega \to \mathbb{R}^2$ denotes the displacement vector. Let $\mathbb{M}_2^+$ denote the set of matrices of order two, with strictly positive determinants. The deformation gradient $[F] \in \mathbb{M}_2^+$ is then defined as $[F] = [\nabla_{\boldsymbol{x}} \boldsymbol{\varphi}]$ and is assumed to be invertible almost everywhere in $\Omega^{\boldsymbol{\varphi}}$. The right Cauchy–Green tensor is defined as $[C] = [F]^T[F]$.

In the case of hyperelastic materials, constitutive modeling is achieved by relating the first Piola–Kirchhoff stress tensor $[P]$ to the deformation gradient $[F]$ through

$$[P(\boldsymbol{x}, [F])] = \frac{\partial W}{\partial [F]}(\boldsymbol{x}, [F]),$$

where $W : \mathbb{M}_2^+ \to \mathbb{R}$ is the stored energy density function (at microscale). Sufficient conditions on $W$ to ensure well-posedness on the associated nonlinear boundary value problem are well established; see, e.g., [22,23]. The strain energy associated with any admissible deformation map $\boldsymbol{\varphi}^* \in \mathcal{F} = \{\boldsymbol{\varphi}^* : \overline{\Omega} \to \mathbb{R}^2,\ \boldsymbol{\varphi}^* = \boldsymbol{\varphi}_0 \text{ on } \partial\Omega\}$ (smoothness requirements are left undefined here for ease of exposition) is defined as $w(\boldsymbol{\varphi}^*) = \int_\Omega W(\boldsymbol{x}, [\nabla \boldsymbol{\varphi}^*])\, d\boldsymbol{x}$. It is well known (see, e.g., Theorem 4.1-2 in [22]) that in the absence of body force and Neumann boundary conditions, the deformation map $\boldsymbol{\varphi}$ satisfying

$$\boldsymbol{\varphi} = \underset{\boldsymbol{\varphi}^* \in \mathcal{F}}{\operatorname{argmin}}\ w(\boldsymbol{\varphi}^*) \tag{1}$$

is the solution to the pure-displacement boundary value problem (BVP) (and reciprocally):

$$\begin{cases} \nabla_{\boldsymbol{x}} \cdot [P] = 0\,, & \forall \boldsymbol{x} \in \Omega\,, \\ \boldsymbol{\varphi} = \boldsymbol{\varphi}_0\,, & \forall \boldsymbol{x} \in \partial\Omega\,. \end{cases} \tag{2}$$

We will leverage this result to formulate the loss function in the operator learning framework presented in Section 2.3.

### 2.2. Homogenization method in nonlinear elasticity

Following [24,25], the macroscopic deformation gradient $[\overline{F}]$ associated with $\boldsymbol{\varphi}$ is defined as

$$[\overline{F}] = \frac{1}{|\Omega|} \int_{\partial\Omega} \boldsymbol{\varphi}(\boldsymbol{x}) \otimes \boldsymbol{n}(\boldsymbol{x})\, ds = \frac{1}{|\Omega|} \int_\Omega [F(\boldsymbol{x})]\, d\boldsymbol{x},$$

where $\boldsymbol{n}(\boldsymbol{x})$ denotes the outward-pointing unit normal vector at point $\boldsymbol{x} \in \Omega$. In a multiscale setting, the macroscopic deformation gradient $[\overline{F}]$ is employed to define *kinematically uniform boundary conditions* (KUBC) on the SVE: $\boldsymbol{\varphi}(\boldsymbol{x}) = [\overline{F}]\boldsymbol{x}\,,\ \boldsymbol{x} \in \partial\Omega$. The *apparent* stored energy density function $\overline{w} : \mathbb{M}_2^+ \to \mathbb{R}$ defining the constitutive model at the coarser scale is then defined as

$$\overline{w}([\overline{F}]) = \min_{[F] \in \mathcal{K}([\overline{F}])} \frac{1}{|\Omega|} \int_\Omega W(\boldsymbol{x}, [F(\boldsymbol{x})])\, d\boldsymbol{x}\,, \tag{3}$$

where $\mathcal{K}([\overline{F}])$ denotes the set of kinematically admissible deformation gradients given by

$$\mathcal{K}([\overline{F}]) = \{[F] \mid \exists \boldsymbol{\varphi}^* \in \mathcal{F} :\ [F] = [\nabla_{\boldsymbol{x}} \boldsymbol{\varphi}^*]\ \forall \boldsymbol{x} \in \Omega\,,\ \boldsymbol{\varphi}^*(\boldsymbol{x}) = [\overline{F}]\boldsymbol{x} \text{ on } \partial\Omega\}\,. \tag{4}$$

The conventional approach to compute the apparent stored energy density function involves solving the nonlinear boundary value problem

$$\begin{cases} \nabla_{\boldsymbol{x}} \cdot [P(\boldsymbol{x}, [\nabla \boldsymbol{\varphi}])] = 0\,, & \forall \boldsymbol{x} \in \Omega\,, \\ \boldsymbol{\varphi}(\boldsymbol{x}) = [\overline{F}]\boldsymbol{x}\,, & \forall \boldsymbol{x} \in \partial\Omega\,, \end{cases} \tag{5}$$

and then calculating the apparent stored energy function using ergodic averaging:

$$\overline{w}([\overline{F}]) = \frac{1}{|\Omega|} \int_\Omega W(\boldsymbol{x}, [F(\boldsymbol{x})])\, d\boldsymbol{x}\,. \tag{6}$$

The apparent stored energy $\overline{w}$ is thus defined as the volume average of the (fine scale) strain energy density function taken at the minimizer of the total energy functional. When the domain $\Omega$ is sufficiently large, the so-called separation of scales holds, meaning that the response of the material becomes independent of the realization of the random media. This convergence is well discussed in the literature of multiscale analysis and will not be studied in this work.

### 2.3. Using a neural operator for homogenization

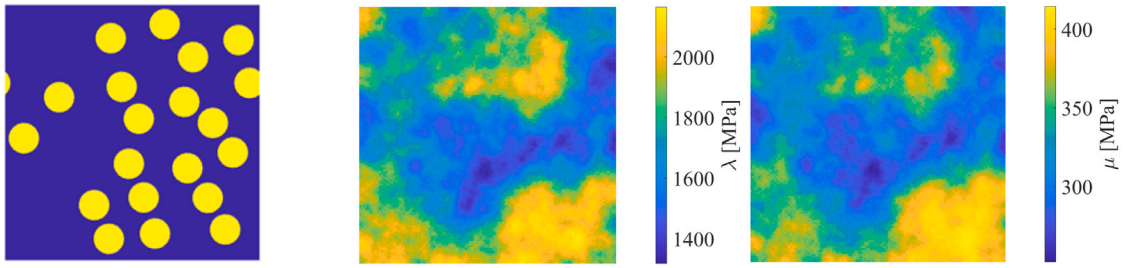
As previously indicated, homogenization is a procedure that aims to determine the apparent stored energy function, given *(i)* a macroscopic deformation gradient $[\overline{F}]$ and *(ii)* a sample of the material at fine scale. This task can be greatly accelerated by using a operator learning technique that approximates the mapping between the input field defining the material at fine scale and the deformation map. Specifically, operator learning seeks to build a surrogate for the map $\mathcal{G}^\dagger : \mathcal{A} \to \mathcal{F}^\circ$, where $\mathcal{A}$ denotes the input function space and $\mathcal{F}^\circ$ is the output function space containing all admissible deformation maps from $\Omega$ to $\mathbb{R}^2$. Note that we do not consider the closure $\overline{\Omega}$ in the learning task, since the deformation map is prescribed on $\partial\Omega$. Our goal is therefore to propose, and assess the relevance of, a training strategy that does not involve solving the nonlinear boundary value problem defined in Eq. (5). In this context, the Fourier neural operator (FNO) framework proposed in [15] is selected. This approach relies on the combination of lifting and projection operators with a flow of trainable operators. More specifically, the Fourier neural operator defines an approximation $\mathcal{G}$ to $\mathcal{G}^\dagger$ as

$$\mathcal{G} = \mathcal{Q} \circ \mathcal{L}_T \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{P}\,, \tag{7}$$

where the input and output layers, denoted by $\mathcal{P}$ and $\mathcal{Q}$, respectively, are lifting and projection operators defined using neural networks. Each element in the sequence of operators $\{\mathcal{L}_i\}_{i=1}^T$ is defined as

$$\mathcal{L}_i\{v\}(\boldsymbol{x}) = \sigma([W_i]v(\boldsymbol{x}) + \boldsymbol{b}_i + \mathcal{K}_i\{v\}(\boldsymbol{x}))\,, \tag{8}$$

where $\sigma$ is an activation function, $\{[W_i], \boldsymbol{b}_i\}$ is a set of coefficients defining an affine transformation (in the $i$th hidden layer), $v : \Omega \to \mathbb{R}^{d_c}$ in a function defining the input for the layer under consideration, with




(a) Scenario 1: Microstructure sample

(b) Scenario 2: Samples of $\lambda$ (left) and $\mu$ (right)

**Fig. 1.** Examples of material samples for the two scenarios.

$d_c$ the so-called channel width (which can be layer-dependent), and $\mathcal{K}_i$ is a convolutional integral operator given by

$$\mathcal{K}_i\{v\}(\boldsymbol{x}) = \int_{\Omega} \kappa_i(\boldsymbol{x}-\boldsymbol{y})v(\boldsymbol{y})\,d\boldsymbol{y}\,, \quad \forall \boldsymbol{x}\in\Omega\,, \tag{9}$$

where $\kappa_i$ is a matrix-valued integral kernel. The operator $\mathcal{K}_i$ is conveniently computed using forward and inverse Fourier transforms, parameterized by trainable Fourier coefficients. The latter, combined with the parameters involved in $\mathcal{P}$ and $\mathcal{Q}$, constitute the tunable parameters for the FNO. In practice, a convergence analysis with respect to the parameters defining the architecture (including the channel width, the number of modes retained in Fourier space, and the depth $T$) must be performed. Interested readers are referred to [15,16,26–28] for further (technical) details about FNO and variations thereof. Examples of applications in computational mechanics can be found in [29,30]. Note that an exhaustive comparison between state-of-the-art operator learning strategies is beyond the scope of this work; see [26,31] for instance.

In this work, and following the discussion in Section 2.1, we propose to use a physics-informed loss function defined as

$$\mathcal{L}(\theta^*) = \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{1}{|\Omega_i|}\int_{\Omega_i} W\left(\boldsymbol{x}, [\nabla_{\boldsymbol{x}}\hat{\boldsymbol{\varphi}}^{(i)}(\theta^*)]\right)\,d\boldsymbol{x}\,, \tag{10}$$

where $\theta^*$ collects all trainable parameters in the neural operator, $N_t$ is the number of samples, and $\hat{\boldsymbol{\varphi}}^{(i)}(\theta^*)$ is the deformation map predicted by the neural operator parameterized by $\theta^*$ for the $i$th input sample $\Omega_i$ of the material at fine scale. Once the above loss function has been minimized with respect to $\theta^*$, with a solution denoted by $\theta$, the apparent stored energy associated with $\Omega_i$ is then approximated as

$$\overline{w} \approx (1/|\Omega_i|)\int_{\Omega_i} W(\boldsymbol{x}, [\nabla_{\boldsymbol{x}}\hat{\boldsymbol{\varphi}}(\theta)])\,d\boldsymbol{x}\,. \tag{11}$$

## 3. Applications

Two scenarios involving different types of regularity are specifically investigated below. In the first scenario, microscopic modeling is realized using a two-phase microstructure with piece-constant Mooney–Rivlin materials. In the second example, we consider mesoscopic modeling where randomness is introduced by representing the material parameters in the Neo-Hookean model as non-Gaussian random fields. These scenarios are described in Section 3.1.

### 3.1. Material descriptions

#### 3.1.1. Constitutive modeling

In this work, we employ the Neo-Hookean model and Mooney–Rivlin model given by [22,32] for the sake of illustration. The Neo-Hookean model reads as

$$W([F]) = \frac{\mu}{2}(\|[F]\|^2 - 2) + \frac{\lambda}{2}(\det([F]) - 1)^2 - \mu \ln\det([F]),$$

where $\lambda$ and $\mu$ are the Lamé parameters. The Mooney–Rivlin model is defined as

$$W([F]) = a(\|[F]\|^2 - 2) + b(\|\mathrm{Cof}([F])\|^2 - 3) + \frac{s_1}{2}(\det([F]) - 1)^2 - s_2 \ln\det([F])$$

where $a$ and $b$ are strictly positive material coefficients such that $2a + 2b = \mu$, $s_1 = \lambda - 4b$, and $s_2 = 2a + 4b$. Following standard notation in nonlinear elasticity, $\|[F]\|^2 = \mathrm{tr}([F][F]^T)$, $\det([A])$ and $\mathrm{Cof}([A])$ are the determinant and cofactor matrix of any matrix $[A]$, respectively.

#### 3.1.2. Description at microscale

In the first scenario, we consider a two-phase random microstructure. Particles are randomly placed within $\Omega$ using the packing algorithm detailed in [33], with periodicity enforced along the boundary $\partial\Omega$. Samples are shown in the left panel in Fig. 1. As previously indicated, the inclusion and matrix phases are modeled as Mooney–Rivlin materials with deterministic properties. Material parameters for the matrix phase are given by $a^{(m)} = 164.0625$ [MPa], $b^{(m)} = 10$ [MPa], $\lambda^{(m)} = 1531.14$ [MPa], while parameters for the inclusion phase are given by $a^{(m)} = 10a^{(m)}$, $b^{(m)} = 2b^{(m)}$, $\lambda^{(f)} = 10\lambda^{(m)}$. This stochastic microstructure is completely defined by the indicator function of any of the phases, say, the inclusion phase. We denote by $\{A(\boldsymbol{x}), \boldsymbol{x}\in\Omega\}$ this random field, with values in $\{0,1\}$.

#### 3.1.3. Description at mesoscale

In the second scenario, we adopt a mesoscopic perspective where random fields of elastic parameters are used to introduce stochasticity. These random fields are analog, in essence, to those obtained from pointwise stochastic homogenization without separation of scales (in a moving-window fashion). Let $\boldsymbol{A}(\boldsymbol{x}) = (\lambda(\boldsymbol{x}), \mu(\boldsymbol{x}))^T$ denote the random vector gathering the Lamé parameters at point $\boldsymbol{x}\in\Omega$. Following previous work by the authors [34–36], a prior model for the non-Gaussian field $\{\boldsymbol{A}(\boldsymbol{x}), \boldsymbol{x}\in\Omega\}$ is considered:

$$\boldsymbol{A}(\boldsymbol{x}) = \mathcal{T}(\boldsymbol{\Xi}(\boldsymbol{x})),$$

where $\mathcal{T}$ is a nonlinear (transport) map and $\{\boldsymbol{\Xi}(\boldsymbol{x}) = (\Xi_1(\boldsymbol{x}), \Xi_2(\boldsymbol{x}))^T, \boldsymbol{x}\in\Omega\}$ is a bivariate normalized stationary Gaussian random field (with statistically independent components). Each component $\{\Xi_i(\boldsymbol{x}), \boldsymbol{x}\in\Omega\}$ is assumed to be isotropic and defined by a Matérn-type covariance function, where the correlation length parameter is chosen as 0.1 and the smoothness parameter is set to 1. The transport map $\mathcal{T}$ is defined using an information-theoretic approach (see [3,37,38] for information-theoretic models in stochastic anisotropic nonlinear elasticity). We specifically define $\lambda$ through

$$\lambda(\boldsymbol{x}) = F^{-1}_{\mathcal{G}(s_1^\lambda, s_2^\lambda)}\left(F_{\mathcal{N}(0,1)}(\Xi_1(\boldsymbol{x}))\right)\,, \tag{12}$$

while the second Lamé parameter is defined as

$$\mu(\boldsymbol{x}) = F^{-1}_{\mathcal{G}(s_1^\mu, s_2^\mu)}\left(\rho F_{\mathcal{N}(0,1)}(\Xi_1(\boldsymbol{x})) + \sqrt{1-\rho^2}F_{\mathcal{N}(0,1)}(\Xi_2(\boldsymbol{x}))\right), \tag{13}$$

where $F$ denotes the cumulative density function of the distribution specified as subscript, $\mathcal{G}$ denotes the Gamma law (with shape and scale

parameters indicated within parentheses), $\mathcal{N}(0,1)$ denotes the normalized Gaussian law, and $\rho$ is the Pearson correlation coefficient between the Lamé parameters (at any location fixed in $\Omega$). The parameters $(s_1^\lambda, s_2^\lambda)$ and $(s_1^\mu, s_2^\mu)$ in the Gamma laws are calibrated such that $\underline{\lambda} = \mathbb{E}(\lambda(\boldsymbol{x})) = 1{,}750$ [MPa], $\delta_\lambda = 0.3$, $\underline{\mu} = \mathbb{E}(\mu(\boldsymbol{x})) = 328.125$ [MPa], and $\delta_\mu = 0.3$, where $\underline{\lambda}$ and $\underline{\mu}$ denote the means of the Lamé parameters, respectively, and $\delta_\lambda$ and $\delta_\mu$ denote their coefficients of variation. The correlation coefficient is chosen as $\rho = 0.9$, following the multiscale-informed results in [39]. Fig. 1 shows a pair of samples of the pointwise correlated non-Gaussian random fields of Lamé parameters. With the elastic contrast considered in this study, the coefficient of variation of the apparent strain energy density is about 2% for scenario 1, and about 3% for scenario 2.

### 3.2. Description of training strategies

In what follows, we denote by $\text{FNO}_{\text{mic}}$ and $\text{FNO}_{\text{mes}}$ the FNOs associated with the first and second scenarios, respectively. In the results presented below, the discretization of the input fields contains 128 elements per edge, yielding $128 \times 128 = 16{,}384$ elements in total. The designed loss function (Eq. (10)) is employed to train the neural operators without PDE data. All FNOs contain four Fourier layers (as in [16]), and convergence results with respect to the number of Fourier modes and width are presented in Section 3.3. The GELU [40] activation function is employed for $\text{FNO}_{\text{mes}}$, while the Leaky-ReLU activation function is used to better capture sharp gradients in $\text{FNO}_{\text{mic}}$ (with a piecewise constant input field). The batch size is chosen as 40. The Adam optimization algorithm [41] is employed to minimize the loss function. The initial learning rate is set to 0.001, and decays by half after every 50 epochs. Each neural operator is trained using a total number of 300 epochs. In the case of $\text{FNO}_{\text{mic}}$, transfer learning involving neural operators calibrated with a lower finite element resolution (i.e., a coarser mesh) is used to facilitate training [7,9].

In order to evaluate accuracy, two metrics are considered. The first metric is a relative $H^1$ error (RHE) involving the displacement field and the deformation gradient:

$$\text{RHE} = \frac{1}{N_v}\sum_{i=1}^{N_v}\left(\frac{\|\boldsymbol{u}^{(i)} - \hat{\boldsymbol{u}}^{(i)}\|_{L^2}^2 + \|[F]^{(i)} - [\hat{F}]^{(i)}\|_{L^2}^2}{\|\boldsymbol{u}^{(i)}\|_{L^2}^2 + \|[F]^{(i)}\|_{L^2}^2}\right)^{1/2}, \tag{14}$$

where the superscript $i$ indicates predictions obtained for the $i$th sample of the input field, variables with and without a hat denote approximated and reference quantities of interest, respectively. The second metric measures the relative error in energy (REE):

$$\text{REE} = \frac{1}{N_v}\sum_{i=1}^{N_v}\left|\frac{\overline{w}^{(i)} - \hat{\overline{w}}^{(i)}}{\overline{w}^{(i)}}\right|. \tag{15}$$

### 3.3. Results

We first investigate the architecture of the FNO models, including the number of retained low-frequency Fourier modes and FNO model width. Recall that in the FNO approach, the input is lifted to a higher-dimensional space, transformed iteratively using linear and nonlinear operators, and then projected back. For each neural operator, $N_t = 1000$ SVE samples are employed to train the FNO, while $N_v = 1000$ unseen samples are used to test using the REE metric. These values were determined from a convergence analysis, with an error decay behaving as $1/\sqrt{N_t}$ (as reported in [19]). Training is realized for each specific choice of the macroscopic Cauchy–Green tensor, here taken as $[\overline{C}] = [1.4, 0; 0, 1.0]$. The macroscopic deformation gradient $[\overline{F}]$ is then calculated using the Cholesky decomposition of $[\overline{C}]$. Results for the accuracy test for the two neural operators are shown in Fig. 2. The number of Fourier modes is increased from 5 to 30, and the channel width is increased from 24 to 80. In the case of $\text{FNO}_{\text{mic}}$, the performance of the neural operator improves with the increase of both the number of Fourier modes and channel width. Best performance for this scenario is obtained with 30 Fourier modes and a width of 64. It is noticeable that the performance of the FNO improves with additional Fourier modes, in contrast with the results presented for Voronoi microstructures in [19]. This behavior may be explained by the fact that the two-phase microstructure considered in this paper can be seen as an intermediate configuration where the consideration of a matrix phase (with constant parameters) introduces additional regularity. In the case of $\text{FNO}_{\text{mes}}$, best performance is realized with 20 Fourier modes and a width of 80. The error increases for a number of modes larger than 15 for $d_c = 48$, and larger than 20 for $d_c \in \{24, 64, 80\}$. This observation is consistent with the results reported in [19] for a "smooth" input field analogous to the mesoscopic case; see [42] for numerical studies on the choice of hyperparameters, as well as [43] for theoretical analysis. In the results presented below, the number of Fourier modes is set as 20 in the two FNOs, and the width in the Fourier layers is set to 64.

The training loss history is shown in Fig. 3. It can be seen that the loss decays rapidly and almost monotonically during the training process. This suggests that the designed loss function provides an efficient path towards *data-free* (in the sense of PDE solving) training for the neural operator in the multiscale setting.

Next, we test the prediction accuracy of the two FNOs using the metrics defined in Section 3.2. The neural operators are tested on different macroscale deformation gradients $[\overline{F}]$ (recall that the neural operators are trained for each case; the important question of generalization with respect to boundary conditions is left for future work). Four different macroscopic loading conditions are employed, including uniaxial and biaxial tensile tests, and mixed tensile-shear loading. In addition, the performance of the proposed formulation is also compared with the performance achieved by using a standard data-driven loss function:

$$\mathcal{L}_{\text{data}} = \frac{1}{N_t}\sum_{i=1}^{N_t}\left(\|\boldsymbol{u}^{(i)} - \hat{\boldsymbol{u}}^{(i)}\|_{L^2}^2 + \|[F]^{(i)} - [\hat{F}]^{(i)}\|_{L^2}^2\right), \tag{16}$$

where $\boldsymbol{u}^{(i)}$ and $[F]^{(i)}$ are the reference data obtained by solving the nonlinear PDEs (see Eq. (5)). Regarding the nonlinear finite element analysis, an optimized in-house C++ implementation, verified by the method of manufactured solutions, was used. The absolute and relative tolerances were set to 1e–8 and 1e–9 in the Newton–Raphson solver, respectively, with a maximum number of nonlinear iterations set to 8.

Prediction errors for the two scenarios and the two loss functions (data-driven and proposed) are shown in Fig. 4(a) for RHE, and in Fig. 4(b) for REE. For RHE, a similar evolution of the prediction error is observed across all four types of boundary conditions, regardless of training loss (with or without training data). Errors for the first scenario involving piece-constant material parameters are larger than for the second scenario, for both the proposed training loss and the data-driven loss (a result consistent with the observations reported in [19]). This may be explained by the smoothness in the solution field induced by mesoscopic modeling. Comparing errors between the two training strategies, errors for the physics-informed loss are slightly larger than the errors obtained with the standard data-driven loss. Specifically, the maximum error for scenario 1 is 2.7% for the proposed loss, and 1.9% for the data-driven loss. In scenario 2, the maximum error is 1.7% for the physics-informed loss, and 0.9% for the data-driven loss.

Consistent results are obtained for REE. The evolution across all boundary conditions is also similar for both training losses. Errors for scenario 1 remain larger than for second scenario 2. However, it is noticeable that for the first scenario (with microscopic modeling), the error estimated with the proposed loss (with a maximum error of 3.7%) is indeed smaller than the one obtained with data (with a maximum error equal to 4.9%). Errors associated with the second scenario are fairly similar for both strategies, regardless of the boundary conditions. The maximum error is obtained with mixed boundary conditions, with error equal to 0.6% and 0.3% for the proposed loss and data-driven loss,

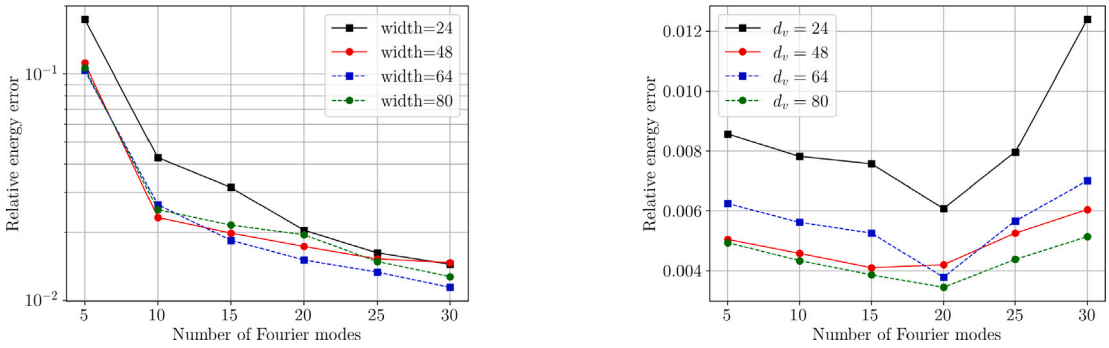



**Fig. 2.** Errors on test datasets for $\text{FNO}_{\text{mic}}$ (left) and $\text{FNO}_{\text{mes}}$ (right), for different numbers of Fourier modes and channel width in the hidden layers.

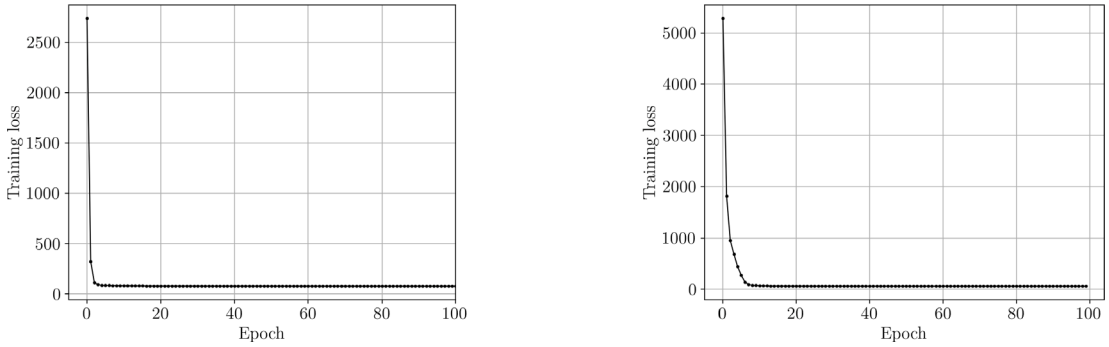



**Fig. 3.** Training loss history for $\text{FNO}_{\text{mic}}$ (left) and $\text{FNO}_{\text{mes}}$ (right).

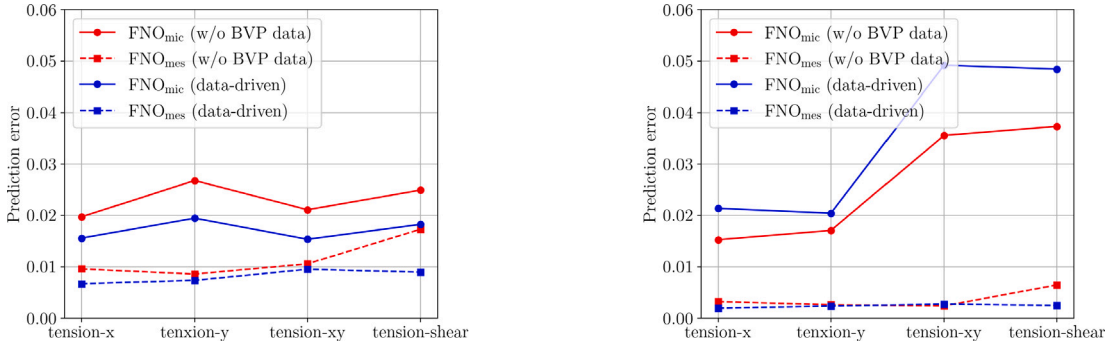


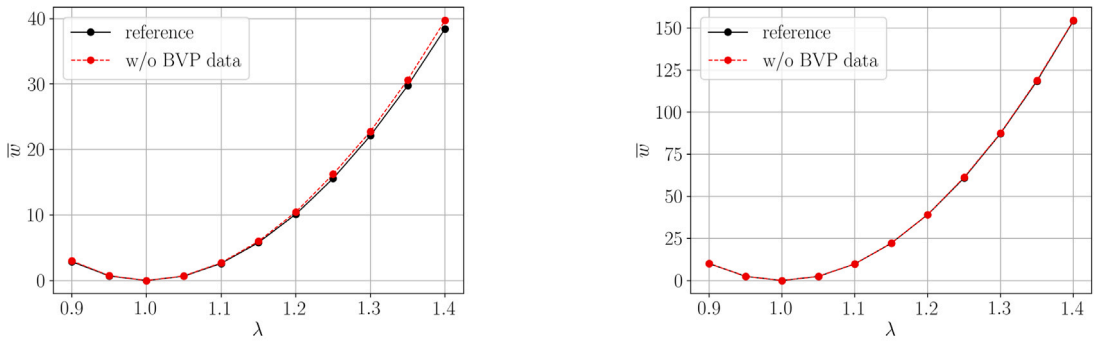

**Fig. 4.** Prediction errors for the two scenarios, using RHE (left) and REE (right).

respectively. The apparent stored energy density function $\overline{w}$ associated with a macroscopic right Cauchy–Green tensor taken as $[\overline{C}] = [\lambda, 0; 0, \lambda]$ (and given material samples) is shown in Fig. 5 for both scenarios. Good agreement is observed with the reference solution obtained by the finite element method, especially for scenario 2.

While results involving RHE and REE are more relevant in the context of multiscale predictions, it is instructive to also compare performance in terms of the displacement solution field (for one sample for each input field). To this end, denote by $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$ the reference solution fields, obtained by solving the nonlinear boundary value problem with the finite element method; standard convergence analyses were carried out and are not reported below for the sake of conciseness. Let $\hat{\boldsymbol{u}}_x$ and $\hat{\boldsymbol{u}}_y$ be the displacement fields predicted by the trained neural operator, for a given choice of the loss function. These displacement fields, as well as pointwise errors, are shown in Figs. 6 and 7 for the two scenarios and mixed tension-shear loading conditions. It is seen that the pointwise error is larger for scenario 1 than for scenario 2, regardless of the training method — an observation that may be explained by the difference in regularity in the two input fields. Unsurprisingly, the data-driven approach delivers slightly more accurate results and exhibits uniform convergence, especially for scenario 1. On the other hand, the pointwise prediction accuracy of the proposed training method remains fairly close to the one obtained with the conventional data-driven approach for both the microscopic and mesoscopic models.

For the microscale problem, training for each epoch takes 2.2 s with our loss function, and 2.1 s for the data-driven approach; for the mesoscale problem, training requires 2.33 s for the proposed loss technique, and 2.3 s for the data-based loss function.

Regarding computational speedup, solving the nonlinear BVP using the finite element method takes 2.93 s and 1.57 s per sample for the microscopic and mesoscopic problems, respectively. The cost of generating the training dataset with $N_t = 1000$ samples (for the data-driven loss) is therefore $\approx 2.93 \times 1000 = 2930$ s (about 49 min) and $\approx 1.57 \times 1000 = 1570$ s (about 26 min) for the two multiscale problems (with sequential execution). This highlights the substantial reduction in the computational cost offered by the proposed training strategy. Finally, the forward evaluation of the solution field takes 0.06 and 0.03 s per sample for the microscopic and mesoscopic problems, respectively, on a single Nvidia RTX A6000 GPU. These computing times are




(a) Scenario 1 (discontinuous input model)

(b) Scenario 2 (smooth input model)

**Fig. 5.** Apparent nonlinear response estimated under biaxial tensile testing. Reference is obtained by the finite element method.

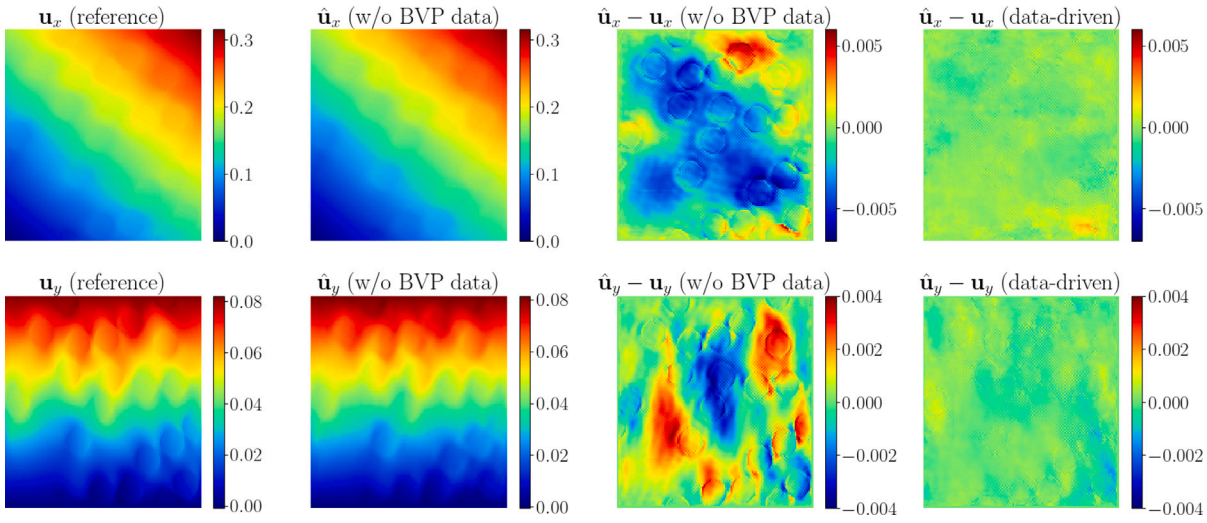



**Fig. 6.** Reference and approximated displacement fields for the FNOs trained using our proposed PDE-data-free method and the data-driven method (scenario 1). The Cauchy–Green tensor is set to $[1.3, 0.2; 0.2, 1.2]$.

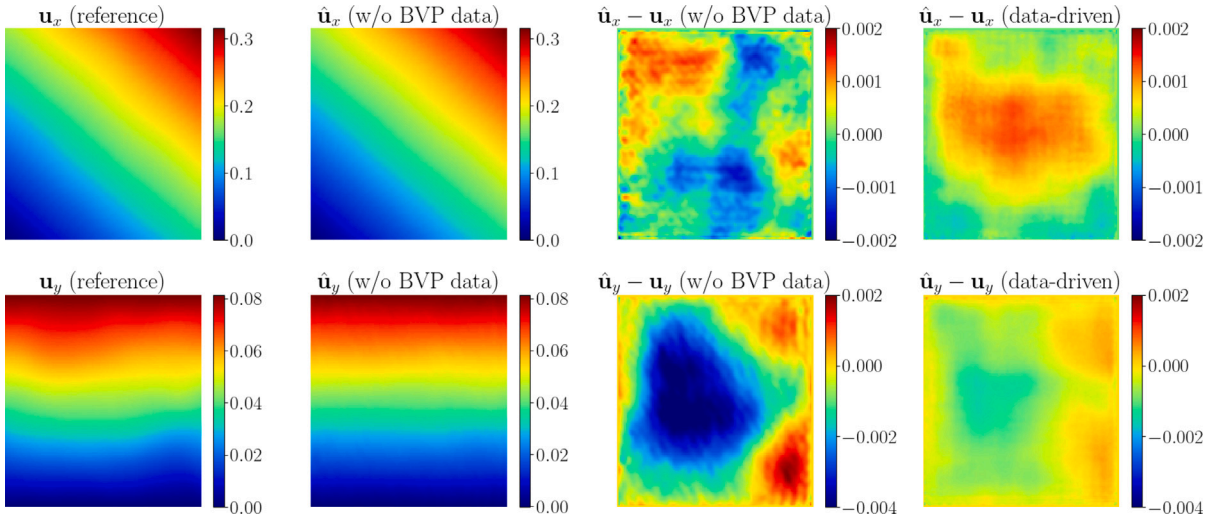



**Fig. 7.** Reference and approximated displacement fields for the FNOs trained using our proposed PDE-data-free method and the data-driven method (scenario 2). The Cauchy–Green tensor is set to $[1.3, 0.2; 0.2, 1.2]$.

about 50 times smaller than those associated with the corresponding nonlinear finite element analyses (with in-house C++ algorithms).

## 4. Conclusion

In this work, we investigated operator learning for stochastic homogenization in hyperelastic random media. The approach involves Fourier neural operators, which are used to approximate the solution map between the input field describing the material at a scale of interest, and the deformation map. We proposed a PDE-data-free training strategy that enables the computation of the effective stored energy function. The approach was tested on material structures presenting piecewise constant or mean-squared continuous fluctuations. Good accuracy in terms of both the solution fields and multiscale results was observed. Numerical results also show that the prediction accuracy of this proposed training method is comparable with the conventional data-driven method. Future directions include *(i)* generalization in terms of macroscopic boundary conditions, *(ii)* the analysis of expressivity for non-smooth input fields, *(iii)* the mathematical analysis of approximation capabilities, and *(iv)* applications to other variational formulations (e.g., for plasticity, strain and stress gradient models for generalized continua, and viscoelasticity).

## CRediT authorship contribution statement

**Hao Zhang:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Johann Guilleminot:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Johann Guilleminot reports financial support was provided by National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgments

The financial support of the National Science Foundation, USA under grant DGE-2022040 is gratefully acknowledged.

## References

[1] Frédéric Feyel, Jean-Louis Chaboche, FE2 multiscale approach for modelling the elastoviscoplastic behaviour of long fibre SiC/Ti composite materials, Comput. Methods Appl. Mech. Engrg. 183 (3–4) (2000) 309–330.
[2] Julien Yvonnet, Q.-C. He, The reduced model multiscale method (R3M) for the non-linear homogenization of hyperelastic media at finite strains, J. Comput. Phys. 223 (1) (2007) 341–368.
[3] B. Staber, J. Guilleminot, Functional approximation and projection of stored energy functions in computational homogenization of hyperelastic materials: A probabilistic perspective, Comput. Methods Appl. Mech. Engrg. 313 (2017) 1–27.
[4] Miguel A Bessa, Ramin Bostanabad, Zeliang Liu, Anqi Hu, Daniel W Apley, Catherine Brinson, Wei Chen, Wing Kam Liu, A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality, Comput. Methods Appl. Mech. Engrg. 320 (2017) 633–667.
[5] Souhail Chaouch, Julien Yvonnet, An unsupervised machine learning approach to reduce nonlinear FE2 multiscale calculations using macro clustering, Finite Elem. Anal. Des. 229 (2024) 104069.
[6] Satyaki Bhattacharjee, Karel Matouš, A nonlinear manifold-based reduced order model for multiscale analysis of heterogeneous hyperelastic materials, J. Comput. Phys. 313 (2016) 635–653.
[7] Chengping Rao, Yang Liu, Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization, Comput. Mater. Sci. 184 (2020) 109850.
[8] Hernan J. Logarzo, German Capuano, Julian J. Rimoli, Smart constitutive laws: Inelastic homogenization through machine learning, Comput. Methods Appl. Mech. Engrg. 373 (2021) 113482.
[9] Wing Tat Leung, Guang Lin, Zecheng Zhang, NH-PINN: Neural homogenization-based physics-informed neural network for multiscale problems, J. Comput. Phys. 470 (2022) 111539.
[10] Ehsan Motevali Haghighi, Seonhong Na, A multifeatured data-driven homogenization for heterogeneous elastic solids, Appl. Sci. 11 (19) (2021) 9208.
[11] Ling Wu, Van-Dung Nguyen, Laurent Adam, Ludovic Noels, An inverse micromechanical analysis toward the stochastic homogenization of nonlinear random composites, Comput. Methods Appl. Mech. Engrg. 348 (2019) 97–138.
[12] Dana Bishara, Yuxi Xie, Wing Kam Liu, Shaofan Li, A state-of-the-art review on machine learning-based multiscale modeling, simulation, homogenization and design of materials, Arch. Comput. Methods Eng. 30 (1) (2023) 191–222.
[13] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, George Em Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, Nat. Mach. Intell. 3 (3) (2021) 218–229.
[14] Kaushik Bhattacharya, Bamdad Hosseini, Nikola B Kovachki, Andrew M Stuart, Model reduction and neural networks for parametric PDEs, SMAI J. Comput. Math. 7 (2021) 121–157.
[15] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, Anima Anandkumar, Fourier neural operator for parametric partial differential equations, 2020, arXiv preprint arXiv:2010.08895.
[16] Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, Anima Anandkumar, Physics-informed neural operator for learning partial differential equations, 2021, arXiv preprint arXiv:2111.03794.
[17] Pau Batlle, Matthieu Darcy, Bamdad Hosseini, Houman Owhadi, Kernel methods are competitive for operator learning, J. Comput. Phys. 496 (2024) 112549.
[18] Burigede Liu, Nikola Kovachki, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, Andrew M Stuart, Kaushik Bhattacharya, A learning-based multiscale method and its application to inelastic impact problems, J. Mech. Phys. Solids 158 (2022) 104668.
[19] Kaushik Bhattacharya, Nikola Kovachki, Aakila Rajan, Andrew M Stuart, Margaret Trautner, Learning homogenization for elliptic operators, 2023, arXiv preprint arXiv:2306.12006.
[20] Christian Miehe, Strain-driven homogenization of inelastic microstructures and composites based on an incremental variational formulation, Int. J. Numer. Methods Eng. 55 (11) (2002) 1285–1322.
[21] Christian Miehe, Computational micro-to-macro transitions for discretized microstructures of heterogeneous materials at finite strains based on the minimization of averaged incremental energy, Comput. Methods Appl. Mech. Engrg. 192 (5–6) (2003) 559–591.
[22] Philippe G. Ciarlet, Mathematical Elasticity: Three-Dimensional Elasticity, SIAM, 2021.
[23] John M. Ball, Convexity conditions and existence theorems in nonlinear elasticity, Arch. Ration. Mech. Anal. 63 (1976) 337–403.
[24] Rodney Hill, On constitutive macro-variables for heterogeneous solids at finite strain, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 326 (1565) (1972) 131–147.
[25] Rodney Hill, J.R. Rice, Elastic potentials and the structure of inelastic constitutive laws, SIAM J. Appl. Math. 25 (3) (1973) 448–461.
[26] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, Anima Anandkumar, Neural operator: Learning maps between function spaces, J. Mach. Learn. Res. 24 (89) (2023) 1–97.
[27] Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, Anima Anandkumar, Fourier neural operator with learned deformations for PDEs on general geometries, 2022, arXiv:2207.05209.
[28] Nikola B. Kovachki, Samuel Lanthaler, Andrew M. Stuart, Operator learning: Algorithms and analysis, 2024, arXiv:2402.15715.
[29] Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, Sally M Benson, U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow, Adv. Water Resour. 163 (2022) 104180.
[30] Huaiqian You, Quinn Zhang, Colton J Ross, Chung-Hao Lee, Yue Yu, Learning deep implicit Fourier neural operators (IFNOs) with applications to heterogeneous material modeling, Comput. Methods Appl. Mech. Engrg. 398 (2022) 115296.
[31] Lu Lu, Xuhui Meng, Shengze Cai, Zhiping Mao, Somdatta Goswami, Zhongqiang Zhang, George Em Karniadakis, A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data, Comput. Methods Appl. Mech. Engrg. 393 (2022) 114778.
[32] Raymond W. Ogden, Non-Linear Elastic Deformations, Courier Corporation, 1997.
[33] Monica Skoge, Aleksandar Donev, Frank H. Stillinger, Salvatore Torquato, Packing hyperspheres in high-dimensional Euclidean spaces, Phys. Rev. E 74 (2006) 041127.

[34] Shanshan Chu, Johann Guilleminot, Stochastic multiscale modeling with random fields of material properties defined on nonconvex domains, Mech. Res. Commun. 97 (2019) 39–45.
[35] Peiyi Chen, Johann Guilleminot, Spatially-dependent material uncertainties in anisotropic nonlinear elasticity: Stochastic modeling, identification, and propagation, Comput. Methods Appl. Mech. Engrg. 394 (2022) 114897.
[36] Shanshan Chu, Johann Guilleminot, Cambre Kelly, Bijan Abar, Ken Gall, Stochastic modeling and identification of material parameters on structures produced by additive manufacturing, Comput. Methods Appl. Mech. Engrg. 387 (2021) 114166.
[37] Brian Staber, Johann Guilleminot, Stochastic modeling of a class of stored energy functions for incompressible hyperelastic materials with uncertainties, C. R. Méc. 343 (9) (2015) 503–514.
[38] Brian Staber, Johann Guilleminot, Stochastic modeling of the ogden class of stored energy functions for hyperelastic materials: the compressible case, ZAMM-J. Appl. Math. Mech./Z. Angew. Math. Mech. 97 (3) (2017) 273–295.
[39] Darith-Anthony Hun, Johann Guilleminot, Julien Yvonnet, Michel Bornert, Stochastic multiscale modeling of crack propagation in random heterogeneous media, Internat. J. Numer. Methods Engrg. 119 (13) (2019) 1325–1344.
[40] Dan Hendrycks, Kevin Gimpel, Gaussian error linear units (GELUs), 2016, arXiv preprint arXiv:1606.08415.
[41] Diederik P. Kingma, Jimmy Ba, Adam: A method for stochastic optimization, 2017, arXiv:1412.6980.
[42] Maarten V. Hoop, Daniel deHuang, Elizabeth ZhengyuQian, Andrew M. Stuart, The cost-accuracy trade-off in operator learning with neural networks, J. Mach. Learn. 1 (3) (2022) 299–341.
[43] Nikola Kovachki, Samuel Lanthaler, Siddhartha Mishra, On universal approximation and error bounds for Fourier neural operators, J. Mach. Learn. Res. 22 (290) (2021) 1–76.