

Reduced-order structure-property linkages for stochastic metamaterials

Hooman Danesh^{1,2,*}, Maruthi Annamaraju², Tim Brepols¹, Stefanie Reese^{1,3} and Surya R. Kalidindi²

¹*Institute of Applied Mechanics, RWTH Aachen University, Mies-van-der-Rohe-Str. 1, 52074 Aachen, Germany*

²*George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 801 Ferst Drive, Atlanta, Georgia 30332, USA*

³*University of Siegen, Adolf-Reichwein-Str. 2a, 57076 Siegen, Germany*



(Received 2 May 2025; accepted 7 July 2025; published 21 July 2025)

The capabilities of additive manufacturing have facilitated the design and production of mechanical metamaterials with diverse unit cell geometries. Establishing linkages between the vast design space of unit cells and their effective mechanical properties is critical for the efficient design and performance evaluation of such metamaterials. However, physics-based simulations of metamaterial unit cells across the entire design space are computationally expensive, necessitating a materials informatics framework to efficiently capture complex structure-property relationships. In this work, principal component analysis of two-point correlation functions is performed to extract the salient features from a large dataset of randomly generated 2D metamaterials. Physics-based simulations are performed using a fast Fourier transform (FFT)-based homogenization approach to efficiently compute the homogenized effective elastic stiffness across the extensive unit cell designs. Subsequently, Gaussian process regression is used to generate reduced-order surrogates, mapping unit cell designs to their homogenized effective elastic constants. It is demonstrated that the adopted workflow enables a high-value low-dimensional representation of the voluminous stochastic metamaterial dataset, facilitating the construction of robust structure-property maps. Finally, an uncertainty-based active learning framework is utilized to train a surrogate model with a significantly smaller number of data points compared to the original full dataset. It is shown that a dataset as small as 0.61% of the entire dataset is sufficient to generate accurate and robust structure-property maps.

DOI: [10.1103/8zzt-4b7z](https://doi.org/10.1103/PhysRevMaterials.9.075201)

I. INTRODUCTION

The manufacturability of metamaterials has significantly advanced because of recent developments in additive manufacturing technologies [1–3]. While this enables a broad range of feasible geometries, it also presents a considerable challenge, as engineers must explore an increasingly complex design space to identify structures that meet specific performance requirements. Although physics-based simulations are capable of predicting the response of a given design with high accuracy, their computational cost becomes prohibitive as the design space expands. This not only limits efficient design space exploration, but also renders inverse design approaches largely impractical. Consequently, there is a critical need for accurate and efficient surrogate models that can establish reliable mappings between the geometry of cellular structures and their effective properties (i.e., structure-property linkages). Such models not only enable real-time parametric studies and sensitivity analyses, but also serve as a foundation for future inverse design and material discovery frameworks, where rapid evaluation of thousands of candidate structures is

required. By drastically reducing computational cost, surrogate models make it feasible to explore, screen, and optimize within large design spaces that would otherwise be intractable using conventional physics-based simulations.

Data-driven machine learning approaches have been widely employed to establish structure-property linkages for various metamaterial families and diverse mechanical properties [4–13]. These approaches range from different ensemble learning algorithms [4–8] to various neural network architectures [9–13]. Focusing on different approaches is beyond the scope of this paper, and interested readers are referred to recently published review articles on the application of machine learning for metamaterials [14–17]. Such data-driven approaches typically require large datasets to train accurate and robust models, as they rely solely on empirical patterns rather than physical knowledge. Although physics-informed approaches [18,19] may reduce the data required, they can introduce challenges, such as the complexity of formulating the underlying physics, training instabilities, or computational overhead [20–23]. Gaussian process regression (GPR) [24–27], as a nonparametric approach that provides uncertainty estimates alongside predictions, offers an alternative to establish robust models with significantly less data. Although the application of GPR for building structure-property relationships in metamaterials remains underexplored, this approach has been demonstrated in a few studies to train models effectively using small datasets [28–31].

When it comes to establishing structure-property maps for metamaterial unit cells, a critical consideration is how the

*Contact author: hooman.danesh@rwth-aachen.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

design geometry is represented as the input of the model. Although many classes of metamaterials can be effectively described using geometric parameters (e.g., strut thickness, void diameter, volume fraction, rotational angle, etc.) [17], such parametrizations inherently constrain the design space to a limited set of idealized configurations. In practice, however, the space of manufacturable unit cells is considerably broader, encompassing stochastic and randomly generated architectures [11,32–34], which are not easily captured by geometric descriptors. Furthermore, fabrication processes inevitably introduce deviations from the intended design, including dimensional inaccuracies and local defects, all of which can significantly affect material behavior [35–39]. Therefore, surrogate models that rely exclusively on idealized geometric parameters may overlook critical factors influencing material performance. This highlights the need for surrogate models that incorporate the spatial geometry of the unit cell, since parametric representations may fail to capture the complexity of random architectures and overlook critical fabrication imperfections.

While metamaterial unit cells can be characterized using various microscopy and tomography techniques [35,39,40], it should be noted that these techniques only provide a single instantiation of the material structure and not the material structure itself. This is because multiple realizations of the material structure can be obtained from a single physical sample, each exhibiting inherent variability in the extracted statistical features [41–43]. Such variability introduces uncertainty arising from both the characterization process and sample preparation. In addition to these uncertainties, there is also an arbitrary aspect in defining a metamaterial unit cell owing to the absence of a natural origin. Any region that repeats periodically to reconstruct the full structure can serve as a valid unit cell, with no single fixed choice [44]. Nevertheless, these different unit cell representations describe the same physical material and should, in principle, yield consistent predictions in structure-property linkages. Together, these sources of variability highlight the need for a stochastic framework for microstructure quantification that considers the material structure as a random process.

The recently developed materials knowledge systems (MKS) framework [41–43,45–54] quantifies material microstructure as a random process using two-point spatial correlations [55–59]. This higher-order statistical representation captures spatial information in a simple yet effective manner and bypasses the need for manual feature selection in establishing structure-property linkages. However, these two-point statistics yield a large set of features that are unwieldy for establishing structure-property maps. Therefore, MKS leverages the principal component analysis (PCA) [60–62] to achieve a reduced set of features for the low-dimensional representation of spatial correlations. Reduced-order structure-property linkages can then be established using a wide range of available regression techniques. However, it has been demonstrated in previous works [46–49] that GPR provides key benefits by employing a nonparametric modeling approach and rigorous treatment of prediction uncertainty.

In most practical applications, generating data for structure-property modeling, such as through high-fidelity

simulations or physical experiments, is expensive and resource intensive, necessitating frameworks that minimize data needs while maintaining model accuracy. Active learning addresses this challenge by iteratively selecting the most informative samples to build effective surrogate models with limited data [63]. Various approaches exist to identify these samples (i.e., sampling strategies) [63], among which uncertainty sampling has proven to be a simple yet effective strategy for GPR models, owing to the inherent uncertainty quantification provided by GPR [64–66]. By prioritizing samples where the model uncertainty is highest, uncertainty sampling ensures that each new training point maximally improves the model, enhancing the fidelity of the surrogate with as little data as possible [63,67].

Although the MKS framework has already been applied to composites with different ranges of stiffness contrast [52–54,68], to the best of our knowledge, it has not yet been used to model the mechanical response of materials with infinite contrast, such as the mechanical metamaterials in the present work. To this end, we employ the MKS framework to establish reduced-order structure-property linkages for an existing stochastic metamaterial dataset [11,69]. The primary quantity of interest in this study is the C_{11} component (C_{1111} in full tensorial notation) of the effective homogenized elastic stiffness, chosen as a representative example, while the approach is generalizable to other components as well. In order to compute this homogenized elastic constant, we perform physics-based simulations using an efficient fast Fourier transform (FFT)-based homogenization method [7,70]. Motivated by previous works [48–50] showcasing higher achievable fidelity by incorporating the statistics of the interface between the phases present in the microstructure, we also extract the solid-void interface for each metamaterial unit cell to investigate its influence on the performance of the surrogate model. The reduced-order representation of the metamaterial unit cells is then obtained by computing the two-point spatial correlations for both the solid and the interface phases, followed by the application of PCA on different combinations of the computed two-point statistics. GPR is subsequently used to establish the structure-property maps by linking the reduced set of features obtained from different combinations of the two-point statistics to the homogenized elastic constant. Our findings show that, although modest, the inclusion of the interface information improves the accuracy of the surrogate when combined with solid-solid statistics. Once the most accurate model is selected, we adopt an uncertainty-based active learning approach to identify the smallest subset of data necessary for generating structure-property maps with optimal predictive performance. We demonstrate that a surrogate model trained on approximately 0.61% of the dataset can achieve nearly the same accuracy as a model trained on the full dataset.

II. METHODOLOGY

A. Stochastic metamaterial dataset

To leverage the capabilities of the MKS framework, a voluminous dataset of stochastic metamaterial unit cells is beneficial. Here, the dataset of periodic 2D stochastic cellular

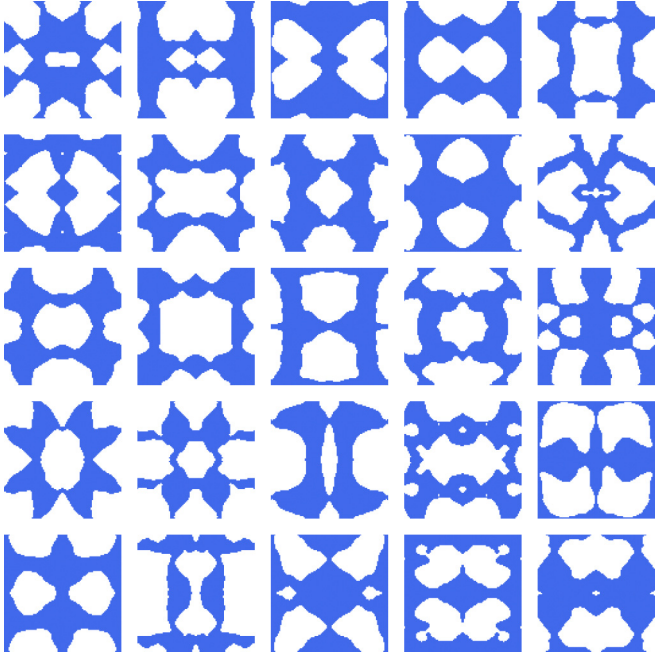


FIG. 1. Example unit cells from the dataset of periodic 2D stochastic metamaterials.

structures generated by [11,69] is employed, consisting of 53 019 microstructures represented on a 96×96 pixel grid. Each structure is created by sampling a Gaussian random field, which is then binarized using a randomly selected threshold to distinguish solid from void regions. To ensure structural connectivity and periodicity, only samples with sufficient material presence along all boundaries are retained. The accepted structures are then sequentially mirrored along both horizontal and vertical directions to produce periodic unit cells. As a pixel-based representation, this dataset is particularly well suited for the MKS framework. A randomly selected subset of 25 unit cells from this large dataset is shown in Fig. 1.

This generation approach, which imposes minimal design constraints apart from periodicity, results in a diverse collection of microstructures. The resulting structures span a broad range of relative densities, from 0.30 to 0.68, further contributing to the dataset geometric diversity. Nevertheless, to quantitatively assess the diversity within the dataset, the average pairwise Euclidean distance between all structures is computed. The resulting average is 66. Given that the maximum possible distance for binary 96×96 images is 96, this yields a normalized diversity score of 0.69 (with 1 being the theoretical maximum). This indicates a relatively high level of dissimilarity between structures, with unit cells differing by approximately 69% on average. To further investigate the diversity of the dataset, k-means clustering with a large number of 500 clusters is applied. It is found that, despite using such a high number of clusters, the coefficient of variation in the cluster sizes is 0.19. This suggests that the dataset is not concentrated in just a few clusters but is instead broadly distributed across the design space, further highlighting the diversity of the considered dataset.

B. Fast Fourier transform (FFT)-based homogenization

The FFT-based homogenization scheme was first introduced by [71,72] to study the homogenized effective response of both linear and nonlinear composites. Reformulating the homogenization problem as the convolution of local strain fluctuations with the Green's function of a reference elastic medium leads to the periodic Lippmann–Schwinger equation. Owing to the availability of the Green's function in closed form in Fourier space, and the advantage of transforming the convolution into algebraic products in the Fourier domain, the Lippmann–Schwinger equation can be efficiently solved using FFT in combination with fixed-point iterations (see, e.g., [73,74] for review).

Because of their matrix-free formulation, FFT-based methods are generally more memory-efficient than the finite element method (FEM), making them well suited for homogenizing large systems. In FFT-based homogenization, periodicity arises naturally from the Fourier transform, eliminating the need for additional treatment often required in FEM. This built-in periodicity also removes the need to ensure symmetric node placement along the boundaries, which is essential in FEM for applying periodic boundary conditions. In terms of performance, FFT-based methods have also proven to be both accurate and computationally efficient. A previous comparative study on the homogenization of metamaterial unit cells [70] showed that FFT becomes competitive with FEM for relative densities above 7%, assuming the same discretization inside the unit cell. For example, at a relative density of around 30% (corresponding to 70% porosity), FFT was found to be four to eight times faster than FEM. Despite this speed-up, the predicted effective elastic properties remained highly consistent with those from FEM, with deviations of less than 0.6%.

Yet, the convergence of the classical FFT-based homogenization method depends on the stiffness contrast between the phases present in the microstructure, making it unsuitable for materials with very high or infinite contrast, such as the metamaterial unit cells investigated in the present work [75]. Among the various approaches developed to address this limitation [76–79], Ref. [70] effectively extended FFT-based homogenization to metamaterials by employing the Galerkin FFT method [80,81], combined with a rotated finite difference grid [82] and the MINRES solver. This framework was also adopted in our previous work [7] to generate high-fidelity datasets by computing the homogenized elastic constants of auxetic unit cells.

The present work employs the same methodology [7] to compute the components of the effective stiffness tensor $\mathbb{C}$ for the large dataset of metamaterial unit cells introduced in Sec. II A. As the full formulation and implementation details are already provided in [7], they are not repeated here. Instead, we briefly recall the core idea of the perturbation-based approach [83,84] used to extract the effective stiffness components. The homogenized elastic stiffness tensor $\mathbb{C}$ is defined as

$$\mathbb{C} = \frac{\partial \bar{\sigma}}{\partial \bar{\epsilon}}, \quad (1)$$

where $\bar{\epsilon}$ and $\bar{\sigma}$ denote the macroscopic average strain and stress tensors, respectively. The components of the

homogenized effective elastic stiffness C_{ijkl} are computed by applying an independent macroscopic strain perturbation $\bar{\epsilon}_{kl} = \beta$, with all other components set to zero. The resulting average stress response $\bar{\sigma}_{ij(kl)}$ is then used to evaluate

$$C_{ijkl} = \frac{\bar{\sigma}_{ij(kl)}}{\beta}, \quad (2)$$

where β is a constant strain perturbation. Considering linear elasticity, the specific value of β is irrelevant, and the unperturbed state is taken as zero strain. While three (in 2D) or six (in 3D) independent perturbations are required to compute all independent components of the effective stiffness tensor, we focus in this work solely on the C_{1111} component (denoted C_{11} in reduced notation), which can be obtained from a single perturbation $\bar{\epsilon}_{11} = \beta$ as

$$C_{11} = C_{1111} = \frac{\bar{\sigma}_{11(11)}}{\beta}. \quad (3)$$

C. Two-point spatial correlations

As discussed in Sec. I and in previous works [41–43], obtaining a low-dimensional representation of the material microstructure (i.e., metamaterial unit cells in our case) requires the computation of two-point spatial correlation functions prior to the application of PCA. To achieve this, the discretized microstructure array m_s^h is defined as the volume fraction of the local state h at the discretized spatial bin $s \in \mathcal{S}$ (i.e., voxel in 3D or pixel in 2D). The spatial index s can be represented as $s = (s_1, s_2)$ in 2D and $s = (s_1, s_2, s_3)$ in 3D, corresponding to the position in the discretized microstructure. Considering that each spatial bin contains only a single distinct local state, the microstructure array m_s^h can take only the binary values 0 or 1. Subsequently, the discretized two-point correlation functions, denoted as $f_r^{hh'}$, quantify the probability of simultaneously observing the local states h and h' at two spatial locations separated by a vector $\mathbf{r} = (r_1, r_2, r_3)$, and are mathematically formulated as [42,43,55–59]

$$f_r^{hh'} = \frac{1}{S_r} \sum_{s \in \mathcal{S}} m_s^h m_{s+\mathbf{r}}^{h'}, \quad (4)$$

where S_r is the total number of trials accessible for the valid placement of the vector $\mathbf{r}$. In this setting, $f_r^{hh'}$ is referred to as auto-correlation when $h = h'$ and as cross-correlation when $h \neq h'$. It is worth noting that for periodic microstructures, such as the cellular structures in the present work, S_r is equal to the total number of spatial bins $|\mathcal{S}|$ (i.e., $S_r = |\mathcal{S}|$) [42,85]. Eventually, taking advantage of the convolution theorem, the two-point statistics for periodic microstructures can readily be computed using FFT as [55,85,86]

$$f_r^{hh'} = \frac{1}{|\mathcal{S}|} \mathcal{F}^{-1}[(\mathcal{F}[m_s^h])^* \mathcal{F}[m_s^{h'}]], \quad (5)$$

where $(\cdot)^*$ denotes the complex conjugate, and $\mathcal{F}[\cdot]$ and $\mathcal{F}^{-1}[\cdot]$ represent the forward and inverse Fourier transforms, respectively. It is also worth noting that, even in the absence of periodicity, FFT algorithms can still be applied to compute two-point correlations by leveraging a padding strategy [42,85].

D. Principal component analysis (PCA)

The computation of the two-point correlations through Eqs. (4) and (5) results in an extensive and unmanageable set of features. It is therefore practical to apply dimensionality reduction techniques to obtain a reduced yet informative feature space. PCA [60–62] has been shown to be effective in capturing a significantly smaller set of salient features ideal for establishing structure-property linkages [47–53]. PCA is a linear distance-preserving transformation that identifies orthogonal linear combinations of the original features, which are ordered by the amount of variance they explain in the dataset, from highest to lowest.

Let $j \in \{1, 2, \dots, J\}$ index the members of the ensemble of microstructures, and let $f_r^{(j)}$, with $r \in \{1, 2, \dots, R\}$, represent the high-dimensional vectorized representation of all two-point statistics deemed important to generate structure-property maps for the j th microstructure, where J and R denote the total number of microstructures and features, respectively. The principal component (PC) representation of this microstructure exemplar can then be written as [41,42,58,59]

$$f_r^{(j)} = \sum_{i=1}^{\min(J-1, R)} \alpha_i^{(j)} \varphi_{ir} + \bar{f}_r, \quad (6)$$

where $\alpha_i^{(j)}$ denotes the i th PC score of the j th microstructure, φ_{ir} are the PC basis vectors, and $\bar{f}_r$ is the ensemble average. As shown in Eq. (6), PCA identifies up to $\min(J-1, R)$ orthogonal basis vectors (i.e., PCs) in the feature space. However, because of the descending order of variance captured by the PC scores, it is possible to objectively truncate the series in Eq. (6) to obtain a low-dimensional representation without significant loss of information. It has been demonstrated in previous studies [47–53] that only a limited number of PC scores are sufficient to achieve a large portion of the explained variance from the entire dataset, resulting in high-fidelity structure-property linkages. Assuming $\hat{R}$ denotes the number of retained components, it typically holds that $\hat{R} \ll \min(J-1, R)$.

E. Gaussian process regression (GPR)

After extracting the salient features of the microstructures and obtaining the low-dimensional representation, the mapping of input features to the desired output can be performed through various model building approaches. Among others, GPR has proven to be a powerful tool to establish structure-property linkages in previous works [46–49]. This nonparametric approach not only predicts the outputs of interest but also provides the uncertainty associated with the predictions [24–27], which offers additional insight and becomes especially useful when employing active learning approaches (see Sec. II F).

Assuming a zero-mean Gaussian process (GP), the output function y is modeled by a multivariate normal distribution $\mathcal{N}$ as [25]

$$y(\mathbf{x}) \sim \mathcal{N}(0, k(\mathbf{x}, \mathbf{x}')), \quad 7$$

where $\mathbf{x}$ denotes the vector of input features with the dimension D , and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function of the GP.

The automatic relevance determination squared exponential (ARD-SE) kernel has proven to be a suitable choice for the GP covariance because of its flexibility in assigning a separate hyperparameter to each input feature. The mathematical expression for the ARD-SE kernel is given by [25]

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}\right) + \sigma_n^2 \delta_{\mathbf{x}\mathbf{x}'}. \quad (8)$$

Here, the scaling factor of the output variance σ_f , the noise factor σ_n , and the characteristic length-scale ℓ_d corresponding to the input variable x_d comprise the set of hyperparameters for the kernel function $k(\mathbf{x}, \mathbf{x}')$, while $\delta_{\mathbf{x}\mathbf{x}'}$ represents the Kronecker delta function. Let $\mathbf{y}$ and $\mathbf{y}_*$ denote the observed training and unseen test output vectors with N and N_* data points, respectively. The joint distribution of the training outputs $\mathbf{y}$ and the test outputs $\mathbf{y}_*$ is then expressed as [25]

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^\top & \mathbf{K}_{**} \end{bmatrix}\right). \quad (9)$$

In Eq. (9), $\mathbf{K} = k(\mathbf{X}, \mathbf{X})$, $\mathbf{K}_* = k(\mathbf{X}, \mathbf{X}_*)$, and $\mathbf{K}_{**} = k(\mathbf{X}_*, \mathbf{X}_*)$, where $\mathbf{X}$ and $\mathbf{X}_*$ denote the $N \times D$ and $N_* \times D$ input matrices for the training and test data, respectively. Leveraging the conditional Gaussian distribution, the mean μ_* and variance Σ_* (i.e., uncertainty) of the predictive distribution at the test points are obtained by [25]

$$\mu_* = \mathbf{K}_*^\top \mathbf{K}^{-1} \mathbf{y}, \quad (10)$$

$$\Sigma_* = \mathbf{K}_{**} - \mathbf{K}_*^\top \mathbf{K}^{-1} \mathbf{K}_*. \quad (11)$$

The predictive accuracy of the GPR model is highly sensitive to the choice of hyperparameters. Let $\boldsymbol{\theta} = (\sigma_f, \sigma_n, \ell_1, \ell_2, \dots, \ell_D)$ denote the vector of all hyperparameters characterizing the GP model. The optimal hyperparameters $\hat{\boldsymbol{\theta}}$ are determined from the training data by minimizing the negative log likelihood function $\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$ as

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} (-\log p(\mathbf{y} | \mathbf{X}, \boldsymbol{\theta})), \quad (12)$$

where the log marginal likelihood $\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$ is given by [25]

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2} \mathbf{y}^\top \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log \det(\mathbf{K}) - \frac{N}{2} \log 2\pi. \quad (13)$$

Finally, it is worth noting that, considering that our main output in the present study is the homogenized stiffness component C_{11} , we employ the mean absolute stiffness error (MAE) on the unseen test data as the main error measure to evaluate the predictive performance of the model. The MAE is defined as

$$\text{MAE} = \frac{1}{N_*} \sum_{n=1}^{N_*} |C_{11}^{(n)} - \hat{C}_{11}^{(n)}|, \quad (14)$$

where $n \in \{1, 2, \dots, N_*\}$ enumerates the samples (i.e., metamaterial unit cells) in the test dataset, C_{11} is the actual value obtained from physics-based simulations, and $\hat{C}_{11}$ denotes the predicted value by the GPR model.

F. Active learning

To establish structure-property linkages, it is essential to label the data, i.e., to compute the property of interest for each metamaterial unit cell and assign it accordingly. However, when the amount of unlabeled data is large or the labeling process is expensive (e.g., because of high computational cost of simulations or time-consuming and costly experiments), it becomes crucial to develop models that can achieve high performance with as little labeled data as possible.

Active learning is an effective approach in such scenarios, where the algorithm interactively queries new data points to be labeled. Several strategies exist for generating candidate data points for labeling, including membership query synthesis, stream-based sampling, pool-based sampling, etc. [63]. The present study adopts the pool-based sampling strategy, as the dataset is readily available. In pool-based sampling, candidate data points are known *a priori*, unlike the other methods where candidates are generated dynamically during the learning process. It is therefore essential that the candidate pool sufficiently covers the entire input space, ensuring that all possible input microstructures are taken into account.

Another key aspect of active learning is the query strategy, which determines how new training data points are selected for labeling. A wide range of strategies has been proposed in the literature, such as random sampling, query-by-committee, and uncertainty sampling, among others [63]. Given that GPR provides an estimate of prediction uncertainty, uncertainty sampling is particularly well suited for this work. In this study, a simple yet effective approach is used: At each active learning iteration, the data point with the highest predictive uncertainty (i.e., maximum variance) is selected from the candidate pool and added to the training set. This criterion has been shown to be effective [64–66], especially when the model noise is relatively uniform across the input space [63].

Finally, it is important to define a clear stopping criterion for the active learning process. Without such a criterion, the algorithm may continue querying new data points unnecessarily, leading to inefficiencies in computational or experimental resources. Typically, active learning iterations continue until a predefined labeling budget is exhausted or a desired level of model accuracy is reached. In practice, stopping criteria can also include convergence of model performance metrics (e.g., error measures) or stabilization of prediction uncertainty, indicating that adding more data no longer yields meaningful improvements. As mentioned in Sec. II E, the MAE is employed as the error metric to evaluate the predictive performance of the model. Therefore, we adopt the following stopping criterion, defined by the relative change in MAE over a sliding window of recent active learning iterations:

$$\frac{1}{Q} \sum_{q=I-Q+1}^I \left| \frac{\text{MAE}(\hat{\boldsymbol{\theta}}_q) - \text{MAE}(\hat{\boldsymbol{\theta}}_{q-1})}{\text{MAE}(\hat{\boldsymbol{\theta}}_{q-1})} \right| < \epsilon. \quad (15)$$

In Eq. (15), $\hat{\boldsymbol{\theta}}_q$ represents the optimized model hyperparameters at the q th active learning iteration, while I denotes the total number of completed iterations. The parameter Q is a user-defined value specifying the number of recent iterations over which the relative change in MAE is averaged. The stopping criterion is met when the average relative change

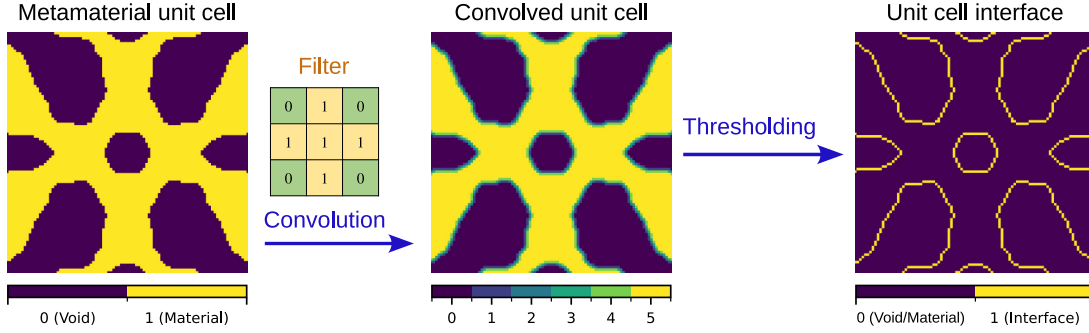


FIG. 2. Interface extraction workflow: the solid microstructure array m_s^1 (left) is convolved with a five-point kernel to produce the filtered field $m_s^{1,\text{conv}}$ (middle), where each pixel value indicates the number of neighboring pixels in the solid phase. The interface microstructure array m_s^2 (right) is obtained by labeling void pixels that are adjacent to at least one solid pixel.

in MAE over the last Q iterations falls below a predefined threshold ϵ , indicating convergence in model performance and diminishing benefit from further sampling.

III. RESULTS AND DISCUSSION

A. Feature engineering

Following the MKS framework, we employ the workflow of two-point correlation functions and PCA, as formulated in Secs. II C and II D, to obtain a reduced set of salient features for the metamaterial dataset described in Sec. II A. To do so, we start by labeling the local states h present in the microstructure. Since we are dealing with metamaterial unit cells, we initially consider two local states $h \in \{0, 1\}$, where the void and solid phases are denoted by $h = 0$ and $h = 1$, respectively. In this setting, m_s^0 represents the volume fraction of the void phase at pixel $\mathbf{s}$, while m_s^1 refers to the volume fraction of the solid phase. As the metamaterial unit cells at hand are eigen microstructures (i.e., only a single local state is present at each spatial bin), the discretized microstructure array m_s^h can only take binary values, 0 or 1.

Previous studies have shown the effect of the interface between different phases in the microstructure on the accuracy of surrogate models for various physical phenomena, such as permeability [48], charge transport [49], and damage initiation [50]. Motivated by these works, we use a 2D convolution operation to extract the interface pixels between the solid and void phases in metamaterial unit cells and use this as a third material state descriptor at the pixel scale. An illustrative example of the interface extraction is shown in Fig. 2. The left plot shows the pixelized microstructure, where pixels with value 1 represent the solid phase and those with value 0 correspond to the void phase. This binary field is convolved with a five-point kernel ϕ_c as

$$m_s^{1,\text{conv}} = \sum_{\mathbf{c} \in \mathcal{C}} \phi_{\mathbf{c}} m_{\mathbf{s}+\mathbf{c}}^1, \quad \phi_{\mathbf{c}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad (16)$$

where $m_s^{1,\text{conv}}$ is the convolved microstructure array, $\mathbf{c}$ denotes relative coordinates in the kernel, and $\mathcal{C} = \{(0, 0), (0, 1), (0, -1), (1, 0), (-1, 0)\}$. The convolution is performed assuming periodic boundary conditions on the microstructure, consistent with the assumptions implicit in the

FFT-based homogenization method described earlier. The resulting convolved field, shown in the middle plot of Fig. 2, contains integer values between 0 and 5, each representing the number of solid phase pixels within a five-point stencil, including the central pixel and its four immediate neighbors. We then define the interface using a simple criterion, based on which a single pixel on the void side adjacent to at least one solid pixel is labeled as an interface pixel. Formally, this is implemented through a thresholding operation,

$$m_s^2 = \chi_{\{m_s^1=0\} \cap \{m_s^{1,\text{conv}} > 0\}}(\mathbf{s}). \quad (17)$$

Here, m_s^2 denotes the interface microstructure array, where $m_s^2 = 1$ identifies the interface pixels, while $m_s^2 = 0$ elsewhere (i.e., void or solid phases). The symbol $\chi_{\mathcal{A}}(\mathbf{s})$ denotes the indicator function of a set $\mathcal{A}$, which returns 1 if $\mathbf{s} \in \mathcal{A}$ and 0 otherwise. Additionally, the operator $\cap$ denotes the intersection of two sets, meaning that $\mathbf{s}$ must satisfy both conditions simultaneously. The right plot in Fig. 2 shows the extracted interface for the example unit cell. This operation introduces a third local state, $h = 2$, representing the interface, thereby extending the set of local states to $h \in \{0, 1, 2\}$.

After defining the local states of the metamaterial unit cells, we compute the most informative two-point spatial correlations. For a microstructure with H local states, prior work [56] has shown that only $H - 1$ of the H^2 possible two-point correlations are independent. However, the optimal combination for establishing high-fidelity linkages is not known *a priori*, as potential nonlinear dependencies among these correlations may not be captured by PCA [48]. For the present metamaterial unit cells with three local states (i.e., void, solid, and interface), only two of the nine possible sets of correlations (three auto-correlations and six cross-correlations) are independent. However, in a previous study [51], it was shown that considering a combination of three sets of two-point correlations is beneficial when dealing with microstructures with three local states, where one of these sets must be a cross-correlation. Building on these observations, we select three sets of correlations for our analysis: the solid-solid auto-correlation $f_{\mathbf{r}}^{11}$, the interface-interface auto-correlation $f_{\mathbf{r}}^{22}$, and the solid-interface cross-correlation $f_{\mathbf{r}}^{12}$. The computed two-point correlations for an example metamaterial unit cell are shown in Fig. 3. A notable feature of the auto-correlation maps is that their value at the zero-shift vector $\mathbf{r} = \mathbf{0}$ reflects

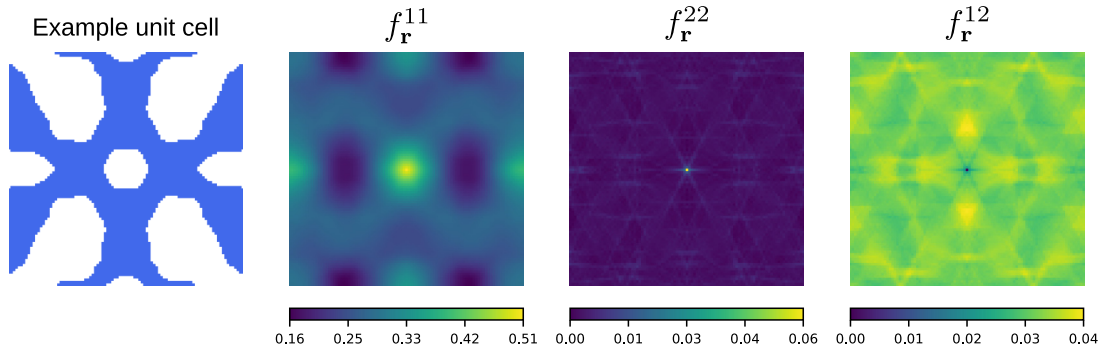


FIG. 3. Two-point spatial correlations for an example unit cell, including the solid-solid auto-correlation f_r^{11} (left), the interface-interface auto-correlation f_r^{22} (middle), and the solid-interface cross-correlation f_r^{12} (right).

the volume fraction of the associated phase. More specifically, in Fig. 3, the values of the central pixels in the solid and interface auto-correlation maps (i.e., f_0^{11} and f_0^{22}) correspond to the volume fractions of the solid and interface phases, respectively.

As mentioned in earlier works [48–51], it is necessary to rescale the computed two-point statistics prior to dimensionality reduction, as PCA gives greater weight to features with higher numerical values owing to their stronger influence on the overall variance. Therefore, inspired by [49], we rescale the two-point correlations in a way that each set exhibits the same variance. To achieve this, we compute the mean and, consequently, the standard deviation of each set of correlations across all spatial bins and unit cells in the ensemble as follows:

$$\mu^{hh'} = \frac{1}{J} \frac{1}{|S|} \sum_{j=1}^J \sum_{\mathbf{r}}^{(j)} f_{\mathbf{r}}^{hh'},$$

$$\sigma^{hh'} = \sqrt{\frac{1}{J} \frac{1}{|S|} \sum_{j=1}^J \sum_{\mathbf{r}}^{(j)} (f_{\mathbf{r}}^{hh'} - \mu^{hh'})^2}, \quad (18)$$

where $\mu^{hh'}$ and $\sigma^{hh'}$ denote the mean and standard deviation across all bins and unit cells corresponding to the two-point correlation set $f_{\mathbf{r}}^{hh'}$. The rescaled two-point correlation of the j th unit cell in the ensemble $^{(j)}\tilde{f}_{\mathbf{r}}^{hh'}$ is then computed as

$$^{(j)}\tilde{f}_{\mathbf{r}}^{hh'} = \left(\frac{\sigma^{11}}{\sigma^{hh'}} \right)^{^{(j)}f_{\mathbf{r}}^{hh'}}. \quad (19)$$

In Eq. (19), σ^{11} denotes the standard deviation corresponding to the solid auto-correlation $f_{\mathbf{r}}^{11}$. In this way, the rescaled correlations $\tilde{f}_{\mathbf{r}}^{22}$ and $\tilde{f}_{\mathbf{r}}^{12}$ are adjusted to have the same standard deviation as that of the solid auto-correlation $f_{\mathbf{r}}^{11}$. Although $f_{\mathbf{r}}^{11}$ remains unchanged during this rescaling, we still use the notation $\tilde{f}_{\mathbf{r}}^{11}$ for consistency. The three computed two-point spatial correlations result in seven possible combinations. However, for the reasons mentioned above, we consider the following three combinations: $\tilde{f}_{\mathbf{r}}^{11}$, $\{\tilde{f}_{\mathbf{r}}^{11}, \tilde{f}_{\mathbf{r}}^{22}\}$, and $\{\tilde{f}_{\mathbf{r}}^{11}, \tilde{f}_{\mathbf{r}}^{22}, \tilde{f}_{\mathbf{r}}^{12}\}$.

Each set of computed two-point correlations, derived from a 2D unit cell of 96×96 pixels, results in $96^2 = 9216$ features. This number becomes doubled (18432) and tripled (27648) when a combination of two or three sets of correlations is considered, respectively. As discussed in

Sec. IID, these high-dimensional sets of unwieldy and unmanageable features can be transformed into a low-dimensional space by applying dimensionality reduction. Therefore, we apply PCA to each of the selected combinations to obtain a low-dimensional representation. Figure 4 illustrates the PC representation of all metamaterial unit cells in the space of the first two PCs, derived from the combination $\{\tilde{f}_{\mathbf{r}}^{11}, \tilde{f}_{\mathbf{r}}^{22}\}$. The PC representation for this specific combination (i.e., $\{\tilde{f}_{\mathbf{r}}^{11}, \tilde{f}_{\mathbf{r}}^{22}\}$) is shown because, as will be demonstrated in Sec. IIIB, it provides the most valuable statistics for establishing structure-property linkages.

In Fig. 4, the data points are colored by three different quantities, including the solid volume fraction f_0^{11} (left), the interface volume fraction $\tilde{f}_0^{22}$ (middle), and the effective elastic constant C_{11} computed from physics-based simulations (right). Although the adopted feature engineering workflow is unsupervised (i.e., feature selection is independent of the output variable), the data points are colored by the output quantity C_{11} to visualize how the features captured by the PCs relate to the mechanical response. In the left plot of Fig. 4, PC1 shows a clear correlation with the solid volume fraction. As PC1 decreases from right to left, the solid volume fraction increases. This visual trend is supported quantitatively by a Pearson correlation coefficient of -0.975 between PC1 and the solid volume fraction, confirming that the first principal component encodes material density. In the middle plot, both PC1 and PC2 exhibit visible correlations with the interface volume fraction, though to different extents. A quantitative analysis reveals that the Pearson correlation coefficients between the interface volume fraction and PC1 and PC2 are 0.803 and 0.585, respectively, indicating that PC1 captures the dominant share of interface features, while PC2 contributes additional but weaker variation. The right plot shows that C_{11} follows a trend similar to the solid volume fraction along PC1, suggesting that the stiffness of the metamaterial is primarily influenced by the amount of solid material, which is intuitive, as an increased amount of material leads to higher stiffness.

To further interpret the physical meaning of PCs, we evaluated the connectivity of the solid phase in each microstructure using the mean coordination number, defined as the average number of neighboring solid pixels (using eight-neighbor connectivity) for all solid pixels in the structure. This metric reflects how well the solid domains are interconnected, with higher values indicating load-bearing solid

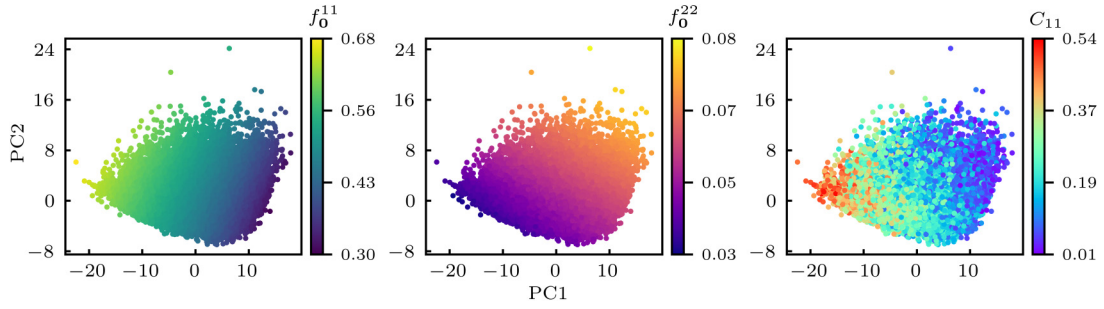


FIG. 4. PC representation of all metamaterial unit cells for the combination $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$ in the space of the first two PCs, colored by different quantities: solid volume fraction f_0^{11} (left), interface volume fraction f_0^{22} (middle), and effective elastic constant C_{11} (right).

regions with high connectivity. We found that the Pearson correlation coefficient between the mean coordination number and PC1 is -0.969 , demonstrating that PC1 not only reflects the solid phase volume fraction but also encodes aspects of phase connectivity, which is a key determinant of mechanical performance.

B. Structure-property linkages

Once the reduced set of features is obtained, GPR is used to establish the structure-property linkages. We use the FFT-based homogenization scheme introduced in Sec. II B to compute the effective elastic constant C_{11} . In this setting, the solid regions are modeled as a linear elastic isotropic material with Young's modulus $E = 1$ GPa and Poisson's ratio $\nu = 0.3$, while the void regions are assigned $E = 0$ GPa and $\nu = 0$, effectively treating them as empty space with no stiffness contribution. It should be emphasized that the specific value chosen for E does not affect the generality of the results. Since Hooke's law is linear in E , and the second phase in the homogenization problem (i.e., the void) has zero contribution to the stiffness, the effective elastic constant C_{11} is linearly proportional to E . Therefore, regardless of the value assigned to E , the resulting output represents a normalized elastic constant C_{11}/E . This normalization allows the predicted output to be rescaled post-prediction to correspond to any desired Young's modulus. For instance, a model trained with $E = 1$ GPa can be readily adapted to predict effective properties for materials with different values of E through a simple scaling operation. In contrast, the effect of Poisson's ratio ν is nonlinear. To construct a predictive model applicable to varying values of ν , one must generate data across a range of Poisson's ratios and include ν as an additional input variable alongside the PC scores (see, e.g., [7]). In this study, however, we restrict ourselves to a fixed value $\nu = 0.3$, although the approach is generalizable. After computing the normalized effective elastic constant C_{11}/E , we filter out samples with values below 0.01, as such low stiffness values are not relevant for practical applications. The original dataset from [11,69] included such low-density designs to explore buckling-prone geometries. Applying this filter reduces the dataset size only slightly, from 53 019 to 52 885 samples, eliminating just a small fraction of impractical structures.

Regarding the input parameters for the model, we consider the first 8 PC scores obtained from different combinations of two-point statistics, including $\tilde{f}_r^{11}$, $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$, and

$\{\tilde{f}_r^{11}, \tilde{f}_r^{22}, \tilde{f}_r^{12}\}$. Each input feature (i.e., PC score) is standardized to have zero mean and unit variance to improve numerical stability and enhance the performance of GPR model training. We gradually increase the number of PCs from 1 to 8 and train a GPR model for each combination of the mentioned two-point statistics to identify the model with the highest fidelity. This results in a total of 24 models. The GPR models are implemented using the GPyTorch library, with hyperparameter optimization performed via the Adam optimizer from the scikit-learn library. To ensure the stability of hyperparameter optimization, we perform multiple random initializations and verify that the optimized values converge to consistent solutions. Additionally, a cosine annealing learning rate schedule is employed to promote smooth convergence during training. Following common practice, an 80/20% train/test split is used to evaluate the predictive performance of the trained models on unseen data. This standard choice offers a well-established balance between having sufficient training data for model fitting and enough test data for robust generalization assessment.

Figure 5 plots the MAE of the test set [computed from Eq. (14)] for models trained on different numbers of PCs obtained from the various two-point statistics combinations. The results reveal a consistent decrease in the prediction error as the number of PCs increases, demonstrating the improved predictive power of models incorporating more microstructural features. The inclusion of interface auto-correlation notably

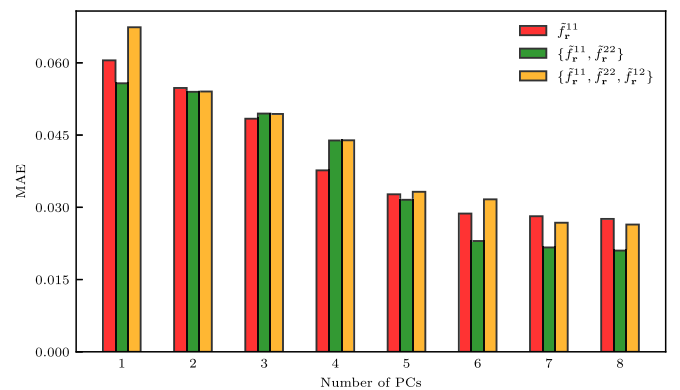


FIG. 5. MAE of the test set for GPR models trained with varying numbers of PC scores (from 1 to 8) obtained from three different combinations of two-point statistics, including $\tilde{f}_r^{11}$, $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$, and $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}, \tilde{f}_r^{12}\}$.

improves model performance compared to using only the solid auto-correlation. On the other hand, adding the solid-interface cross-correlation does not appear to further improve the model accuracy. For instance, when using 6 PCs, the MAE decreases to 0.0230 for the combination including interface auto-correlation (i.e., $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$), compared to 0.0287 for the model trained on PCs obtained from the solid auto-correlation alone (i.e., $\tilde{f}_r^{11}$). However, it increases to 0.0317 when the cross-correlation is also incorporated (i.e., $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}, \tilde{f}_r^{12}\}$). A saturation in the predictive performance of the models is observed as more PCs are included. For the solid auto-correlation $\tilde{f}_r^{11}$ and the combination $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$, the models reach a stable level of accuracy with 6 or more PCs. The same trend is observed for the combination $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}, \tilde{f}_r^{12}\}$, but its performance saturates at 7 PCs or more. These results suggest that the solid and interface auto-correlations (i.e., the combination $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$) with 6 PC scores form a compact and informative feature set, while the solid-interface cross-correlation does not contribute meaningful additional information for the regression task. This observation can be rationalized primarily through statistical redundancy and limited mechanical relevance. On the one hand, since interface voxels are explicitly defined as neighbors of solid voxels, the solid-interface cross-correlation $\tilde{f}_r^{12}$ inherently contains spatial variations already captured by the solid and interface auto-correlations, $\tilde{f}_r^{11}$ and $\tilde{f}_r^{22}$, leading to significant statistical redundancy and thus minimal new information. On the other hand, the effective stiffness primarily depends on the volume fraction, connectivity, and spatial arrangement of the solid phase, already comprehensively captured by $\tilde{f}_r^{11}$ and $\tilde{f}_r^{22}$. Interface voxels, being adjacent to voids, do not directly carry stress, rendering their spatial cross-correlation with solid voxels of limited mechanical significance. Consequently, the inclusion of $\tilde{f}_r^{12}$ does not enhance predictive accuracy. It is also worth noting that although the inclusion of interface auto-correlation reduces the MAE by approximately 20% (from 0.0287 to 0.0230), the importance of such interface statistics is expected to be even more pronounced in cases involving plasticity or damage, where phase boundaries play a more critical role in mechanical responses such as plastification or damage initiation.

The parity plot, illustrating the predicted versus actual values of the normalized effective elastic constant (i.e., $\hat{C}_{11}/E$ versus C_{11}/E) for both the training and test datasets, is shown in Fig. 6 for the optimal model, i.e., the model trained with 6 PC scores obtained from the combination $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$. The resulting R^2 score of approximately 0.92 is competitive given the lower complexity and parameter efficiency of the GPR model, which achieves strong performance with a small number of kernel hyperparameters compared to the millions of parameters in neural-network-based approaches. While neural networks may achieve higher prediction accuracy (e.g., R^2 scores up to 0.99 as reported in [10,13]), such models typically require significantly larger datasets to optimize their extensive parameter space. In contrast, as will be demonstrated in Sec. III C, our GPR approach achieves acceptable saturated accuracy levels with substantially fewer training samples. A detailed comparison with deep learning approaches previously developed for other metamaterials is provided in Sec. III D.

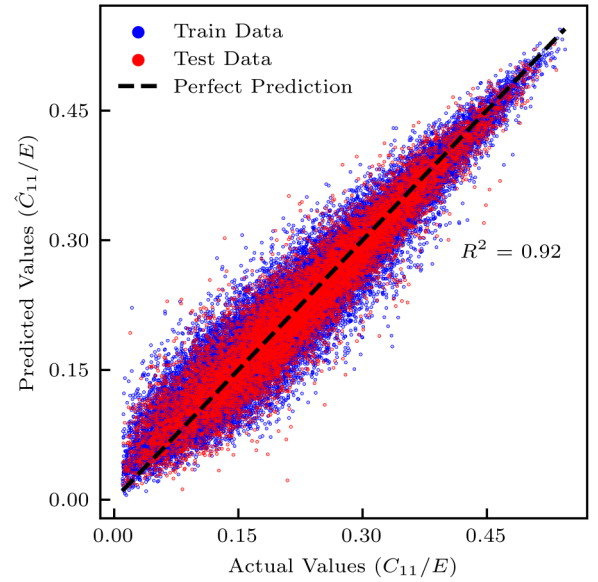


FIG. 6. Parity plot showing the predicted vs actual values of the normalized effective elastic constant (i.e., $\hat{C}_{11}/E$ vs C_{11}/E) for the model trained with 6 PC scores obtained from the combination $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$. Results are shown for both the training and test datasets and the diagonal line represents the ideal case of perfect prediction.

C. Uncertainty-based active learning

In this section, we employ the uncertainty-based active learning framework described in Sec. II F to identify the minimum number of labeled data points required to establish robust structure-property linkages comparable to those trained on the full dataset. As demonstrated in Sec. III B, the model trained with 6 PCs obtained from the combination $\{\tilde{f}_r^{11}, \tilde{f}_r^{22}\}$ yields the most accurate predictions. Therefore, we focus solely on this model and disregard other combinations of two-point correlations and numbers of PCs.

We begin by training a GPR model on 10 randomly selected data points, which is then used to predict the output for the remaining points, referred to as the candidate pool. Subsequently, the point with the highest predictive variance is selected from the candidate pool and added to the training set to retrain an updated GPR model. This iterative procedure is repeated until a specified labeling budget is exhausted or a defined stopping criterion is met. Here, we apply the stopping criterion defined in Eq. (15) to identify the minimum number of observations required for saturated model performance.

Since all data points are labeled using the efficient FFT-based homogenization scheme, we are able to evaluate the prediction error using MAE on the entire candidate pool, rather than a hold-out test set, which would typically be the case in real-world applications. Figure 7 plots the MAE over the candidate pool versus the number of labeled observations throughout the active learning process. This curve represents the average of 25 repetitions of the active learning workflow, each with a different random initialization. For comparison, we also include the performance of a GPR model trained on the entire dataset using an 80/20% train/test split, averaged over 25 random splits. Although only 80% of the data is used

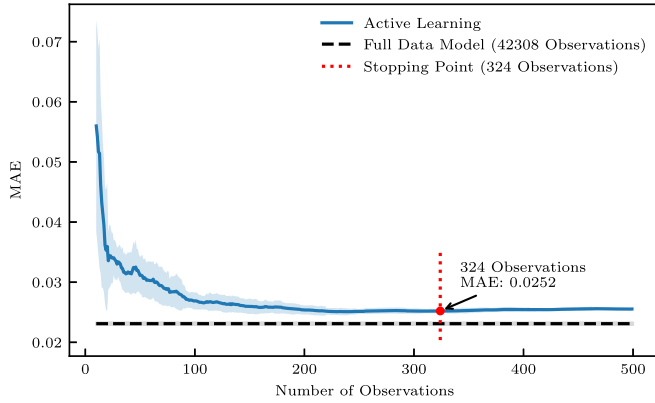


FIG. 7. Active learning curve showing MAE over the candidate pool vs the number of labeled observations. The active learning curve corresponds to the mean of 25 independent runs with different random initializations, and the shaded region indicates ± 1 standard deviations. For comparison, the performance of a model trained on the full dataset with an 80/20% train/test split is also shown, averaged over 25 random splits with its own ± 1 standard deviation shaded region.

for training in this case, we refer to this model as the *full data model* for simplicity.

It can be seen that the error decreases sharply during the first 90 iterations, dropping from approximately 0.0560 for the initial model with 10 observations to around 0.0269 for the model with 100 observations. As active learning continues, the MAE is further reduced to around 0.025, after which no substantial improvements are observed. By adopting the stopping criterion in Eq. (15) with $Q = 5$ and $\epsilon = 0.0001$, we find that a total of 324 observations (i.e., 10 initial points and 314 active learning iterations) is sufficient for the model to reach saturated performance, with a final MAE of 0.0252. At this point, the average relative change in MAE over the last five iterations falls below 0.0001. The resulting error is only 0.0021 higher than that of the model trained on the entire dataset (i.e., 0.0252 versus 0.0231), which is acceptable given the significantly smaller training size (324 versus 42 308 points). This number of observations constitutes only about 0.61% of the full dataset, underscoring the effectiveness of the proposed active learning framework.

Finally, we can also investigate the prediction uncertainty in the candidate pool by plotting the maximum predictive standard deviation within the pool at each iteration in Fig. 8. It is evident that, as the training set is expanded by iteratively adding the most uncertain point, the maximum predictive standard deviation decreases from approximately 0.12 to about 0.04. The standard deviation then reaches a plateau near 0.04, confirming the efficiency of the uncertainty sampling strategy in reducing prediction uncertainty with a relatively small number of observations. Beyond this point, adding more data points does not yield further benefit. This highlights that the adopted active learning workflow enables training a surrogate model with significantly fewer data points, achieving not only predictive accuracy comparable to a model trained on the full dataset but also a substantial reduction in prediction uncertainty.

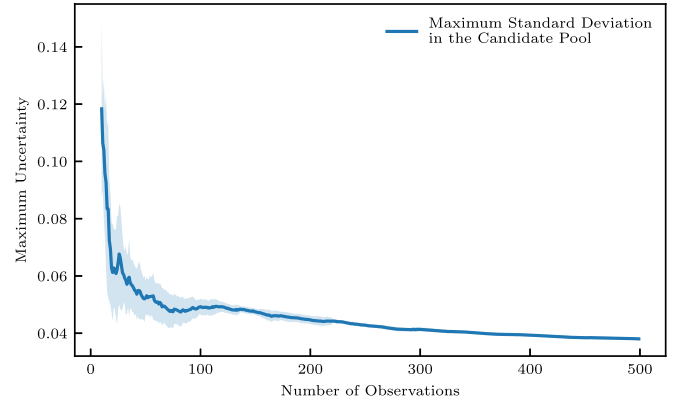


FIG. 8. Maximum predictive standard deviation in the candidate pool vs the number of labeled observations throughout the active learning process. The curve corresponds to the mean of 25 independent runs with different random initializations, and the shaded region indicates ± 1 standard deviations.

D. Comparison with deep learning models

As a final step of this study, we compare the present workflow with alternative neural-network-based deep learning approaches reported in the literature, in terms of training data efficiency, model complexity, and predictive performance. In particular, we consider the graph neural network (GNN) and the GoogLeNet architecture enhanced with a convolutional block attention module (CBAM), both proposed by [13], as well as the modified crystal graph convolutional neural network (mCGCNN) developed by [10]. The GNN and GoogLeNet_CBAM models were trained on 6531 truss-based lattice structures to predict the elastic modulus. In contrast, the mCGCNN model was trained on 43 505 shell-lattice microstructures to predict a range of effective properties, including thermal, electrical, and magnetic properties, as well as mechanical properties such as the elastic constant C_{11} reported here.

Since the range of the output variables differs between the works considered for comparison and the present study, we normalize the MAE [i.e., Eq. (14)] by the range of the output variable in each dataset (i.e., the difference between the maximum and minimum output values), and refer to the resulting quantity as the normalized mean absolute error (nMAE). For a consistent comparison, Table I summarizes the dataset sizes, the fraction of dataset size used for training (training set fraction), model complexities in terms of number of parameters, and the nMAE achieved by the different models. We also include our models trained on the full dataset and achieved through active learning. Notably, the dataset used in the present study comprises 52 885 samples, making it the largest among the compared studies. Although the nMAE of our active learning model (4.7%) is slightly higher than the errors achieved by the neural-network-based models, it is important to highlight that our model attains this level of accuracy by using only 0.61% of the entire dataset for training, whereas the other approaches used 90% of their datasets. Furthermore, our GPR model is remarkably compact, involving only eight model parameters: an output variance factor, a noise factor, and 6 characteristic length

TABLE I. Comparison with different available deep learning models.

Model	Dataset size	Train set fraction	No. of Parameters	nMAE	Predicted property
GNN [13]	6531	0.9	44 401	2.2%	Elastic modulus
GoogLeNet_CBAM [13]	6531	0.9	1 038 755	1.3%	Elastic modulus
mCGCNN [10]	43 505	0.9	~230 000	2.2%	C_{11}
Present (Full data model)	52 885	0.8	8	4.3%	C_{11}
Present (Active learning model)	52 885	0.0061	8	4.7%	C_{11}

scales corresponding to 6 PC scores, i.e., $\theta = (\sigma_f, \sigma_n, \ell_1, \ell_2, \dots, \ell_6)$. In contrast, the other deep learning approaches require tens or hundreds of thousands of parameters to achieve such high levels of predictive accuracy. These results underscore the efficiency, simplicity, and strong predictive capabilities of the proposed framework, particularly when combined with active learning for data selection.

IV. CONCLUSIONS

In this work, we employed the MKS framework to establish structure-property linkages for a large dataset of periodic 2D stochastic metamaterial unit cells. An FFT-based homogenization scheme was used to efficiently compute the effective elastic constant C_{11} (i.e., the output of interest) for the entire dataset. A low-dimensional representation of the microstructures was obtained by first computing two-point spatial correlations and subsequently applying PCA. To investigate the role of the solid-void interface, we extracted the interface for each unit cell via a 2D convolution operation and included its two-point statistics in combination with the correlations of the solid phase to assess its effect on model fidelity. Different combinations of solid and interface correlations were then considered to train GPR models. Each model was evaluated across varying numbers of PC scores to identify the most accurate surrogate with the fewest input parameters. It was found that the combination of solid and interface auto-correlations with 6 PC scores yielded the most accurate model. However, the inclusion of the solid-interface cross-correlation did not lead to noticeable improvements in model accuracy, suggesting it provides limited additional information for the regression task. Subsequently, an uncertainty-based active learning workflow was employed to determine the minimum number of labeled data points required to train a high-fidelity surrogate. We demonstrated that using only approximately 0.61% of the entire dataset, it is possible to achieve predictive performance comparable to that of a model trained on the full dataset. This highlights the data efficiency of the proposed approach, which achieves competitive accuracy compared to more data-intensive neural-network-based methods. Moreover, the use of GPR naturally provides prediction uncertainty, which is not inherently available in neural-network-based models and typically requires additional treatment.

Although this study focused on linear elastic, two-phase, 2D microstructures, the MKS framework has been successfully extended to far more demanding scenarios, including multiscale models [54], high-resolution 3D

microstructures [48,50], multiphase systems [49,52], time-dependent microstructural evolution [46,51], and nonlinear mechanical response [47,53]. In all such cases, the primary computational bottlenecks involve the evaluation of large two-point statistics and high-rank PCA. These challenges, however, are effectively mitigated by advances in high-performance computing, including GPU-accelerated algorithms and incremental or randomized dimensionality reduction techniques.

For future studies, the proposed workflow can be extended to more complex material behaviors, such as plasticity and damage [87,88]. As demonstrated in previous works [50,89], the influence of interface statistics on the fidelity of surrogate models is expected to be even more pronounced in these regimes, further motivating continued investigation. One future avenue for extending the current framework is to generalize the prediction from a single elastic constant to the full set of stiffness components, using multi-output GP models [65,66]. While the unsupervised feature engineering workflow remains unchanged, the regression model must be adapted to capture correlations among multiple outputs, for instance through the linear model of coregionalization (LMC) [65,66], which introduces shared latent processes. This leads to moderate increases in both training time and memory usage. Additionally, the active learning query strategy must be revised: In the single-output case, the scalar predictive variance directly guides sample selection, whereas in the multi-output setting, the predictive uncertainty is a matrix and scalar acquisition metrics such as its determinant (i.e., the generalized variance) are typically used [65]. Given the importance of multiscale modeling in metamaterials [75,90,91], a potential direction for future work is to develop a two-scale approach in which FEM is employed for the macroscale problem and the MKS framework is used to model microscale unit cells, as previously adopted for composite materials [54]. Moreover, the MKS framework has already been extended to inverse stochastic microstructure design and applied to woven composites [92], motivating the development of an inverse design approach to identify metamaterial unit cell configurations that yield target mechanical responses. Eventually, although localization has been explored for materials with varying stiffness contrast using neural networks [93–95] or the MKS localization framework [96–98], challenges in high-contrast scenarios necessitates studying localization for metamaterial unit cells with infinite stiffness contrast. Such an investigation would seek to accurately predict not only homogenized properties but also full-field solutions, thereby enhancing the predictive capabilities of the present methodology.

ACKNOWLEDGMENTS

H.D., T.B., and S.R. gratefully acknowledge the funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 956401 (XS-Meta). M.A. and S.R.K. were sponsored by the Army Research Laboratory and was

accomplished under Cooperative Agreement No. W911NF-22-2-0106.

DATA AVAILABILITY

The data that support the findings of this article are openly available [99].

- [1] C. Lu, M. Hsieh, Z. Huang, C. Zhang, Y. Lin, Q. Shen, F. Chen, and L. Zhang, Architectural design and additive manufacturing of mechanical metamaterials: A review, *Engineering* **17**, 44 (2022).
- [2] Z. Chen, Y.-T. Lin, H. Salehi, Z. Che, Y. Zhu, J. Ding, B. Sheng, R. Zhu, and P. Jiao, Advanced fabrication of mechanical metamaterials based on micro/nanoscale technology, *Adv. Engn. Mater.* **25**, 2300750 (2023).
- [3] J. Chen, J. Chen, H. Wang, L. He, B. Huang, S. Dadbakhsh, and P. Bartolo, Fabrication and development of mechanical metamaterials via additive manufacturing for biomedical applications: A review, *Int. J. Extreme Manuf.* **7**, 012001 (2025).
- [4] A. H. Tajalsir, K. B. Mustapha, and T. Ibn-Mohammed, Numerical and random forest modelling of the impact response of hierarchical auxetic structures, *Mater. Today Commun.* **31**, 103797 (2022).
- [5] S. Luan, E. Chen, J. John, and S. Gaitanaros, A data-driven framework for structure-property correlation in ordered and disordered cellular metamaterials, *Sci. Adv.* **9**, eadi1453 (2023).
- [6] I. Ben-Yelun, G. Gómez-Carano, F. J. San Millán, M. Á. Sanz, F. J. Montáns, and L. Saucedo-Mora, GAM: General auxetic metamaterial with tunable 3D auxetic behavior using the same unit cell boundary connectivity, *Materials* **16**, 3473 (2023).
- [7] H. Danesh, D. Di Lorenzo, F. Chinesta, S. Reese, and T. Brepols, FFT-based surrogate modeling of auxetic metamaterials with real-time prediction of effective elastic properties and swift inverse design, *Mater. Des.* **248**, 113491 (2024).
- [8] A. Mishra, LatticeML: A data-driven application for predicting the effective young modulus of high temperature graph based architected materials, *Int. J. Interact. Des. Manuf.* **19**, 3899 (2025).
- [9] L. Wang, Y.-C. Chan, F. Ahmed, Z. Liu, P. Zhu, and W. Chen, Deep generative modeling for mechanistic-based learning and design of metamaterial systems, *Comput. Methods Appl. Mech. Eng.* **372**, 113377 (2020).
- [10] P. P. Meyer, C. Bonatti, T. Tancogne-Dejean, and D. Mohr, Graph-based metamaterials: Deep learning of structure-property relations, *Mater. Des.* **223**, 111175 (2022).
- [11] J.-H. Bastek and D. M. Kochmann, Inverse design of nonlinear mechanical metamaterials via video denoising diffusion models, *Nat Mach Intell* **5**, 1466 (2023).
- [12] L. Xiao, G. Shi, and W. Song, Machine learning predictions on the compressive stress-strain response of lattice-based metamaterials, *Int. J. Solids Struct.* **300**, 112893 (2024).
- [13] Y. Liu, W. Huang, Z. Wang, J. Zhang, and J. Liu, Machine learning-based mechanical performance prediction and design of lattice structures, *Int. J. Mech. Sci.* **294**, 110230 (2025).
- [14] X. Zheng, X. Zhang, T.-T. Chen, and I. Watanabe, Deep learning in mechanical metamaterials: from prediction and generation to inverse design, *Adv. Mater.* **35**, 2302530 (2023).
- [15] G. Cerniauskas, H. Sadia, and P. Alam, Machine intelligence in metamaterials design: A review, *Oxf. Open Mater.* **4**, itae001 (2024).
- [16] J. Song, J. Lee, N. Kim, and K. Min, Artificial intelligence in the design of innovative metamaterials: A comprehensive review, *Int. J. Precis. Eng. Manuf.* **25**, 225 (2024).
- [17] D. Lee, W. Chen, L. Wang, Y.-C. Chan, and W. Chen, Data-driven design for metamaterials and multiscale systems: A review, *Adv. Mater.* **36**, 2305254 (2024).
- [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, *J. Comput. Phys.* **378**, 686 (2019).
- [19] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, *Nat. Rev. Phys.* **3**, 422 (2021).
- [20] A. S. Krishnapriyan, A. Gholami, S. Zhe, R. M. Kirby, and M. W. Mahoney, Characterizing possible failure modes in physics-informed neural networks, in *Proceedings of the 35th International Conference on Neural Information Processing Systems (NIPS '21)* (Curran Associates Inc., Red Hook, NY, USA, 2021), pp. 26548–26560.
- [21] S. Wang, X. Yu, and P. Perdikaris, When and why PINNs fail to train: A neural tangent kernel perspective, *J. Comput. Phys.* **449**, 110768 (2022).
- [22] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, Scientific machine learning through physics-informed neural networks: Where we are and what's next, *J. Sci. Comput.* **92**, 88 (2022).
- [23] A. Farea, O. Yli-Harja, and F. Emmert-Streib, Understanding physics-informed neural networks: Techniques, applications, trends, and challenges, *AI* **5**, 1534 (2024).
- [24] M. Seeger, Gaussian processes for machine learning, *Int. J. Neural Syst.* **14**, 69 (2004).
- [25] C. K. I. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, 2006).
- [26] E. Schulz, M. Speekenbrink, and A. Krause, A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions, *J. Math. Psychol.* **85**, 1 (2018).
- [27] J. Wang, An intuitive tutorial to gaussian process regression, *Comput. Sci. Eng.* **25**, 4 (2023).
- [28] Y. Wang, W. Zhao, G. Zhou, and C. Wang, Analysis and parametric optimization of a novel sandwich panel with double-V auxetic structure core under air blast loading, *Int. J. Mech. Sci.* **142-143**, 245 (2018).
- [29] Y. Wang, W. Zhao, G. Zhou, Q. Gao, and C. Wang, Optimization of an auxetic jounce bumper based on Gaussian process metamodel and series hybrid GA-SQP algorithm, *Struct. Multidiscip. Optim.* **57**, 2515 (2018).

- [30] H. Song, E. Park, H. J. Kim, C.-I. Park, T.-S. Kim, Y. Y. Kim, and S. Ryu, Free-form optimization of pattern shape for improving mechanical characteristics of a concentric tube, *Mater. Des.* **230**, 111974 (2023).
- [31] A. Challapalli, J. Konlan, and G. Li, Inverse machine learning discovered metamaterials with record high recovery stress, *Int. J. Mech. Sci.* **244**, 108029 (2023).
- [32] R. Bostanabad, Y.-C. Chan, L. Wang, P. Zhu, and W. Chen, Globally approximate gaussian processes for big data with application to data-driven metamaterials design, *J. Mech. Des.* **141**, 111402 (2019).
- [33] L. Wang, Y.-C. Chan, Z. Liu, P. Zhu, and W. Chen, Data-driven metamaterial design with laplace-beltrami spectrum as “shape-dna”, *Struct. Multidiscip. Optim.* **61**, 2613 (2020).
- [34] T. Xue, T. J. Wallin, Y. Menguc, S. Adriaenssens, and M. Chiaramonte, Machine learning generative models for automatic design of multi-material 3D printed composite solids, *Extreme Mech. Lett.* **41**, 100992 (2020).
- [35] C. Morris, L. Bekker, M. R. Haberman, and C. C. Seepersad, Design exploration of reliably manufacturable materials and structures with applications to negative stiffness metamaterials and microstereolithography, *J. Mech. Des.* **140**, 111415 (2018).
- [36] A. du Plessis, I. Yadroitsava, and I. Yadroitsev, Effects of defects on mechanical properties in metal additive manufacturing: A review focusing on x-ray tomography insights, *Mater. Des.* **187**, 108385 (2020).
- [37] L. Xiao, X. Xu, W. Song, and M. Hu, A multi-cell hybrid approach to elevate the energy absorption of micro-lattice materials, *Materials* **13**, 4083 (2020).
- [38] K. Dastani, M. R. Movahhedy, H. Yu, S. Khodaygan, L. Zhang, and M. Y. Wang, Effect of geometric deviations on the strength of additively manufactured ultralight periodic shell-based lattices, *Eng. Fail. Anal.* **150**, 107328 (2023).
- [39] E. Cuan-Urquizo and R. G. Silva, Fused filament fabrication of cellular, lattice and porous mechanical metamaterials: A review, *Virtual Phys. Prototyping* **18**, e2224300 (2023).
- [40] L. Germain, C. A. Fuentes, A. W. van Vuure, A. des Rieux, and C. Dupont-Gillain, 3D-printed biodegradable gyroid scaffolds for tissue engineering applications, *Mater. Des.* **151**, 113 (2018).
- [41] S. R. Kalidindi, S. R. Niezgoda, and A. A. Salem, Microstructure informatics using higher-order statistics and efficient data-mining protocols, *JOM* **63**, 34 (2011).
- [42] S. R. Kalidindi, *Hierarchical Materials Informatics: Novel Analytics for Materials Data* (Elsevier, Amsterdam, 2015).
- [43] S. R. Kalidindi, Feature engineering of material structure for AI-based materials knowledge systems, *J. Appl. Phys.* **128**, 041103 (2020).
- [44] S. Li and E. Sitnikova, *Representative Volume Elements and Unit Cells: Concepts, Theory, Applications and Implementation* (Woodhead Publishing, Duxford, 2019).
- [45] D. B. Brough, D. Wheeler, and S. R. Kalidindi, Materials knowledge systems in python—A data science framework for accelerated development of hierarchical materials, *Integr. Mater. Manuf. Innovation* **6**, 36 (2017).
- [46] Y. C. Yabansu, V. Rehn, J. Hötzer, B. Nestler, and S. R. Kalidindi, Application of gaussian process autoregressive models for capturing the time evolution of microstructure statistics from phase-field simulations for sintering of polycrystalline ceramics, *Modell. Simul. Mater. Sci. Eng.* **27**, 084006 (2019).
- [47] A. Marshall and S. R. Kalidindi, Autonomous development of a machine-learning model for the plastic response of two-phase composites from micromechanical finite element models, *JOM* **73**, 2085 (2021).
- [48] Y. C. Yabansu, P. Altschuh, J. Hötzer, M. Selzer, B. Nestler, and S. R. Kalidindi, A digital workflow for learning the reduced-order structure-property linkages for permeability of porous membranes, *Acta Mater.* **195**, 668 (2020).
- [49] S. Hashemi, B. Ganapathysubramanian, S. Casey, J. Su, and S. R. Kalidindi, Feature engineering for microstructure-property mapping in organic photovoltaics, *Integr. Mater. Manuf. Innovation* **11**, 352 (2022).
- [50] D. M. de Oca Zapiain, E. Popova, F. Abdeljawad, J. W. Foulk, S. R. Kalidindi, and H. Lim, Reduced-order microstructure-sensitive models for damage initiation in two-phase composites, *Integr. Mater. Manuf. Innovation* **7**, 97 (2018).
- [51] Y. C. Yabansu, P. Steinmetz, J. Hötzer, S. R. Kalidindi, and B. Nestler, Extraction of reduced-order process-structure linkages from phase-field simulations, *Acta Mater.* **124**, 182 (2017).
- [52] M. I. Latypov and S. R. Kalidindi, Data-driven reduced order models for effective yield strength and partitioning of strain in multiphase materials, *J. Comput. Phys.* **346**, 242 (2017).
- [53] M. I. Latypov, L. S. Toth, and S. R. Kalidindi, Materials knowledge system for nonlinear composites, *Comput. Methods Appl. Mech. Eng.* **346**, 180 (2019).
- [54] H. F. Al-Harbi, G. Landi, and S. R. Kalidindi, Multi-scale modeling of the elastic response of a structural component made from a composite material using the materials knowledge system, *Modell. Simul. Mater. Sci. Eng.* **20**, 055001 (2012).
- [55] D. T. Fullwood, B. L. Adams, and S. R. Kalidindi, A strong contrast homogenization formulation for multi-phase anisotropic materials, *J. Mech. Phys. Solids* **56**, 2287 (2008).
- [56] S. R. Niezgoda, D. T. Fullwood, and S. R. Kalidindi, Delineation of the space of 2-point correlations in a composite material system, *Acta Mater.* **56**, 5285 (2008).
- [57] D. T. Fullwood, S. R. Niezgoda, B. L. Adams, and S. R. Kalidindi, Microstructure sensitive design for performance optimization, *Prog. Mater. Sci.* **55**, 477 (2010).
- [58] S. R. Niezgoda, Y. C. Yabansu, and S. R. Kalidindi, Understanding and visualizing microstructure and microstructure variance as a stochastic process, *Acta Mater.* **59**, 6387 (2011).
- [59] S. R. Niezgoda, A. K. Kanjarla, and S. R. Kalidindi, Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data, *Integr. Mater. Manuf. Innovation* **2**, 54 (2013).
- [60] C. Suh, A. Rajagopalan, X. Li, and K. Rajan, The application of principal component analysis to materials science data, *Data Sci. J.* **1**, 19 (2002).
- [61] R. Bro and A. K. Smilde, Principal component analysis, *Anal. Methods* **6**, 2812 (2014).
- [62] I. T. Jolliffe and J. Cadima, Principal component analysis: A review and recent developments, *Phil. Trans. R. Soc. A* **374**, 20150202 (2016).
- [63] B. Settles, Active learning literature survey (2009), <http://digital.library.wisc.edu/1793/60660>.
- [64] H. Liu, B. Yucel, B. Ganapathysubramanian, S. R. Kalidindi, D. Wheeler, and O. Wodo, Active learning for regression of structure-property mapping: The importance of sampling and representation, *Digit. Discov.* **3**, 1997 (2024).

- [65] M. O. Buzzy, D. M. de Oca Zapiain, A. P. Generale, S. R. Kalidindi, and H. Lim, Active learning for the design of polycrystalline textures using conditional normalizing flows, *Acta Mater.* **284**, 120537 (2025).
- [66] O. Ozbayram, D. Olsen, M. Annamaraju, A. E. Robertson, A. Venkatraman, S. R. Kalidindi, M. Zhou, and L. Graham-Brady, Batch active learning for microstructure-property relations in energetic materials, *Mech. Mater.* **205**, 105308 (2025).
- [67] D. D. Lewis, A sequential algorithm for training text classifiers: Corrigendum and additional data, *Acm Sigir Forum* **29**, 13 (1995).
- [68] P. Fernandez-Zelaia, Y. C. Yabansu, and S. R. Kalidindi, A comparative study of the efficacy of local/global and parametric/nonparametric machine learning methods for establishing structure-property linkages in high-contrast 3D elastic composites, *Integr. Mater. Manuf. Innovation* **8**, 67 (2019).
- [69] J.-H. Bastek and D. M. Kochmann, Inverse-design of nonlinear mechanical metamaterials via video denoising diffusion models: dataset and model checkpoints, ETHZ Research Collection (2023), URL: <https://doi.org/10.3929/ethz-b-000629716>.
- [70] S. Lucarini, L. Cobian, A. Voitus, and J. Segurado, Adaptation and validation of FFT methods for homogenization of lattice based materials, *Comput. Methods Appl. Mech. Eng.* **388**, 114223 (2022).
- [71] H. Moulinec and P. Suquet, A fast numerical method for computing the linear and nonlinear mechanical properties of composites, *C. R. Acad. Sci. Paris* **318**, 1417 (1994).
- [72] H. Moulinec and P. Suquet, A numerical method for computing the overall response of nonlinear composites with complex microstructure, *Comput. Methods Appl. Mech. Eng.* **157**, 69 (1998).
- [73] M. Schneider, A review of nonlinear FFT-based computational homogenization methods, *Acta Mech.* **232**, 2051 (2021).
- [74] S. Lucarini, M. V. Upadhyay, and J. Segurado, FFT based approaches in micromechanics: Fundamentals, methods and applications, *Modell. Simul. Mater. Sci. Eng.* **30**, 023002 (2022).
- [75] H. Danesh, T. Brepols, and S. Reese, Challenges in two-scale computational homogenization of mechanical metamaterials, *PAMM* **23**, e202200139 (2023).
- [76] J. C. Michel, H. Moulinec, and P. Suquet, A computational scheme for linear and non-linear composites with arbitrary phase contrast, *Int. J. Numer. Methods Eng.* **52**, 139 (2001).
- [77] S. Brisard and L. Dormieux, Fft-based methods for the mechanics of composites: A general variational framework, *Comput. Mater. Sci.* **49**, 663 (2010).
- [78] V. Monchieta and G. Bonnet, A polarization-based fft iterative scheme for computing the effective properties of elastic composites with arbitrary contrast, *Int. J. Numer. Methods Eng.* **89**, 1419 (2012).
- [79] M. Schneider, F. Ospald, and M. Kabel, Computational homogenization of elasticity on a staggered grid, *Int. J. Numer. Methods Eng.* **105**, 693 (2016).
- [80] J. Vondřejc, J. Zeman, and I. Marek, An FFT-based galerkin method for homogenization of periodic media, *Comput. Math. Appl.* **68**, 156 (2014).
- [81] J. Vondřejc, J. Zeman, and I. Marek, Guaranteed upper–lower bounds on homogenized properties by FFT-based Galerkin method, *Comput. Methods Appl. Mech. Eng.* **297**, 258 (2015).
- [82] F. Willot, Fourier-based schemes for computing the mechanical response of composites with accurate local fields, *C.R. Mec.* **343**, 232 (2015).
- [83] M. El Fallaki Idrissi, F. Praud, V. Champaney, F. Chinesta, and F. Meraghni, Multiparametric modeling of composite materials based on non-intrusive pgd informed by multiscale analyses: Application for real-time stiffness prediction of woven composites, *Compos. Struct.* **302**, 116228 (2022).
- [84] C. Gierden, J. Kochmann, J. Waimann, B. Svendsen, and S. Reese, A review of Fe-FFT-based two-scale methods for computational modeling of microstructure evolution and macroscopic material behavior, *Arch. Comput. Methods Eng.* **29**, 4115 (2022).
- [85] A. Cecen, T. Fast, and S. R. Kalidindi, Versatile algorithms for the computation of 2-point spatial correlations in quantifying material structure, *Integr. Mater. Manuf. Innovation* **5**, 1 (2016).
- [86] D. T. Fullwood, S. R. Niezgoda, and S. R. Kalidindi, Microstructure reconstructions from 2-point statistics using phase-recovery algorithms, *Acta Mater.* **56**, 942 (2008).
- [87] T. Brepols, S. Wulfinghoff, and S. Reese, Gradient-extended two-surface damage-plasticity: Micromorphic formulation and numerical aspects, *Int. J. Plast.* **97**, 64 (2017).
- [88] T. Brepols, S. Wulfinghoff, and S. Reese, A gradient-extended two-surface damage-plasticity model for large deformations, *Int. J. Plast.* **129**, 102635 (2020).
- [89] A. Venkatraman, D. M. de Oca Zapiain, and S. R. Kalidindi, Reduced-order models for ranking damage initiation in dual-phase composites using Bayesian neural networks, *JOM* **72**, 4359 (2020).
- [90] D. M. Kochmann, J. B. Hopkins, and L. Valdevit, Multiscale modeling and optimization of the mechanics of hierarchical metamaterials, *MRS Bull.* **44**, 773 (2019).
- [91] H. Danesh, L. Heußen, F. J. Montáns, S. Reese, and T. Brepols, A two-scale computational homogenization approach for elastoplastic truss-based lattice structures, *Results Eng.* **25**, 103976 (2025).
- [92] A. P. Generale, A. E. Robertson, C. Kelly, and S. R. Kalidindi, Inverse stochastic microstructure design, *Acta Mater.* **271**, 119877 (2024).
- [93] Z. Yang, Y. C. Yabansu, D. Jha, W.-k. Liao, A. N. Choudhary, S. R. Kalidindi, and A. Agrawal, Establishing structure-property localization linkages for elastic deformation of three-dimensional high contrast composites using deep learning approaches, *Acta Mater.* **166**, 335 (2019).
- [94] A. Harandi, H. Danesh, K. Linka, S. Reese, and S. Rezaei, SPiFOL: A spectral-based physics-informed finite operator learning for prediction of mechanical behavior of microstructures, *J. Mech. Phys. Solids* **203**, 106219 (2025).
- [95] C. Kelly and S. R. Kalidindi, Thermodynamically-informed iterative neural operators for heterogeneous elastic localization, *Comput. Methods Appl. Mech. Eng.* **441**, 117939 (2025).
- [96] T. Fast and S. R. Kalidindi, Formulation and calibration of higher-order elastic localization relationships using the MKS approach, *Acta Mater.* **59**, 4595 (2011).

- [97] Y. C. Yabansu, D. K. Patel, and S. R. Kalidindi, Calibrated localization relationships for elastic response of polycrystalline aggregates, *Acta Mater.* **81**, 151 (2014).
- [98] D. M. de Oca Zapiain and S. R. Kalidindi, Localization models for the plastic response of polycrystalline materials using the material knowledge systems framework, *Modell. Simul. Mater. Sci. Eng.* **27**, 074008 (2019).
- [99] H. Danesh, M. Annamaraju, T. Brepols, S. Reese, and S. Kalidindi, Reduced-order structure-property linkages for stochastic metamaterials (dataset), Zenodo (2025), <https://doi.org/10.5281/zenodo.15302946>.