Acta Materialia 271 (2024) 119877

Contents lists available at ScienceDirect

Acta Materialia

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

Full length article

Inverse stochastic microstructure design

Adam P. Generale [a], Andreas E. Robertson [a], Conlain Kelly [b], Surya R. Kalidindi [a,b,*]

[a] George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
[b] School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

Keywords:
Inverse design
Generative modeling
Uncertainty quantification
Bayesian inference
Computational materials design
Microstructure

ABSTRACT

Inverse Microstructure Design problems are ubiquitous in materials science; for example, property-driven microstructure design requires the inversion of a structure–property linkage. However, prior frameworks have struggled to address this problem's unique combination of challenges: the high dimensionality and stochasticity of microstructures, under sampled initial datasets, and ill-conditioning of the inversion. In this work, we propose a computational framework for Inverse Microstructure Design problems using a Bayesian methodology. We construct this framework from three modular components, enabling flexible extension and re-use. First, we define a low-dimensional, informative microstructure prior by integrating domain knowledge (i.e., statistical continuum mechanics) into a distributional learning scheme. This scheme includes multiple latent representations which address the challenges inherent to representing microstructures. Second, we define a property-specific likelihood using a multi-output Gaussian process regression surrogate model. Finally, we efficiently learn the conditional posterior density for a given target property, and generate samples using deep variational inference. We demonstrate our proposed method for solving stochastic microstructure design problems by identifying woven ceramic matrix composites matching target anisotropic thermal conductivities. Through this example, we analyze the integral role of each component in the inversion framework.

1. Introduction

The *inverse problem* is an archetypical mathematical challenge arising in countless disciplines, including atmospheric [1] and oceanographic sciences [2], quantum mechanics [3], astrophysics [4], medical imaging [5], and compressed sensing [6,7]. In its prototypical form, an inverse problem requires one to estimate the unknown state, $\boldsymbol{x}$, associated to an observed signal, $\boldsymbol{y}$, by a forward model, $f(\cdot)$.

$$\boldsymbol{y} = f(\boldsymbol{x}) \tag{1}$$

The forward model is an arbitrarily general mapping (e.g., a learned surrogate model, a physics-based simulation, or an actual physical system) [8–10] from causal information, $\boldsymbol{x}$, to observations, $\boldsymbol{y}$. Despite their simple formulation, inverse problems can prove extremely challenging to solve for several reasons, including corruption in the forward model, noninvertability, and high sensitivity of the inversion [8–10]. For example, *if* a suitable $\boldsymbol{x}$ exists at all, it may be non-unique or highly sensitive to variations in $\boldsymbol{y}$. When modeling real-world systems, a number of additional uncertainties can arise, including both corrupted output observations and stochastic forward models (i.e., both aleatoric and epistemic uncertainties in forward models [11]). This scenario is formally known as a *stochastic inverse problem* and often requires additional knowledge (via regularization or constraints) in order to regain tractability [8–10]. As such, solving a stochastic inverse problem requires both an (approximate) solution, $\hat{\boldsymbol{x}}$, and an estimated uncertainty in $\hat{\boldsymbol{x}}$.

Stochastic inverse problems are prolific within materials science. They most commonly occur when fusing experimental data and computational simulations, such as calibration of constitutive model parameters [12–16], the design of materials and processing conditions to meet target properties [17], and non-destructive examination (NDE) of constituent properties in manufacturing quality assurance [18–20]. One particularly laborious, but extremely useful class of stochastic inverse problem arises when the input signal, $\boldsymbol{x}$, is the material's microstructure itself. Here, the central goal is to identify the microstructure(s) displaying specific properties or performance targets. This inverse problem class is particularly challenging for several reasons, including the extremely high dimensionality and the irregular distribution of the microstructure space [21,22], the natural stochasticity of microstructures [23], and stochasticity in the structure–property (SP) linkage [24–26].

In this paper, we present a novel framework for efficiently solving inverse stochastic microstructure design problems that addresses the challenges described above. We formulate the design problem

* Corresponding author at: George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.
E-mail address: surya.kalidindi@me.gatech.edu (S.R. Kalidindi).

https://doi.org/10.1016/j.actamat.2024.119877
Received 18 September 2023; Received in revised form 19 February 2024; Accepted 28 March 2024
Available online 1 April 2024
1359-6454/© 2024 Acta Materialia Inc. Published by Elsevier Ltd. All rights reserved.

as one of Bayesian inversion, which provides a complete statistical methodology for approaching stochastic inverse problems and naturally accounts for forward model uncertainty alongside degeneracy in the inverse problem. We employ dual latent spaces – composed of a simple 2-point spatial correlations representation as well as a compact machine-learned encoding. This combination leverages the strengths of both statistical approaches and learning methods while covering the weaknesses of each to support both efficient forward models and tractable inversion. In addition to outlining algorithms to efficiently implement the proposed framework, we demonstrate its application to the design of woven composites given a target orthotropic thermal conductivity. In doing so, we provide statistical analysis of the framework's performance in a variety of salient scenarios, including interpolation within and extrapolation beyond the training distribution.

2. Background

2.1. Previous work on inverse microstructure design

A variety of frameworks have been applied to the inverse problem of microstructure design, differing in their choice of microstructure representation, optimality conditions, and prediction methodology. Topology optimization (TO) methods perform microstructure design by directly optimizing the spatial arrangement of salient microstructure features to meet target functional requirements. They have enjoyed some success at identifying optimal designs at the extremes of various properties (e.g., Poisson's ratio [27,28], shear and bulk moduli [29,30], thermal expansion [31], and thermal and electrical conductivities [32]) with volume fraction constraints. However, these methods tend to incur extremely high computational costs when they directly employ finite element (FE) schemes as forward models in their optimization framework. While this can be somewhat ameliorated through the use of surrogate forward models, TO remains an uncomfortable fit for microstructure inverse problems because it searches for a single optimal microstructure; it ignores the microstructure's natural stochasticity. Specifically, we are referring to the common observation that visually distinct, but statistically identical microstructures display similar properties [23,33–35]. As a result, the ideal inverse solution to a microstructure design problem would identify an optimal set of microstructures (i.e., a stochastic microstructure function or microstructure random process) instead of an optimal individual instance. This relaxation is also practically valuable because it allows inclusion of manufacturability constraints, since it is rarely economical or feasible to exactly manufacture a specified microstructure without tolerance.

Microstructure statistics are commonly adopted to quantify stochastic microstructure functions. Now, the inverse problem is redefined to identify microstructure statistics instead of individual instances. For example, the Microstructure-Sensitive Design (MSD) framework – a set of methods reminiscent of TO – employs statistical homogenization theories [36] as efficient forward models. In these theories, the properties of interest are represented as Neumann series expansions containing the microstructure's n-point spatial statistics [23,35]. In theory, such methods can identify optimal microstructure statistics, to arbitrary order – via direct optimization. However, due to the deleterious dimensionality explosion of higher-order statistics and difficulties performing optimization in high-dimensional spaces, the only tractable statistics used in these models have been overly simplified microstructure measures (mostly 1-point statistics).

Bayesian Optimization [37,38] (BO) frameworks have gained popularity as efficient optimization schemes which minimize the number of necessary forward model calls and account for uncertainty in the forward model [37]. They have been successfully used to identify optimal features of microstructures for properties such as ultimate tensile strength [39] and elastic [40] and bulk [41] moduli. However, efficient Bayesian Optimization schemes are limited to relatively low-dimensional design spaces, precluding a rigorous description of microstructure (i.e., the "curse of dimensionality"). Therefore, as with the MSD framework, these schemes have been practically limited to simple first-order microstructure statistics. This is a significant limitation in identifying potential design candidates as such statistics are intimately connected to the behavior of materials.

Recent computational advances in the development of distributional deep learning, such as Generative Adversarial Networks (GAN) [42], variational autoencoders (VAE) [43], normalizing flows (NF) [44,45], and diffusion-based models [46–50], have opened up new avenues in the design of heterogeneous microstructures. In particular, these efforts leverage data-driven latent representations of microstructures as extremely information-dense descriptors. These low-dimensional learned encodings have provided compact design spaces for inverse design of dual phase steels [39,51], metamaterials [52], band gap engineering in atomistics [53], and various composites [54–57]. In comparison with the MSD and traditional BO frameworks, these approaches provide significantly more expressive microstructure descriptors and better evaluation-time computational efficiency; in exchange, they require an a priori dataset on which to train the learning model. Combining these ideas, Deshwal and Doppa [58] propose a BO scheme in a learned latent space, leveraging the expressiveness of deep learning frameworks with natural handling of stochasticity in the forward model.

However, deep learning methods also have major limitations. In the majority of proposed deep learning frameworks for microstructure design, the goal is to deterministically predict a single microstructure; therefore, they do not account for microstructure stochasticity. Additionally, the surrogate forward models for property prediction are usually trained directly from the latent space. In this manner, the quality of the latent space is inherently tied to the training of the forward model and the encoder. As a result, learned latent spaces lack any particular domain knowledge or physics-supported topology [59], limiting their ability to represent novel or unknown structures. Finally, this non-stochastic treatment of material microstructure forces additional model complexity and data requirements during training as the network must needlessly learn to distinguish between statistically-identical structures. In the already data-scarce applications in materials informatics, this limitation has constrained prior works to consider simple microstructures (either 2-phase or gray-scale representations) and relatively small spatial domains (e.g., $\sim 64 \times 64$ 2D images). There is also growing evidence that these models struggle with more complicated material local states [60].

While learned latent spaces have largely been used for deterministic optimization to date, they also open the possibility to addressing *microstructure inverse design problems* stochastically. For example, Fung et al. [53] use a flow-based generative model to identify optimal distributions of design statistics for band-gap engineering of atomistics as a starting point for downstream optimization.

2.2. 2-point spatial correlations

Microstructure quantification lies at the heart of integrated computational materials engineering and materials informatics efforts. Since it is a standard practice to adopt microstructure statistics as condensed descriptors, a wide variety of statistics have been proposed; these include descriptors such as volume fractions, distributions of grain size, chord length, and tow/fiber position [36,61], amongst many others. All of these descriptors rest upon the idea that the true microstructure itself is a stochastic function, where individual observed microstructures are just singular samples from the governing stochastic microstructure function [23,62]. This theoretical treatment naturally handles the inherent stochasticity displayed by material microstructures. It connects the stochastic function itself, not individual instances, to the homogenized response. It also provides a methodology for understanding the underlying behavior between visually dissimilar microstructures which behave similarly in their homogenized response.

N-point spatial microstructure correlations (i.e., n-point statistics) are a rigorous physics-based mathematical formalization of this conceptualization [23,32,63,64]. In the generalized sense, n-point spatial correlations denote the correlation among a collection of material local states and randomly placed set of vectors, $\boldsymbol{\tau}_i$. They are formally defined as

$$f^{\alpha\beta\ldots\upsilon}(\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_{n-1}) = \mathbb{E}[M^{\alpha}(\boldsymbol{x})M^{\beta}(\boldsymbol{x}+\boldsymbol{\tau}_1)\ldots M^{\upsilon}(\boldsymbol{x}+\boldsymbol{\tau}_{n-1})] \tag{2}$$

where $M^{\alpha}(\boldsymbol{x})$ [23] denotes the stochastic microstructure function for the material local state α at spatial location $\boldsymbol{x}$, and $\mathbb{E}[\cdot]$ is the expectation operator. The n-point spatial correlation is only dependent upon the difference vectors denoted as $\boldsymbol{\tau}_i$ because the stochastic microstructure function is assumed to be stationary in space [23,35]. In practice, the ergodic hypothesis (i.e., spatial averaging within a sample is equivalent to averaging a single location across samples) is then leveraged to formulate this intractable expectation as an integral over a single sufficiently large sample volume [23,24,34,35,63–65]:

$$f^{\alpha\beta\ldots\upsilon}(\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_{n-1}) = \frac{1}{\Omega}\int_{\Omega} m^{\alpha}(\boldsymbol{x})m^{\beta}(\boldsymbol{x}+\boldsymbol{\tau}_1)\ldots m^{\upsilon}(\boldsymbol{x}+\boldsymbol{\tau}_{n-1})d\boldsymbol{x}. \tag{3}$$

Here, $m^{\alpha}(\boldsymbol{x})$ is an instantiation of $M^{\alpha}(\boldsymbol{x})$, and Ω defines the microstructural domain considered, frequently taken to be the volume of a representative volume element (RVE) [24].

Due to inherent discretization in microscopy techniques, a discrete approximation is the only practical approach for this computation [63,64]. A discrete microstructure function is then defined as m_s^{α}, representing the individual sample from the stochastic function, $M^{\alpha}(\boldsymbol{x})$, with s representing the discretized spatial voxel. A voxelized version of Eq. (3) for 2-point spatial correlations can then be written as follows [66]:

$$f_r^{\alpha\beta} = \frac{1}{S}\sum_{s=1}^{S} m_s^{\alpha}m_{s+r}^{\beta}. \tag{4}$$

Here, r indexes a permissible shift from one voxel to another in the microstructure, and s indexes the voxels in the discretized microstructural domain.

Discretized 2-point spatial correlations[1] are leveraged in the Materials Knowledge System (MKS) framework as highly expressive microstructure quantifiers for learning SP linkages [25,26,65,67–70], establishing mappings from microstructure spatial arrangements to their associated effective properties. As the extreme dimensionality of the complete set of 2-point spatial correlations precludes any direct incorporation in downstream SP model building efforts, dimensionality reduction within the MKS framework is achieved through Principal Component Analysis (PCA). PCA offers a robust dimensionality reduction technique for the set of spatial correlations considered in this work [71]. This method linearly transforms the original data into a new frame that is arranged to maximize the variance captured with each orthonormal basis vector, where the transformation can be mathematically described as

$$\boldsymbol{f}^{(j)} = \sum_{k=1}^{\min((J-1),R)} \alpha_k^{(j)}\boldsymbol{\varphi}_k + \overline{\boldsymbol{f}} \tag{5}$$

where $\boldsymbol{f}^{(j)}$ is a vectorized representation[2] of the spatial correlations of the jth microstructure, J is the total number of microstructures in the ensemble, R is the number of dimensions in the original dataset, $\alpha_k^{(j)}$ are the principal component weights (i.e., PC scores) of the jth microstructure, $\boldsymbol{\varphi}_k$ are the principal components (i.e., basis vectors of the transformed space), and $\overline{\boldsymbol{f}}$ represents the expected value of $\boldsymbol{f}^{(j)}$ across the ensemble. Frequently, the number of principal components to be retained is determined through evaluation of the explained variance by each basis vector as a portion of the total variance in the full dataset. Prior work [26,67,68,70,73,74] has demonstrated the ability to extract salient microstructure information dictating effective properties within the first few terms. Conversely, remaining information is scattered near-uniformly among the numerous remaining terms. It is this stable, interpretable reduction in dimensionality which enables the development of high-fidelity surrogate SP linkages in the forward direction. Briefly, we note that, in an inverse problem, we anticipate that a significantly higher number of retained terms will be required in order to capture the smaller variations in 2-point spatial correlations [75].

2.3. Sparse variational multi-output Gaussian process regression

Gaussian Process Regression is a frequent architecture chosen for building surrogate models in the MKS framework due to its flexibility, low data requirement, and natural uncertainty quantification. Gaussian processes ($\mathcal{GP}$s) [76] can be viewed as probability distributions over functions, providing important properties related to Bayesian analysis [77,78]. This relationship is denoted as a $\mathcal{GP}$, i.e., $f(\cdot) \sim \mathcal{GP}(\upsilon(\cdot), k(\cdot,\cdot'))$, which is uniquely determined through a mean function $\upsilon(\cdot)$ and a covariance function $k(\cdot,\cdot')$ parameterized by hyperparameters, $\boldsymbol{\theta}$. Often, the mean function is taken to be $\upsilon \equiv 0$ without loss of generality. Given a training dataset, $\{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$ of N corrupted observations with an assumed Gaussian noise $\xi_i \sim \mathcal{N}(0,\sigma_y^2)$, the collection of all training inputs can be denoted as $\boldsymbol{X}$, the vector of all outputs as $\boldsymbol{y}$, and $\boldsymbol{f}$ the infinite dimensional process latent function values. The covariance function used in this work is the spectral mixture (SM) kernel [79] defined as

$$k(\boldsymbol{x},\boldsymbol{x}') = \sum_{q=1}^{Q} w_q \frac{|\boldsymbol{L}_q|^{\frac{1}{2}}}{(2\pi)^{\frac{D}{2}}} \exp\left(-\frac{1}{2}\|\boldsymbol{L}_q^{\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{x}')\|^2\right)\cos\langle \boldsymbol{x}-\boldsymbol{x}', 2\pi\boldsymbol{\mu}_q\rangle \tag{6}$$

where Q represents the number of Gaussians considered in the mixture, w_q is the weight specifying the relative contribution of each mixture component, $\boldsymbol{\mu}_q = (\mu_q^{(1)},\ldots,\mu_q^{(P)})$ represents the mean vector and $\boldsymbol{L}_q = \mathrm{diag}(\upsilon_q^{(1)},\ldots,\upsilon_q^{(P)})$ the covariance matrix for the qth mixture component, with P denoting the dimensionality of the input $\boldsymbol{x}$. For this kernel, the resulting set of hyperparameters is $\boldsymbol{\theta} = \{\boldsymbol{\mu}_q, \boldsymbol{L}_q, w_q\}$. This kernel function has been shown to automatically discover short and long-range patterns, and as a result, an ability to extrapolate outside of the training dataset [79]. Impressive extrapolatory abilities have been demonstrated in diverse tasks such as predicting CO_2 concentrations [79], image inpainting [80], and feature extraction from electrophysiological signals [81]. $\boldsymbol{K}(\boldsymbol{X},\boldsymbol{X}')$ represents the constructed covariance matrix using the covariance function established in Eq. (6). For legibility, this will be abbreviated as $\boldsymbol{K}_{ff}$ to denote the covariance matrix constructed with the available training dataset, defining the latent process.

In low-dimensional datasets, $\mathcal{GP}$ regression provides a powerful and flexible method for the learning of scalar outputs while providing uncertainty measures on its predictions. A core limitation preventing its widespread use is the necessity of computing $\left[\boldsymbol{K}_{ff}+\sigma_y^2\boldsymbol{I}\right]^{-1}$, whose computational complexity of $\mathcal{O}(N^3)$ limits its practical use when confronted with even modestly large datasets. Sparse global approximations aim to address this limitation through the use of M inducing inputs $\boldsymbol{Z} = \{\boldsymbol{z}_m\}_{m=1}^{M}$ independent of $\boldsymbol{X}$, $M \leq N$, providing a low rank approximation to the full covariance $\boldsymbol{K}_{ff}$. An elegant method to accomplish this task is the Variational Free Energy (VFE) approximation [82]. The covariance function hyperparameters and inducing inputs are jointly optimized by maximizing a lower bound to the exact marginal likelihood, where the inducing input locations are selected through minimization of the Kullback–Leibler (KL) divergence (denoted as D_{KL}) between the variational $\mathcal{GP}$ and the true posterior $\mathcal{GP}$. Importantly, the quantity of inducing points included provides a method of regularization on the solution, avoiding overfits [82,83]. This lower bound to the exact marginal likelihood can be approximated by first placing an assumption on the variational posterior as $q(\boldsymbol{f}) =$

[1] 2-point spatial correlations are also referred to as 2-point correlation functions or 2-point spatial correlations.

[2] Scaling factors may be applied to sets of flattened spatial correlations within this vector [26,69,72].

$q(\boldsymbol{u}, \boldsymbol{f}_{\neq \boldsymbol{u}}|\boldsymbol{\theta}) = p(\boldsymbol{f}_{\neq \boldsymbol{u}}|\boldsymbol{u}, \boldsymbol{\theta})q(\boldsymbol{u})$. The VFE bound can then be optimized in closed form as

$$\begin{aligned}\mathcal{F}(q^*, \boldsymbol{\theta}) &= \log p(\boldsymbol{y}|\boldsymbol{\theta}) - D_{KL}(q(\boldsymbol{f})\|p(\boldsymbol{f}|\boldsymbol{y}, \boldsymbol{\theta})) \\ &= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log\left|\hat{\boldsymbol{K}}_{ff}\right| - \frac{1}{2}\boldsymbol{y}^\top \hat{\boldsymbol{K}}_{ff}^{-1}\boldsymbol{y} - \frac{1}{2\sigma_y^2}\mathrm{tr}(\boldsymbol{K}_{ff} - \boldsymbol{Q}_{ff})\end{aligned} \tag{7}$$

where $\hat{\boldsymbol{K}}_{ff} = \boldsymbol{Q}_{ff} + \sigma^2\boldsymbol{I}$ and $\boldsymbol{Q}_{ff} = \boldsymbol{K}_{fu}\boldsymbol{K}_{uu}^{-1}\boldsymbol{K}_{uf}$ provides the low rank approximation.

Predictions of the sparse variational model can be expanded to handle multi-output functions in a similar manner to the scalar output case, through expansion of the covariance matrix to express correlations between related outputs [84]. Such Multioutput Gaussian processes (MOGP) learn a multioutput function $f(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}^P$ with the input space $\mathcal{X} \in \mathbb{R}^D$. The pth output of $f(\boldsymbol{x})$ is expressed as $f_p(\boldsymbol{x})$, with its complete representation given as $f = \{f_p(\boldsymbol{x})\}_{i=1}^P$. MOGPs are similarly completely defined by their covariance function (assuming $\upsilon \equiv 0$), resulting in a covariance matrix $\boldsymbol{K} \in \mathbb{R}^{NP\times NP}$. Under this scenario, the advantages of a sparse representation become increasingly apparent, reducing the complexity from $\mathcal{O}(N^3P^3)$ to $\mathcal{O}(M^3P)$ with efficient inference schemes utilizing interdomain approximations [85]. In this work, the multi-output covariance matrix is constructed through the Linear Model of Coregionalization (LMC) [84,86]. This model represents a method of constructing the multi-output function from a linear transformation $W \in \mathbb{R}^{P\times L}$ of L independent functions $g(\boldsymbol{x}) = \{g_l(\boldsymbol{x})\}_{l=1}^L$. Each function is constructed as an independent $\mathcal{GP}$, $g_l(\boldsymbol{x}) \sim \mathcal{GP}(0, k_l(\boldsymbol{x}, \boldsymbol{x}'))$, each with its own covariance function, resulting in the final expression $f(\boldsymbol{x}) = \boldsymbol{W}g(\boldsymbol{x})$. The multi-output covariance function described by this model is then expressed as

$$k(\{\boldsymbol{x}, p\}, \{\boldsymbol{x}', p'\}) = \sum_{l=1}^{L} W_{pl}k_l(\boldsymbol{x}, \boldsymbol{x}')W_{p'l} \tag{8}$$

which can be seen to encode correlations between output dimensions. Due to the form of the covariance function in Eq. (8), the MOGP provides for an incredibly expressive model.

2.4. Bayesian inference

The ubiquity of stochastic inverse problems has given rise to numerous solution approaches with varying mechanisms to either uniquely select between or quantify the degenerate solutions [8,10,87–89]. One of the most popular solution approaches is Bayesian inference, since it allows incorporation of prior knowledge and provides a complete statistical inferential methodology for addressing stochastic inverse problems. In such schemes, both independent inputs $\mathbf{x}$ and observations $\mathbf{y}$ are treated as random variables, which are coupled via a *likelihood* $p(\boldsymbol{y}|\boldsymbol{x})$. The likelihood directly incorporates the system's forward model. Additionally, Bayesian inference requires a *prior* distribution over input states, $p(\boldsymbol{x})$. This distribution is used to describe prior knowledge of the input space, enabling a tractable solution through regularization driven by the practitioner. Once these two are specified, the solution to the inverse problem is itself a conditional *posterior* density $p(\boldsymbol{x}|\boldsymbol{y})$, which can be recovered through the application of Bayes' Theorem.

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{\int_X p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}} \tag{9}$$

Importantly, this density provides a theoretically compact representation of all degenerate solutions (as samples from the posterior $p(\boldsymbol{x}|\boldsymbol{y})$). In practice, however, the solution quality and tractability often depend heavily on the form of prior and likelihood. The normalization constant provided by the integral in the denominator is in general intractable, leading to the need for approximate solution and sampling schemes, such as importance sampling [77,90] or Markov Chain Monte Carlo [77,91].

2.5. Variational autoencoder

Variational Autoencoders (VAE) [43] are a class of distributional learning model which are used to approximate data distributions (e.g., $p(\boldsymbol{x})$). This classic architecture is an extension of the traditional autoencoder [92,93]; it combines two networks that are trained to map samples between a simple, latent distribution, traditionally a unit variance multivariate Gaussian, and samples from the data distribution. Each network supports one direction of the transformation. In cases where the latent distribution is lower dimensional than the data itself, the learned transformations are interpreted as a probabilistic encoding and decoding, taking as an input the higher dimensional $\boldsymbol{x}$, and providing a nonlinear transformation of the information into a descriptive latent space. The model provides the major benefit of transforming a complex, irregular target distribution into a well-behaved latent distribution. The well-structured latent encodings often act as features for downstream tasks. In Variational Autoencoders, the relationship between these two random variables is fully defined in a Bayesian setting. Mathematically, the observable random variable $\boldsymbol{x} \in \mathbb{R}^n$ becomes the result of a causal latent random variable $\boldsymbol{z} \in \mathbb{R}^m$, often with $m \ll n$. The transformation, for example, from the latent space to the real data space is defined as

$$p(\boldsymbol{x}) = \int p_\psi(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})d\boldsymbol{z}. \tag{10}$$

In both the forward and the reverse transformation, the unknown likelihood – in this case, $p_\psi(\boldsymbol{x}|\boldsymbol{z})$ – is approximated using a neural network. A prior latent distribution is arbitrarily selected and imposed, often a unit variance multivariate Gaussian, $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. The parameters of the forward and reverse transformations, $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ and $p_\psi(\boldsymbol{x}|\boldsymbol{z})$, respectively, are optimized by minimizing a composite loss function, including the reconstruction error from the latent space, and a KL-divergence term between the encoding and prior unit-Gaussian of the latent space as follows.

$$\mathcal{L}_{VAE}(\psi, \phi) = \mathbb{E}_{z\sim q_\phi(z|x)}\log p_\psi(\boldsymbol{x}|\boldsymbol{z}) - D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})\|p(\boldsymbol{z})). \tag{11}$$

The first term enforces that the forward and reverse transformations are approximate inverses, while the second enforces that the latent space is unit-Gaussian. The β-VAE [94] is a slight extension of this foundational framework which allows the user to tune training emphasis between these two objectives, promoting disentanglement of the latent representation when $\beta > 1$.

$$\mathcal{L}_{VAE}(\psi, \phi; \beta) = \mathbb{E}_{z\sim q_\phi(z|x)}\log p_\psi(\boldsymbol{x}|\boldsymbol{z}) - \beta D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})\|p(\boldsymbol{z})). \tag{12}$$

The β parameter additionally promotes the learning of the latent unit-Gaussian structure, which may be important in downstream tasks. At first glance, it might be expected that the additional hyperparameter β, exerting additional pressure on the latent representation to match the prior, would necessarily result in a degradation of reconstruction performance. Paradoxically, this is not always the case depending on the value selected for β: for some datasets, a more efficient encoding can be obtained through forcing the learning of a disentangled latent representation without significant degradation to reconstruction error [94,95]. It is readily apparent that by emphasizing the KL-divergence term, latent spaces with $\beta > 1$ will tend towards the specified prior $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$; unfortunately, for any given problem the optimal choice of β cannot be determined *a priori*.

2.6. Deep variational inference

Traditional methods for sampling posterior distributions in Bayesian inference – that is, drawing samples from $p(\boldsymbol{x}|\boldsymbol{y})$ via Eq. (9) – are often prohibitively slow to converge and scale poorly to larger state spaces [96,97]. The recent work of Sun et al. [98,99] offers an alternative that addresses these limitations by training a variational

approximation of the posterior using a normalizing flow-based generative model instead of aiming to directly generate samples [98,99]. Once trained, this generative neural network can be sampled almost instantly (via sampling a Gaussian and a neural network forward pass). Following the standard stochastic variational inference, the normalizing flow model is trained to minimize the KL-divergence between the variational distribution $q_\theta(\boldsymbol{x})$ and the desired posterior $p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$. Optimal generative model parameters are given by the following expression.

$$\begin{aligned}\theta^* &= \underset{\theta}{\operatorname{argmin}}\, D_{KL}(q_\theta(\boldsymbol{x})\|p(\boldsymbol{x}|\boldsymbol{y})) \\ &= \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{\boldsymbol{x}\sim q_\theta(\boldsymbol{x})}\left[-\log p(\boldsymbol{y}|\boldsymbol{x}) - \log p(\boldsymbol{x}) + \log q_\theta(\boldsymbol{x})\right]\end{aligned} \tag{13}$$

Here, the expectation can be approximated through Monte Carlo estimation. The variational distribution itself is constructed by combining a reference distribution, often $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, with a learned transformation mapping between the reference and the data distribution. A single invertible network, $f_\theta(\cdot)$, is used to learn a bijective map $f : \boldsymbol{u} \to \boldsymbol{x}$ between the two distributions. In this sense, this construction is simpler than a variational autoencoder since only one network is used, however, the invertability requirement often makes training more difficult. The primary benefit of adopting a normalizing flow with invertible coupling layers is that it enables the *exact* specification of the variational sample likelihood $q_\theta(\boldsymbol{x})$ through the change of variable formula [44,100,101]. Practically, then, training is performed by estimating Eq. (13) as follows.

$$\theta^* \approx \underset{\theta}{\operatorname{argmin}}\, \frac{1}{N}\sum_{k=1}^{N}\left[-\log p(\boldsymbol{y}|\boldsymbol{x}) - \lambda \log p(\boldsymbol{x}) + \log\left|\det \frac{\partial f_\theta(\boldsymbol{u})}{\partial \boldsymbol{u}}\right|\right] \tag{14}$$

In Sun's work, the mapping is approximated using a normalizing flow constructed with RealNVP [101] affine coupling layers. λ, usually taken to be between 0 and 1, is a regularization hyperparameter modulating the importance of the prior, $p(\boldsymbol{x})$, in the Bayes' Rule computation [99]. Low values of λ will allow wider exploration of the X space – in our case, potentially identifying more novel inputs x, but also increase the likelihood of producing infeasible values.

3. Stochastic microstructure design

3.1. Framework

In this section, we develop the computational scaffolding necessary to treat the Stochastic Microstructure Design problem through Bayesian inference. Fig. 1 visually summarizes the overarching strategy. For this problem, we redefine variable names to correspond to the task of microstructure design, where inputs of the forward model $\boldsymbol{x}$ are replaced by our microstructure representation $\boldsymbol{\alpha}$ (i.e., principal component (PC) scores of spatial correlations), and outputs $\boldsymbol{y}$ are replaced by the target effective property set $\boldsymbol{k}$.

The proposed framework combines the respective strengths of several deep generative model architectures to be interpretable, modular, and robust. It is compromised of 3 main components each addressing a salient element of the Bayes' Rule computation described in Eq. (9).

1. A forward model $\mathcal{H} : \boldsymbol{\alpha} \to \boldsymbol{k}$. This is used to approximate the likelihood term, $p(\boldsymbol{k}|\boldsymbol{\alpha})$, mapping a microstructure representation, $\boldsymbol{\alpha}$, to effective properties, $\boldsymbol{k}$.
2. A β-VAE based microstructure prior, $p(\boldsymbol{\alpha})$. This crucial mechanism derives a usable prior distribution over $\boldsymbol{\alpha}$ that regularizes the inverse problem, provides data compression alongside efficient sampling.
3. Finally, a flow-based generative model trained on a specific target property $\boldsymbol{k}^*$ which, when combined with the β-VAE's decoder, estimates the posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{k}^*)$.

We emphasize that the first two components only incur a one time upfront training cost (e.g., over all possible $\boldsymbol{\alpha}, \boldsymbol{k}$), while the inference of the posterior $p(\boldsymbol{\alpha}|\boldsymbol{k}^*)$ (the third component) must be repeated for each target property $\boldsymbol{k}^*$. An amortized approach [102] would certainly be possible, but is outside the scope of this work.

In this work, we adopt the MKS framework for Materials Informatics and quantify material microstructures using their 2-point spatial correlations. To simplify downstream computations, we compress these statistics into a primary lower-dimensional design manifold using PCA [70,103–105]. This first latent space compactly represents microstructures, and, more specifically, their 2-point spatial correlations, via their principal components, $\boldsymbol{\alpha}$. We emphasize that the ultimate goal of the presented framework is to recover the distribution of 2-point spatial correlations corresponding to a desired property set. The usage of correlations removes the notion of an origin and provides a translation-invariant representation of microstructures (which is desirable for predicting global properties). The PCA process removes the notion of space at all, exchanging high-dimensional spatial fields for dense coefficients corresponding to an orthonormal set of spatial basis functions. This provides a low-rank representation of microstructure which is interpretable, analytically well-founded, and extremely cheap to compute.

However, we adopt a much larger PC representation than usual: in the presented case studies we use ∼1000 PC scores instead of the ∼10 in most previous MKS efforts [25,26,65,67–70]. This switch reflects the difference in goals between this work and previous efforts. These remaining principal components define finer microstructure details; while they are unnecessary for building forward models because such details have little effect on resulting properties, they are often necessary for downstream tasks, such as generation of microstructure instances [33,47,75] corresponding to the inverse solution identified by the proposed framework. Therefore, retaining additional PC scores is necessary to produce a robust solution to the inverse problem. We emphasize that this is still a significant compression when compared to the 109^3 voxels of the original microstructures. This quantity of retained PC scores was selected as it denoted the threshold past which greater than 99.99% of the variance was explained in our demonstration dataset.

3.2. A surrogate likelihood

It is unavoidably necessary to evaluate Bayes' Rule an extremely large number of times in most practical implementations of Bayesian inference.[3] This requirement poses a significant challenge as the structure-to-property mapping, needed in the likelihood computation, is often an expensive, high-fidelity numerical simulation [18]. Fortunately, the construction of reduced-order surrogate models alleviates this computational bottleneck.

In this work, we use Sparse Variational Multi-Output Gaussian Process Regression (SV-MOGP) with a Spectral Mixture (SM) kernel to approximate the forward model, $\mathcal{H}$, from our principal component microstructure representation to property. These GP surrogates not only provide probabilistic predictions of the output, but also inform the computation of the likelihood. The expressive SM kernel was selected in this work due to its previously observed stability during extrapolation [79]. The stability of this surrogate model is explored in more detail in Appendix E. Such stability is often important because existing microstructure data rarely covers the design space uniformly and is often clustered into subregions [70]. In the present case, this stability was critical to reaching beyond the existing microstructure clusters and facilitating the ultimate goal of materials discovery: identifying valuable, novel microstructures.

The surrogate modeling framework has several parameters that must be tuned to facilitate successful performance. Appendices A and

[3] In the absence of model conjugacy.

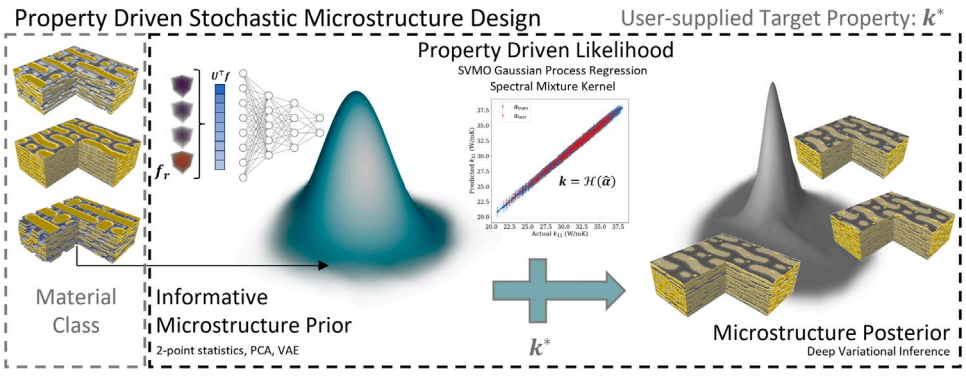

Fig. 1. Depiction detailing the overall framework. First, a generalized microstructure prior is derived to represent a target microstructure class. A surrogate likelihood encapsulates the requisite SP linkage for this class. The combination of the microstructure prior, the likelihood, and a user-supplied target property value produces a property conditioned posterior distribution – the solution to the inverse problem. Note that the microstructure diversity in the posterior is far smaller than the prior. The distributions shown in this figure are directly extracted from the first case study presented in this work.

E offer a much more thorough analysis of the specifics. Briefly, the parameter with the most theoretical impact to the inverse problem is the number of utilized principal components. In practice, extensive prior efforts have demonstrated that only the first few principal components are necessary to successfully train surrogate forward models in the MKS framework [26,67,68,70,73,74]. With respect to the training process, this is quite useful because it means that lightweight models trained on small datasets are usually sufficient. However, this also highlights the primary source of complexity surrounding microstructure inverse problems: the problem is highly under-constrained. Through the likelihood, a given property can only inform the value of the first few principal components. For example, in the case studies below, the MOGP is fed only the first 16 PC scores; the remaining principal components must be identified by the prior.

3.3. A microstructure prior

A well-crafted prior is an essential component to the inversion process. While this is generally important for any Bayesian inference framework, it is especially important in our case because the likelihood can only inform the PC scores it takes as input. As a result, the microstructure prior plays the triple role of informing the remaining PC scores by identifying nonlinear interdependencies between principal components, encoding prior domain knowledge or constraints, and efficiently sampling the high-dimensional microstructure space.

Unfortunately, the complex topology and high dimensionality of the principal component space makes it complicated to estimate a useful prior. As will be shown later, mapping randomly-generated microstructures into principal component space generally produces multiple clusters and large cavities (regions where microstructures have not yet been instantiated). To overcome this, we introduce a second latent space by utilizing a β-VAE as an additional nonlinear embedding of the PC scores. By design this latent space affords a natural prior: the unit-Gaussian $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Therefore, we perform Bayesian inference in the β-VAE's latent space.

The β-VAE architecture was selected because the β hyperparameter provides a means of trading consistency of the unit-Gaussian latent structure against reconstruction performance of the encoding/decoding. This directly provides a means of tightening the prior to adjust performance of the inversion. We trained the β-VAE to further compress the primary PCA latent space instead of taking the full 2-point correlations as inputs because it significantly simplified the compression that the autoencoder must learn. In practice, we observed that this combined process allowed us to achieve superior performance with much smaller models. In essence, PCA acts as an orthogonal linear transformation, the first layer in a larger augmented-VAE mapping 2-point spatial correlations to $\boldsymbol{z}$ and back. Specifics regarding the β-VAE architecture and training can be found in Appendix B.

Finally, we retained the usage of two latent spaces in the framework – the primary PC space and the secondary β-VAE space – instead of transitioning to just the single β-VAE space where we perform Bayesian inference because we found that constructing linkages from the β-VAE's latent space resulted in significantly degraded performance in comparison to construction solely based on the latent PC space. For the SV-MOGP model utilized in the presented case studies, training on the β-VAE latent space resulted in worsened mean predictions alongside higher marginal posterior uncertainty. Therefore, the likelihood $p(\boldsymbol{k}^*|\boldsymbol{z})$ used in Bayesian inference is then defined through the use of both the decoder and the SV-MOGP forward model (i.e., $p(\boldsymbol{k}^*|\boldsymbol{z}) = \int p(\boldsymbol{k}^*|\boldsymbol{\alpha}) p_\psi(\boldsymbol{\alpha}|\boldsymbol{z}) d\boldsymbol{\alpha}$). Additional details regarding this comparison can be found in Appendix D.

3.4. Inference

The two components described above combine to fully define Bayes' Rule in the β-VAE based latent space.

$$p(\boldsymbol{z}|\boldsymbol{k}^*) \propto \left(\int p(\boldsymbol{k}^*|\boldsymbol{\alpha}) p_\psi(\boldsymbol{\alpha}|\boldsymbol{z}) d\boldsymbol{\alpha} \right) p(\boldsymbol{z}) \tag{15}$$

The likelihood, given by the integral in Eq. (15), directly incorporates the forward SV-MOGP alongside the decoder of the β-VAE. The prior is the unit-Gaussian, $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ – the heavily imposed prior distribution on the β-VAE's latent space. Finally, deep variational inference is performed to solve the microstructure inverse problem using a flow-based generative model to estimate the target property conditioned posterior, $q_{\theta^*}(\boldsymbol{z}) \approx p(\boldsymbol{z}|\boldsymbol{k}^*)$. The model is trained for a specific target property value, $\boldsymbol{k}^*$, with associated learned weights θ^*. After training, we can generate individual solutions by sampling our approximate posterior and, subsequently, transforming them into 2-point spatial correlations by applying the β-VAE decoder and then the inverse PCA transformation.

$$p(\boldsymbol{\alpha}|\boldsymbol{k}^*) \approx \int p_\psi(\boldsymbol{\alpha}|\boldsymbol{z}) q_{\theta^*}(\boldsymbol{z}|\boldsymbol{k}^*) d\boldsymbol{z} \tag{16}$$

Additional background and details regarding the specifics of the flow-based generative model implementation can be found in Appendix C.

4. Results and discussion

In order to validate the proposed Stochastic Microstructure Design framework, we present a stress test consisting of three increasingly challenging test cases. These cases consist of {microstructure, property} pairs selected from a dataset [26] that was previously used to study orthotropic thermal conductivities of layered woven composites. Collectively, the three tests are designed to scrutinize the framework's performance across increasingly under sampled regions of the dataset, resulting in increasingly challenging test cases.

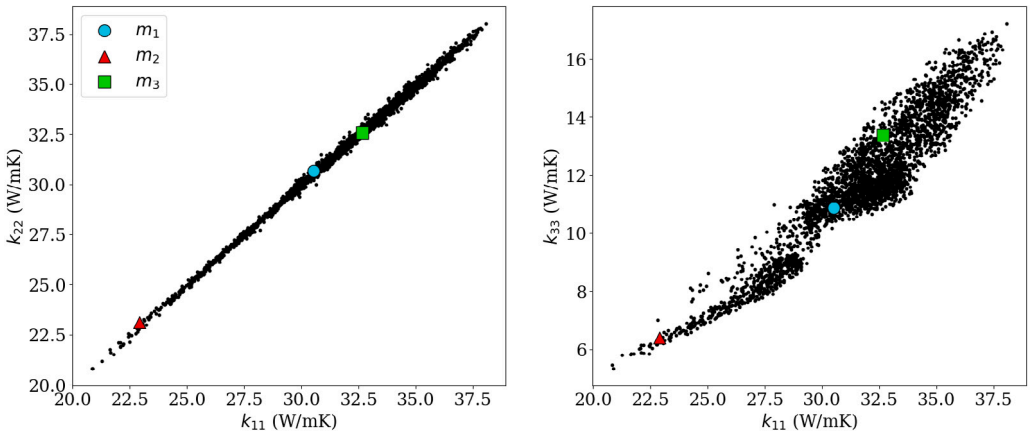

Fig. 2. Property closure of the microstructure ensemble considered in this work. Select microstructures are identified as test cases for the proposed stochastic microstructure design framework.

4.1. Microstructure dataset

The dataset contains 3125 instantiations of an 8-ply stack of 5-harness satin (5HS) weave ceramic matrix composites (CMCs) from prior work [26]. Each 3D microstructure in this dataset is comprised of 109 × 109 × 109 voxels, each of which are occupied by one of three possible distinct material local states: tow, matrix, or pore. The microstructure ensemble was generated by varying the tow major and minor axes, tow spacing, ply spacing, and matrix thickness. This resulted in significant microstructural diversity and, consequently, effective orthotropic thermal conductivity. Additional details regarding the generative procedure and numerical simulation inputs can be found in prior work [26].

Fig. 2 visually summarizes the resulting property closure for this dataset.[4] Each microstructure is statistically represented by a subset of its 2-point spatial correlations; we computed $\{f_r^{00}, f_r^{11}, f_r^{22}, f_r^{02}\}$ using Eq. (4), where the material local state indexes $\{0, 1, 2\}$, correspond to tow, matrix, and pore, respectively. An exemplar set of spatial correlations (corresponding to the first test case discussed later) can be seen in Fig. 3. Fig. 3b highlights the $X - Z$ cross-section displaying the uniformly deposited matrix surrounding individual tows, and the residual porosity from the approximated chemical vapor deposition process. Fig. 3a displays the base 5HS weave architecture, with the relative locations of cross-over points in the weave demarcated by arrows. The secondary peaks highlighted in Fig. 3d for the tow auto-correlation f_r^{00}, coincide with cross-over locations of the 5HS weave, with tertiary peaks across the domain reflecting the periodic nature of the microstructure itself. Importantly, the central peaks in the set of auto-correlations (i.e., $\{f_r^{00}, f_r^{11}, f_r^{22}\}$), reflect the volume fractions of each local state, and necessarily sum to one. A deeper analysis and interpretation of the spatial correlation maps for this class of microstructures can be found in our prior work [26].

As discussed, we reduced the dataset's dimensionality using PCA on the flattened spatial correlations. The full set of 2-point spatial correlations for each microstructure instantiation is represented by a feature vector of length $4 \times 109^3 = 5,180,116$. When performing PCA, scaling factors were applied to the subsets of spatial correlations as $\{f_r^{00}, 9.83 f_r^{11}, 21.58 f_r^{22}, 8.44 f_r^{02}\}$ such that the variance of each subset were equivalent (i.e., no particular subset dominated the transformation). Fig. 4 illustrates projections of the microstructure ensemble in the first 3 PCs as well as cross-sections of the test microstructures. We kept 1024 PC scores in order to overwhelmingly capture any salient variations in the dataset and preserve higher-order connections between PCs. Fig. 4 highlights the motivation for selecting these particular test cases, which are arranged in increasingly sparse regions in PC space. Consequently, this represents increasing levels of difficulty in identifying the correct solution to the inverse problem. In particular, it can be observed that case #1 corresponds to a dense region in the microstructure ensemble, and case #3 is locally isolated, with case #2 set to an intermediate location on the extremity of the target effective properties.

4.2. Case study 1: Microstructure identification in dense regions

In each case study, we jointly select a target property and microstructure pair, $\{m_1, \boldsymbol{k}_1^*\}$ from the training dataset for the prior and likelihood. The first case study tests the framework's performance on a modestly challenging inverse problem in which a valid solution posterior (conditioned on $\boldsymbol{k}_1$) should be readily identifiable; moreover, this posterior should contain the known microstructure m_1. We select the pair from a densely-sampled region in both the property closure and the microstructure ensemble. The primary challenge that arises in this situation is degeneracy in the solution – a wide number of microstructures can produce a narrow range of properties. Therefore, in addition to benchmarking the framework's performance, this situation tests its capacity to identify a wide and likely multimodal posterior.

The posterior $p(\boldsymbol{z}|\boldsymbol{k}_1^*)$ was identified through the procedure laid out in Section 3, enabling us to draw conditioned samples over the latent space of the β-VAE. Analyzing this space is complicated because the β-VAE latent dimensions do not inherently possess any ranking corresponding to information (as opposed to PC scores, which are ordered by explained variance). Due to these limitations, we begin by identifying the maximally informative dimensions of the full 64-dimensional posterior distribution. The dimensions of $\boldsymbol{z}$ exhibiting the largest deviations from unit Gaussian were identified by ranking in descending order the absolute value of their marginal higher-order moments,[5] summarized in Fig. 5. As can be seen in Fig. 5b, only four (of the 64 total) latent dimensions of $\boldsymbol{z}$ display strong sensitivity to this specific target property.

Fig. 5a displays projections of the posterior's most informative dimensions. Their distribution is non-Gaussian, capturing the complex interdependencies between these parameters and the property in this nonlinear design space. This is most prominently displayed in the off-diagonal projections. Importantly, the framework passes our simple initial test: the Maximum a Posteriori (MAP) location of the instantiating microstructure for this property is readily located within the

[4] The projection of k_{22} and k_{33} was omitted due to the similarity of in-plane properties.

[5] Dimensions whose moments match the initial Gaussian distribution are necessarily uninformative.

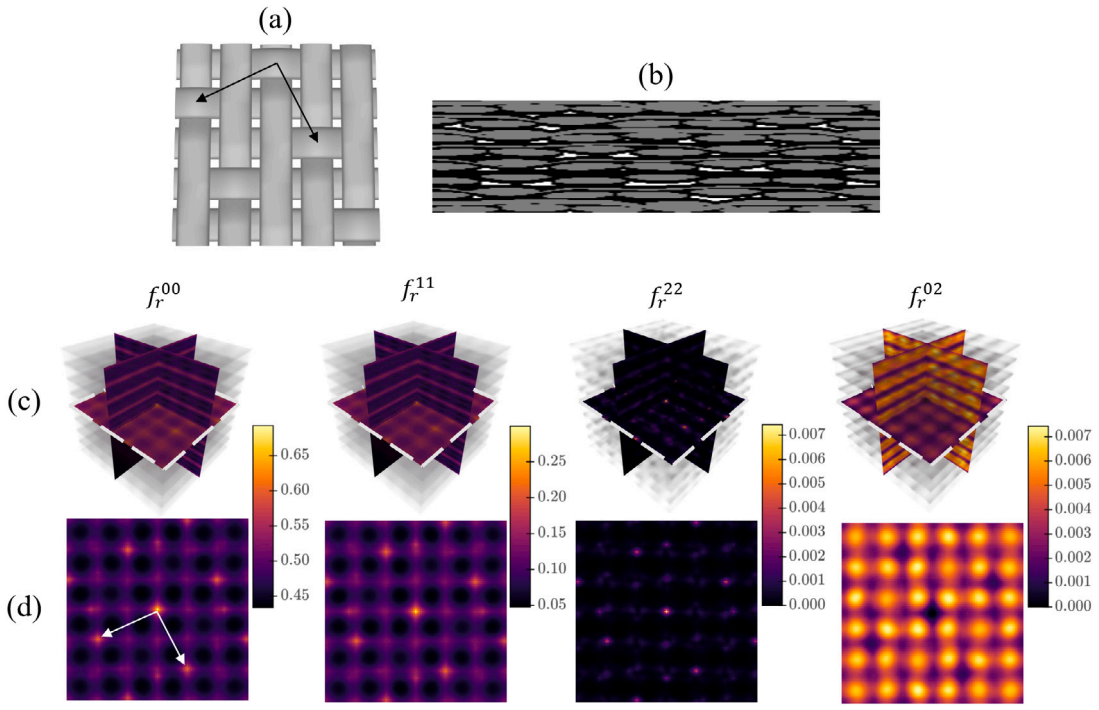

Fig. 3. Exemplar microstructure (corresponding to the first microstructure test case) and its spatial correlation maps comprising $\{f_r^{00}, f_r^{11}, f_r^{22}, f_r^{02}\}$. (a) 5HS weave displaying relationship between cross-over points, (b) $X-Z$ cross-section of the exemplar microstructure, with tows depicted in gray, matrix in black, and porosity in white. (c) Set of 2-point spatial correlations associated with the microstructure in (b) with the central planes of the 3D field highlighted for clarity. (d) $X-Y$ central plane from (c).

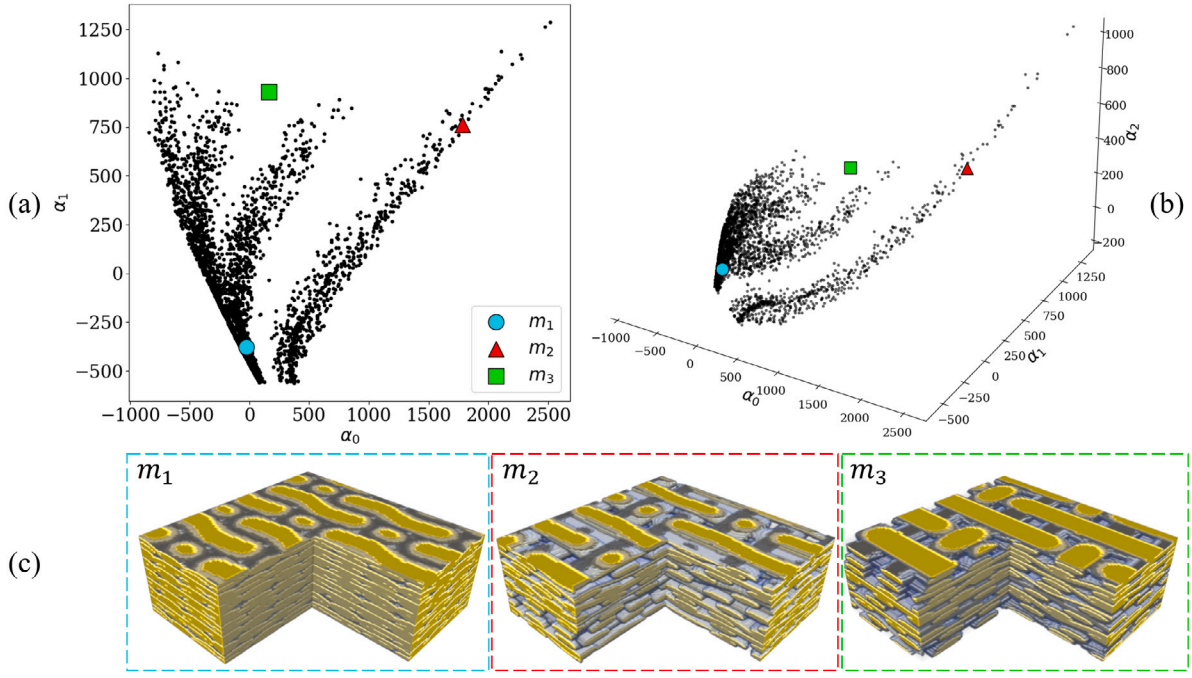

Fig. 4. PCA representation of the ensemble of microstructures along with cross-sections of the identified test-case microstructures. (a) Projected view of the synthetic microstructure ensemble in the α_0 *vs.* α_1 PC subspace, (b) projected view of the synthetic microstructure ensemble in the PC subspace spanned by $\alpha_{0:2}$ and (c) cut-away views of the test-case microstructures, with yellow corresponding to the tows, gray to matrix, and porosity left transparent.

posterior distribution in all dimensions – including those not displayed. Excitingly, the posterior is able to recover more than just the test microstructure. The identified posterior is relatively wide (as expected). Given the diversity of microstructures near the selected m_1 in both the PC and property spaces of the training set, we would expect to find a wide range of spatial correlations satisfying such a target property set. The framework autonomously identifies this degeneracy and quantifies it neatly in the returned posterior. This is a primary benefit of posing the problem of inverse microstructure design as a stochastic inverse problem. By utilizing probabilistic tools, the framework can estimate uncertainty through the inversion and characterize the variability of potential microstructures matching a design target.

We next visualize posterior samples in the primary PC space (i.e., samples from $p(\boldsymbol{\alpha}|\boldsymbol{k}_1^*)$ obtained through use of the VAE decoder $p_\psi(\boldsymbol{\alpha}|$

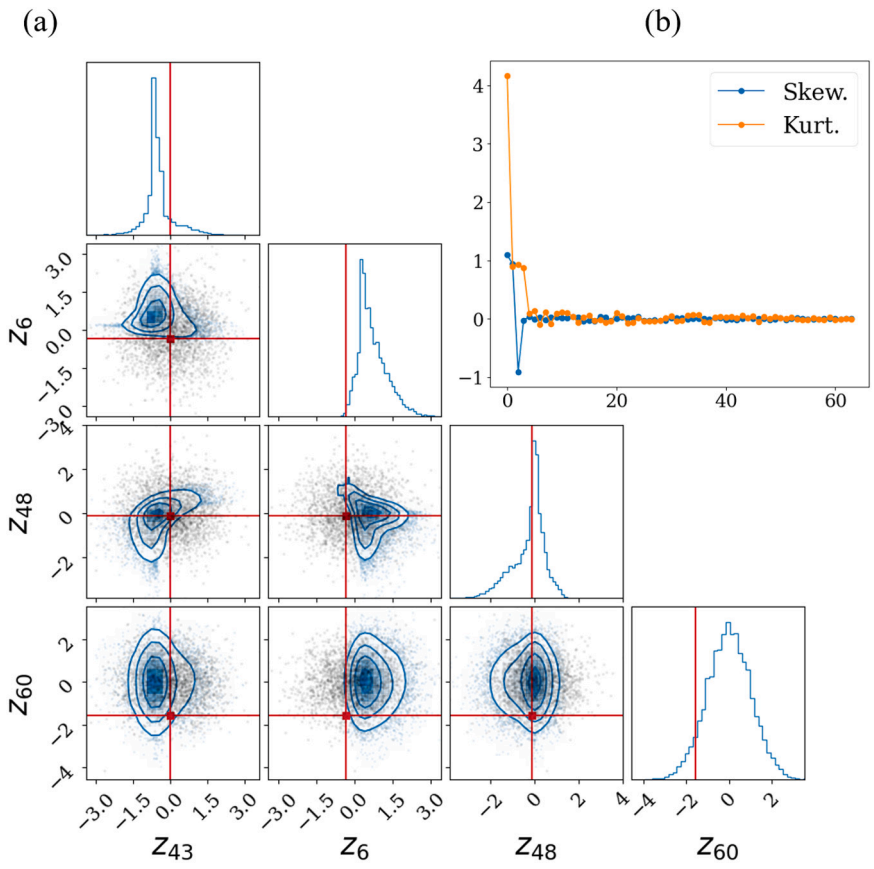

Fig. 5. (a) Posterior distribution $p(\boldsymbol{z}|\boldsymbol{k}_1^*)$ in the latent space of the β-VAE for the dimensions with the highest absolute skewness values in the marginalized distributions. The posterior is identified in blue marginalized histograms on the main diagonal and 2D projections on the lower triangle, with the MAP instantiated microstructure position demarcated in red, and the initial ensemble in black. Contours denote the 0.5σ, 1.0σ, 1.5σ, and 2.0σ confidence intervals. (b) Ranked decaying absolute skewness values across the 64 dimensions of $\boldsymbol{z}$ alongside corresponding Fisher kurtosis values [106].

$\boldsymbol{z}$)), shown in Fig. 6. We reemphasize that this is our primary space of interest because of the poor topology of the β-VAE latent space [107–109]. Contrasting the distributions in the two latent spaces, we see several important benefits of the bi-level inference framework. The non-linear decoding transformation warps the originally-unimodal posterior $p(\boldsymbol{z}|\boldsymbol{k}_1^*)$ into a complex, multimodal, disconnected distribution in PC space. Additionally, performing inference in the VAE latent space ameliorates several known limitations of the stochastic variational inference methodology, namely, zero-forcing effects of the KL-divergence [99,110], and spurious connectivity in the target density of normalizing flows [44,98,111]. The compacted nature of the space over which inference is being performed significantly reduces the chances of an exact target density displaying discontinuities or multimodal behavior.

These disconnected components of the posterior in PC space can be directly observed in the marginal distributions in Fig. 6, especially for α_3. Importantly, in this space, the highest-probability regions of the identified posterior *surround* the original microstructure m_1, with increasing variability in higher-order PC terms. This is expected due to the intimate link between the highest PC scores and resulting effective property [25,26,65,68–70,103–105,112]. Moreover, we observe that the posterior seems to bridge cavities in the design space, most clearly seen in the projected view $\{\alpha_0, \alpha_1\}$. Sets of 2-point spatial correlations falling inside these design cavities demonstrates the framework's ability to extend beyond the training set and highlight novel microstructures with similar properties to m_1.

Next, we interrogate how well the posterior predictions match the desired target property set. To answer this we push posterior samples $\boldsymbol{\alpha}^{(i)} \sim p(\boldsymbol{\alpha}|\boldsymbol{k}_1^*)$ through the surrogate forward model and compare the (average) predicted property of each generated sample with the target property $\boldsymbol{k}_1^*$.

For comparison, we also consider microstructures from the training dataset which lie within the convex hull (along the first 6 PC dimensions) of $\boldsymbol{\alpha}^{(i)}$ generated from the posterior. A total of 77 structures were identified meeting this criterion. An enlarged view of the posterior distribution in the first 3 dimensions can be seen in Fig. 7a, with the structures for which FE results are available highlighted in black. Interestingly, the inverse solution posterior was able to identify structures both from the region of the instantiating microstructure, as well as potential structures bridging the design cavity, and highlighting additional structures with vastly different values of α_1. The push-forward of the posterior distribution through the forward model can be seen in Fig. 7b alongside the target set of orthotropic thermal conductivity. For each property, the distribution of mean predictions of the SV-MOGP encompasses the design target, and the distribution of FE-simulated results overlaps the entire posterior of the SV-MOGP. Select FE results can be seen to lie outside of the push-forward $\pm 2\sigma$ of the posterior for k_{33}, although we believe this is due to the limitations of this selection criterion. Due to computational overhead in identifying the convex hull [113] of the posterior samples across the high-dimensional space of $\boldsymbol{\alpha}$, it was only feasible to search in the first 6 PC-dimensions; therefore these property-outlier FE results could be caused by deviations in their higher-order PC scores. Additional evidence for this theory was noted during construction of the forward model, where improved performance was found with the inclusion of up to 16 PC scores.

Finally, we inspect the solution posterior in the ultimately-desired space of 2-point spatial correlations. To do so, samples of $p(\boldsymbol{\alpha}|\boldsymbol{k}_1^*)$ are simply mapped back to 2-point correlations using the PC basis. A number of visualizations for this distribution are presented in Fig. 8. One-dimensional slices (along the central X axis) of sampled correlations are shown in Fig. 8a, and means and variances of 2D slices

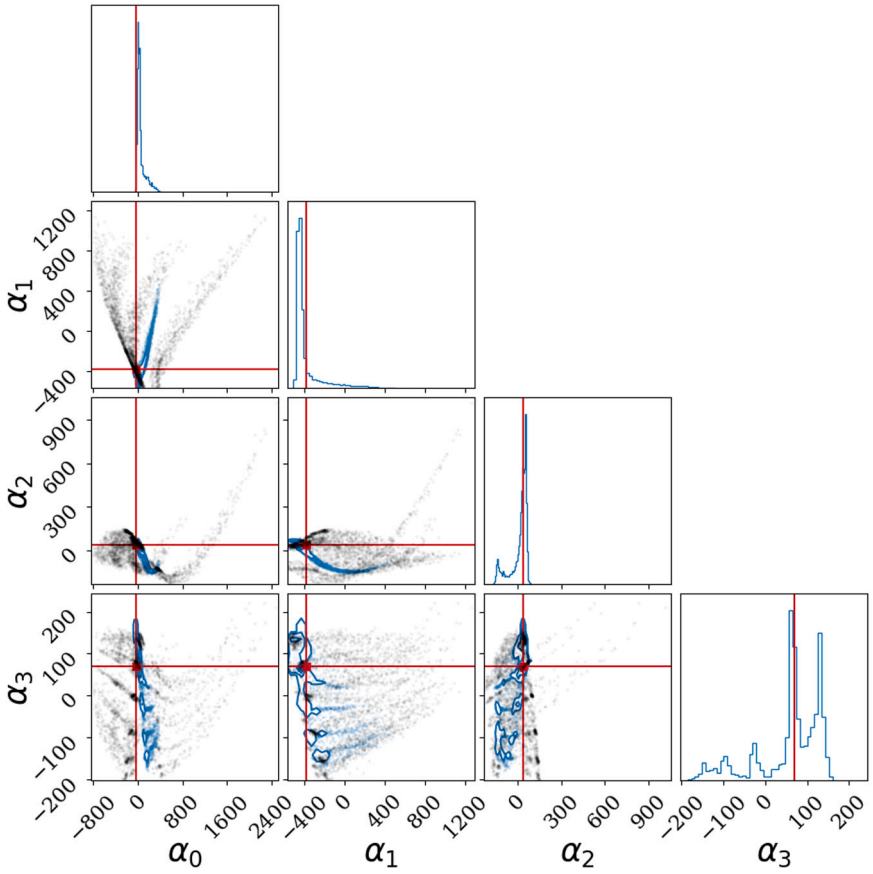

Fig. 6. Posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{k}_1^*)$ transformed through the decoder $p_\varphi(\boldsymbol{\alpha}|\boldsymbol{z})$. The posterior is identified in blue marginalized histograms on the main diagonal and 2D projections on the lower triangle, with the instantiated microstructure position demarcated in red, and the initial ensemble in black. Contours denote the 0.5σ, 1.0σ, 1.5σ, and 2.0σ confidence intervals.

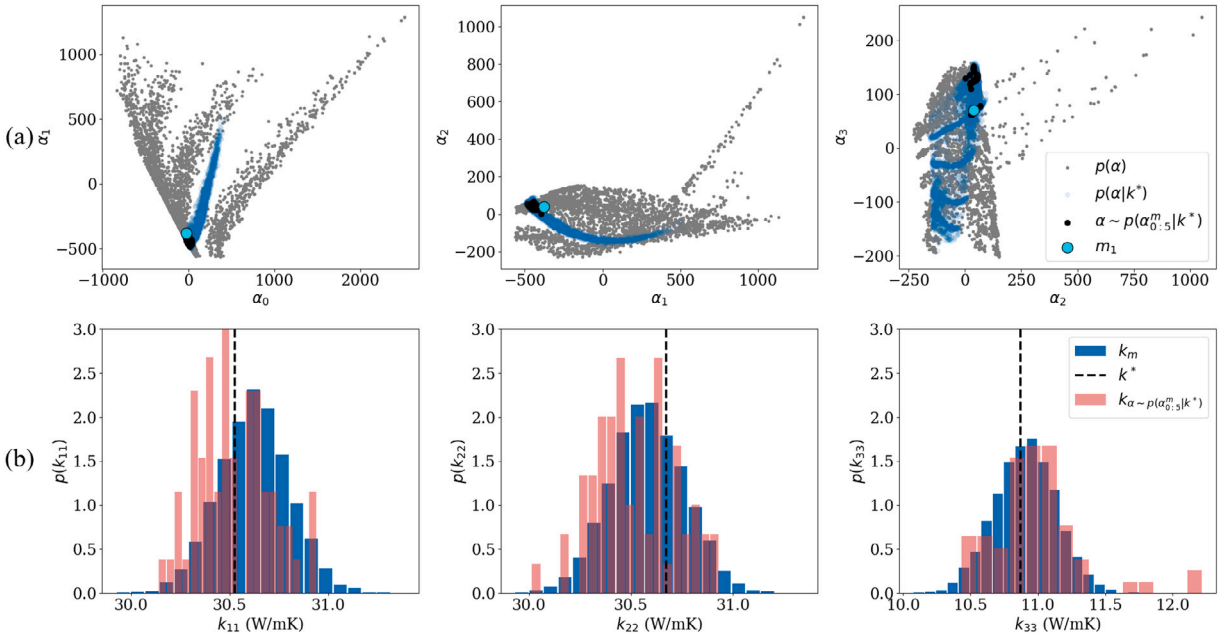

Fig. 7. Validation of the push-forward of the posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{k}_1^*)$ through the SV-MOGP. (a) Posterior distribution in blue overlaid upon the prior distribution of the microstructure ensemble, the instantiating microstructure highlighted in aqua, and sampled points $\boldsymbol{\alpha} \sim p(\boldsymbol{\alpha}_{0:5}^m|\boldsymbol{k}_1^*)$ from $p(\boldsymbol{\alpha})$ identified in black. (b) Histograms of 1000 posterior samples of the SV-MOGP prediction in blue, FE-simulated results in red, and the target $\boldsymbol{k}_1^*$ as a dashed vertical black line.

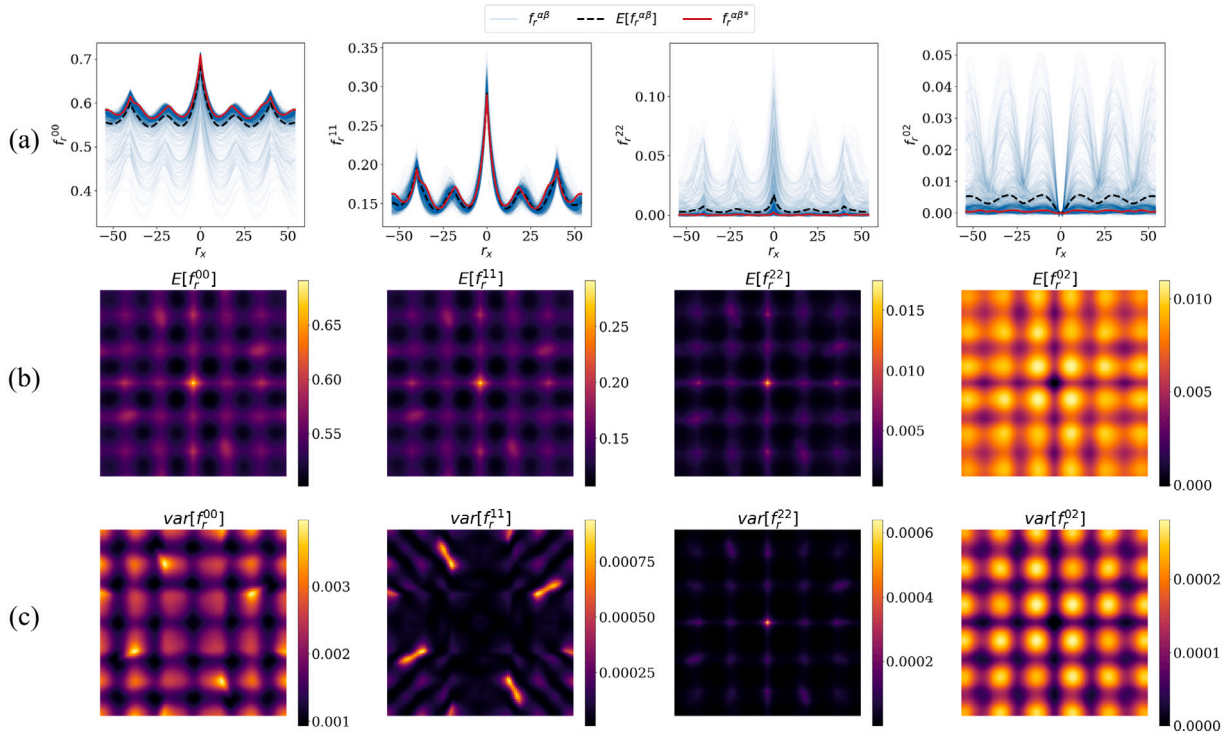

Fig. 8. Posterior over the set of 2-point spatial correlations $\{f_r^{00}, f_r^{11}, f_r^{22}, f_r^{02}\}$ computed using 1000 samples. (a) 1D cuts through the central plane $X-Y$ of the 2-point spatial correlations with samples displayed in blue, the dashed black line corresponds to the estimated mean, and the instantiating microstructure's 2-point spatial correlations are displayed in red. (b) The mean of posterior on central $X-Y$ plane, and (c) the variance of posterior on central $X-Y$ plane.

(along the central $X-Y$ plane) can be seen in Fig. 8b. In particular, 8a demonstrates that the posterior exhibits very little variation in the matrix autocorrelations, f_r^{11}, but is dispersed in the tow and porosity autocorrelations, f_r^{00} and f_r^{22}.

Fig. 8b,c provide a more information-dense view of the identified posterior. In particular, the mean 2-point spatial correlations in Fig. 8b display a blurring effect of the secondary and tertiary peaks, most prominently observed in comparison with those of a single training microstructure (Fig. 3d). This effect can also be seen to increase in intensity with $\|\boldsymbol{r}\|_2$. In other words, the posterior has identified a set of microstructures with *variable* tow spacing and tow width which still satisfy the target property set. The variance of this central plane across the set of correlations in Fig. 8c only reinforces this conclusion. The streaking exhibited in var$[f_r^{00}]$, var$[f_r^{11}]$, and var$[f_r^{22}]$ further highlights the variability in the position of 5HS cross-over locations as a result of shifting tow spacing and tow width. These are all useful observations for engineering activities such as quality control: large trade-off between the volume fraction of pores and tows, as well as perturbations in tow positioning or cross-section, do not significantly change the material performance for the targeted properties (under the uncertainty of the forward model). The interdependency between the tow and matrix 2-point correlations can be derived through the normalization constraints on the local material volume fractions (i.e., central peaks in 2-point auto-correlations).

The results of this first case study demonstrate a promising framework capable of identifying interdependencies between design parameters across exceedingly high-dimensional domains. When asked to identify solutions over the regularized space of structures to match a given target property, it was not only able to identify the region in space from which the instantiating microstructure was located, but simultaneously explored the design space and identified novel microstructures. Overall, this demonstrates the ability of the framework to explore an exceptionally high-dimensional design space and provide estimate uncertainty over this domain.

4.3. Case study 2: Microstructure identification in sparse regions

The aim of the second case study is analyze the proposed framework's performance on extremal property values. Beyond being a natural use case of this framework, this situation is important because we expect candidate microstructures to exist on sparse boundaries of the design space. Our test pair $\{m_2, \boldsymbol{k}_2^*\}$ is selected such that both property and microstructure reside in sparse regions at the extremities of their respective spaces.

Once again we fit a posterior distribution $p(\boldsymbol{z}|\boldsymbol{k}_2^*)$ and apply a similar validation process. This time, however, our analysis focuses solely on the $\boldsymbol{\alpha}$ design space. The posterior in this space is visualized across the first 4 dimensions of $\boldsymbol{\alpha}$ in Fig. 9. Compared to both the first case study and the entire microstructure ensemble, this posterior is significantly more compact. This is particularly evident in the $\{\alpha_0, \alpha_1\}$ projection. Excitingly, the benchmark microstructure m_2 (depicted in red) is neatly located near the center of the posterior. This provides an early indication that the framework successfully identified inverse solutions in this sparse region. We attribute this success to the design of the likelihood surrogate model, and more specifically its Spectral Mixture kernel. Even in sparse regions, the SV-MOGP provides stable and informative predictions. However, it also demonstrates the importance of the informative microstructure prior – provided by the β-VAE. The implicit consistency and normalization constraints in the decoder preclude arbitrary exploration of the design space and reduce the likelihood of generating artifacts or "hallucinations".

As before, a set of posterior samples were pushed forward through the SV-MOGP into property space and compared against the conditioning target $\boldsymbol{k}_2$ and nearby FE results. Due to the instantiating microstructure's extremal location within the initial dataset, only 10 training microstructures fell within the convex hull of the posterior. These highlighted points, alongside an enlarged view of the posterior in PC space, can be seen in Fig. 10a, with the corresponding push-forward in Fig. 10b. Excitingly, both the push-forward predictions

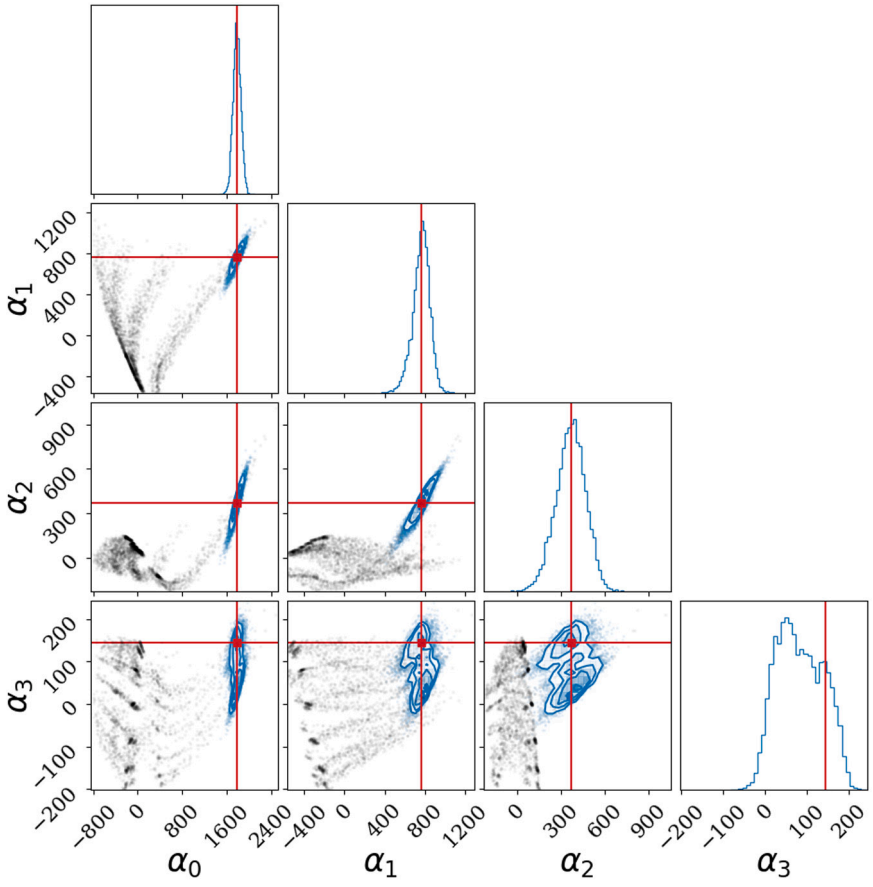

Fig. 9. Posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{k}_2^*)$ identified as transformed through the decoder $p_\varphi(\boldsymbol{\alpha}|\boldsymbol{z})$. The posterior is identified in blue marginalized histograms on the main diagonal and 2D projections on the lower triangle, with the instantiated microstructure position demarcated in red, and the initial ensemble in black. Contours denote the 0.5σ, 1.0σ, 1.5σ, and 2.0σ confidence intervals.

and the ground truth simulations are in excellent agreement, and the property distribution of FE results is similar to the uncertainty bounds of the posterior itself. In other words, the scatter in posterior samples appropriately matches uncertainty in the structure–property mapping itself.

Subsections from the posterior set of 2-point spatial correlations are depicted in Fig. 11, visualized as 1D slices of samples along the central X axis. For reference we also display the 2-point spatial correlations of the reference microstructure m_2. Here, the posterior's compactness is on clear display: there is much less variation in the identified 2-point spatial correlations than in the previous case study. For example, the variance in each correlation is approximately two orders of magnitude lower than the scale of the correlation themselves (i.e, the maximum variance is 0.0025 compared to a maximum expectation of 0.45). Clearly, the framework has determined that a tighter microstructural specificity is necessary to achieve this desired property (Fig. 11a versus Fig. 8a). This conclusion matches existing literature: extremal properties are achieved by highly optimized microstructures [39–41].

4.4. Case study 3: Microstructure extrapolation

Finally, the last case study analyzes the framework's capacity to interpolate *between* clusters in the dataset. As discussed, even when one is interested in studying a broad class of microstructures, such as woven composites, one is often limited to repeatedly generating only a few distinct subclasses of microstructures when building a dataset. This is particularly true for experimentally-curated datasets, but also arises in synthetically-generated datasets due to the limited expressability of parametric generating strategies [33,47,114]. As a result, discontinuities in the dataset's PC distribution (for example, between the tendril arms and within the large internal cavity, Fig. 4) are exceedingly common [70]. Because of the topological smoothness and convexity of 2-point spatial correlations and PC space, cavities in this design space within the dataset's convex hull are still expected to contain valid and interesting microstructures of the broader microstructure class. Therefore, the framework should be able to stably explore these regions.

To analyze this situation, we apply our methodology to an instantiating microstructure m_3 (and its properties $\boldsymbol{k}_3$) located directly in the center of a design cavity, as seen in Fig. 4. In this situation, the instantiating microstructure is located in an extremely under sampled region of PC space while its properties lie in a dense region of the property hull. By virtue of this discrepancy, we expect to find a variety of novel heterogeneous microstructures throughout the design space meeting this target property set.

Again, we apply the Stochastic Microstructure Design framework to approximate the density $p(\boldsymbol{z}|\boldsymbol{k}_3^*)$. The first 4 dimensions of the transformed version in PC-space ($p(\boldsymbol{\alpha}|\boldsymbol{k}_3^*)$) are shown in Fig. 12. As expected, the generated posterior is quite wide; like the previous cases, it covers portions of the training distribution. This reflects the fact that the property value is not uncommon in the training dataset itself. However, the posterior distribution also extends out into the low density cavities. This is especially visible in the $\{\alpha_0, \alpha_1\}$ projection, where it proposes a number of new 2-point spatial correlations. Interestingly, the posterior is able to rediscover the target microstructure m_3, but places it in low-density regions of the first 4 PC dimensions. This implies that either the higher PC dimensions play a strong role here, or our original microstructure m_3 was not a high-probability choice to produce properties $\boldsymbol{k}_3$ (*a posteriori*).

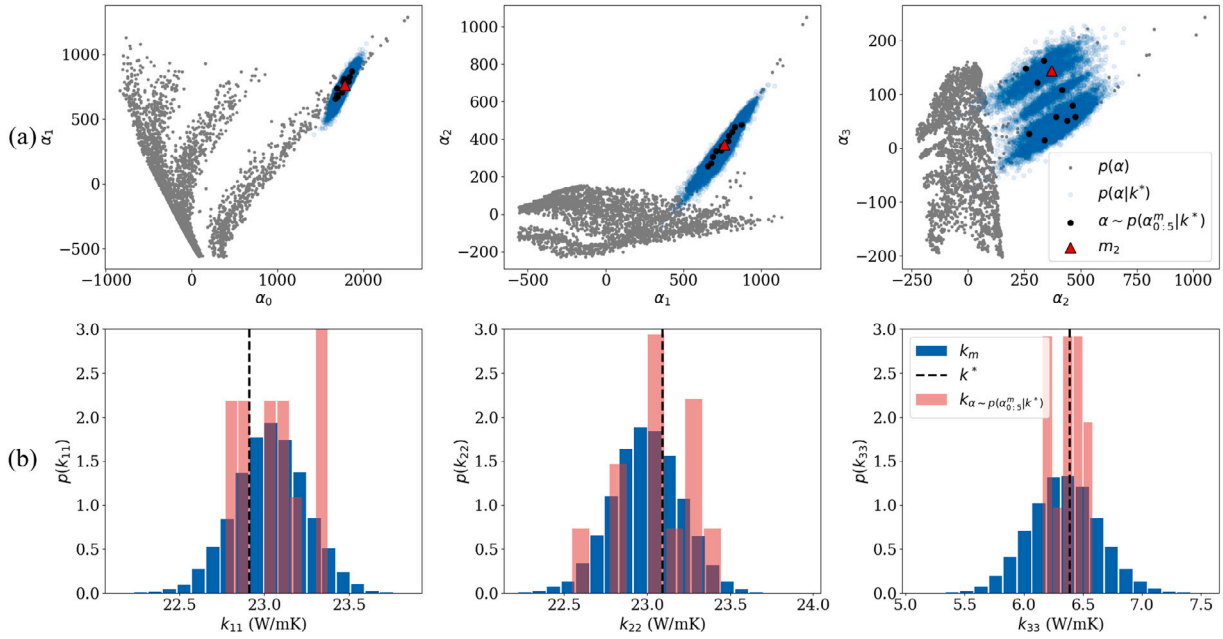

Fig. 10. Validation of the push-forward of the posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{k}_2^*)$ through the SV-MOGP. (a) Posterior distribution in blue overlaid upon the prior distribution of the microstructure ensemble, the instantiating microstructure highlighted in aqua, and sampled points $\boldsymbol{\alpha} \sim p(\boldsymbol{\alpha}^m_{0:5}|\boldsymbol{k}_2^*)$ from $p(\boldsymbol{\alpha})$ identified in black. (b) Histograms of 1000 posterior samples of the SV-MOGP prediction in blue, FE-simulated results in red, and the target $\boldsymbol{k}_2^*$ as a dashed vertical black line.

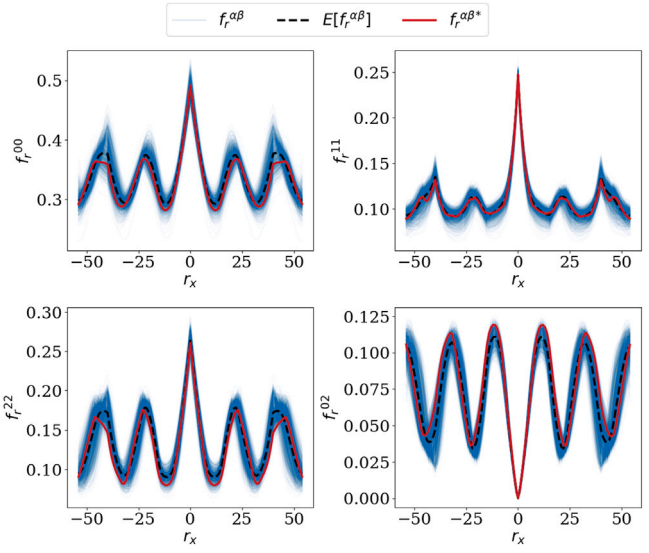

Fig. 11. Posterior over the set of 2-point spatial correlations $\{f_r^{00}, f_r^{11}, f_r^{22}, f_r^{02}\}$ computed using 1000 samples, presented along the central X axis. 2-point spatial correlations samples are displayed in blue, the dashed black line corresponds to the estimated mean, and the instantiating microstructure's 2-point spatial correlations are displayed in red.

This capacity to stably expand and explore new regions highlights the importance of the SM kernel for the SV-MOGP and demonstrates that the prior learns to include these internal regions. Bridging across these cavities requires that the forward model be capable of some degree of extrapolation, something not possible with many popular stationary kernel functions (e.g., Squared Exponential, Matérn [79]). The use of the SM kernel means that reasonable property predictions are obtained even for points from the latent space of the β-VAE which correspond to the inside of such cavities.

Importantly, the generated posterior, even in cavities of the initial dataset, is composed of *realistic* statistics displaying, to the best of our capability to validate, the desired property value. Fig. 13b displays the push-forward of the posterior through the forward model. The predicted distribution tightly encircles the desired target property value even though the posterior distribution is quite wide, Fig. 13a. As validation, we see that the few nearby training microstructures (black points in Fig. 13) display FE-predicted properties in alignment with the predicted push-forward distribution (red distribution in Fig. 13b).

Unfortunately, a more thorough validation process would require an advanced generative model for instantiating fiber composites from the 2-point spatial correlations, which is outside the scope of this paper. However, analyzing the generated two-point statistics lends a degree of confidence. Fig. 14 compares the posterior in the space of 2-point spatial correlations against the set of correlations of the instantiating microstructure. The instantiating microstructure falls both well within distribution and in close proximity to the posterior mean across all sets of 2-point spatial correlations, in agreement with the PC plots. Even though the posterior distribution primarily contains newly fabricated 2-point spatial correlations, their features are generally consistent with those seen in previous case studies. Specifically, the immediately characteristic periodic valleying of the statistics – a fingerprint of the 5HS weave architecture – is still present. This validates our original assumption: these low density regions are still representative of the general microstructure class. In a similar manner as in the first case study, Fig. 14a demonstrates that the target property set can be attained even with fluctuations between the spatial arrangement of tows and pores within the microstructure, although variability in the matrix spatial arrangement is more limited. Inspection of the differing variance between the four reconstructed 2-point spatial correlations clearly demonstrates that this property set is most sensitive to the amount and position of the deposited matrix material.

5. Conclusions

We present a framework for addressing inverse stochastic microstructure design problems. Our methodology presents several key strategies for overcoming persistent challenges in microstructure

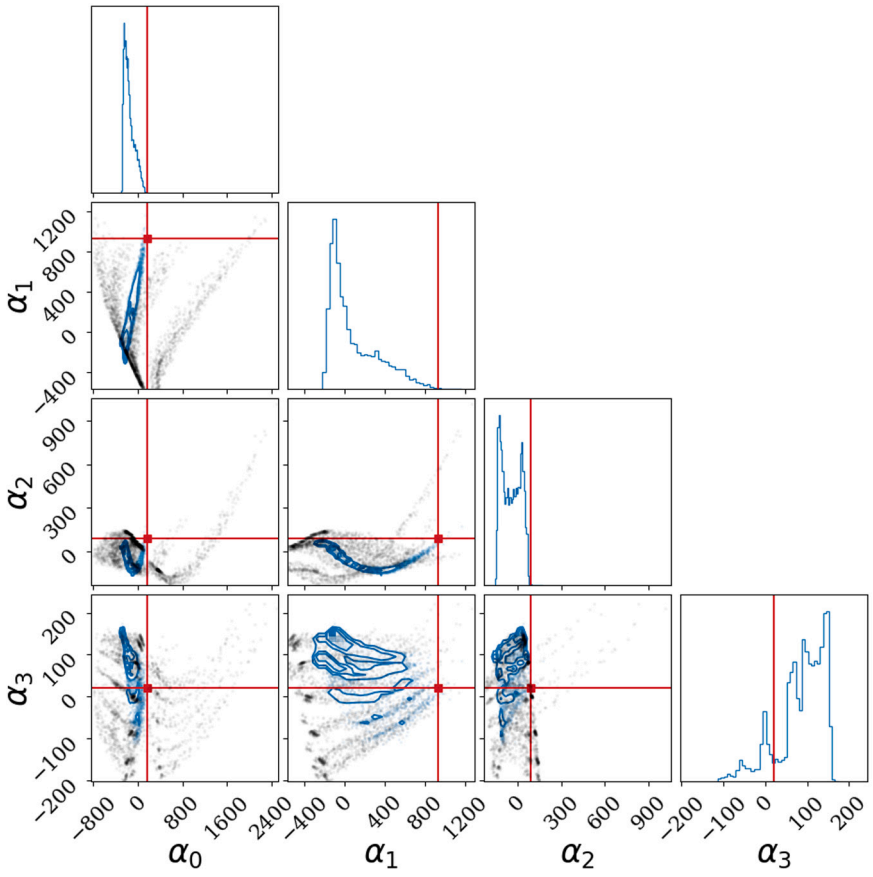

Fig. 12. Posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{k}_3^*)$ identified as transformed through the decoder $p_\varphi(\boldsymbol{\alpha}|\boldsymbol{z})$. The posterior is identified in blue marginalized histograms on the main diagonal and 2D projections on the lower triangle, with the instantiated microstructure position demarcated in red, and the initial ensemble in black. Contours denote the 0.5σ, 1.0σ, 1.5σ, and 2.0σ confidence intervals.

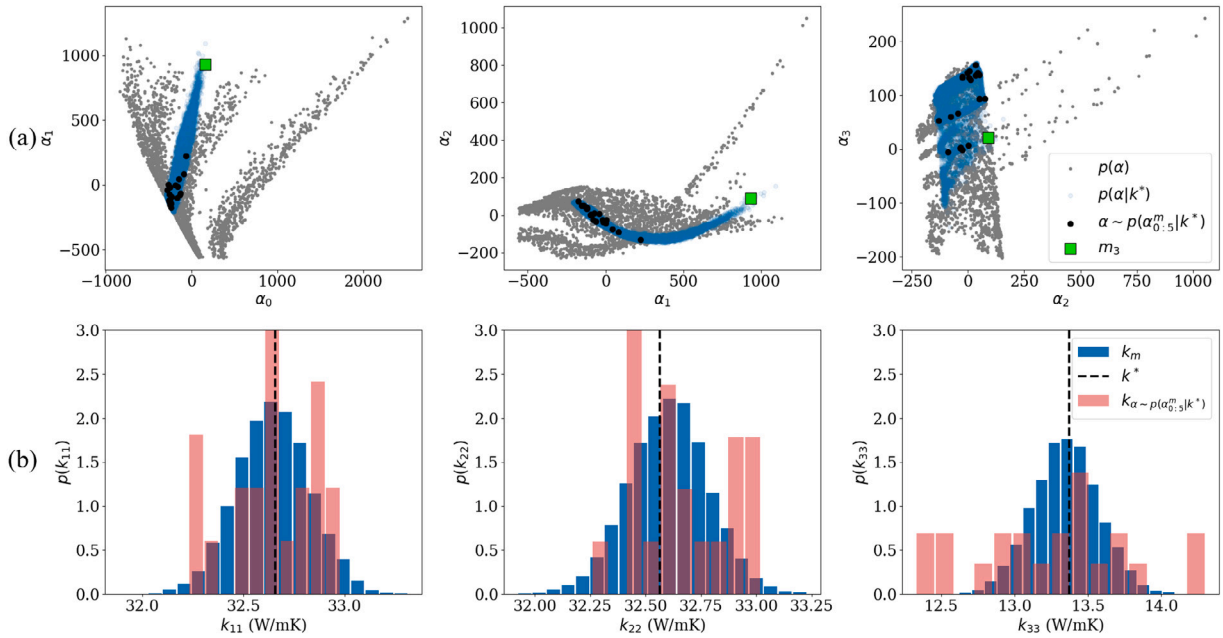

Fig. 13. Validation of the push-forward of the posterior distribution $p(\boldsymbol{\alpha}|\boldsymbol{k}_3^*)$ through the SV-MOGP. (a) Posterior distribution in blue overlaid upon the prior distribution of the microstructure ensemble, the instantiating microstructure highlighted in aqua, and sampled points $\boldsymbol{\alpha} \sim p(\boldsymbol{\alpha}^m_{0:5}|\boldsymbol{k}_3^*)$ from $p(\boldsymbol{\alpha})$ identified in black. (b) Histograms of 1000 posterior samples of the SV-MOGP prediction in blue, FE-simulated results in red, and the target $\boldsymbol{k}_3^*$ as a dashed vertical black line.

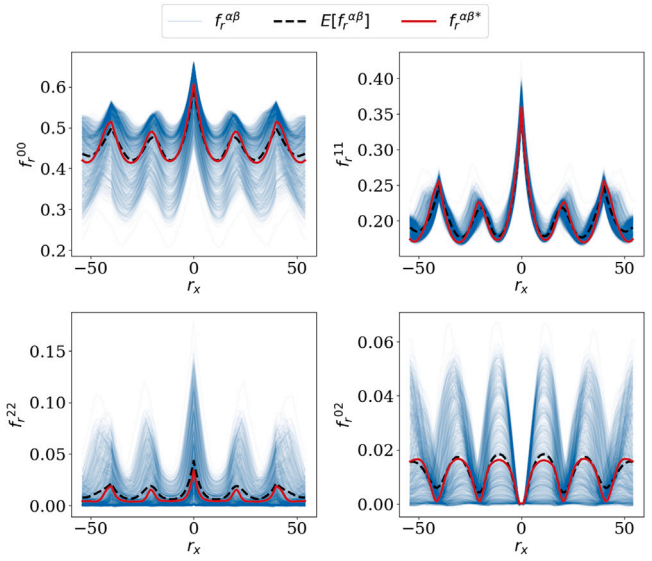

Fig. 14. Posterior over the set of 2-point spatial correlations $\{f_r^{00}, f_r^{11}, f_r^{22}, f_r^{02}\}$ computed using 1000 samples, presented along the central X axis. 2-point spatial correlations samples are displayed in blue, the dashed black line corresponds to the estimated mean, and the instantiating microstructure's 2-point spatial correlations are displayed in red.

design, including the high-dimensionality and irregular distribution of the microstructure space, the stochasticity of microstructures and the structure–property linkage, and the ill-posed nature of the inversion. The proposed framework is highly modular and reusable, readily able to incorporate already existing SP linkages for a wide variety of phenomena [25,26,65,68–70,115].

The first component is a generalized microstructure prior. This informative prior facilitates practical inversion of the problem by disentangling the microstructure solution space. Moreover, it is constructed in a property-independent manner and composed of nested pre-existing computational tools. This nesting defines a set of hierarchical latent spaces, each of which addresses a specific challenge in solving microstructure design problems. The first latent space is extracted using 2-point spatial correlations and, subsequently, PCA. It addresses the portion of the solution's degeneracy arising from the microstructure process's natural stochasticity [23,33,62]. Additionally, by leveraging physics-defined statistics, this latent space is naturally topologically well-structured, making it highly efficient for predicting arbitrary material properties [23,34,70]. The second latent space is extracted by further distilling PC scores using a β-VAE. Solution of the stochastic inverse problem is possible in this latent space because the β-VAE enforces a unit-Gaussian (prior) distribution. This compression and Gaussianization enables the seamless application of a prior during downstream Bayesian inference. Importantly, the entire prior is constructed without specializing to a specific material property. Therefore, it can be efficiently reused, without retraining, for microstructure inverse problems involving arbitrary new target properties. It only requires that the properties of interest be predictable using 2-point spatial correlations [23,24].

The second component of the framework is a property-driven likelihood. We surrogate the structure–property linkage using a Sparse Variational Multi-Output Gaussian Process with a Spectral Mixture kernel. The use of a SM kernel in our forward model makes for a stable surrogate model which can be robustly trained even in the presence of nonuniformly-distributed training datasets. Lastly, the two components are combined using Bayes' Rule, and sampled using a flow-based generative model to solve the inverse problem: to produce a distribution of microstructures matching a target property. In contrast to prior approaches, the proposed framework is interpretable, modular, and reusable.

As a whole, the presented framework permits the transformation of an initially intractable problem, namely, the probabilistic design of complex, heterogeneous materials, into one which is feasible. While we believe the presented framework represents a significant effort towards unlocking this problem, several continuing research areas remain open. One major limitation is the need to retrain the deep variational inference algorithm for every new value of the conditioning property. Although the process is relatively efficient, an amortized result would be ideal [102]. Additionally, better methods for validating the generated posterior distributions must be developed. Such developments would be especially valuable for guiding model selection between the large library of available deep distributional learning models.

Declaration of competing interest

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

Acknowledgments

This research acknowledges funding by various sources. Adam P. Generale: Pratt & Whitney; Andreas E. Robertson: ONR N00014-18-1-2879; Conlain Kelly: NSF 2027105, NSF Graduate Research Fellowship DGE-1650044, and ONR N00014-18-1-2879. Surya R. Kalidindi acknowledges funding from all previously listed funding sources. This work utilized computing resources and services provided by the Partnership for an Advanced Computing Environment (PACE) and the Hive computing cluster at the Georgia Institute of Technology, Atlanta, Georgia, USA. A.E. Robertson also acknowledges the continuing support of the Jack Kent Cooke Foundation, and A.P. Generale that of the Alfred P. Sloan Foundation.

Appendix A. Forward model

The SV-MOGP Structure–Property surrogate forward model was implemented with the package *GPyTorch* [116], and constructed with $L = 2$ latent processes, $M = 50$ inducing points, and $Q = 4$ Gaussian mixtures in the SM kernel function. It was trained for 20,000 epochs with the Adam optimizer [117] using an initial learning rate of 1e–2 and cosine annealing of the learning rate with warm restarts every 10,000 epochs. We used a minibatch size of 1024. The first 16 PC scores, $\boldsymbol{\alpha} \in \mathbb{R}^{16}$, were used as inputs. We observed negligible improvements in model performance after the first 16 PC scores. We note that the choice of which PC scores to use is independent of the size of the PC space because the PC basis vectors are importance ordered with respect to variation in the space. Error metrics for the model performance were evaluated through an 80/20 train/test split, with the resulting values displayed in Table 1. The consistency between the two datasets indicates a successful, robust training procedure which did not overfit. Fig. A.1 displays parity plots of the model outputs on both datasets. Again, we see strong agreement between the train and test performance. Additionally, the fact that error is uniformly distributed across the outputs – instead of being localized – further indicates successful training. Consistent with the reported error, improved model performance can be observed in the predictions of k_{11} and k_{22}, representing in-plane values for effective thermal conductivity, as compared to k_{33}.

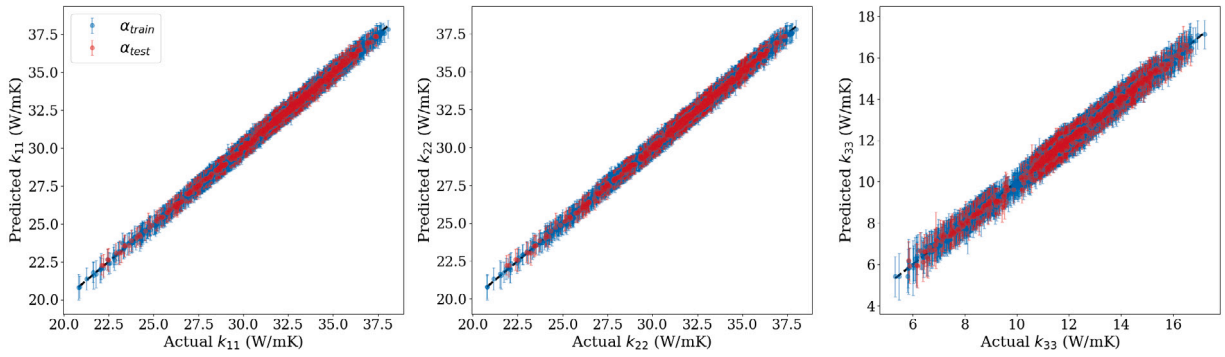

Fig. A.1. Parity plot of SV-MOGP forward model performance on the available microstructure ensemble. Predicted $\pm 2.0\sigma$ confidence intervals are displayed alongside mean predictions.

Table 1
Summary mean and standard deviation of error metrics of SV-MOGP forward model performance.

	MAE (W/m K)		NMAE (%)	
	Train	Test	Train	Test
k_{11}	0.107 ± 0.085	0.111 ± 0.088	0.338 ± 0.267	0.349 ± 0.276
k_{22}	0.109 ± 0.084	0.112 ± 0.085	0.344 ± 0.264	0.352 ± 0.269
k_{33}	0.106 ± 0.089	0.113 ± 0.098	0.898 ± 0.756	0.955 ± 0.825

Appendix B. β-VAE

Specifics regarding β-VAE model training and the selection of hyperparameters and architecture can be found in this section. We performed a set of experiments to tune the architecture size and the β hyperparameter.

As discussed in Section 3.3, the β-VAE architecture's β parameter is fundamental to our intended usage; it controls the importance of the reconstruction error against the KL-divergence on the latent space. As we will discuss momentarily, in this work, we place greater emphasis upon the KL-divergence as long as reconstruction errors were reasonable. This promotes a unit-variance Gaussian, $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, latent space which we use as a prior in downstream inference. As such, the β-VAE plays the dual role of both providing additional nonlinear compression to our PC representation, but more importantly, enables the specification of this analytic prior over microstructure space. In balancing these competing objectives, we aim to select the β-VAE to balance these roles, identifying a model which generalizes, but enables this prior specification with minimal increase in model complexity.

The individual dimensions of the input feature vectors (i.e. the PC scores, α_i) were not rescaled even though they often exist on varying length scales. We avoided rescaling in order to preserve their ranking importance in the β-VAE training, resulting in the reconstruction error term more heavily rewarding positive performance on the more important (i.e., larger) PC scores. All models considered during this experimentation process were similarly trained for 600,000 epochs with the Adam optimizer [117], an initial learning rate of 1e–3, and cosine annealing of the learning rate with warm restarts every 50,000 epochs. Cyclical annealing of the hyperparameter β [107] was also applied with warm restarts every 100,000 epochs in order to further aid the construction of an informative β-VAE latent space. In the training of each model, the available dataset was partitioned into a 80/20 train/test split.

8 model architectures were considered across a search domain with increasing complexity of the encoder and decoder. Each architecture is adapted from the architecture displayed in Fig. C.1 by modulating the size of the dense blocks. We found that a final latent space size of 64 was the minimum necessary for stable compression, independent of the architecture size. Fig. B.1a,b,c summarize the results of the architecture study. Two opposite trends are observable: the D_{KL} decreases (i.e., adherence of the latent space to the unit-Gaussian improves) with increasing parameter count, while testing error increases with parameter count. We adopted the model architecture with 622,208 parameters as it provides the best combination of reconstruction error on the test set and minimization of the KL-divergence.

Once a particular architecture was selected, the hyperparameter β was swept from 1–200. The results of which can be seen in Fig. B.1d,e,f. $\beta = 100$ in particular resulted in an 84.5% reduction in the KL-divergence while incurring only a small increase in the testing error: a test MAE comparable with the ~1 million parameter count model architecture. The final architecture, Fig. C.1, is composed of three dense blocks: 1024 to 256 to 128 to 64, and uses a leakyReLU activation. The final error metrics for this selected model architecture and β hyperparameter combination were a MAE of 0.615 and 0.844, and a NMAE of 2.194% and 9.017%, for train and test sets, respectively.

Appendix C. DVI

Each individual component of the proposed modular framework as well as a user identified target material property value is combined together using Bayes' Rule to estimate a target property conditioned posterior density over the space of microstructures, $p(\boldsymbol{\alpha}|\boldsymbol{k}^*)$. This posterior density is sampled using a third deep learning model: a normalizing flow-based generative model. The three components of the proposed framework and their interplay can be visualized in Fig. C.1. This Appendix outlines this combined process and the training of this final model. Altogether, we refer to this process as Deep Variational Inference.

The framework for inferring this density requires the utilization of 3 different latent spaces associated with various models (PCA, β-VAE, normalizing flow). Let $\boldsymbol{\alpha}$ represent the PC latent space, $\boldsymbol{z}$ the latent space of the β-VAE, $\boldsymbol{u}$ the latent space of the normalizing flow neural network. Additionally, let $\boldsymbol{k}^*$ be the target orthotropic thermal conductivity, $\bar{\boldsymbol{k}}$ the SV-MOGP mean prediction, and var$[\bar{\boldsymbol{k}}]$ the variance in the SV-MOGP prediction.

We approximate the target-property conditioned microstructure posterior through a flow-based generative model. Normalizing flow based deep generative models are a class of generative model widely used to approximate complex density functions. Such models build flexible variational distributions, $q_\theta(\cdot)$, using invertible neural networks (INNs). The neural network provides a bijective transformation map from a well understood latent random variable $\boldsymbol{u}$ to a random variable we are interested in approximating, in this application the target property conditioned microstructure $\boldsymbol{z}$, as

$$\boldsymbol{z} = f_\theta(\boldsymbol{u}), \ \boldsymbol{u} = f_\theta^{-1}(\boldsymbol{z}) \tag{C.1}$$

where $f_\theta(\cdot)$ represents the invertible deep neural network parameterized by θ. Importantly, this invertible transformations enables efficient

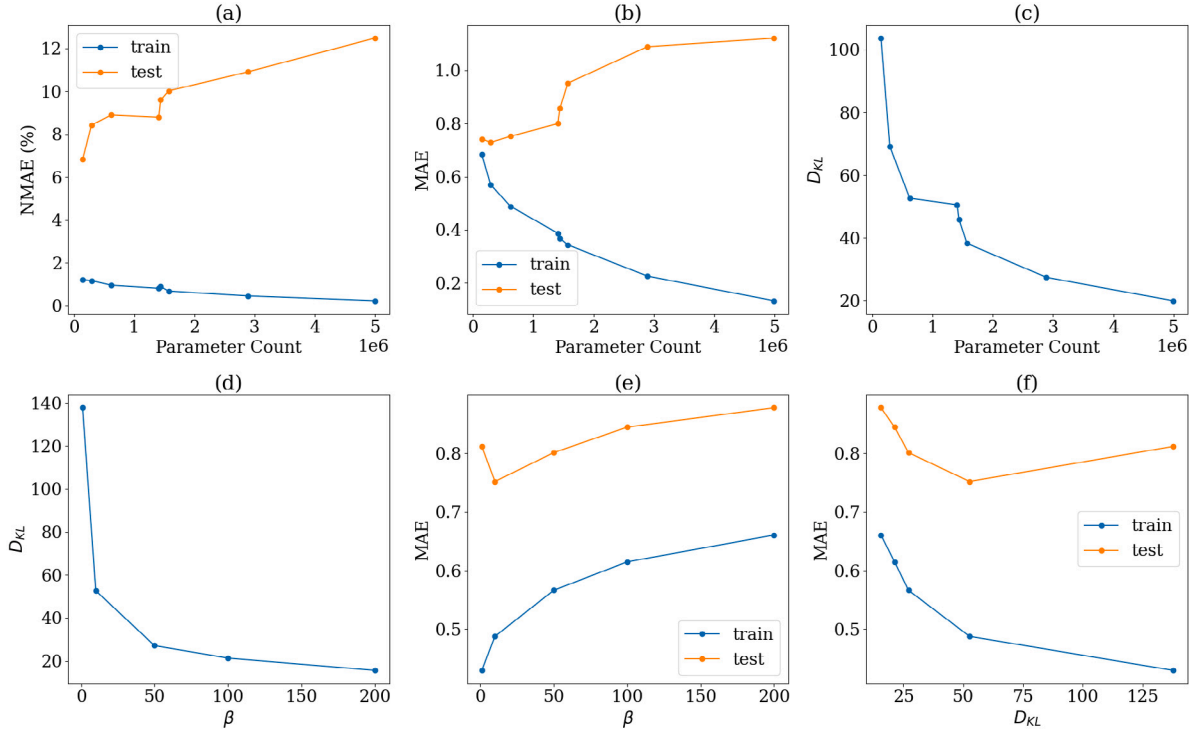

Fig. B.1. Criterion for evaluation of β-VAE model architecture involving the evaluation of reconstruction error metrics and KL-divergences over the latent space.

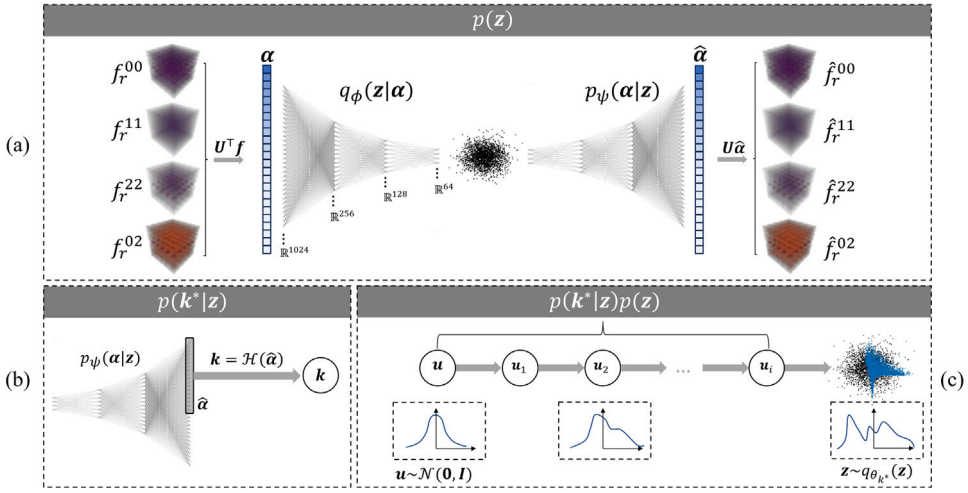

Fig. C.1. Depiction detailing the main computational components required by the stochastic inverse microstructure design framework. (a) Depiction of PCA performed on the set of 2-point spatial correlations, and the resulting PC scores taken as inputs to the generative β-VAE model to produce the microstructure prior, (b) The likelihood included in the Bayesian inference methodology, which incorporates the pretrained decoder, $p_\psi(\boldsymbol{\alpha}|\boldsymbol{z})$ and forward model, $\mathcal{H}$. The forward model $\mathcal{H}$ only takes in a subset of the decoded $\hat{\boldsymbol{\alpha}}$ vector in predictions of $\boldsymbol{k}$. (c) The NF for performing inference, where its parameters are learned through the objective in Eq. (14), incorporating the likelihood $p(\boldsymbol{k}^*|\boldsymbol{z})$ and prior $p(\boldsymbol{z})$.

sampling, as well as an exact expression for the log-likelihood. The log-likelihood of an individual sample can then be computed based on the change of variable theorem as

$$\log q_\theta(\boldsymbol{z}) = \log p(\boldsymbol{u}) - \log \left| \det \frac{\partial f_\theta(\boldsymbol{u})}{\partial \boldsymbol{u}} \right| \tag{C.2}$$

where $\left|\det \frac{\partial f_\theta(\boldsymbol{u})}{\partial \boldsymbol{u}}\right|$ is the absolute value of the determinant of the Jacobian. The neural network is defined to represent an invertible transformation through restricting its parameterization to model forms such as NICE [100], Real-NVP [101], Glow [118], Flow++ [119], amongst many others. These architectures simplify the calculation of the Jacobian by producing matrices for which the determinant can be computed efficiently, such as Jacobian's with lower triangular structure.

As described in Section 3.4, the parameters of INN could then be learned by stochastically optimizing the expression in Eq. (14) to tune the variational distribution to the desired posterior. The forward output of the INN, $\boldsymbol{x}$ in Eq. (14), is the latent space of the β-VAE, $\boldsymbol{z}$. In this expression, the prior is taken to be the unit-Gaussian due to the structure of the β-VAE's latent space while the likelihood is derived from the MOGP forward model. Training of the normalizing flow was performed by sampling a minibatch of $N = 1024$ points from the latent space of the normalizing flow as $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and evaluating the forward mapping to $\boldsymbol{z}$. To evaluate the likelihood, the push-forward of these points into the β-VAE latent space are once more pushed through the decoder, to result in a 1024-dimensional vector $\boldsymbol{\alpha}$. From this reconstruction, the first 16 components are taken as an input to

the SV-MOGP such that the likelihood $p(\boldsymbol{k}^*|\boldsymbol{\alpha})$ could be evaluated as

$$p(\boldsymbol{k}^*|\boldsymbol{z}) = \frac{1}{(2\pi)^{D/2}\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}\left(\boldsymbol{k}^* - \bar{\boldsymbol{k}}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{k}^* - \bar{\boldsymbol{k}}\right)\right] \tag{C.3}$$

where $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{K})$ and $\boldsymbol{K}$ denotes the complete covariance matrix of the SV-MOGP, $\bar{\boldsymbol{k}}$ the mean prediction, and D the dimensionality of $\boldsymbol{k}$. As a result of the well ordered topological space of PC scores of 2-point spatial correlations, any components past this dimensionality of $\boldsymbol{\alpha}$ held little additional informative knowledge on the underlying microstructure property, as previously discussed in Appendix A.

The optimization problem specified in Eq. (14) was adjusted by including simulated annealing training to avoid the normalizing flow from prematurely stopping exploration in early training epochs, and converging to a poorly identified posterior [120]. The annealed objective function is then specified as

$$\theta^* \approx \underset{\theta}{\text{argmin}} \frac{1}{N}\sum_{k=1}^{N}\left[-\log p(\boldsymbol{k}^*|\boldsymbol{z}) - \lambda \log p(\boldsymbol{z}) + \tau_i \log\left|\det \frac{\partial f_\theta(\boldsymbol{u})}{\partial \boldsymbol{u}}\right|\right] \tag{C.4}$$

where i is the epoch number, and τ a variable which evolved according to $\tau_i = \max\{1, \tau_0 \exp(-\frac{i}{\tau})\}$, where τ is the decay period and τ_0 the initial annealing value. This additional weighting on the log-determinant term serves to encourage higher entropy in the initial stages of training, when values of the prior and likelihood are likely to dominate [98,120]. In this work the initial annealing value was $\tau_0 = 1000$, the decay period was $\tau = 1000$ and training was performed over 40,000 epochs. The generative model contained 48 RealNVP [101] affine coupling layers. The scale-shift function in each layer was modeled as a fully connected network with a width of 512. The full effect of the prior was maintained for all case studies by setting $\lambda = 1$.

In this rather traditional forward flow originating from a continuous unit-variance Gaussian, it should be highlighted that a core limitation lies in the fact that any learned transformation of this density will necessarily also be a continuous density. This leads to high probability regions in the transformed density that will be incapable of identifying disconnected modes without the presence of a bridge between such regions [98,99,121]. In this work, we anticipate that the effects of this deleterious phenomena will be mitigated through performing inference in the well structured latent space of the β-VAE, although we leave refinements to the divergence measure in Eq. (13) to directly address this issue to future work.

Appendix D. Latent spaces for SP linkages

Recent works investigating microstructure inverse design frameworks have often advocated for constructing forward SP linkages directly utilizing the latent space of deep generative models as the input [39,51–53,57,122]. However, we found that the construction of the SP linkage in this manner decreases performance as compared to the established MKS approach (i.e., using PC scores extracted from 2-point spatial correlations as the input). We hypothesize that the decrease in performance is due to the deep learning model's (i.e., β-VAE's) latent space having a suboptimal topology. Unlike the MKS latent space which is topologically well ordered with respect to general homogenized material properties due to its connection to statistical continuum mechanics [35,123], deep generative models have been architected to promote latent spaces with strong ordering with respect to human vision and perception due to their development in computer vision [94,95]. Therefore, we argue that utilizing their latent space increases the complexity of the learned forward model as compared to the MKS approach by requiring it to disentangle the complex nonlinear embedding of the deep generative model.

A comparison between these two approaches was performed using the SV-MOGP architecture detailed in Section 2.3. The MKS based model is the approach described in Appendix A; it takes the first 16 PC scores, $\boldsymbol{\alpha} \in \mathbb{R}^{16}$, as input. This is more than a sufficient number to optimally capture the structure–property linkage as is established in the original work [26]. We note that we can always pick the initial 16 PC scores independent of the size of the whole PC based latent space because the PC basis vectors are importance ordered. The β-VAE approach takes all 64 dimensions of the latent space as input. Because the expressiveness and ordering of the latent space is intimately connected to the size of the β-VAE, we repeat the experiments presented in Appendix B and compare against surrogates built with varying β-VAE parameterizations to ensure a fair comparison between the two methods.

Each SV-MOGP was constructed with $M = 50$ inducing points, and $Q = 4$ Gaussian mixtures in the SM kernel function. The MKS based model was trained for 20,000 epochs with the Adam optimizer [117]. The β-VAE models proved more unstable; they were trained for 40,000 epochs of the same optimizer. Each model's initial learning rate was set to 1e–2, subsequently the learning rate was scheduled using cosine annealing and warm restarts every 10,000 epochs, and a minibatch size of 1024.

Fig. D.1a,b summarizes the variation in the error metrics for the β-VAE based SV-MOGP model's performance on the test set with increasing β-VAE parameterization. Fig. D.1c summarizes the Evidence Lower BOund (ELBO)[6] value – the maximization target for the MOGP's training optimization – at the end of training. The MKS-based approach outperforms in both cases. The maximal ELBO attained for SV-MOGPs build upon the β-VAE latent space was found to be 1.02, while the MKS-based approach yielded a superior ELBO of 1.37, far outperforming any of the alternate forward models. Similarly, the minimum train MAE achieved by the $\beta - VAE$ based models were 0.275, 0.271, and 0.256 greater than the 0.107, 0.109, and 0.106 for the MKS-based model, for k_{11}, k_{22}, and k_{33} respectively. A similar disparity was observed in test error metrics of 0.285, 0.289, and 0.255 for the $\beta - VAE$ based model, and 0.111, 0.112, and 0.113 for the MKS-based model.

A parity plot comparison between the SV-MOGP model utilized in this work, and one which was built upon the latent space of a 5-million parameter count β-VAE can be seen in Fig. D.2. Upon visual inspection, the predictive model uncertainty in the β-VAE based model is significantly higher. This result is somewhat surprising at first since the SV-MOGP built upon the latent space of the β-VAE had access to a compressed latent representation of the equivalent of 1024 PC scores, in comparison to the model built solely off of the first 16. However, this access to a greater quantity of embedded microstructural information was ultimately inaccessible to the SV-MOGP due to the complexity of the nonlinear transformation, enabling the model built upon the well structured PC space to ultimately display better performance.

Appendix E. Forward model stability

The stability of the Sparse Variational Multi-Output Gaussian Process (SV-MOGP) is further discussed in this Appendix. To demonstrate the kernel's ability to learn underlying patterns in the dataset, we have devised two test cases; (A) one in which microstructures at the extremities of the dataset in the low-dimensional PC space are excluded from training, and (B) one in which microstructures nearest the origin of this space are excluded. The models performance on the first excluded region will demonstrate its capacity to stably extrapolate while its performance on the second will demonstrate its stability over irregularities or nonuniformities in the dataset. The dataset was partitioned in the two cases according to the following expressions.

$$\begin{aligned}\{\boldsymbol{\alpha}_A\} &= \{\boldsymbol{\alpha} \in \boldsymbol{D} : \|\boldsymbol{\alpha}\|_2 > 0.5(\|\boldsymbol{\alpha}_{max}\|_2 - \|\boldsymbol{\alpha}_{min}\|_2) + \|\boldsymbol{\alpha}_{min}\|_2\} \\ \{\boldsymbol{\alpha}_B\} &= \{\boldsymbol{\alpha} \in \boldsymbol{D} : \|\boldsymbol{\alpha}\|_2 < 0.1(\|\boldsymbol{\alpha}_{max}\|_2 - \|\boldsymbol{\alpha}_{min}\|_2) + \|\boldsymbol{\alpha}_{min}\|_2\}\end{aligned} \tag{E.1}$$

$\boldsymbol{\alpha}_{max}$ is a vector whose elements are the maximum PC value over the entire dataset for the corresponding dimension, while $\boldsymbol{\alpha}_{min}$ is the equivalent vector but for minimum values. $\boldsymbol{D}$ is the complete 5HS

[6] Also referred to as the variational free energy (VFE) [82].

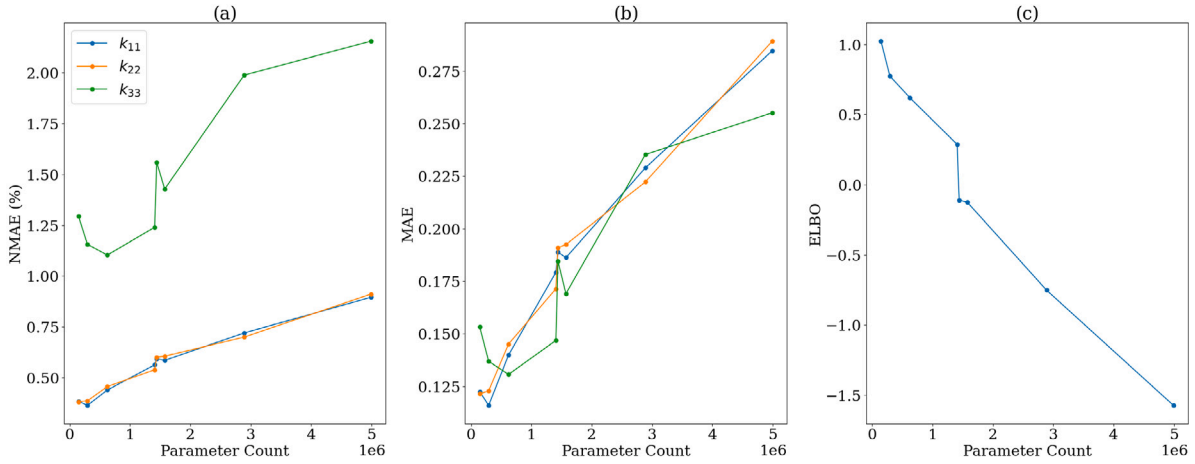

Fig. D.1. Comparison of SV-MOGP forward model performance evaluated against increasingly complex parameterization of $q_\phi(\boldsymbol{z}|\boldsymbol{\alpha})$ and $p_\theta(\boldsymbol{\alpha}|\boldsymbol{z})$. (a) Test set MAE of orthotropic thermal conductivity against β-VAE parameterization, (b) test set NMAE of orthotropic thermal conductivity against β-VAE parameterization, and (c) Evidence Lower BOund (ELBO) at the end of training with increasing β-VAE parameterization.

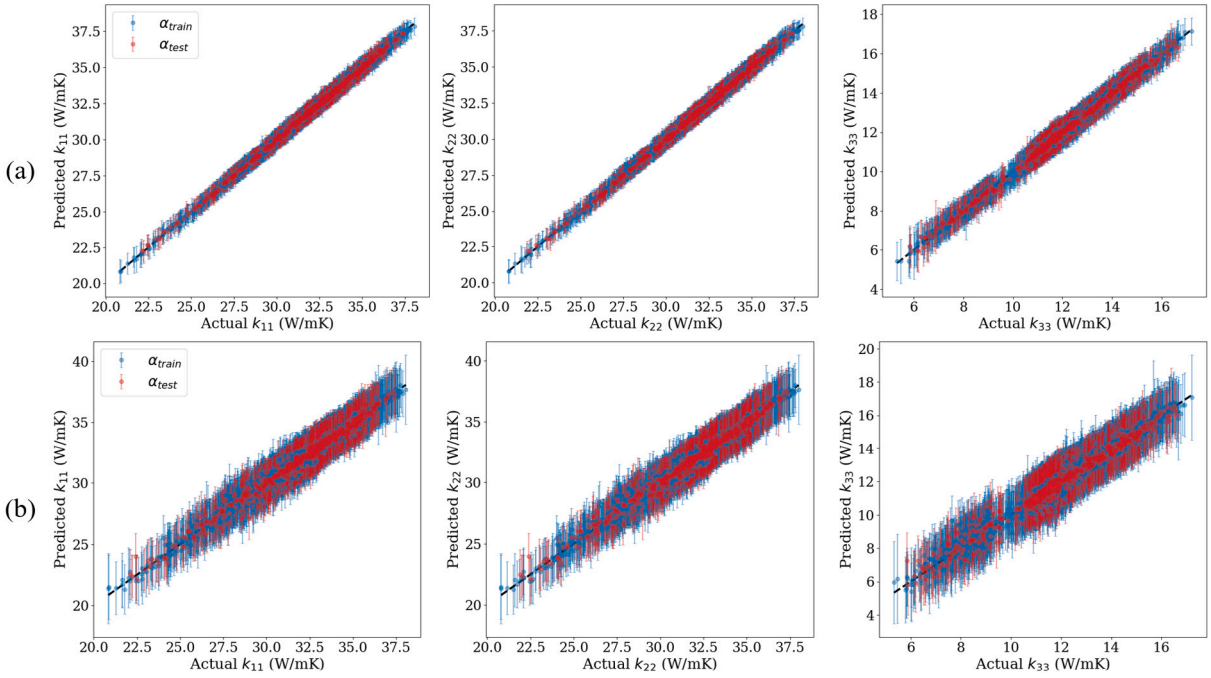

Fig. D.2. Parity plots of SV-MOGP forward models build upon (a) the PC latent space $\boldsymbol{\alpha}$ and (b) the β-VAE latent space $\boldsymbol{z}$. Predicted $\pm 2.0\sigma$ confidence intervals are displayed alongside mean predictions.

Table 2
Error metrics of SV-MOGP model performance in extrapolation test cases.

		(A)			(B)		
		k_{11}	k_{22}	k_{33}	k_{11}	k_{22}	k_{33}
NMAE (%)	Train	0.372 ± 0.285	0.373 ± 0.289	1.195 ± 0.973	0.382 ± 0.300	0.372 ± 0.289	1.133 ± 0.969
	Test	0.359 ± 0.287	0.379 ± 0.301	1.220 ± 0.987	0.378 ± 0.305	0.358 ± 0.294	1.140 ± 0.963
	Ext.	0.930 ± 0.797	0.879 ± 0.741	5.418 ± 5.522	0.419 ± 0.302	0.414 ± 0.299	1.121 ± 0.934
MAE	Train	0.119 ± 0.091	0.119 ± 0.092	0.143 ± 0.116	0.119 ± 0.093	0.116 ± 0.089	0.131 ± 0.112
	Test	0.114 ± 0.091	0.121 ± 0.096	0.145 ± 0.116	0.118 ± 0.095	0.111 ± 0.092	0.131 ± 0.111
	Ext.	0.216 ± 0.186	0.204 ± 0.172	0.356 ± 0.363	0.137 ± 0.098	0.134 ± 0.097	0.136 ± 0.114

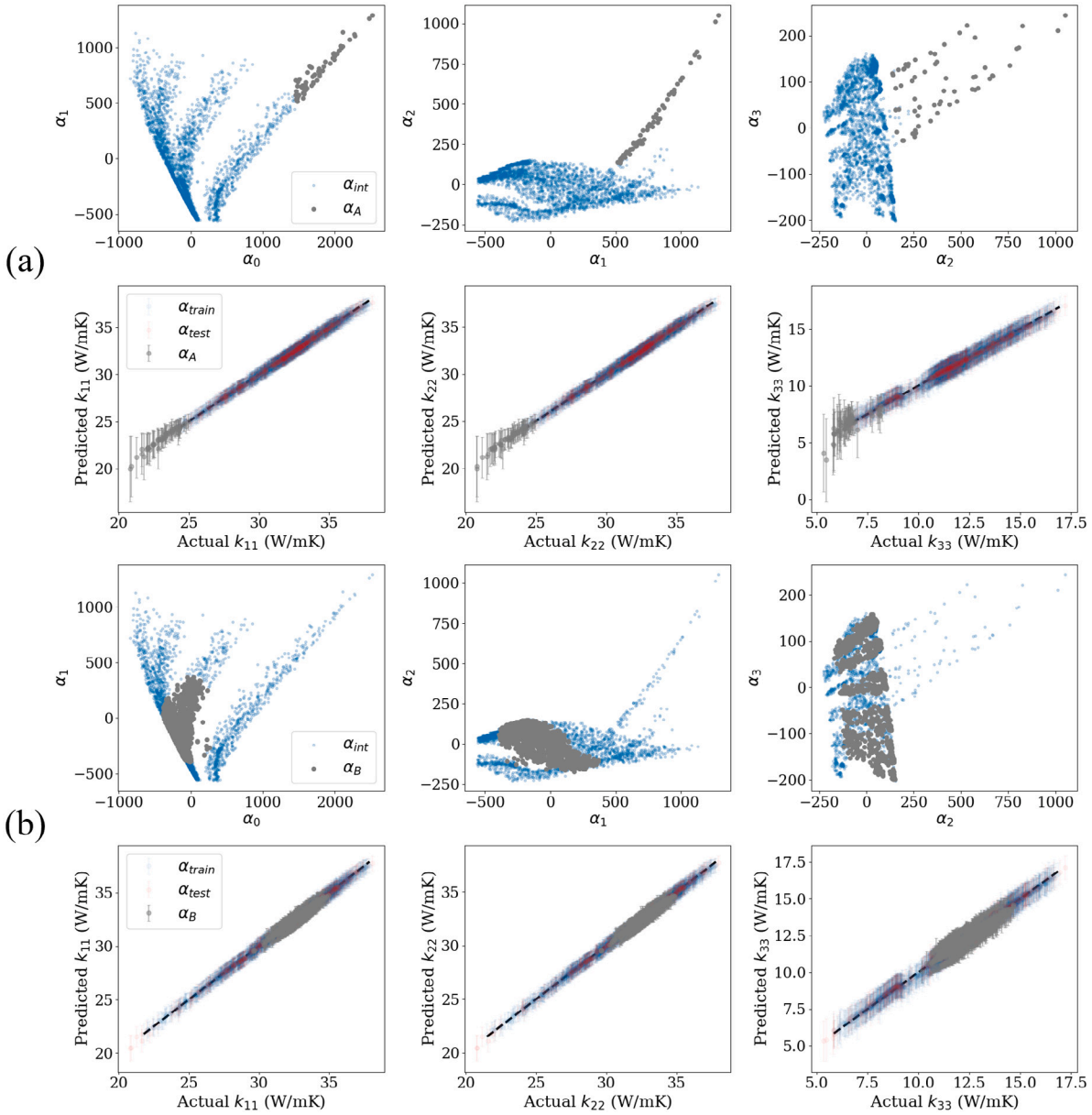

Fig. E.1. Case studies demonstrating extrapolatory ability of SM kernel in SV-MOGP. (a) Projections of first components of $\boldsymbol{\alpha}$ showing external grouping, and parity plots of model performance for the first case, (b) projections of first components of $\boldsymbol{\alpha}$ showing external grouping, and parity plots of model performance for the second case.

CMC dataset. In the first case, 101 microstructures were identified by splitting the dataset in half, and in the second case, 1295 microstructures were identified in the dense region closest to the origin. The complement in each case (i.e., the internal region, $\{\boldsymbol{\alpha}_{int}\}$) were then split randomly in a traditional 80/20 train/test split, $\{\boldsymbol{\alpha}_{train}\}$ and $\{\boldsymbol{\alpha}_{test}\}$. We emphasize that the union of these three sets, $\{\boldsymbol{\alpha}_A\} \cup \{\boldsymbol{\alpha}_{train}\} \cup \{\boldsymbol{\alpha}_{test}\} = \boldsymbol{D}$.

Separate SV-MOGP models were trained with the same parameters previously described in Section 3.2 and Appendix A. Projections of the first four PC components highlighting the external portion of the dataset removed by Eq. (E.1) during training, alongside parity plots demonstrating the respective model's performance on each external dataset can be seen in Fig. E.1. Summary error metrics for the trained models on each of the partitioned sets can be found in Table 2. In both cases, the performance remains stable. Case (A) is pure extrapolation; the extracted set is located out of distribution for both the inputs (see the first row of Fig. E.1) and the outputs (second row, Fig. E.1). In this case, the MOGP model continues to predict stably. Its expected predictions maintain relatively good agreement with the parity line. The NMAE roughly triples in comparison to the in distribution testing set – at worst, 5.418% versus 1.220% for k_{33}. However, even this increase is relatively small. The primary response from the model is a clear increase in the predicted variance. As quantified by the uncertainty bars on the parity plot, the model stably represents its increased uncertainty as it extrapolates further away from the training data. In the second case, (B), we again note that strong performance is retained. In this case, the agreement with the parity line is even better. We see almost no change in the NMAE: at worst, 0.384% versus 0.358% for the extrapolation set and the testing set, respectively. This demonstrates that the selected surrogate architecture is able to stably fill in any existing gaps or nonuniformity in the dataset as well as extrapolate stably to the surrounding regions. Both situations arise frequently in materials inverse problems, see Sections 4.3 and 4.4.

References

[1] E. Kalnay, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, ISBN: 9780511802270, 2002, http://dx.doi.org/10.1017/CBO9780511802270, URL https://www.cambridge.org/highereducation/books/atmospheric-modeling-data-assimilation-and-predictability/C5FD207439132836E85027754CE9BC1A.

[2] A.F. Bennett, Inverse Modeling of the Ocean and Atmosphere, Cambridge University Press, Cambridge, 2002, http://dx.doi.org/10.1017/CBO9780511535895, URL https://www.cambridge.org/core/books/inverse-modeling-of-the-ocean-and-atmosphere/5CFE830AEB412531F6419D17DB620B90.

[3] C. Lubich, From quantum to classical molecular dynamics: Reduced models and numerical analysis, 2008, http://dx.doi.org/10.4171/067, ISBN: 9783037190678 9783037195673 URL https://ems.press/books/zlam/55.

[4] J. Chen, D. Kipping, Probabilistic forecasting of the masses and radii of other worlds, Astrophys. J. 834 (1) (2016) 17, http://dx.doi.org/10.3847/1538-4357/834/1/17, URL https://iopscience.iop.org/article/10.3847/1538-4357/834/1/17.

[5] J. Adler, O. Oktem, Learned primal-dual reconstruction, IEEE Trans. Med. Imaging 37 (6) (2018) 1322–1332, http://dx.doi.org/10.1109/TMI.2018.2799231.

[6] E. Candes, M. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 21–30, http://dx.doi.org/10.1109/MSP.2007.914731, URL http://ieeexplore.ieee.org/document/4472240/.

[7] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289–1306, http://dx.doi.org/10.1109/TIT.2006.871582, URL http://ieeexplore.ieee.org/document/1614066/.

[8] A.M. Stuart, Inverse problems: A Bayesian perspective, Acta Numer. 19 (2010) 451–559, http://dx.doi.org/10.1017/S0962492910000061, URL https://www.cambridge.org/core/product/identifier/S0962492910000061/type/journal_article.

[9] J.P. Kaipio, E. Somersalo, Statistical and computational inverse problems, in: Applied Mathematical Sciences, vol. 160, Springer, New York, NY, 2005, http://dx.doi.org/10.1007/b138659, URL https://link.springer.com/10.1007/b138659.

[10] S. Arridge, P. Maass, O. Öktem, C.-B. Schönlieb, Solving inverse problems using data-driven models, Acta Numer. 28 (2019) 1–174, http://dx.doi.org/10.1017/S0962492919000059, Publisher: Cambridge University Press. URL https://www.cambridge.org/core/journals/acta-numerica/article/solving-inverse-problems-using-datadriven-models/CE5B3725869AEAF46E04874115B0AB15.

[11] J. Mullins, Y. Ling, S. Mahadevan, L. Sun, A. Strachan, Separation of aleatory and epistemic uncertainty in probabilistic model validation, Reliab. Eng. Syst. Saf. 147 (2016) 49–59, http://dx.doi.org/10.1016/j.ress.2015.10.003, URL https://www.sciencedirect.com/science/article/pii/S0951832015002859.

[12] E. Adeli, B. Rosić, H.G. Matthies, S. Reinstädler, D. Dinkler, Comparison of Bayesian methods on parameter identification for a viscoplastic model with damage, Metals 10 (7) (2020) 876, http://dx.doi.org/10.3390/met10070876, URL https://www.mdpi.com/2075-4701/10/7/876.

[13] A.P. Generale, R.B. Hall, R.A. Brockman, V.R. Joseph, G. Jefferson, L. Zawada, J. Pierce, S.R. Kalidindi, Bayesian calibration of continuum damage model parameters for an oxide-oxide ceramic matrix composite using inhomogeneous experimental data, Mech. Mater. 175 (2022) 104487, http://dx.doi.org/10.1016/j.mechmat.2022.104487, URL https://linkinghub.elsevier.com/retrieve/pii/S0167663622002514.

[14] C.C. Pacheo, G.S. Dulikravich, M. Vesenjak, M. Borovinšek, I.M.A. Duarte, R. Jha, S.R. Reddy, H.R.B. Orlande, M.J. Colaço, Inverse parameter identification in solid mechanics using Bayesian statistics, response surfaces and minimization, Tech. Mech. - Eur. J. Eng. Mech. 36 (1–2) (2016) 120–131, http://dx.doi.org/10.24352/UB.OVGU-2017-014, Number: 1-2. URL https://journals.ub.ovgu.de/index.php/techmech/article/view/695.

[15] E. Zhang, J. Chazot, J. Antoni, Parametric identification of elastic modulus of polymeric material in laminated glasses, IFAC Proc. Vol. 45 (16) (2012) 422–427, http://dx.doi.org/10.3182/20120711-3-BE-2027.00222, URL https://linkinghub.elsevier.com/retrieve/pii/S1474667015379878.

[16] A. Gallina, L. Ambrozinski, P. Pieczonka, T. Uhl, W. Staszewski, Bayesian parameter identification of orthotropic composite materials using Lamb waves dispersion curves measurement - Alberto Gallina, Lukasz Ambrozinski, Pawel Packo, Lukasz Pieczonka, Tadeusz Uhl, Wieslaw j Staszewski, 2017, J. Vib. Control 23 (16) (2015) URL https://journals.sagepub.com/doi/full/10.1177/1077546315619264.

[17] D.L. McDowell, Microstructure-sensitive computational structure-property relations in materials design, in: D. Shin, J. Saal (Eds.), Computational Materials System Design, Springer International Publishing, Cham, 2018, pp. 1–25, http://dx.doi.org/10.1007/978-3-319-68280-8_1.

[18] J. Rossin, P. Leser, K. Pusch, C. Frey, S.P. Murray, C.J. Torbet, S. Smith, S. Daly, T.M. Pollock, Bayesian inference of elastic constants and texture coefficients in additively manufactured cobalt-nickel superalloys using resonant ultrasound spectroscopy, Acta Mater. 220 (2021) 117287, http://dx.doi.org/10.1016/j.actamat.2021.117287, URL https://www.sciencedirect.com/science/article/pii/S1359645421006674.

[19] C. Mandache, Overview of non-destructive evaluation techniques for metal-based additive manufacturing, Mater. Sci. Technol. 35 (9) (2019) 1007–1015, http://dx.doi.org/10.1080/02670836.2019.1596370, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02670836.2019.1596370.

[20] M. Kobayashi, Ultrasonic nondestructive evaluation of microstructural changes of solid materials under plastic deformation—Part i. Theory, Int. J. Plast. 14 (6) (1998) 511–522, http://dx.doi.org/10.1016/S0749-6419(98)00005-9, URL https://www.sciencedirect.com/science/article/pii/S0749641998000059.

[21] S.R. Niezgoda, D.T. Fullwood, S.R. Kalidindi, Delineation of the space of 2-point correlations in a composite material system, Acta Mater. 56 (18) (2008) 5285–5292, http://dx.doi.org/10.1016/j.actamat.2008.07.005, URL http://www.sciencedirect.com/science/article/pii/S1359645408004886.

[22] N.H. Paulson, M.W. Priddy, D.L. McDowell, S.R. Kalidindi, Data-driven reduced-order models for rank-ordering the high cycle fatigue performance of polycrystalline microstructures, Mater. Des. 154 (2018) 170–183, http://dx.doi.org/10.1016/j.matdes.2018.05.009, URL http://www.sciencedirect.com/science/article/pii/S0264127518303861.

[23] S. Torquato, Random heterogeneous materials: Microstructure and macroscopic properties, in: Interdisciplinary Applied Mathematics, Springer-Verlag, New York, 2002, http://dx.doi.org/10.1007/978-1-4757-6355-3, URL https://www.springer.com/gp/book/9780387951676.

[24] S.R. Kalidindi, Hierarchical Materials Informatics: Novel Analytics for Materials Data, first ed., Butterworth-Heinemann, Amsterdam, 2015.

[25] A. Marshall, 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 (7) (2021) 2085–2095, http://dx.doi.org/10.1007/s11837-021-04696-w.

[26] A.P. Generale, S.R. Kalidindi, Reduced-order models for microstructure-sensitive effective thermal conductivity of woven ceramic matrix composites with residual porosity, Compos. Struct. 274 (2021) 114399, http://dx.doi.org/10.1016/j.compstruct.2021.114399, URL https://linkinghub.elsevier.com/retrieve/pii/S0263822321008618.

[27] A. Clausen, F. Wang, J.S. Jensen, O. Sigmund, J.A. Lewis, Topology optimized architectures with programmable Poisson's ratio over large deformations, Adv. Mater. 27 (37) (2015) 5523–5527, http://dx.doi.org/10.1002/adma.201502485, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.201502485. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.201502485.

[28] P. Vogiatzis, S. Chen, X. Wang, T. Li, L. Wang, Topology optimization of multi-material negative Poisson's ratio metamaterials using a reconciled level set method, Comput. Aided Des. 83 (2017) 15–32, http://dx.doi.org/10.1016/j.cad.2016.09.009, URL https://www.sciencedirect.com/science/article/pii/S0010448516301154.

[29] L.V. Gibiansky, O. Sigmund, Multiphase composites with extremal bulk modulus, J. Mech. Phys. Solids 48 (3) (2000) 461–498, http://dx.doi.org/10.1016/S0022-5096(99)00043-5, URL https://www.sciencedirect.com/science/article/pii/S0022509699000435.

[30] A. Radman, X. Huang, Y. Xie, Topological optimization for the design of microstructures of isotropic cellular materials, Eng. Optim. 45 (11) (2013) 1331–1348, http://dx.doi.org/10.1080/0305215X.2012.737781, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0305215X.2012.737781.

[31] O. Sigmund, S. Torquato, Design of materials with extreme thermal expansion using a three-phase topology optimization method, J. Mech. Phys. Solids 45 (6) (1997) 1037–1067, http://dx.doi.org/10.1016/S0022-5096(96)00114-7, URL https://www.sciencedirect.com/science/article/pii/S0022509696001147.

[32] S. Torquato, S. Hyun, A. Donev, Optimal design of manufacturable three-dimensional composites with multifunctional characteristics, J. Appl. Phys. 94 (9) (2003) 5748–5755, http://dx.doi.org/10.1063/1.1611631, URL http://aip.scitation.org/doi/10.1063/1.1611631.

[33] A.E. Robertson, S.R. Kalidindi, Efficient generation of anisotropic N-field microstructures from 2-point statistics using multi-output Gaussian random fields, Acta Mater. 232 (2022) 117927, http://dx.doi.org/10.1016/j.actamat.2022.117927, URL https://www.sciencedirect.com/science/article/pii/S1359645422003093.

[34] S. Kalidindi, S. Niezgoda, G. Landi, S. Vachhani, T. Fast, A novel framework for building materials knowledge systems, Comput. Mater. Continua 17 (2010) 103–125.

[35] E. Kröner, Statistical continuum mechanics, in: CISM International Centre for Mechanical Sciences, vol. 92, Springer Vienna, Vienna, 1971, http://dx.doi.org/10.1007/978-3-7091-2862-6, URL http://link.springer.com/10.1007/978-3-7091-2862-6.

[36] D.T. Fullwood, S.R. Niezgoda, B.L. Adams, S.R. Kalidindi, Microstructure sensitive design for performance optimization, Prog. Mater. Sci. 55 (6) (2010) 477–562, http://dx.doi.org/10.1016/j.pmatsci.2009.08.002, URL http://www.sciencedirect.com/science/article/pii/S0079642509000760.

[37] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. de Freitas, Taking the human out of the loop: A review of Bayesian optimization, Proc. IEEE 104 (1) (2016) 148–175, http://dx.doi.org/10.1109/JPROC.2015.2494218, Conference Name: Proceedings of the IEEE.

[38] X. Wang, Y. Jin, S. Schmitt, M. Olhofer, Recent advances in Bayesian optimization, 2022, http://dx.doi.org/10.48550/arXiv.2206.03301, arXiv:2206.03301 [cs, math]. URL http://arxiv.org/abs/2206.03301.

[39] J. Jung, J.I. Yoon, H.K. Park, H. Jo, H.S. Kim, Microstructure design using machine learning generated low dimensional and continuous design space, Materialia 11 (2020) 100690, http://dx.doi.org/10.1016/j.mtla.2020.100690, URL https://linkinghub.elsevier.com/retrieve/pii/S2589152920301071.

[40] A. Solomou, G. Zhao, S. Boluki, J.K. Joy, X. Qian, I. Karaman, R. Arróyave, D.C. Lagoudas, Multi-objective Bayesian materials discovery: Application on the discovery of precipitation strengthened NiTi shape memory alloys through micromechanical modeling, Mater. Des. 160 (2018) 810–827, http://dx.doi.org/10.1016/j.matdes.2018.10.014, URL https://www.sciencedirect.com/science/article/pii/S026412751830769X.

[41] A. Talapatra, S. Boluki, T. Duong, X. Qian, E. Dougherty, R. Arróyave, Autonomous efficient experiment design for materials discovery with Bayesian model averaging, Phys. Rev. Mater. 2 (11) (2018) 113803, http://dx.doi.org/10.1103/PhysRevMaterials.2.113803, Publisher: American Physical Society. URL https://link.aps.org/doi/10.1103/PhysRevMaterials.2.113803.

[42] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, 2014, http://dx.doi.org/10.48550/arXiv.1406.2661, arXiv:1406.2661 [cs, stat]. URL http://arxiv.org/abs/1406.2661.

[43] D.P. Kingma, M. Welling, Auto-encoding variational Bayes, 2013, arXiv:1312.6114 [cs, stat]. URL http://arxiv.org/abs/1312.6114.

[44] D.J. Rezende, S. Mohamed, Variational inference with normalizing flows, 2016, arXiv:1505.05770 [cs, stat]. URL http://arxiv.org/abs/1505.05770.

[45] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E.W. Pellegrini, R.S. Klessen, L. Maier-Hein, C. Rother, U. Köthe, Analyzing inverse problems with invertible neural networks, 2019, http://dx.doi.org/10.48550/arXiv.1808.04730, arXiv:1808.04730 [cs, stat]. URL http://arxiv.org/abs/1808.04730.

[46] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, 2020, http://dx.doi.org/10.48550/arXiv.2006.11239, arXiv:2006.11239 [cs, stat]. URL http://arxiv.org/abs/2006.11239.

[47] A.E. Robertson, C. Kelly, M. Buzzy, S.R. Kalidindi, Local–global decompositions for conditional microstructure generation, Acta Mater. 253 (2023) 118966, http://dx.doi.org/10.1016/j.actamat.2023.118966, URL https://www.sciencedirect.com/science/article/pii/S1359645423002975.

[48] C. Düreth, P. Seibert, D. Rücker, S. Handford, M. Kästner, M. Gude, Conditional diffusion-based microstructure reconstruction, Mater. Today Commun. 35 (2023) 105608, http://dx.doi.org/10.1016/j.mtcomm.2023.105608, URL https://www.sciencedirect.com/science/article/pii/S2352492823002982.

[49] K.-H. Lee, G.J. Yun, Microstructure reconstruction using diffusion-based generative models, Mech. Adv. Mater. Struct. (2023) 1–19, http://dx.doi.org/10.1080/15376494.2023.2198528.

[50] N.N. Vlassis, W. Sun, Denoising diffusion algorithm for inverse design of microstructures with fine-tuned nonlinear material properties, Comput. Methods Appl. Mech. Engrg. 413 (2023) 116126, http://dx.doi.org/10.1016/j.cma.2023.116126, URL https://www.sciencedirect.com/science/article/pii/S0045782523002505.

[51] Z. Pei, K.A. Rozman, Ö.N. Doğan, Y. Wen, N. Gao, E.A. Holm, J.A. Hawk, D.E. Alman, M.C. Gao, Machine-learning microstructure for inverse material design, Adv. Sci. 8 (23) (2021) 2101207, http://dx.doi.org/10.1002/advs.202101207, URL https://onlinelibrary.wiley.com/doi/10.1002/advs.202101207.

[52] L. Wang, Y.-C. Chan, F. Ahmed, Z. Liu, P. Zhu, W. Chen, Deep generative modeling for mechanistic-based learning and design of metamaterial systems, Comput. Methods Appl. Mech. Engrg. 372 (2020) 113377, http://dx.doi.org/10.1016/j.cma.2020.113377, URL https://linkinghub.elsevier.com/retrieve/pii/S0045782520305624.

[53] V. Fung, J. Zhang, G. Hu, P. Ganesh, B.G. Sumpter, Inverse design of two-dimensional materials with invertible neural networks, npj Comput. Mater. 7 (1) (2021) 200, http://dx.doi.org/10.1038/s41524-021-00670-x, URL https://www.nature.com/articles/s41524-021-00670-x.

[54] H.J. Lim, K.-H. Lee, G.J. Yun, Microstructure design of multifunctional particulate composite materials using conditional diffusion models, 2023, http://dx.doi.org/10.48550/arXiv.2301.09051, arXiv:2301.09051 [cond-mat, physics:physics]. URL http://arxiv.org/abs/2301.09051.

[55] Y. Kim, H.K. Park, J. Jung, P. Asghari-Rad, S. Lee, J.Y. Kim, H.G. Jung, H.S. Kim, Exploration of optimal microstructure and mechanical properties in continuous microstructure space using a variational autoencoder, Mater. Des. 202 (2021) 109544, http://dx.doi.org/10.1016/j.matdes.2021.109544, URL https://www.sciencedirect.com/science/article/pii/S0264127521000976.

[56] L. Xu, N. Hoffman, Z. Wang, H. Xu, Harnessing structural stochasticity in the computational discovery and design of microstructures, Mater. Des. 223 (2022) 111223, http://dx.doi.org/10.1016/j.matdes.2022.111223, URL https://www.sciencedirect.com/science/article/pii/S0264127522008450.

[57] R.K. Tan, N.L. Zhang, W. Ye, A deep learning–based method for the design of microstructural materials, Struct. Multidiscip. Optim. 61 (4) (2020) 1417–1438, http://dx.doi.org/10.1007/s00158-019-02424-2, URL http://link.springer.com/10.1007/s00158-019-02424-2.

[58] A. Deshwal, J. Doppa, Combining latent space and structured kernels for Bayesian optimization over combinatorial spaces, in: Advances in Neural Information Processing Systems, Vol. 34, Curran Associates, Inc., 2021, pp. 8185–8200, URL https://proceedings.neurips.cc/paper/2021/hash/44e76e99b5e194377e955b13fb12f630-Abstract.html.

[59] C. Hu, S. Martin, R. Dingreville, Accelerating phase-field predictions via recurrent neural networks learning the microstructure evolution in latent space, Comput. Methods Appl. Mech. Engrg. 397 (2022) 115128, http://dx.doi.org/10.1016/j.cma.2022.115128, URL https://www.sciencedirect.com/science/article/pii/S0045782522003097.

[60] S. Kench, S. Cooper, Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion, Nat.: Mach. Intell. 3 (2021) 299–305, http://dx.doi.org/10.1038/s42256-021-00322-1.

[61] B.N. Cox, H.A. Bale, M. Begley, M. Blacklock, B.-C. Do, T. Fast, M. Naderi, M. Novak, V.P. Rajan, R.G. Rinaldi, R.O. Ritchie, M.N. Rossol, J.H. Shaw, O. Sudre, Q. Yang, F.W. Zok, D.B. Marshall, Stochastic virtual tests for high-temperature ceramic matrix composites, Annu. Rev. Mater. Res. 44 (1) (2014) 479–529, http://dx.doi.org/10.1146/annurev-matsci-122013-025024, URL http://www.annualreviews.org/doi/10.1146/annurev-matsci-122013-025024.

[62] S.R. Niezgoda, Y.C. Yabansu, S.R. Kalidindi, Understanding and visualizing microstructure and microstructure variance as a stochastic process, Acta Mater. 59 (16) (2011) 6387–6400, http://dx.doi.org/10.1016/j.actamat.2011.06.051, URL https://www.sciencedirect.com/science/article/pii/S1359645411004654.

[63] S.R. Kalidindi, Materials, Data, and Informatics, Elsevier, 2015, http://dx.doi.org/10.1016/B978-0-12-410394-8.00001-1, URL https://linkinghub.elsevier.com/retrieve/pii/B9780124103948000011.

[64] B.L. Adams, S.R. Kalidindi, D.T. Fullwood, Microstructure-sensitive design for performance optimization, in: Microstructure Sensitive Design for Performance Optimization, Elsevier, 2013, p. i, http://dx.doi.org/10.1016/B978-0-12-396989-7.01001-7, URL https://linkinghub.elsevier.com/retrieve/pii/B9780123969897010017.

[65] S.R. Kalidindi, Feature engineering of material structure for AI-based materials knowledge systems, J. Appl. Phys. 128 (4) (2020) 041103, http://dx.doi.org/10.1063/5.0011258, Publisher: American Institute of Physics. URL https://aip.scitation.org/doi/10.1063/5.0011258.

[66] A. Cecen, T. Fast, S.R. Kalidindi, Versatile algorithms for the computation of 2-point spatial correlations in quantifying material structure, Integr. Mater. Manuf. Innov. 5 (1) (2016) 1–15, http://dx.doi.org/10.1186/s40192-015-0044-x.

[67] A. Gupta, A. Cecen, S. Goyal, A.K. Singh, S.R. Kalidindi, Structure–property linkages using a data science approach: Application to a non-metallic inclusion/steel composite system, Acta Mater. 91 (2015) 239–254, http://dx.doi.org/10.1016/j.actamat.2015.02.045, URL http://www.sciencedirect.com/science/article/pii/S1359645415001603.

[68] M.I. Latypov, L.S. Toth, S.R. Kalidindi, Materials knowledge system for nonlinear composites, Comput. Methods Appl. Mech. Engrg. 346 (2019) 180–196, http://dx.doi.org/10.1016/j.cma.2018.11.034, URL http://www.sciencedirect.com/science/article/pii/S0045782518305930.

[69] Y.C. Yabansu, P. Altschuh, J. Hötzer, M. Selzer, B. Nestler, S.R. Kalidindi, A digital workflow for learning the reduced-order structure-property linkages for permeability of porous membranes, Acta Mater. 195 (2020) 668–680, http://dx.doi.org/10.1016/j.actamat.2020.06.003, URL http://www.sciencedirect.com/science/article/pii/S1359645420304274.

[70] N.H. Paulson, M.W. Priddy, D.L. McDowell, S.R. Kalidindi, Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics, Acta Mater. 129 (2017) 428–438, http://dx.doi.org/10.1016/j.actamat.2017.03.009, URL http://www.sciencedirect.com/science/article/pii/S135964541730188X.

[71] S.R. Niezgoda, A.K. Kanjarla, S.R. Kalidindi, Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data, Integr. Mater. Manuf. Innov. 2 (1) (2013) 54–80, http://dx.doi.org/10.1186/2193-9772-2-3.

[72] A.E. Robertson, S.R. Kalidindi, Digital representation and quantification of discrete dislocation structures, JOM 73 (7) (2021) 2143–2158, http://dx.doi.org/10.1007/s11837-021-04669-z.

[73] Y.C. Yabansu, P. Steinmetz, J. Hötzer, S.R. Kalidindi, B. Nestler, Extraction of reduced-order process-structure linkages from phase-field simulations, Acta Mater. 124 (2017) 182–194, http://dx.doi.org/10.1016/j.actamat.2016.10.071, URL http://www.sciencedirect.com/science/article/pii/S1359645416308485.

[74] A.P. Generale, S.R. Kalidindi, Uncertainty quantification and propagation in the microstructure-sensitive prediction of the stress-strain response of woven ceramic matrix composites, Comput. Struct. 286 (2023) 107110, http://dx.doi.org/10.1016/j.compstruc.2023.107110, URL https://linkinghub.elsevier.com/retrieve/pii/S0045794923001402.

[75] G.H. Harrington, C. Kelly, V. Attari, R. Arroyave, S.R. Kalidindi, Application of a chained-ANN for learning the process-structure mapping in $Mg_2Si_xSn_{1-x}$ spinodal decomposition, Integr. Mater. Manuf. Innov. 11 (3) (2022) 433–449, http://dx.doi.org/10.1007/s40192-022-00274-3.

[76] C.E. Rasmussen, C.K.I. Williams, Gaussian processes for machine learning, in: Adaptive Computation and Machine Learning, MIT Press, Cambridge, Mass, 2006, OCLC: ocm61285753.

[77] C.M. Bishop, Pattern recognition and machine learning, in: Information Science and Statistics, Springer, New York, 2006.

[78] K.P. Murphy, Machine learning: a probabilistic perspective, in: Adaptive Computation and Machine Learning Series, MIT Press, Cambridge, MA, 2012.

[79] A.G. Wilson, R.P. Adams, Gaussian process kernels for pattern discovery and extrapolation, 2013, http://dx.doi.org/10.48550/arXiv.1302.4245, arXiv:1302.4245 [cs, stat]. URL http://arxiv.org/abs/1302.4245.

[80] A.G. Wilson, E. Gilboa, A. Nehorai, J.P. Cunningham, Fast kernel learning for multidimensional pattern extrapolation, in: Advances in Neural Information Processing Systems, Vol. 27, Curran Associates, Inc., 2014, URL https://proceedings.neurips.cc/paper/2014/hash/77369e37b2aa1404f416275183ab055f-Abstract.html.

[81] K. Ulrich, D.E. Carlson, W. Lian, J.S. Borg, K. Dzirasa, L. Carin, Analysis of brain states from multi-region LFP time-series, in: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS '14, MIT Press, Cambridge, MA, USA, 2014, pp. 2483–2491.

[82] M. Titsias, Variational learning of inducing variables in sparse Gaussian processes, in: Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics, PMLR, 2009, pp. 567–574, ISSN: 1938-7228. URL https://proceedings.mlr.press/v5/titsias09a.html.

[83] A.G.d.G. Matthews, J. Hensman, R. Turner, Z. Ghahramani, On sparse variational methods and the Kullback-Leibler divergence between stochastic processes, in: Artificial Intelligence and Statistics, PMLR, 2016, pp. 231–239.

[84] M.A. Alvarez, L. Rosasco, N.D. Lawrence, Kernels for vector-valued functions: A review, 2012, Number: arXiv:1106.6251 arXiv:1106.6251 [cs, math, stat]. URL http://arxiv.org/abs/1106.6251.

[85] M. Lázaro-Gredilla, A. Figueiras-Vidal, Inter-domain Gaussian processes for sparse inference using inducing features, in: Advances in Neural Information Processing Systems, Vol. 22, Curran Associates, Inc., 2009, URL https://proceedings.neurips.cc/paper/2009/hash/5ea1649a31336092c05438df996a3e59-Abstract.html.

[86] A.G. Journel, C.J. Huijbregts, Mining Geostatistics, Google-Books-ID: Id1GAAAAYAAJ, Blackburn Press, 2003.

[87] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (3) (2001) 425–464, http://dx.doi.org/10.1111/1467-9868.00294, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00294. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00294.

[88] M. Van Oijen, J. Rougier, R. Smith, Bayesian calibration of process-based forest models: bridging the gap between models and data, Tree Physiol. 25 (7) (2005) 915–927, http://dx.doi.org/10.1093/treephys/25.7.915.

[89] T.G. Trucano, L.P. Swiler, T. Igusa, W.L. Oberkampf, M. Pilch, Calibration, validation, and sensitivity analysis: What's what, Reliab. Eng. Syst. Saf. 91 (10) (2006) 1331–1357, http://dx.doi.org/10.1016/j.ress.2005.11.031, URL https://www.sciencedirect.com/science/article/pii/S0951832005002437.

[90] S.T. Tokdar, R.E. Kass, Importance sampling: a review, WIREs Comput. Stat. 2 (1) (2010) 54–60, http://dx.doi.org/10.1002/wics.56, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wics.56. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.56.

[91] S. Brooks, A. Gelman, G. Jones, X.-L. Meng, Handbook of Markov Chain Monte Carlo, Google-Books-ID: qfRsAIKZ4rIC, CRC Press, 2011.

[92] D.H. Ballard, Modular learning in neural networks, in: Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1, AAAI '87, AAAI Press, Seattle, Washington, 1987, pp. 279–284.

[93] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Netw. 61 (2015) 85–117, http://dx.doi.org/10.1016/j.neunet.2014.09.003, arXiv:1404.7828 [cs]. URL http://arxiv.org/abs/1404.7828.

[94] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, A. Lerchner, Beta-VAE: Learning basic visual concepts with a constrained variational framework, 2016, URL https://openreview.net/forum?id=Sy2fzU9gl.

[95] C.P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, A. Lerchner, Understanding disentangling in β-VAE, 2018, arXiv:1804.03599 [cs, stat]. URL http://arxiv.org/abs/1804.03599.

[96] A.B. Owen, Monte Carlo theory, methods and examples, 2013, URL https://artowen.su.domains/mc/.

[97] D. Huijser, J. Goodman, B.J. Brewer, Properties of the affine invariant ensemble sampler in high dimensions, 2017, http://dx.doi.org/10.48550/arXiv.1509.02230, physics, stat. URL http://arxiv.org/abs/1509.02230.

[98] H. Sun, K.L. Bouman, P. Tiede, J.J. Wang, S. Blunt, D. Mawet, α-Deep probabilistic inference (α-DPI): Efficient uncertainty quantification from exoplanet astrometry to black hole feature extraction, Astrophys. J. 932 (2) (2022) 99, http://dx.doi.org/10.3847/1538-4357/ac6be9, Publisher: The American Astronomical Society.

[99] H. Sun, K.L. Bouman, Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging, Proc. AAAI Conf. Artif. Intell. 35 (3) (2021) 2628–2637, http://dx.doi.org/10.1609/aaai.v35i3.16366, Number: 3. URL https://ojs.aaai.org/index.php/AAAI/article/view/16366.

[100] L. Dinh, D. Krueger, Y. Bengio, NICE: Non-linear independent components estimation, 2015, http://dx.doi.org/10.48550/arXiv.1410.8516, arXiv:1410.8516 [cs]. URL http://arxiv.org/abs/1410.8516.

[101] L. Dinh, J. Sohl-Dickstein, S. Bengio, Density estimation using Real NVP, 2017, arXiv:1605.08803 [cs, stat]. URL http://arxiv.org/abs/1605.08803.

[102] S.T. Radev, U.K. Mertens, A. Voss, L. Ardizzone, U. Köthe, BayesFlow: Learning complex stochastic models with invertible neural networks, IEEE Trans. Neural Netw. Learn. Syst. 33 (4) (2022) 1452–1466, http://dx.doi.org/10.1109/TNNLS.2020.3042395, Conference Name: IEEE Transactions on Neural Networks and Learning Systems.

[103] A. Khosravani, A. Cecen, S.R. Kalidindi, Development of high throughput assays for establishing process-structure-property linkages in multiphase polycrystalline metals: Application to dual-phase steels, Acta Mater. 123 (2017) 55–69, http://dx.doi.org/10.1016/j.actamat.2016.10.033, URL https://www.sciencedirect.com/science/article/pii/S135964541630800X.

[104] M.I. Latypov, S.R. Kalidindi, Data-driven reduced order models for effective yield strength and partitioning of strain in multiphase materials, J. Comput. Phys. 346 (2017) 242–261, http://dx.doi.org/10.1016/j.jcp.2017.06.013, URL https://www.sciencedirect.com/science/article/pii/S0021999117304588.

[105] A. Iskakov, Y.C. Yabansu, S. Rajagopalan, A. Kapustina, S.R. Kalidindi, Application of spherical indentation and the materials knowledge system framework to establishing microstructure-yield strength linkages from carbon steel scoops excised from high-temperature exposed components, Acta Mater. 144 (2018) 758–767, http://dx.doi.org/10.1016/j.actamat.2017.11.024, URL https://www.sciencedirect.com/science/article/pii/S1359645417309679.

[106] R.A. Fisher, The moments of the distribution for normal samples of measures of departure from normality, Proc. Royal Soc. Lond. Ser. A, Contain. Pap. Math. Phys. Character 130 (812) (1930) 16–28, http://dx.doi.org/10.1098/rspa.1930.0185, Publisher: Royal Society. URL https://royalsocietypublishing.org/doi/10.1098/rspa.1930.0185.

[107] H. Fu, C. Li, X. Liu, J. Gao, A. Celikyilmaz, L. Carin, Cyclical annealing schedule: A simple approach to mitigating KL vanishing, 2019, arXiv:1903.10145 [cs, stat]. URL http://arxiv.org/abs/1903.10145.

[108] T. Spinner, J. Körner, J. Görtler, O. Deussen, Towards an interpretable latent space : An intuitive comparison of autoencoders with variational autoencoders, 2018, URL https://kops.uni-konstanz.de/handle/123456789/43657.

[109] M. Connor, G. Canal, C. Rozell, Variational autoencoder with learned latent structure, in: Proceedings of the 24th International Conference on Artificial Intelligence and Statistics, PMLR, (ISSN: 2640-3498) 2021, pp. 2359–2367, URL https://proceedings.mlr.press/v130/connor21a.html.

[110] C. Zhang, J. Bütepage, H. Kjellström, S. Mandt, Advances in variational inference, IEEE Trans. Pattern Anal. Mach. Intell. 41 (8) (2019) 2008–2026, http://dx.doi.org/10.1109/TPAMI.2018.2889774, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[111] G. Papamakarios, E. Nalisnick, D.J. Rezende, S. Mohamed, B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, 2021.

[112] A. Castillo, S.R. Kalidindi, A Bayesian framework for the estimation of the single crystal elastic parameters from spherical indentation stress-strain measurements, Front. Mater. 6 (2019) 136, http://dx.doi.org/10.3389/fmats.2019.00136, URL https://www.frontiersin.org/article/10.3389/fmats.2019.00136/full.

[113] Qhull code for convex hull, delaunay triangulation, voronoi diagram, and halfspace intersection about a point. URL http://www.qhull.org/.

[114] H. Lin, L.P. Brown, A.C. Long, Modelling and simulating textile structures using TexGen, Adv. Mater. Res. 331 (2011) 44–47, http://dx.doi.org/10.4028/www.scientific.net/AMR.331.44, URL https://www.scientific.net/AMR.331.44.

[115] A.R. Castillo, V.R. Joseph, S.R. Kalidindi, Bayesian sequential design of experiments for extraction of single-crystal material properties from spherical indentation measurements on polycrystalline samples, JOM 71 (8) (2019) 2671–2679, http://dx.doi.org/10.1007/s11837-019-03549-x.

[116] J. Gardner, G. Pleiss, K.Q. Weinberger, D. Bindel, A.G. Wilson, GPyTorch: Blackbox Matrix-Matrix Gaussian process inference with GPU acceleration, in: Advances in Neural Information Processing Systems, Vol. 31, Curran Associates, Inc., 2018, URL https://proceedings.neurips.cc/paper/2018/hash/27e8e17134dd7083b050476733207ea1-Abstract.html.

[117] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017, http://dx.doi.org/10.48550/arXiv.1412.6980, arXiv:1412.6980 [cs]. URL http://arxiv.org/abs/1412.6980.

[118] D.P. Kingma, P. Dhariwal, Glow: Generative flow with invertible 1x1 convolutions, in: Advances in Neural Information Processing Systems, Vol. 31, Curran Associates, Inc., 2018, URL https://papers.nips.cc/paper/2018/hash/d139db6a236200b21cc7f752979132d0-Abstract.html.

[119] J. Ho, X. Chen, A. Srinivas, Y. Duan, P. Abbeel, Flow++: Improving flow-based generative models with variational dequantization and architecture design, 2019, http://dx.doi.org/10.48550/arXiv.1902.00275, arXiv:1902.00275 [cs, stat]. URL http://arxiv.org/abs/1902.00275.

[120] C.-W. Huang, S. Tan, A. Lacoste, A. Courville, Improving explorability in variational inference with annealed variational objectives, 2018, arXiv:1809.01818 [cs, stat]. URL http://arxiv.org/abs/1809.01818.

[121] R. Gao, E. Nijkamp, D.P. Kingma, Z. Xu, A.M. Dai, Y.N. Wu, Flow contrastive estimation of energy-based models, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, 2020, pp. 7515–7525, http://dx.doi.org/10.1109/CVPR42600.2020.00754, URL https://ieeexplore.ieee.org/document/9157612/.

[122] S. Kumar, S. Tan, L. Zheng, D.M. Kochmann, Inverse-designed spinodoid metamaterials, npj Comput. Mater. 6 (1) (2020) 73, http://dx.doi.org/10.1038/s41524-020-0341-6, URL https://www.nature.com/articles/s41524-020-0341-6.

[123] D.T. Fullwood, B.L. Adams, S.R. Kalidindi, A strong contrast homogenization formulation for multi-phase anisotropic materials, J. Mech. Phys. Solids 56 (6) (2008) 2287–2297, http://dx.doi.org/10.1016/j.jmps.2008.01.003, URL https://www.sciencedirect.com/science/article/pii/S0022509608000057.