

Full length article

Local–Global Decompositions for Conditional Microstructure Generation

Andreas E. Robertson^a, Conlain Kelly^b, Michael Buzzy^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:

Conditional microstructure generation

2-Point statistics

Score based denoising diffusion

Data-efficient training

Efficient generation

ABSTRACT

Conditional microstructure generation tools offer an important, inexpensive pathway to constructing statistically diverse datasets for Integrated Computational Materials Engineering and Materials Informatics efforts. To provide this utility in practice, an ideal generative framework must be able to efficiently, systematically, and robustly generate microstructures corresponding to selected spatial statistics (e.g., 1- and 2-point statistics) while also producing realistic local features (e.g., shapes and sizes of individual phase constituents). Because of the austerity of these requirements, generative frameworks often target either statistical conditioning or visual realism, but not both. In this paper, we propose to bridge these two approaches by approximating a microstructure's generating process (i.e., its stochastic microstructure function) using a two layer semi-directed probabilistic graphical model. The first layer – a Gaussian Random Field (GRF) – provides direct control of the 1- and 2-point statistics. The second layer – a Score Based Generative Deep Learning model – postprocesses GRF predictions to refine local features while preserving global patterning. To understand and evaluate our proposed framework, we apply it to generate statistically equivalent N-phase microstructures from experimental references, including a 2-Phase Nickel-based Super Alloy and a 3-phase $\alpha - \beta$ Titanium alloy. Through these two case studies, we demonstrate that our framework successfully matches both lower-order (1- and 2-point statistics) and several salient higher-order statistics. Additionally, we briefly explore the capacity of these models to extrapolate outside of their training data by varying the input 2-point statistics. We discuss the value of this ability towards systematically generating diverse microstructure datasets.

1. Introduction

Fueled by national initiatives such as the Materials Genome Initiative [11], there is a rapidly growing interest in synthetic microstructure generation [12–17]. Combined with limited, expensive experimental data, such tools promise a pathway for systematically constructing the large, diverse microstructure datasets necessary for design [18–21], discovery [22,23], and manufacturing [24] tasks central to Integrated Computational Materials Engineering (ICME) and Materials Informatics (MI) efforts. Therefore, to be useful to these efforts, it is important that generative frameworks maintain their compatibility with the larger ICME and MI paradigms for studying engineering materials.

Foundational to many ICME and MI frameworks is the statistical treatment¹ of materials and their microstructures, which provides a robust theoretical platform for the rigorous analysis of their response to external stimuli [2,25–28]. Fundamentally, this worldview proposes that instances of material microstructures that we observe (through,

for example, experimental techniques such as Scanning Electron Microscopy (SEM) or during computer simulations) are sampled instances from abstract stochastic processes called Stochastic Microstructure Functions [9,12,25–27]. Over the past few decades, researchers have leveraged the n-point statistics [2,25–27] that identify these stochastic microstructure functions as powerful descriptive features for tasks such as unsupervised quantification and analysis of material microstructures [29,30], construction of advanced homogenization models for identifying effective properties [4,5,19,31], microstructure sensitive design [20,21,32], the construction of process–structure linkages to describe the transformation of material microstructures in response to manufacturing conditions [24,33–35], and in inverse problems for non-destructive testing [36,37]. Additionally, these ideas have been used to form localization models for approximately or exactly solving partial differential equations over heterogeneous microstructures [38–43].

* 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).

¹ Here, we make reference to the general class of theories that approach the quantitative study of materials and their microstructures from the perspective of continuum statistical mechanics and probability theory. Prominent examples are Kroner's Statistical Continuum Theories [1,2], Torquato's Strong Contrast Homogenization Theories [3–6], and Kalidindi's Materials Knowledge System [7–10].

Within ICME and MI efforts, microstructure generation tools are commonly used to systematically generate microstructures targeting specific salient microstructure statistics [19,21,44,45]. For example, because several statistical measures (e.g., Orientation Distribution Functions, Grain Size Distributions) can be controlled in the generating process [46,47], Dream3D and other similar tools [48] have been successfully applied to many important ICME applications (e.g., design [21, 44], manufacturing [24]). The capacity to systematically control the statistics of the generating process (hereafter referred to as conditional microstructure generation) is fundamental to the generator's broader usefulness.

Keeping in mind the desired applications, useful conditional microstructure generation tools should, ideally, meet the following five requirements.

1. **Control:** The generative model should be conditioned by microstructure statistics important to broader engineering problems; in many studies this takes the form of first- and second-order statistics [19,24,31,36].
2. **Realism:** Although we only require explicit control over lower-order statistics, the generative framework should produce microstructures whose higher order statistics are representative of real engineering materials. Because the number of individual statistics involved increases exponentially [25,26,49] with increasing order of the spatial correlations (i.e., n -point statistics), explicit conditioning with higher-order statistics, while ideal, is an extremely challenging requirement.
3. **Robustness:** The generative quality should remain stable across the input domain since these tools are usually used to systematically explore large ranges of the input microstructure statistics. This requirement is especially important when employing deep learning models. For engineering purposes, a generative model whose function is restricted to producing new microstructures that are statistically similar to its training set has very limited practical utility. By definition, we would already have ample examples of microstructures that it can produce.
4. **Data Efficiency:** If training is required, training should be data efficient. Given the well documented data scarcity prevalent in Materials Informatics [50–53], significant training data is often an unachievable luxury.
5. **Computational Efficiency:** An ideal generative model is computationally efficient to sample.

We recently proposed an initial model for synthetic generation of vector-valued microstructures conditioned on 2-point statistics. This was achieved by constructing a computationally tractable approximation of the Stochastic Microstructure Function using multi-output Gaussian Random Fields (GRFs) [12]. The GRF model met many of the previously described criteria. It can be efficiently sampled at $O(NS \ln S)$ computational cost for S spatial pixels and N microstructure states (Requirement Five). Additionally, the model does not require training data (Requirement Four) and as such displays excellent robustness (Requirement Three). Finally, it directly incorporates and statistically matches a desired set of 1- and 2-point statistics (Requirement One). However, in exchange for these strengths, the model is limited by its inability to incorporate and correctly match higher-order statistics. As a result, the GRF-generated microstructures are heavily biased towards semi-connected, pseudo-amorphous features. Obviously such features are rarely representative for the vast majority of engineering materials; the model fails to satisfy Requirement Two. We note that the importance of this requirement depends heavily on the desired application. For example, in low-contrast composites, extensive research has demonstrated that effective elastic and simple plastic properties are sufficiently defined by 1- and 2-point statistics [4,19,31,54]. In such cases, we need not worry about matching the higher-order statistics. However, it remains unclear in which situations the higher-order statistics play an important role [49,55–57].

In this paper, we propose a refined approximation to the Stochastic Microstructure Function. The resulting generating process retains explicit conditioning on lower-order statistics (1- and 2-point statistics), while also offering strong agreement with salient higher-order statistics. Theoretically, we achieve this by expanding on the local-global hypothesis we presented at the end of our prior work [12]: for many engineering materials' microstructures, lower-order statistics sufficiently describe the salient global spatial patterns, as well as average shape, of microstructure features, while higher-order statistics primarily characterize the features' individual shapes and their spatially localized interactions.

Mathematically, we argue that the conceptual split between local and global implies a natural decomposition of the dependency structure² of stochastic microstructure functions. As a higher-order approximation, we propose a two-layer semi-directed Probabilistic Graphical Model³ [59,60]. The first layer, containing a latent variable for every pixel, incorporates the global information of the 1- and 2-point statistics. Here, the graph is densely connected and the resulting distribution is parameterized using the multi-output Gaussian Random Field model [12]. The second layer, containing visible variables for every pixel, accounts for higher-order, local information using a Score-Based Generative model. Connections from the first layer and within the second layer are spatially limited to compact neighborhoods. In the context of previous generation efforts, the proposed framework is a unification of global and local approximation models leveraging the strengths of both.

The remainder of the paper is structured as follows. First, we present a discussion of existing generation frameworks in the context of the local-global split. Next, we introduce Score-Based Generative models as a valuable tool for learning conditional distributions. Subsequently, we present a complete discussion of the proposed generative model and accompanying training methodology. Finally, we specialize the framework to the N -phase generation problem and provide several case studies aimed at exploring its strengths and limitations. In the first case study, we benchmark the expressiveness of the model by generating 2-phase microstructures derived from a segmented SEM micrograph of a Ni-based super alloy [61]. In the second case study, we demonstrate the application of the proposed frameworks to a system with larger numbers of phases and more diverse local features. Specifically, we use a $\alpha - \beta$ Ti microstructure [62] as the reference. In both case studies, the framework's performance is evaluated both qualitatively – in its ability to construct realistic features – and quantitatively – in its ability to match the statistics of the underlying microstructure process. Additionally, we utilize these case studies to facilitate an important discussion regarding architecture design and learning strategies for effectively training the local distribution model (details provided in Section 4 and Appendix C). In particular, we emphasize training within Materials Informatics' characteristic data-starved environments and the necessary considerations for producing reliable deep learning models for this application. In the first case study, we also briefly explore the model's capacity to extrapolate outside of its training data.

To facilitate communication we briefly present the notation adopted throughout the paper. Vector-valued quantities are demarcated in bold, $\mathbf{a}$. Quantities with spatial dependency, such as spatially resolved functions, are demarcated using a subscript for discretely sampled quantities or a spatial dependency for continuous quantities: a_s and $a(\mathbf{x})$,

² Here, we are referring to the existence and nature of inter-variable dependencies within a distribution [58,59].

³ In this formalism, a distribution is represented as a graph over a node set defined by the distribution's random variables (in this case, the vector-valued microstructure state across all pixels). A directed edge between variables represents conditional dependency along the edge. This graph-based formalism offers a quantitative language for constructing and studying the structure of probabilistic models and provides a means of converting domain intuition (e.g., separation of scales) into testable models (such as ours).

respectively. Components of vector-valued quantities are indexed using a superscript in parenthesis, $a^{(\beta)} = \mathbf{a} \cdot \mathbf{e}^\beta$, where $\mathbf{e}^\beta$ is the β -basis vector. Finally, summations will always be written explicitly using the summation operator and are never implied by repeated indices.

2. Background

Conditional synthetic microstructure generation remains an open research question due to the myriad relevant features in a microstructure. Out of practical necessity, generative frameworks are usually constructed to only target specific subsets of these features. For example, many research efforts have proposed frameworks that generate microstructures matching 1- and 2-point statistics [12,14,16,17,63–67]. However, in exchange for strong agreement on the target statistic, this specialization produces highly unrealistic artifacts. For example, the problem of synthetically generating fiber composites from 1- and 2-point statistics is an ubiquitous benchmark in microstructure generation because it clearly delineates important strengths and limitations of proposed methods [12,15,68]. Synthetic microstructures generated by models designed to match the 1- and 2-point statistics of a fiber composite display realistic long range spatial arrangements, but lack the characteristic uniform fiber shapes. Instead, they display significant connectivity in the fiber phase. Similar issues appear in other applications; Jiao et al. present the generation of synthetic Fontainebleau sandstone. Unlike the original micrograph, they note that the synthetic microstructures lack the original's characteristic sharp phase boundaries [63].

Amongst methods specialized for 1- and 2-point statistics, optimization-based generative models (such as Jiao et al.'s method [16, 63]) have partially addressed these limitations. In exchange for long generation times, instead of directly approximating the stochastic microstructure function, optimization-based methods generate new microstructures displaying the desired statistics by minimizing a predefined loss [14,16,63–66]. A deeper analysis of this class of methods has been recently presented [12]. To improve generation quality and capture these fine details, Jiao et al. argued that higher-order spatial statistics are necessary [68]. By including just 3-point statistics in their expanded loss function, they demonstrate significant improvement in the realism of their synthetic microstructures. This approach was recently analyzed and expanded by Cheng et al. who noted that only a small subset of the 3-point statistics – those associate with small vectors ($\mathbf{r}_1$ and $\mathbf{r}_2$) – contribute to the improved realism [15]. This result is not surprising: the subset describes features of the local interfaces present in the microstructure [49]. Therefore, their inclusion reduces the likelihood of local artifacts. As an aside, these observations give weight to the assumed separation between local and global features discussed previously.

While efficient to sample, probabilistic approximation models lack an equivalent to the flexibility of the optimization objective. This makes it challenging to address these local aberrations by including higher-order statistics. As a result, previous research efforts have remained divided: focused on either incorporating 1- and 2-point statistics [12, 17,69] or successfully synthesizing realistic local features. For clarity, we will refer to approximation methods that focus on generating the large-scale patterns observable in material microstructures by incorporating 1- and 2-point statistics in the approximation as global approximation methods. In contrast, we will refer to methods that focus on generating realistic local features – implicitly incorporating local higher-order statistics – as local approximation methods.

2.1. Global approximation methods

Gaussian Random Field (GRF) models are the most common example of global approximation methods [12]. GRFs are a family of simple stochastic functions parameterized by mean functions and covariance

kernels [70,71]. Additionally, the family is defined by the characteristic that the function value associated to any finite set of points in the domain is distributed as a multivariate normal distribution (see Eq. (2)). Stationarity of the stochastic family is a common simplifying assumption; for stationary GRFs the mean function is constant within the domain, $\mu(\mathbf{x}) = \mu$ and the covariance kernel Σ depends only on the difference between spatial locations, $\Sigma(\mathbf{x}_1, \mathbf{x}_2) = \Sigma(\mathbf{x}_1 - \mathbf{x}_2)$.

$$f(\mathbf{x}) \sim \mathcal{GRF}(\mu, \Sigma(\mathbf{r} = \mathbf{x}_i - \mathbf{x}_j)) \quad (1)$$

$$\begin{bmatrix} f^{(a)}(\mathbf{x}_1) \\ \vdots \\ f^{(a)}(\mathbf{x}_S) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu^{(a)} \\ \vdots \\ \mu^{(a)} \end{bmatrix}, \begin{bmatrix} \Sigma^{(aa)}(\mathbf{x}_1 - \mathbf{x}_1) & \dots & \Sigma^{(aa)}(\mathbf{x}_S - \mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \Sigma^{(aa)}(\mathbf{x}_1 - \mathbf{x}_S) & \dots & \Sigma^{(aa)}(\mathbf{x}_S - \mathbf{x}_S) \end{bmatrix} \right) \quad (2)$$

In a recent article [12], we demonstrated an equivalence between a microstructure's anisotropic 1- and 2-point statistics and the mean and covariances of a GRF. Via this connection, we built a second-order (that is, Gaussian) approximation of any N-field stochastic microstructure function. Additionally, for periodic vector-valued microstructures, we proposed an efficient algorithm to produce samples at $O(NS \ln S)$ computational cost for N phases and S pixels. This model can efficiently generate synthetic microstructures corresponding to arbitrary, periodic, stationary 1- and 2-point statistics. Because the model avoids the necessity of training, it is widely generalizable to any observed set of 1- and 2-point statistics. However, its local features are systematically biased; the approximation of the Gaussian Random Field produces pseudo-amorphous structures.

2.2. Local approximation methods

In contrast to global approximation methods, there are numerous, seemingly disparate, approaches which can be naturally classified as local approximation methods. Perhaps the most widely adopted is Dream3D, which generates synthetic polycrystalline microstructures with specified 1-point statistics (e.g., Orientation Distribution Functions [27]) [22,46,47]. To generate individual grains, Dream3D adopts a simple parametric grain shape: an ellipsoid [46]. After sampling the ellipsoid parameters from a grain size distribution, the grains are packed spatially and then refined using a coarsening process. Finally, grain descriptors (e.g., orientations) are sampled and added. In this model, the grain size distribution, orientation distribution, and the ellipsoid model together act as a local feature distribution. Although a number of alternative models use more complex representations, they share an important characteristic with Dream3D: a local, approximate feature distribution is used to synthesize a much larger microstructure. As a result, all of these models display excellent local detail but lack global patterning, making them well-suited (and widely used) for globally homogeneous microstructures such as homogeneous polycrystals [22] and Oolitic Limestone [13].

In many practical applications, the local feature distribution is too complex to parameterize analytically. In such cases, one may instead extract a library of candidate local features from real microstructures. Cecen et al. present a clear example of this in the context of optimization based methods [72]. Using a reference segmented SEM micrograph, they extract a library of local features and use a congruent concept to Dream3D to generate new microstructures. They first sample a feature from the library, then place it to minimize pre-defined objectives using a filter-based loss function.

Markov Random Field models [73–79] build a computationally tractable approximation of the stochastic microstructure function by imposing a local Markov assumption: the state of a material point (i.e., pixel) $\mathbf{x}$ is only dictated by its direct neighbors (or a small neighborhood [76]), $N(\mathbf{x})$. Via this assumption, the approximation is constructed using a local conditional neighborhood distribution, $p(\mathbf{x}|N(\mathbf{x}))$.

To generate microstructures, Markov Random Field models raster over an initially-blank structure and repeatedly sample $p(\mathbf{x}|N(\mathbf{x}))$. At each step, for each neighborhood, they identify the set of $\mathbf{x}$ that best match the neighborhood under a handcrafted metric [75,76]. Randomly selecting from the top K best fits, they add this region to the new microstructure. Like Cecen et al.'s method, MRF methods employ a discrete library approximation of the conditional local feature distribution. As a result, MRF methods often struggle with extremely high computational cost. This occurs for two reasons. First, each sampling of the conditional distribution requires exhaustively searching the discrete library. Second, MRF methods are notoriously unstable. Therefore, they often must be restarted multiple times to successfully produce a sample [76].

Finally, deep distributional learning has emerged as a powerful class of methods for approximating distributions (for example, Generative Adversarial Networks (GAN) [80–84], Normalizing Flows [85], Stochastic Diffusion Models [86,87], Variational Autoencoders [88], Style Transfer methods [89,90]). Recently, many research efforts have employed these models as extremely expressive approximations of microstructure distributions [13,73,82,90–101]. In theory, distributional learning models could be used as a direct approximation of the entire generating process, acting as both local and global approximators. Although some efforts have sought to do so by collecting sufficiently large microstructure datasets [82,94,100,101], widespread application of this monolithic approach is limited by the data-scarcity characteristic to Materials Informatics efforts. Instead, most studies use these models as a natural extension of discrete library frameworks, where a continuous approximation of the local feature distribution is learned implicitly from a dataset composed of small patches cropped from a single (or a few) experimental micrographs [13,92,93,97,102,103]. In such frameworks, large microstructures are synthesized by either directly using the patch-trained model on a larger domain to synthesize a large microstructure⁴ [13,92,93] or by sampling and subsequently stitching together a large set of individual patches [97]. Like other local approximation methods, although such models produce extremely realistic local features, it is well documented that they are unable to generate spatial patterns whose length-scale exceeds the patch size [13, 90].

In addition to requiring large datasets, Deep Learning frameworks are incapable of extrapolating to generate microstructures that are statistically dissimilar to their training data. Therefore, they are generally limited to generating microstructures similar to microstructures one already has. It is worth noting that several attempts have been made to address this by training conditional deep learning models [92, 96,100,102,103]; however, such models are often limited by an even higher data requirement. Furthermore, even when applicable, they have only displayed the ability to interpolate to highly statistically similar structures [92]. While data remains scarce, it remains to be seen whether such models can be diversified.

2.3. Hybrid probabilistic approaches

In prior work, we presented a simple extension of the Gaussian Random Field model to address the fiber composite benchmark problem [12]. In brief, we employ the output of the Gaussian Random Field as a guide to the iterative placement of fiber features (following Cecen et al.'s approach [72]). From the perspective of probabilistic approximation, we interpret this blended generative model as a rudimentary unification of Global and Local Approximation Methods. In effect, the stochastic microstructure function is estimated with higher precision by augmenting the collected lower-order statistics with a local

feature distribution — in this case, approximated by a single circular cross-section of the fiber reinforcement.⁵ Altogether, this hybrid approximation allowed one to generate realistic fiber features while controlling 1- and 2-point statistics.

The extended GRF model can be interpreted of as a kind of two layer probabilistic model [59]. The GRF, parameterized by the 1- and 2-point statistics, acts as the first layer. The iterative feature placement, parameterized by the feature distribution and conditioned on the GRF output, acts as the second layer. This hybridization is exciting because it maintains the stability, speed, and parameterization of the Gaussian Random Field model — important for practical applications — while adding the high local feature quality of local approximation methods. However, this approach only worked because the fiber-reinforced composite exhibits extremely homogeneous local features; its local feature distribution could be approximated using a single example. For more complicated material systems encountered in most engineering applications, the feature distribution cannot be adequately represented in such a simplistic manner.

2.4. Score-based generative models

Denosing diffusion-based deep generative models [86,87,104–106] are a category of deep distributional learning models [88] that have been the source of significant revived interest due to several new theoretical improvements [86,87,107–110] as well as demonstrated success in a variety of engineering problems (such as image generation [111–113], text-image generation [114], video generation [115, 116], molecular design [117,118], waveform generation [119], microstructure generation [100,101]). During training, these models learn a function undoing an imposed noise process (often the iterative addition of Gaussian noise to training images). At sampling, generation is performed by iteratively applying the learned function to samples from a known distribution (often the unit normal).

Score-Based Denosing Generative (SBG) models [87,104,109] are a subcategory⁶ of denosing-based deep generative models that attempt to learn the Stein score of a data distribution, $\nabla_{\mathbf{x}} \log p(\mathbf{x})$, as a denosing function. Theoretically, learning the Stein-score alone is sufficient for generation. However, to overcome training instability between modes and at the fringes of the data-distribution [109,120], SBG models are usually trained and sampled via an annealing process. Each annealing level is characterized by a noise parameter, σ_i , which is used to smooth the data distribution via convolution with a symmetric Gaussian with variance σ_i^2 . As a result, the Stein score is augmented to incorporate the noise level: $\nabla_{\mathbf{x}} \log p_{\sigma}(\mathbf{x})$. Once trained, sampling is performed by iteratively updating samples while slowly decreasing the noise parameter, σ . In its simplest form, this process is referred to as Annealed Langevin Dynamics and produces the following update equation (where subscripts denote a step or iteration in the denosing process).

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \alpha s_{\theta}(\mathbf{x}_{t-1}, \sigma) + \sqrt{2\alpha} \mathbf{z}_t \quad (3)$$

$$\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (4)$$

Here, $s_{\theta}(\mathbf{x}, \sigma)$ is the learned SBG model with trainable parameters θ . After many iterations, the result is $\mathbf{x} \sim p_{\text{data}}(\mathbf{x})$. The original authors recommend performing a single Langevin step over a significant number

⁵ Of course, even this is an approximation because it ignores the short range correlations between neighboring fibers.

⁶ More rigorously, SBG models [87] and denosing diffusion models [86, 106,107] are alternate derivations of a single mathematical framework [104] for conceptualizing and constructing stochastic denosing models. Although both derivations afford similar end results (i.e., nearly identical loss functions and sampling algorithms), each paradigm has its own conceptual strengths. For example, conditional sampling is more naturally addressed in the score-based interpretation [87,113].

⁴ In general, the ability to do this depends significantly on the type of deep learning model and its architecture. For example, a fully convolutional model is absolutely necessary.

of noise levels, along with a geometrically decaying noise schedule,⁷ for optimal performance [121]; that work also presents numerous details regarding implementation and training for this class of models. At the continuous limit of infinitesimal steps in σ , Song et al. [87] demonstrated that the denoising process is equivalent to solving a Stochastic Differential Equation (SDE). This observation initiated the development of SBG models which accept continuous values of σ and are sampled using numerical SDE solvers [87]. For very complex distributions, using exotic solvers can improve sample quality and reduce sampling time [87,122,123]. Application of a simple forward Euler–Murray numerical solver [87] returns the Annealed Langevin Dynamics update equation described above.

Training [121] of denoising diffusion and SBG models is simpler than alternative approaches. They specifically avoid the competitive training (and associated instabilities [124]) required in GANs [80] and the invertible layers of Conditional Normalizing Flow models [125]. Specifically, training SBG models is performed by minimizing a single, unified loss function [87,109,121]. Amongst the proposed alternatives [126], the Denoising Score Matching loss [108] (Eq. (5)) has been widely adopted because of its low computational burden.

$$\frac{1}{2L} \sum_{i=1}^L \sigma_i^2 \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \mathbb{E}_{p_{\sigma_i}(\tilde{\mathbf{x}}|\mathbf{x})} \left[\left\| s_{\theta}(\tilde{\mathbf{x}}, \sigma_i) + \frac{\tilde{\mathbf{x}} - \mathbf{x}}{\sigma_i^2} \right\|_2^2 \right] \quad (5)$$

Here, L is the total number of noise levels. Both expectations are approximated via Monte Carlo integration. The second expectation is approximated over the σ_i perturbed distribution $p(\tilde{\mathbf{x}}|\mathbf{x}) \sim \mathcal{N}(\tilde{\mathbf{x}}; \mathbf{x}, \sigma_i^2 \mathbf{I})$.

SBG models have two characteristics of particular importance to applied engineering problems. First, they offer simple conditioning. In general, standard denoising models approximate unconditional distributions. Sampling them conditionally requires either a learned transformation to the denoising path [114–117,119,127,128] or perturbation approximations of posterior distributions [106,129]. These approaches are complicated, expensive, and can require large training sets and complex models. In contrast, the direct relationship between Stein scores and probability density functions allows one to condition SBG models directly using an additive form of Bayes Rule [87,111–113].

$$\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y}) = \nabla_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) + s_{\theta}(\mathbf{x}, \sigma) \quad (6)$$

Here, $\nabla_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x})$ is the Stein score function of the likelihood. Depending on the application, this can either be learned separately [112,113] or analytically prescribed [86,87,111]. This relationship offers more flexibility in applying a trained SBG model to problems where conditional distributions are necessary without incurring higher training demands.

Second, Song et al. [87] demonstrated that trained continuous SBG models can be used to estimate log likelihoods [87]. Although we will not employ this characteristic in this paper, we briefly note it here as an interesting potential pathway for future research.

3. Proposed framework

3.1. Decomposition of the stochastic microstructure function

In this section, we formally develop our proposed higher-order generative framework – a generalization to the extended Gaussian Random Field (GRF) method [12]. As with all probabilistic approaches, this development amounts to systematically building a tractable approximation to the stochastic microstructure function of interest. If we transition to the discrete setting by adopting a pixelized description for microstructures [12,25,26], then our goal is to approximate the

parameterized (i.e., statistically conditioned) joint distribution over the local states of each pixel.

$$p(\mathbf{m}_1, \dots, \mathbf{m}_S; \boldsymbol{\mu}, \mathbf{f}_r) \quad (7)$$

Here, $\mathbf{m}_s$ refers to the possibly vector valued microstructure state at pixel s (out of S pixels in total). We use the term “pixel” since our case studies focus on two-dimensional structures, however, the framework itself is general and (given appropriate training data) is unchanged for three-dimensional voxelized structures as well. $\boldsymbol{\mu}$ and $\mathbf{f}_r$ are the desired vector-valued mean of the local states and their discrete, vector-valued 2-point statistics function, respectively. Variables appearing after the semicolon denote parameterization of the distribution.

Our fundamental assumption in this work is that a separation exists between the roles of lower-order (i.e., 1- and 2-point statistics) and higher-order statistical information in the generation process. Specifically, lower-order information captures long range correlations such as the spatial arrangement and general shape of important features. In contrast, the impact of higher-order statistical information is localized; it describes features such as interfaces between phases. We argue that this assumption mathematically manifests as a factorized dependency structure for the joint distribution in Eq. (7). This decomposition draws on a rich body of existing mathematical frameworks for formulating and simplifying the structure of joint distributions, such as Copulas [17, 130] and Probabilistic Graphical Models [59,60]. In the following paragraphs, we develop our proposed approximation, incorporating the assumption described previously, in three informative progressions.

As an initial construction, the approximation of the complete joint distribution should capture long range correlations. Because we assume that the correlation between sufficiently separated pixels is solely dependent on 1- and 2-point statistics, the joint distribution is well described by a Gaussian approximation:

$$p^{GRF}(\mathbf{m}_1, \dots, \mathbf{m}_S; \boldsymbol{\mu}, \mathbf{f}_r) = \mathcal{N}(\mathbf{m}_1, \dots, \mathbf{m}_S; \boldsymbol{\mu}, \mathbf{f}_r) \quad (8)$$

Imposing this assumption regardless of the separation of the two pixels of interest recovers the Gaussian Random Field model [12]. However, this approximation is unrealistic if the separation between pixels is small.

As a second step, consider a set of small, spatially-compact neighborhoods in the microstructure. Mathematically, we identify the i th neighborhood as $\mathcal{X}_i$ – a set of pixel locations. Correspondingly, we define N_i as the set of local state assignments, $N_i = \{\mathbf{m}_j : j \in \mathcal{X}_i\}$. Additionally, N_i^c , the compliment of the neighborhood, is the set of local states for all pixels outside of the neighborhood. Here, the length scale of the neighborhood is small enough such that its salient features (e.g., the curvature of phase boundaries) depend heavily on higher-order, short-range microstructure statistics. In this case, the joint distribution of pixels within N_i is strongly non-Gaussian and parameterized by higher-order statistics $\Phi^{(3,\dots)}$ (e.g., 3-point statistics and up).

$$p^{neigh}(N_i) = p^{neigh}(\{\mathbf{m}_j : j \in \mathcal{X}_i\}; \Phi^{(3,\dots)}) \quad (9)$$

Furthermore, stationarity of the stochastic microstructure function requires that all neighborhoods obey the same joint distribution (i.e., $p^{neigh}(N_i) = p^{neigh}(N_j) \forall i, j$). This expression corresponds to an unconditional local feature distribution; like the GRF, it is also unrealistic. Sampling would produce neighborhoods independent of any nearby neighborhoods and without any global context. However, as we will later argue, the structure of this distribution makes it an important intermediate, well-suited for training deep learning models.

We can account for continuity between adjacent neighborhoods and global context by conditioning the local neighborhood distribution. In this case, our local features, N_i , are conditioned on both initial latent predictions (supplied by the GRF and denoted $\hat{\mathbf{m}}$ with i th neighborhood assignments $\hat{N}_i$) as well as on the values of the closest pixels just outside of the neighborhood. For notational simplicity, we overestimate

⁷ This noise schedule is referred to as variance exploding [87] and is more commonly adopted in SBG implementations. Other schedules, such as variance preserving [86,87,104], are also common and display similar performance.

these pixels using the neighborhood's complement, N_i^c . However, we strongly emphasize that this is largely redundant. Once conditioned by the GRF, the dependency between neighborhoods is near-Markovian: only the closest pixels are necessary to ensure continuity. Assuming that all other regions of the domain are fixed, we define the following conditional neighborhood distribution.

$$p^{cond}(N_i|N_i^c) = p^{neigh}(N_i|\hat{N}_i, N_i^c; \Phi^{(3,\dots)}) \mathcal{N}(\hat{N}_i; \mu, \mathbf{f}_r) \quad (10)$$

Importantly, in this construction, the Gaussian distribution continues to provide global context and direct control over the 1- and 2-point statistics. Therefore, ideally, the conditional correction should not impact the lower order statistics. Note that “|” denotes *conditioning*; that is, $\hat{N}_i$ and N_i^c are also random variables. Finally, we multiply the conditional distributions of each neighborhood to approximate the full joint distribution.

$$\begin{aligned} p(\mathbf{m}_1, \dots, \mathbf{m}_S; \mu, \mathbf{f}_r) &= \prod_{i=1}^K p^{cond}(N_i|N_i^c) \\ &= \prod_{i=1}^K p^{cond}(N_i|\hat{N}_i, N_i^c; \Phi^{(3,\dots)}) \mathcal{N}(\hat{N}_i; \mu, \mathbf{f}_r) \\ &= \mathcal{N}(\hat{\mathbf{m}}_1, \dots, \hat{\mathbf{m}}_S; \mu, \mathbf{f}_r) \prod_{i=1}^K p^{cond}(N_i|\hat{N}_i, N_i^c; \Phi^{(3,\dots)}) \end{aligned} \quad (11)$$

K is the total number of neighborhoods. Note that the neighborhoods are not necessarily disjoint sets; this factorization can be achieved by any choice of neighborhoods so long as their union is the full domain ($\bigcup_{i=1}^K \mathcal{X}_i = \Omega$). Since the GRF is jointly Gaussian over any subset of pixels, it can be pulled outside of the product. In contrast, because of the explicit conditioning, the second term remains localized to each neighborhood.

Eq. (11) is the central conceptual contribution and backbone of this paper. It supports conditional microstructure generation frameworks that meet the five requirements previously discussed. Via the GRF, it offers microstructure generation explicitly conditioned on required 1- and 2-point statistics (Requirement One). Furthermore, the conditional local neighborhood distribution incorporates realistic local features (Requirement Two). Finally, this formulation is relatively robust (Requirement Three). In the remainder of the paper, we substantiate these claims and demonstrate that the framework can be implemented computationally efficiently and in extremely data-starved regimes (Requirements Four and Five). To practically implement Eq. (11), we require expressive and tractable models for the global normal distribution and the neighborhood conditional distribution. The first is simply the Multi-Output Gaussian Random Field model [12]. An analytic expression for the second would be ideal, but is impossible (or at least infeasible) due to the extreme complexity of defining an analytic distribution incorporating higher-order microstructure statistics. Instead, we propose to use a Score-Based Denoising Generative (SBG) model to approximate the conditional neighborhood distribution.

3.2. SBG models for neighborhood distributions

SBG models offer a number of advantages for the problem at hand. In practice, learning the conditional neighborhood distribution in Eq. (10) is extremely challenging for several reasons. Primarily, we cannot acquire the necessary data triplets, $(N_i, \hat{N}_i, N_i^c)$, for training. This would require access not only to multiple complete structures, but also to “noisy” versions resulting from a Gaussian truncation. The ability to maintain the 1- and 2-point statistics imposed by the GRF poses an additional constraint on the model form. Fortunately, we can avoid requiring unacquirable training data and satisfy conditioning requirements using SBGs.

Instead of learning a conditional distribution – Eq. (10) – directly, we propose to approximate the unconditional neighborhood distribution in Eq. (9). Learning this simpler distribution is practically achievable because it only requires examples of real neighborhoods (see Section 4.1). Subsequently, we show how to perform conditional sampling in Eq. (11) – without additional training of the unconditional model – via two analytic transformations to the sampling process.

First, we transform the score function of the unconditional neighborhood prior into that of a volume-fraction-conditioned posterior using Bayes' Rule. This posterior form allows us to draw samples with minimal changes to the 1-point statistics introduced by the PGM's first layer (the GRF). For vector-valued microstructures, the likelihood term is applied to each dimension independently.

$$\nabla_{\mathbf{m}} \log p(\mathbf{m}|v_f; \theta, \sigma, \sigma_\mu) = -\frac{1}{S\sigma_\mu^2} \left(\frac{\mathbf{m}^T \mathbf{1}}{S} - v_f \right) \mathbf{1} + s_\theta(\mathbf{m}, \sigma) \quad (12)$$

Here we redefine $\mathbf{m}$ to be a vectorized representation of the microstructure over the entire spatial domain; θ are the learned model parameters, S is the number of pixels and σ is the noise level for the diffusion model. σ_μ is the standard deviation of the volume fraction likelihood, and dictates the acceptable uncertainty in the volume fraction. A derivation of the posterior score function is given in Appendix A. An ablation study demonstrating the importance of this term is given in Appendix B.

In addition to augmenting the score function, we employ a truncated noise schedule to further condition⁸ samples from the neighborhood posterior on the output of the Gaussian Random Field. This procedure was first recommended in Meng et al. [111] and has shown success for a number of problems [86]. Instead of beginning sampling with white noise and iteratively proceeding through every noise level [109], conditional sampling is achieved by using the output of the Gaussian Random Field as an initial iterate and starting the iterative denoising at an intermediate noise level $\sigma_c \in [\sigma_f, \sigma_i]$. Of course, this parameter must be tuned. Although application-specific tuning would provide the best results, for N-phase microstructure generation, we found that an intermediate noise level between 0.5 and 1.0 offers acceptable results in general (we used 0.75 in both case studies).

Together, these two transformations allow us to sample neighborhoods conditioned on the output of the Gaussian Random Field. Importantly, this procedure is designed to minimize deviations from the GRF's explicitly controlled 1- and 2-point statistics (Requirement One). Due to strong coupling between orders of n-point statistics,⁹ even slight errors in the lower-order statistics can cascade into significant errors in the higher-order statistics. The capacity to preserve lower-order statistics while modifying higher-order statistics is another major driving factor in our decision to use score-based denoising models to learn the conditional distribution.

In addition to making it feasible to practically implement the proposed framework, SBGs produce high quality and visually diverse images [87]. For our application, this is especially important for microstructures with multi-modal local neighborhood distributions (see Section 5.3). The precise reason for this high quality is presently an open question. However, a growing body of research has repeatedly observed that iterative learning methods offer better and more stable results, in exchange for slightly slower inference speeds, compared to one-step learning methods [43,131–136]. In the context of Materials Informatics and physics-informed deep learning, it is possible that

⁸ We note that a direct conditioning can be achieved using Bayes Rule [112]. The process would be similar to conditioning on the volume fraction (Appendix A). However, previous research has demonstrated that, in this case, the L_2 norm is not an effective measure for the likelihood. Instead, these efforts have used learned norms. In the interest of simplicity and conserving model parameters, we forgo this approach.

⁹ For example, for large shift distances the 2-point statistics converge to the squared volume fraction [25–27].

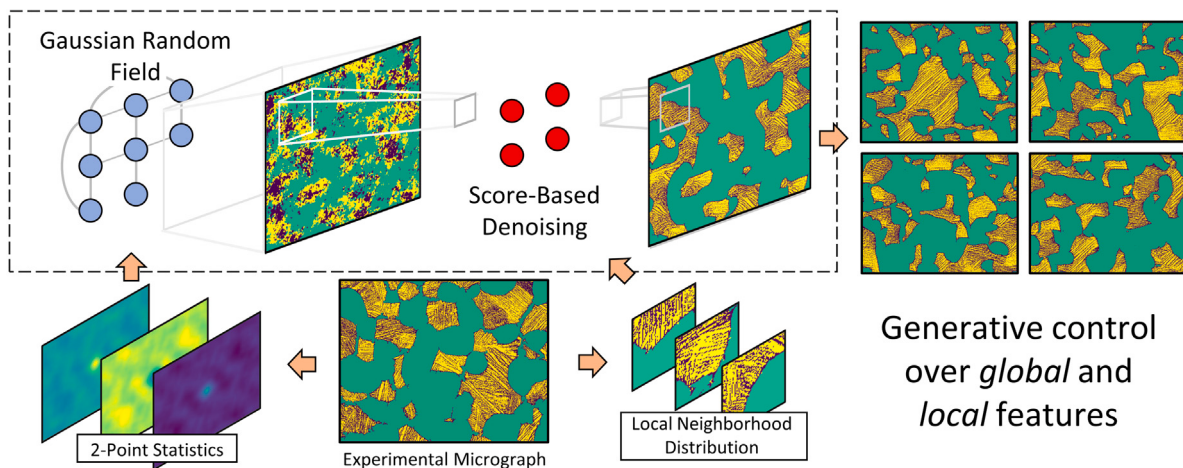


Fig. 1. Visual summary of the proposed framework. The final elements of the proposed generative framework are contained within the dashed gray rectangle.

this improvement is due to the close relationship of these schemes with classical operator splittings and decomposition methods for Partial Differential Equations [137,138] and the Lippmann–Schwinger equation [1,3,40,139].

3.3. Implementation

Altogether, training, constructing and sampling the proposed model is simple. The proposed framework is visually summarized in Fig. 1. To begin, the SBG is trained on small microstructure patches extracted from an experimental reference to approximate the unconditional neighborhood distribution. The GRF is analytically parameterized by the desired set of 1- and 2-point statistics. The statistics can, but need not, be sourced from the reference, Section 5.2. Once training is complete, sampling a synthetic microstructure is performed via two steps. First, the GRF produces an initial second-order approximation. Subsequently, a final synthetic microstructure is produced by performing iterative Langevin sampling on the entire GRF sample using the trained SBG and the conditioning transformations discussed above. Because both the computation of the 2-point statistics and the GRF's sampling algorithm assume periodicity, the initial second-order approximations are periodic. We recommend employing an embedding strategy similar to Bostanabad et al.'s periodic embedding strategy [73] to maintain this periodicity during the SBG refinement stage.¹⁰ The embedding should be reinitialized after each Langevin step to reflect any updates. Importantly, the extended region should be excluded from the volume fraction estimates. We recommend and adopted this strategy over the use of periodic convolutions in the SBG because the SBG model is trained on nonperiodic patches. More generally, it is important to note that two spatial domains are involved with the SBG model; it is trained on small, spatially-compact neighborhoods, but evaluated at sampling time on the entire spatial domain. This is achieved by using a fully-convolutional SBG network with a carefully-chosen receptive field. The impact of this dual-scale usage is further discussed in Section 4.2 and Appendix C. Finally, with respect to neighbor continuity, because the SBG model acts in a strided fashion, it implicitly satisfies local continuity between neighborhoods at each iteration. Since many conditional passes are made during denoising, there is no hard Markov assumption; instead, information diffuses between neighborhoods in a process reminiscent of message passing [59].

¹⁰ We periodically extend the microstructure in all directions by a constant number of pixels. In practice, we found that 16 pixels were sufficient to maintain continuity of local features across the boundaries.

4. Training methodology

4.1. Constructing the training dataset

While the GRF component of our framework requires no pre-existing training data, the score-based denoising model requires a dataset of real neighborhoods to estimate the score function. Adopting standard practice from previous microstructure generation frameworks [13,92,93,97], we construct a large training dataset of local patches by cutting many individual patches from our single reference image. We use the word “patch” to refer to finite regions cut from a larger microstructure; a single patch contains partial neighborhoods for many pixels.

Like all deep learning frameworks, the size and quality of the dataset is paramount [140]. Because our problem is not amenable to traditional data augmentation techniques such as contrast-shifting and rotation, the achievable size and quality of the dataset is strictly defined by the sizes of the reference image and the individual patches. This highlights one major benefit of the two level decomposition adopted in the proposed framework. Because the analytic Gaussian Random Field handles long range patterns in the generated microstructures, the necessary patch size can be small – on the length scale of the smallest features in the microstructure. This allows us to curate acceptable datasets within the strict data scarcity characteristic to microstructure generation problems. Only a single reference microstructure was used to generate training data for each case study. Importantly, the patch size must be chosen to balance several considerations: it must be sufficiently large to contain important local features, but small enough to realistically train and test with available computational resources. We rely on microstructure statistics, such as chord length distributions, to choose an appropriate patch size.

4.2. Model architecture

Recent advances in distributional deep learning are characterized by deep, exotic architectures [86,125,141]. Modern denoising-based approaches (such as score-based denoising generation [87]) are no exception, often employing extremely large models containing millions [86,87] to billions [114,127] of trainable parameters. Such exotic architectures are a direct product of the wealth of training data available to the designer. In contrast, previous materials informatics efforts have achieved better stability in the low-data regime via smaller, lightweight architectures [32,35,43,51,142]. Following suit, we developed a lightweight architecture inspired by the U-Net used in DDPM [86], Fig. 2.

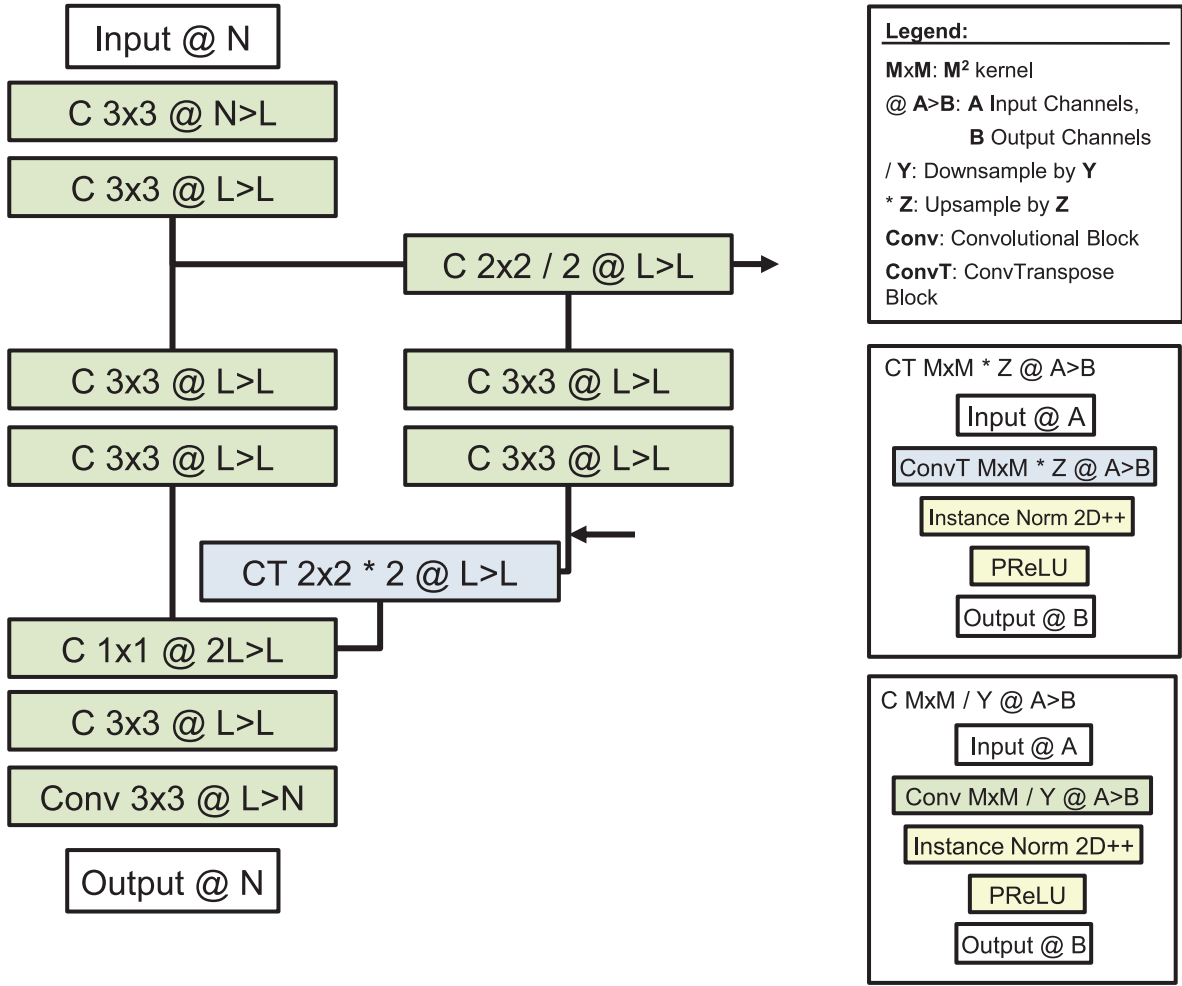


Fig. 2. Architecture for the lightweight SBG model. Only one down/up-sampling layer is shown in the depiction. Additional resolutions are added by repeating the depicted down/up-sampling unit (including the upsampling block and the 1×1 convolution necessary to maintain the latent dimension, L) at the two arrows.

The U-Net archetype combines information from various length scales through the use of up- and down-sampling and has found success in both mechanics of materials applications [43,143,144] and traditional image processing [86,87,115,116,145,146]. For systematic tuning, we compactly hyperparameterize the architecture in terms of a lift dimension and the number of down-/up-sampling layers. The lift dimension refers to the constant number of channels maintained throughout the network's internal operations. In contrast, the number of down-sampling layers dictates how many spatial resolutions the network considers, and thus its receptive field.¹¹ In practice, we found that a receptive field similar to the patch size was extremely important for maintaining stable performance (see Appendix C).

To stabilize training, we added instance normalization blocks after each convolutional layer except for the last one. Such blocks have previously been used extensively in score-based denoising models to improve their stability and generation quality [87]. Furthermore, we employ a parametric rectified linear unit (PReLU) activation function. In our experiments, other activation functions with discontinuous slopes, such as the ReLU and leaky ReLU, also performed well. Although, smooth

activations such as SWISH [148] have performed well in previous studies [86,87], we found they performed very poorly for our problem.¹² This failure might be a consequence of the sharp discontinuities in N-phase microstructures and merits further investigation.

Our proposed architecture further simplifies DDPM and its variants by removing some of the most costly layers. First, we adopt Song's recommended direct parameterization [121]: using a convolutional neural network to approximate the score function $s(\mathbf{m}) = \nabla_{\mathbf{m}} \log p(\mathbf{m})$ directly, and then scaling the output by the noise parameter σ :

$$s_{\theta}(\mathbf{m}, \sigma) = \frac{s_{\theta}(\mathbf{m})}{\sigma}$$

This approach avoids the accompanying dense, fully connected embedding layers required by Fourier embedding [86,87,141]. In experiments, including the Fourier embedding mechanism produced minimal differences in model performance, especially for 2-phase structures.

Second, we removed all attention mechanisms from the architecture. Unlike in traditional image generation applications [86,87,106,114,115], where the attention mechanism has the critical role of controlling non-local context [149] (i.e., learning large spatial patterns), in the proposed framework, non-local context management is performed

¹¹ The receptive field of a convolutional neural network refers to spatial width of the local region of output pixels impacted by a single input pixel [147]. For example, a single 3×3 convolutional filter has a receptive field of 3. When many filters are stacked sequentially, the size of the receptive field increases rapidly.

¹² For architectures with comparable numbers of learnable parameters, we were unable to even generate microstructures whose pixel values were numerically similar to the training data – i.e., were near 0 or 1 – when using smooth activation functions.

by the Gaussian Random Field model. Therefore, a spatially-compact fully-convolutional architecture is sufficient. This removal also has tremendous practical benefits. Attention mechanisms are notoriously expensive [150]. For example, the full spatial attention mechanism used in the DDPM model [86,87,149] incurs an $O(S^2)$ memory cost. Here, S can be as high as the total number of pixels. This cost would be unacceptable in our framework as the trained model is applied to the complete domain during generation. Therefore, foregoing these mechanisms offers the dual practical benefits of lowering the computational cost of a forward pass of the model and affording linear memory scaling to larger domains.

In the presented Case Studies, we find that these changes decrease sampling time (from minutes to seconds) and model size (from ≥ 50 million parameters [86] to between 100,000 and 1 million) without diminishing generative performance.

4.3. Metrics for successful training

Although denoising models boast a loss function that can be optimized via stochastic minimization, the training and hyperparameter tuning processes – especially for engineering applications – are not trivial. This difficulty has two sources: their composite loss function and the visual imperceptibility of their errors. The loss function for denoising models (Eq. (5)) is a sum of losses at each noise level, comprising 1000 terms for our model. In practice, this extreme averaging means that models with similar losses can have wildly varying generation quality. For example, error in early iterations can be overcome by the intermediate iterations, but error in the intermediate iterations is unrecoverable. For example, during our experimentation with deeper network architectures, we observed models that obtained identical composite losses as our final models, but produced pure white noise. The score-based denoising loss provides good gradients for training, but it is a poor measure for evaluating models.

Instead, we quantify model performance (both during hyperparameter tuning and on the final model) by contrasting salient microstructure statistics between the training dataset and a synthetic patch dataset generated *unconditionally* by a model post-training. This choice is aligned with the primary goal of this project: creating a generation framework that matches salient microstructure statistics. In total, we compare three important measures between the training and generated datasets: volume fraction, chord length distribution along the X-direction, and principal component (PC) projections of patch 2-point statistics.¹³ The PC basis is constructed using just the training data. Therefore, this comparison tests the SBG model's capacity to generate patches belonging to the same latent statistical space as the training data. Notably, this collection of metrics contains both lower and higher-order statistics. These metrics are solely used to quantify model performance after training and to select hyperparameters; they are not directly incorporated into the training process.

4.4. Training parameters

To train the SBG model, we used many of the hyperparameter settings recommended by Song et al. [87,120,121]. We selected the starting noise level to be slightly larger than the maximum L_2 distance between patches in the training dataset. We used a final noise level of $\sigma = 0.01$. Training was performed on 1000 noise levels between

the initial and final noise levels following a variance exploding schedule¹⁴ [87]. We used the Adam optimizer with $\beta_1 = 0.90$ and $\beta_2 = 0.999$. In general, we found that between 400,000 and 850,000 training iterations¹⁵ were necessary to train the models (we used 500,000 and 750,000 iterations in Case Studies 1 and 2, respectively). We used a cosine one-cycle learning rate schedule,¹⁶ along with exponential moving averaging with a rate of 0.999 to improve the robustness of the training process to noise.

5. Case studies in generation of N-phase microstructures

We present two case studies exploring the merits and limitations of the proposed framework. The first case study demonstrates our model's generative ability, and additionally explores the framework's capacity to approximate stochastic microstructure functions other than the reference's without the need for retraining. The second case study tests the framework's capacity to be extended to systems with larger numbers of material local states and more complex local morphologies. In both, focus is restricted to synthetically generating N-phase microstructures. Defining the probability of finding phase i in pixel s as $m_s^{(i)}$, an N-phase structure is locally binary: $m_s^{(i)} \in \{0, 1\}^N$; $\sum_i m_s^{(i)} = 1$. Our framework is not inherently restricted to N-phase microstructures. However, extensions (e.g., to polycrystalline microstructures, discrete dislocation systems, etc.) require further treatment which, in general, is not trivial. We leave this to future work.

Generating discrete N-phase microstructures requires a few simple adjustments to the proposed framework. During the iterative generation process, the Langevin dynamic update steps output unbounded, continuous variables $m_s^{(i)} \in \mathbb{R}$. In practice, the model learns to restrict these outputs to $[0, 1]$ early in the generating process and, by the end, produces outputs (within a small margin) of the set $\{0, 1\}$. At the end of the generating process, to obtain discrete microstructures, we segment the final output by naively assigning each pixel to the phase with maximum value in that pixel. In addition, rather than computing the 1-point statistics for Eq. (12) directly during generation, we compute 1-point statistics over a segmented version of the candidate microstructure. This improves the ability to condition on volume fraction by providing better estimates of the post-generation volume fraction. For microstructures with more than 2-phases, a pixel s is excluded from this estimate if the value $1 - \sum_{i=1}^N m_s^{(i)}$ is greater than the value of any individual phase, $m_s^{(i)}$ – that is, pixels with high uncertainty are masked out. However, these uncertain pixels only occur early in the iterative denoising process. Throughout our experiments we have observed that the alternative volume fraction estimate improved performance, especially for microstructures with higher numbers of phases or imbalanced volume fractions across phases. For easier tasks (e.g., 2-phase generation), a simpler, direct averaging and final segmentation – using [12] – strategy generally performed similarly.

To facilitate salient analysis of the framework's performance, in both case studies, we primarily generate synthetic microstructures that are statistically equivalent to a previously observed reference microstructure. We use experimentally collected, segmented SEM micrographs as references. We implement the framework in 2D to simplify acquisition of the experimental data as well as visualization. We emphasize that implementation in 3D is conceptually identical; it simply requires the usage of 3D convolutions in the SBG model architecture.

¹³ Here, patch 2-point statistics refers to the set of nonperiodic 2-point statistics [151] computed for each individual patch. We used a cutoff distance of half the patch size (e.g., 20 pixels for Case Study 1) to ensure sufficient sampling of the 2-point statistics. Following standard procedures [29], we extracted salient low dimensional statistics from the set of computed patch 2-point statistics using Principal Component Analysis. We kept 50 PC scores.

¹⁴ We preferred this schedule over available alternatives because it focuses training on small noise levels. As a result, the model is preferentially trained for the noise levels that are actually employed during conditional generation.

¹⁵ Mimicking previous work [86,87], here, each training iteration was taken to be a single mini-batch gradient step.

¹⁶ We used an initial rate of 0.001, a maximum rate of 0.01, and a final rate of 0.00001. Transition from the initial to the maximum rate was completed in 30% of the training steps.

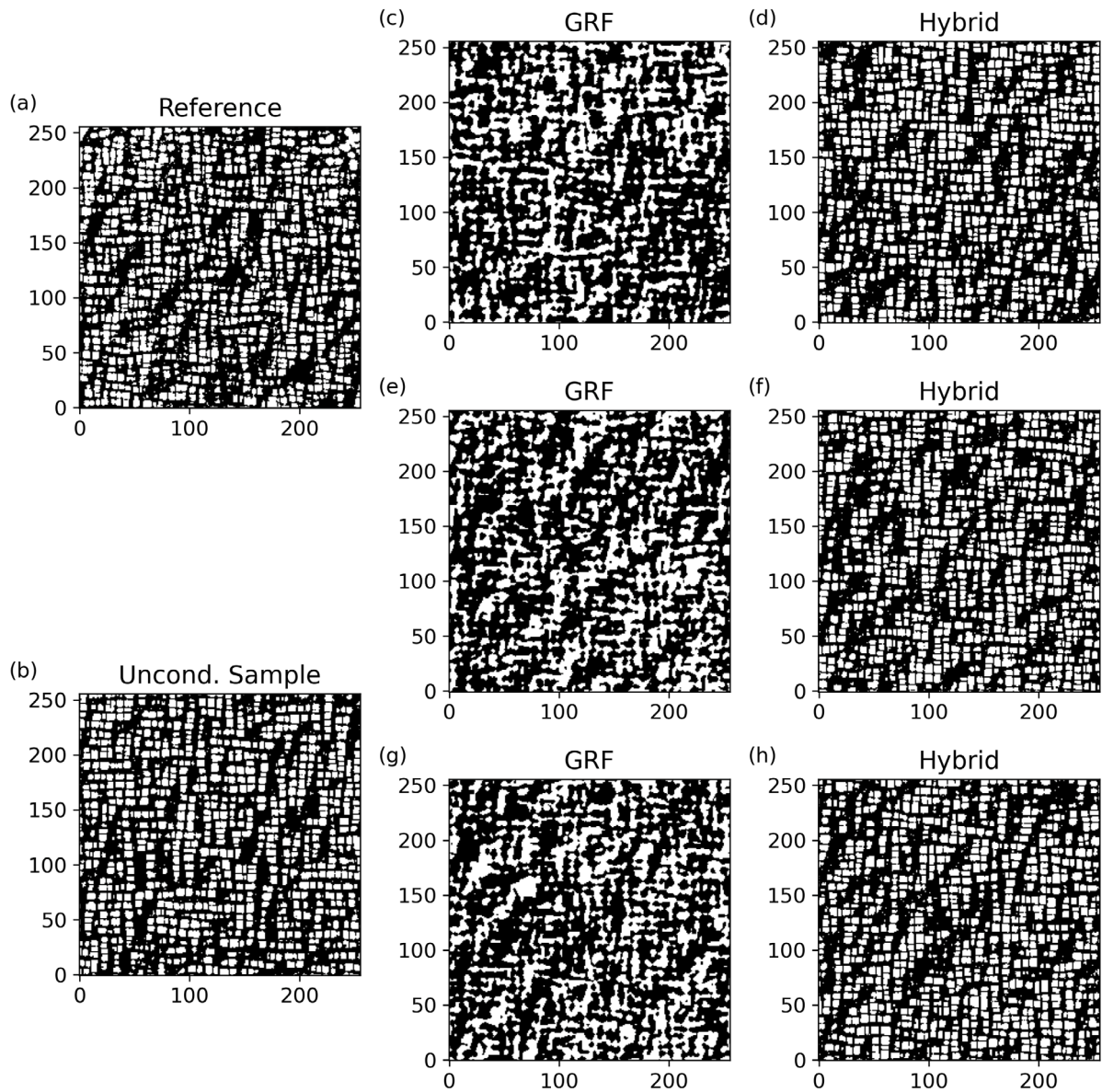


Fig. 3. Images contrasting (a) the original reference microstructure, (b) a microstructure sampled unconditionally from just the trained SBG model, and (d,f,h) samples from the hybrid bayes network model. Figures (c,e,g) display the intermediate samples from the Gaussian Random Field model taken after the first layer of the hybrid model.

However, 3D implementations might experience data challenges. In addition to the greater rarity of 3D experimental data, 3D convolutions require an additional order more trainable parameters. The framework's localized training on individual neighborhoods will help minimize this problem, but the problem's extent remains an open question. For generality, we purposefully ignore any material-specific context and simply treat the micrographs as sources of N-phase reference microstructures. While superior generative performance could likely be achieved by specializing a generative framework to the specific material system and its inherent symmetries, such a construction would serve a different purpose to the generalized framework we are proposing. We provide a discussion of architecture and dataset design in [Appendices C and D](#).

5.1. Case study 1: Statistical analysis of 2-phase generation

Our first case study uses a segmented SEM micrograph of γ' precipitates in a Ni-based superalloy microstructure [61] as the reference, shown in [Fig. 3a](#). Since nothing in our framework is specialized to this

specific material system, we simply refer to the γ' phase as the white phase and the matrix phase as the black phase. This microstructure is particularly challenging to generate because of its highly-geometric local features.

A total of 4000 patches sized 40×40 pixels were extracted from the 256×256 reference image and used to train an SBG model with a latent dimension of 40 and just one down-/up- sampling layer (totaling $\approx 100,000$ trainable parameters). We tuned hyperparameters using the microstructural statistics described in [Section 4.3](#). We initialize the shortened noise schedule (for conditional sampling) at a reduced noise level of $\sigma = 0.75$. See [Appendix C](#) for an in-depth discussion of the choice of patch size, the training process, and the quantitative identification of the reduced noise level.¹⁷

¹⁷ That section also contains a thorough discussion of the consequences that patch size has on the architecture and pertinent considerations for applying these SBG models to materials informatics topics. Practically, such discussion is especially important because it highlights several necessary deviations from the current trends in the general literature on denoising distributional learning.

5.1.1. Statistical analysis of generated microstructures

With all the requisite models trained and hyperparameters identified, we next construct the complete hybrid model and benchmark the performance of the proposed framework, using the GRF model as an informative baseline. When pertinent, we also compare the hybrid model's performance against just the SBG model sampled unconditionally.¹⁸ To facilitate statistical analysis, we generate an ensemble of 200 synthetic samples from all three approaches. While the GRF required a fraction of a second per sample, the hybrid model took 1.9 ± 0.6 seconds per sample on an Nvidia Tesla V100 GPU. However, this minor slowdown produced major improvements in visual realism and statistical fidelity. Fig. 3 contrasts the original reference (Fig. 3a) with representative samples from the proposed hybrid model¹⁹ (Fig. 3d,f,e), the intermediate outputs of the Gaussian Random Field (Fig. 3c,e,g), and a representative sample from a large unconditional sampling of the local neighborhood distribution (Fig. 3b).

Contrasting the four sets, two important observations are immediately evident. First, samples from the hybrid framework display significantly more realistic local features than those of the GRF. Specifically, the local geometric details of Nickel-based superalloys are clearly visible, and thin separation between individual features is achieved. This contrasts sharply with the amorphous patterning and persistent connectivity of the GRF samples. Second, the conditioning from the GRF improves global coherence in the hybrid samples compared to the unconditional sample from the SBG alone. For example, identically located long diagonal black phase regions are clearly visible in corresponding GRF-hybrid pairs (e.g., at $(x = 100, y = 100)$ in Fig. 3d,h). Qualitatively, this indicates that the SBG model successfully transforms local, higher-order features, while retaining the GRF's global, lower-order features, Section 3. Importantly, these global patterns appear not only in the original image, but also samples from the GRF and hybrid model; they are characteristics of the underlying generating process. In contrast, these long range spatial patterns are strikingly absent in the unconditional sample. There are still discontinuities in the regularity of the white phase features, but they do not take on a regular diagonal pattern. This supports our hypothesis that the GRF provides global context to the generative framework that would not be captured by just a local approximation model. As part of our further analysis of this hypothesis, a more dramatic example of the GRF's role in controlling long-range patterns is presented in Section 5.2. There, via the GRF, we systematically change the present long range patterns (introducing large spatial heterogeneities).

Contrasting the statistics of the reference with those of the generated ensembles, we see quantitative support for the qualitative conclusions discussed above. Both the GRF and the hybrid model approximations perform well to first and second order, matching the reference's 1- and 2-point statistics. In contrast, the unconditional SBG ('uncond.') model only matches localized subsets of the lower-order statistics. The samples from the hybrid model have a volume fraction of 0.379 ± 0.001

(mean $\pm 1\sigma$), compared to the GRF²⁰ and reference's volume fraction, 0.379 exactly. Clearly the hybrid model's conditional sampling process retains the correct 1-point statistics with high precision. Samples from the unconditional SBG model have a volume fraction of 0.377 ± 0.004 .

Fig. 4 contrasts two central slices of the 2-point statistics (specifically, the white autocorrelation). Both segments include the zero-shift 2-point statistic at $(x = 0, y = 0)$. Only 1D subsections are depicted to elucidate the minor variances between samples arising from the probabilistic nature of all approximations. The ensemble average statistics are depicted in bold. Observing the 2-point statistics, we see strong agreement between the reference, the GRF, and the hybrid model. For the unconditional SBG model, similar agreement holds only for specific subsets. Interestingly, the relative performance between the three models varies between different regions of the 2-point statistics. At small separations, there is general agreement between all four sets; we observe a strong center peak flanked by secondary peaks ($x = \pm 8.5$), Fig. 4a. While the GRF displays small inconsistencies in both sets of peaks due to implicit biases [12], the hybrid and unconditional SBG models maintain excellent agreement with the reference, Fig. 4b. This local improvement is a direct result of the localized correction introduced by the denoising model. In contrast, only the GRF and hybrid model agree with the reference at large separations. The GRF performs best; it tends to capture minuscule fluctuations in the 2-point statistics, Fig. 4d. The long range 2-point statistics of the unconditional SBG model are completely flat demonstrating that the model has a very short correlation length and that generated samples lack reliable long range patterning. This is consistent with the previous qualitative observations. The hybrid model seems to generally approximate the reference and GRF; it captures general trends and softer fluctuations in the autocorrelations.

Fig. 5 illustrates the full 2D absolute error maps in the ensemble average white autocorrelation for all three models. The SBG's patch size is clearly outlined in each figure using a red box to demarcate the subset of the 2-point statistics that are significantly impacted by the denoising model. Consistent with Fig. 4, the error in hybrid model's long range statistics (i.e., outside the red box) is slightly larger than that of the GRF. However, like the GRF error map, the hybrid model's error map is randomly spatially distributed demonstrating that the hybrid model successfully captures the dominant trends in the 2-point statistics. In contrast, the unconditional SBG model displays systemic patterns in its long range error. Primarily, diagonal, striped regions of low and high error are visible. These are the statistical fingerprints of the absent diagonal bands noted previously. Additionally, as a result of the model's short correlation length, the error displays regular vertical and horizontal patterns that further emphasize the unconditional SBG model's inability to generate consistent long range structure. Additionally, the error in the statistics within the red boxes is consistent with the previous analysis. The hybrid and unconditional SBG models display similar error patterns which are also smaller in magnitude than that of the GRF. Again, this reflects the learned localized correction. Altogether, the achieved lower-order agreement between the reference and hybrid ensemble supports the conclusion that the hybrid model meets Requirement One. Furthermore, the inability of the unconditional SBG model to match long-range lower-order statistics quantitatively demonstrates the importance of the GRF in the proposed framework. The agreement between the GRF and the hybrid model demonstrates that our design of the SBG conditional sampling scheme successfully minimizes the correction's impact to the first GRF layer's correct lower-order statistics.

The hybrid model also shows great promise in matching higher-order statistics (Requirement Two). Here, we consider only an informative subset of the infinitely-large and unwieldy set of possible statistics.

¹⁸ Unconditional SBG samples were generated *without* GRF conditioning (i.e., by utilizing the entire denoising process) but *with* volume fraction conditioning. Additionally forgoing volume fraction conditioning produced visually indistinguishable microstructures with a wider distribution of volume fractions (0.377 ± 0.004 and 0.370 ± 0.007 for with and without volume fraction conditioning, respectively. See Appendix B.). Therefore, the unconditional samples *with* volume fraction conditioning represent the optimal performance of our framework on statistically conditioned sampling if only the local generative model is used (e.g., similar to [73]).

¹⁹ To acquire the outputs from the hybrid model, the raw outputs from the Gaussian Random Field are post-processed using the original methodology [12] and then further processed using the trained SBG. Although we achieved comparable results without the original postprocessing, we, qualitatively, observed that the microstructures were less noisy when postprocessing was applied before the SBG model.

²⁰ The GRF prescribes the volume fraction exactly.

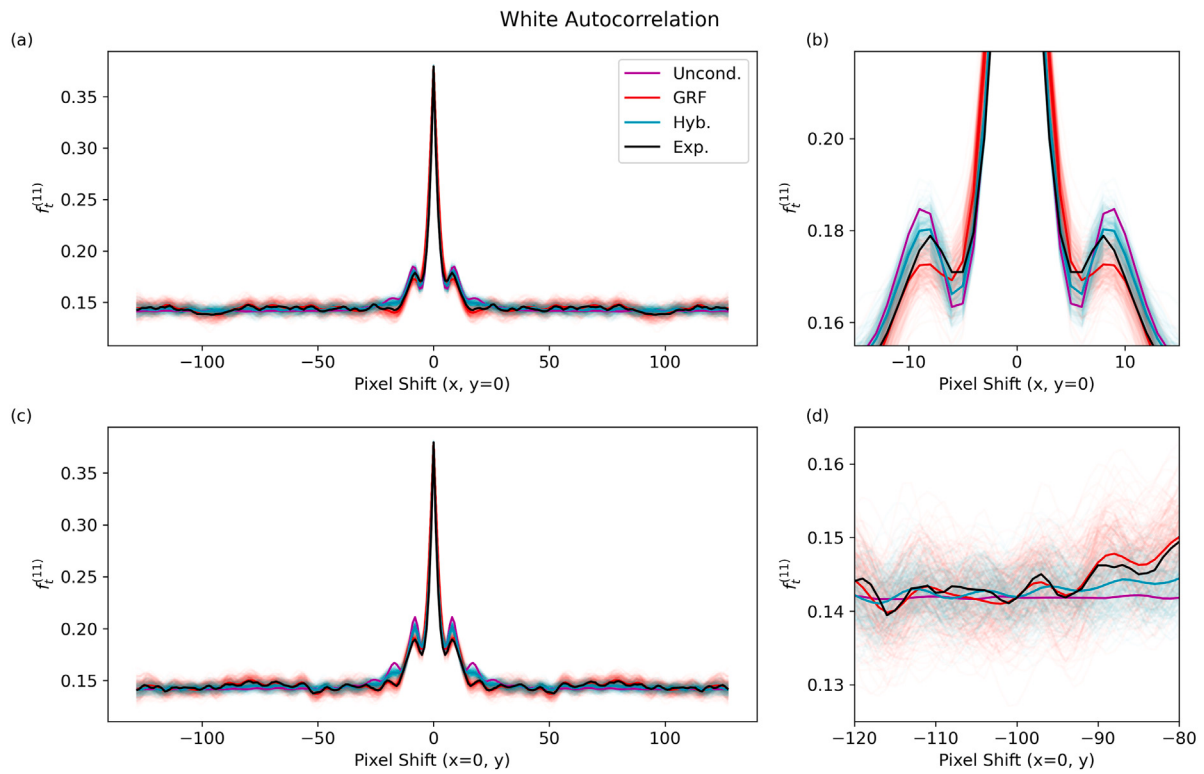


Fig. 4. Plots contrasting representative 1-Dimensional (1D) segments of the reference's white autocorrelations with those from an ensemble of 200 samples from the GRF and the hybrid approximation. The ensemble average is depicted in bold. The legend shown in (a) applies to all images. (a) A central segment of the white phase autocorrelation taken along the x-axis. The slice corresponds to 2-point statistics with zero shift in the y direction. (b) Close up of the central 2-point statistics from (a). (c) A central segment of the white phase autocorrelation taken along the y-axis corresponding to 2-point statistics with zero shift in the x direction. (d) Close up of long range 2-point statistics from (c).

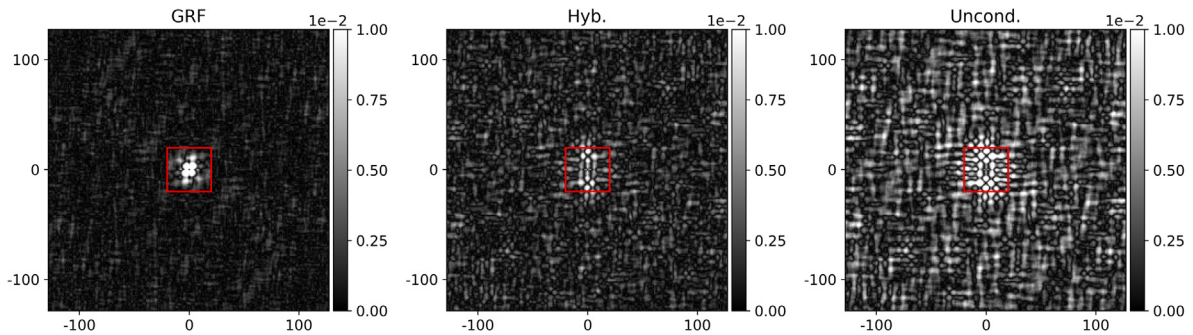


Fig. 5. Maps of the absolute value of the difference in the ensemble average white autocorrelation for the three models (GRF, Hybrid, and Unconditional SBG "Uncond") compared to the experimental reference. The central red box depicted in all three figures outlines the SBG patch size (40 × 40 pixels). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 6 visually summarizes the results. Beginning with the Chord Length Distribution (CLD) [12], we can clearly see the statistical fingerprints of many of the qualitative observations made originally. Studying the CLDs of samples from the GRF (red), we see a clear long tail extending to larger chord lengths. This is a direct result of the pseudo-connected, amorphous featuring evident in Fig. 3c,e,g. In contrast, the CLDs from the hybrid model have a sharp cutoff at higher chord lengths, consistent with the CLD of the reference. The improved agreement between the chord length distributions is a significant strength of the proposed framework.

While the hybrid model generally produces realistic features, we do observe a notable, systematic difference between its samples and the reference. In general the SBG model seems to preferentially produce less noisy features; synthetic feature phase boundaries have fewer randomly extruding pixels. In fact, these small features seem to be artifacts caused by segmentation and/or spatial undersampling and lack

appreciable patterns. Statistically, this discrepancy is most pronounced for small chord lengths. The reference's artificial small chords result in a slight relative shift between the average's of the hybrid and reference chord length distributions, Fig. 6c,d. However, the hybrid model still captures the average feature size,²¹ Fig. 6e. While a more powerful SBG network (accompanied by more training data) might produce closer statistical matching, it would also be more prone to overfitting these high-frequency artifacts.

In addition to displaying the reference's straight phase boundaries, the hybrid samples qualitatively contain its recognizable geometric feature shapes. Unlike the chord length distributions, 3-point statistics

²¹ For N-phase composites, the radially averaged slope of the main peak is the average feature size [27,152,153].

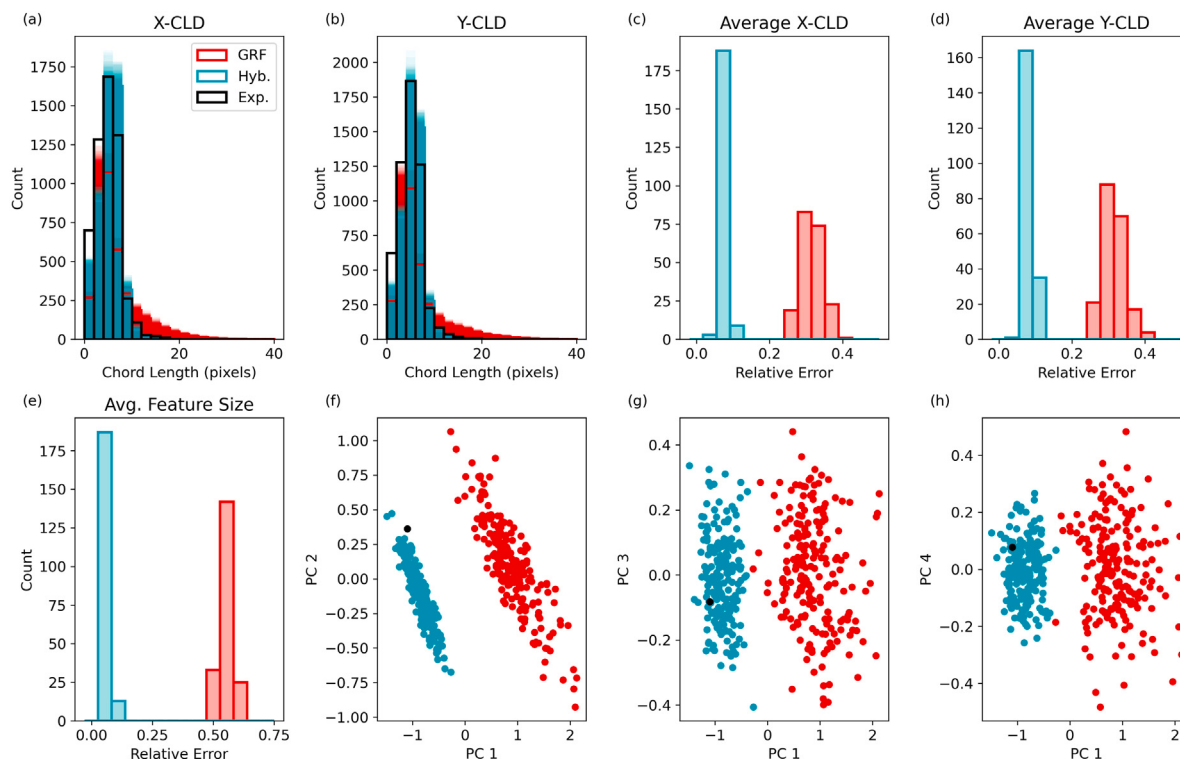


Fig. 6. Visual summary of the analysis of higher-order microstructure statistics. In each image, the statistics of the reference image (black) are contrasted against the statistics of an ensemble of 200 microstructures drawn from the GRF (red) and the proposed hybrid model (cyan). (a, b) The Chord Length Distributions in pixels computed along the X-axis (a), the Y-axis (b). (c, d) The relative error in the average X-, Y-CLD, respectively. (e) The relative error in the average feature size [152]. (f-h) Two dimensional principal component projections of a subset of the 3-point statistics. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

need not lie on a consecutive line. Therefore, these statistics also quantify characteristic nonlinear features (such as corners) in the ensemble. We restrict ourselves to the set of 3-point statistics in which the first displacement has a magnitude less than or equal to 1 pixel²² [49,154]. This specific subset was selected due to their success in previous optimization-based efforts [15], where they encourage generation of visually realistic features. Therefore, agreement in these statistics would be a significant improvement for probabilistic generative methods. Following standard procedures, we extract salient low-rank representations of the 3-point statistics using Principal Component Analysis [155] and present the results in Fig. 6f-h.

Although microstructures from the GRF, the hybrid model, and the reference were indistinguishable after the fourth principal component (PC) score, a clear separation appears in the first three PC dimensions. This reflects the unsurprising conclusion that the two models represent different approximations of a stochastic microstructure function. Notably, the 3-point statistics of the reference are clearly projected into the hybrid distribution,²³ even though the training process did not optimize for 3-point statistics directly and the PC projection was computed independently of the reference. This demonstrates that the hybrid

model is an effective higher-order approximation of the reference's Stochastic Microstructure Function.

5.2. A modular framework

Our proposed framework requires both lower-order statistics (parameterizing the first layer) and a trained denoising model (the second layer). Until now, both elements were acquired from a single source: the reference image. However, this is not necessary; we are free to mix and match elements from different sources to actively control local and global patterns. In this section, we demonstrate this capacity and highlight some of its strengths and current limitations.

We create a set of alternative input 2-point statistics for the hybrid model by interpolating between two known sets of 2-point statistics, Fig. 7a,b. To help illustrate the distinct global patterns implied by each autocorrelation, we include the corresponding microstructure used to compute them, Fig. 7c,d. Unlike the relatively uniform global arrangement of features visible in Fig. 7c and evident in the largely globally homogeneous autocorrelation map (Fig. 7a), the large high probability peaks in the second autocorrelation – at $(x = 0, y = 0)$ and $(x = \pm 90, y = 0)$ – as well as the low probability regions between them indicate the existence of global layering with large black phase regions. This layering is clearly visible in the reference microstructure, Fig. 7d. Importantly, we are only interested in the global patterns implied by the long-range 2-point statistics. Because the pretrained SBG model will impose its learned local neighborhood, the impact of any local 2-point statistics will be minimized.

To test the model's ability to extend to previously-unseen microstructure functions we create six “interpolated” autocorrelations from convex combinations of the two references. Each of these is used as an input to the pretrained hybrid model, the results of which are shown in Fig. 8. We emphasize that we did not perform any additional

²² The 3-point statistic for each initial displacement can be thought of as the 2-point crosscorrelation between the original microstructure and an expanded state produced by multiplying each pixel with one of its nearest neighbors [49].

²³ For this example, however, we note that a perfect agreement is not achieved. Beyond statistical scatter and unavoidable imperfections in the training process of deep learning models, we hypothesize that this slight mismatch represents a trade-off between maintaining the 2-point statistics and producing the learned higher-order statistics when tuning the conditioning process (Fig. C.4).

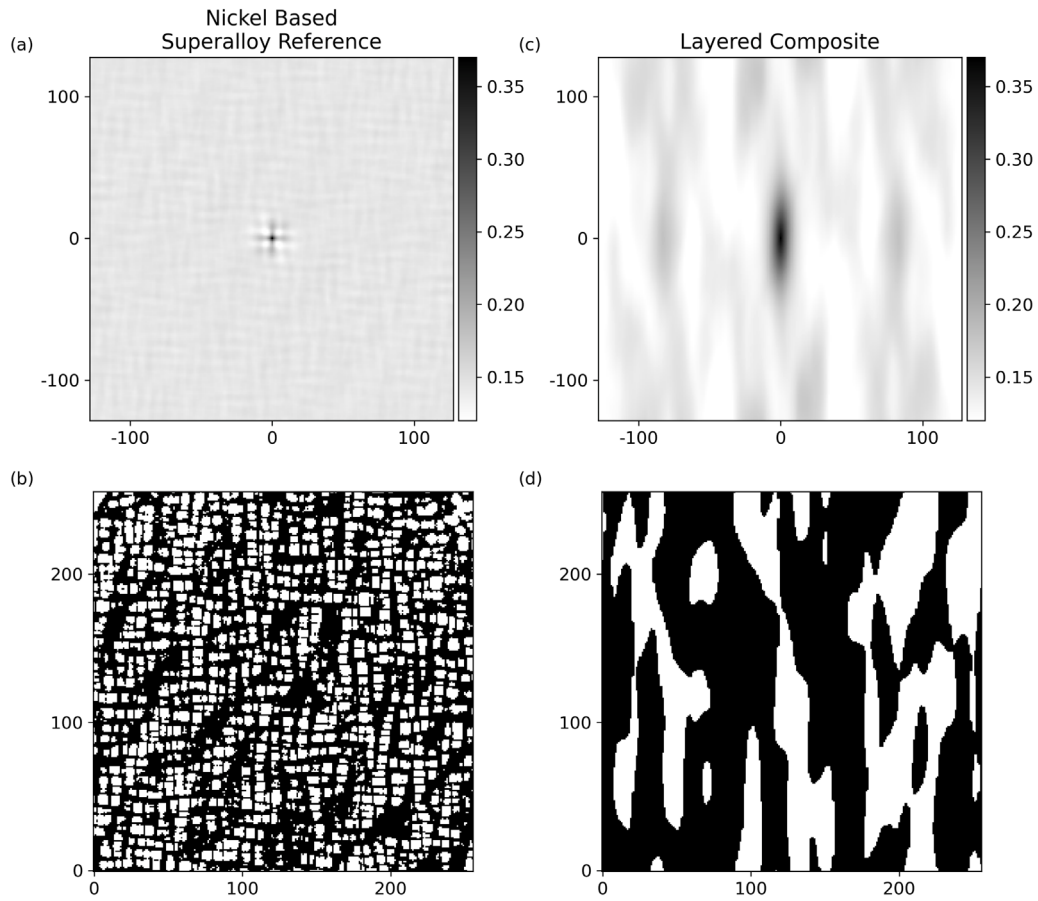


Fig. 7. Two autocorrelation maps ((a) and (b)) used to generate a set of new inputs for the hybrid model trained in Section 5.1.1. For reference, images (c) and (d) depict the microstructures used to compute the 1- and 2-point statistics in (a) and (b), respectively.

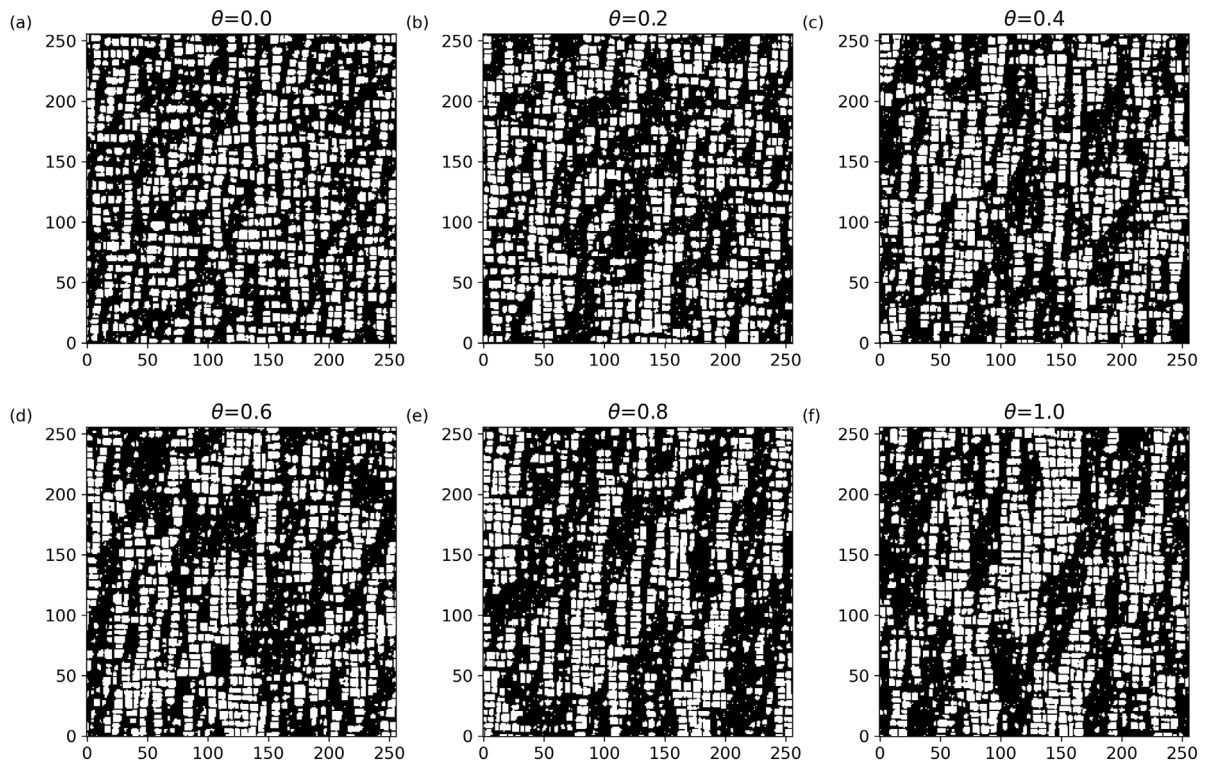


Fig. 8. Example samples synthesized using the convex combination of the autocorrelation maps depicted in Fig. 7 as an alternative input to the hybrid model from Section 5.1.1. The image correspond to a $1 - \theta$, θ mixture of Fig. 7a,b with $\theta = (0.0, 0.2, 0.4, 0.6, 0.8, 1.0)$ for (a)–(f), respectively.

training on the SBG denoising model after training on the superalloy dataset. We clearly see the impact of changing the 2-point statistics. As we begin to incorporate the statistics of the layered system, the relatively uniform patterning of the white phase features is broken up, Fig. 8b. Larger regions of high black phase density emerge but retain the diagonal structuring identified previously. Once the statistics of the layered system take over, each phase takes on vertically elongated arrangements forming the clear global layering implied by the second autocorrelation, Fig. 8f. Importantly, the local features generally retain their highly geometric structure even as the global spatial pattern changes significantly. Importantly, simply the existence of the framework's capacity to systematically translate varying inputted 2-point statistics into distinct long-range patterning emphasizes the value of the GRF-SBG hybridized framework over utilizing just the SBG. Because the output of the SBG does not depend on the 2-point statistics, we argue that this capacity can only arise as a result of the argued connection between the GRF and global patterns. We emphasize that this capacity does not require any additional training data or retraining for the new microstructures because the GRF is an analytic model.

This result is particularly exciting because it indicates that the generative model can be used to stably extrapolate and generate previously unseen microstructures. In materials design and discovery, this capacity could be especially valuable as a means of generating microstructure datasets and exploring the documented impact of global feature arrangements on properties [19,61,156] and processing behavior [24]. While the first sample, Fig. 8a, corresponds to generation within the training distribution, the remainder push successively further outside of the training domain. This stability highlights an important benefit of the framework's assumed two-layer decomposition. Just the analytic Gaussian Random Field uses the input 1- and 2-point statistics (to define global context). Therefore, when the generative framework is asked to extrapolate, the more robust GRF seems to shield the SBG model from having to extrapolate outside of its training dataset. Thus the hybrid framework is relatively robust (Requirement 3) even with the incorporated deep learning model.

While reflecting on this result, it is important to recognize that this capacity to extrapolate is not without its limitations and consequences. First, as we transition away from the 2-point statistics of the reference image used to train the SBG model, the capacity to precisely control the lower-order statistics diminishes slightly. As illustrated in Fig. 8f, we retain the capacity to match long-range patterns. However, the previously documented strong agreement for very short-range 2-point statistics is weakened. This is a natural consequence of the strong coupling between lower and higher-order statistics [14,157]. The pre-trained denoising model imposes a strong bias towards the superalloy's short-range 2-point statistics, which conflict with those of the layered composite. Therefore, synthesized structures will display the desired input statistics up to this correction (i.e., similar average behavior with superimposed minor fluctuations indicating the geometric features). Second, as we extrapolate away from the superalloy's 2-point statistics, the synthesized structures are more prone to noise. This results in situations where the Gaussian Random Field (parameterized by the extrapolated 2-point statistics) introduces neighborhoods outside of the training domain of the SBG model. For example, in Fig. 8f, we see several neighborhoods with volume fractions below those present in the training data, Fig. C.2b. In these neighborhoods, the model is more likely to introduce noise or other artifacts. However we argue that in many practical applications these are acceptable limitations and are outweighed by the usefulness of extrapolation. We simply note them here as potential inspiration for future improvements to the framework.

5.3. Case study 2: Generation of 3-phase microstructures

While the previous Case Study was restricted to 2-phase microstructures, we now challenge the generative performance of the proposed framework on a more complex reference microstructure. As before, the

goal is to generate statistically similar synthetic microstructures to a reference, in this case a segmented SEM micrograph of a Bi-Modal $\alpha-\beta$ Titanium microstructure [62]. The reference microstructure is depicted in Fig. 9a; as before we reference each phase by its color: yellow, teal, and dark blue.

With respect to the generation problem, this system has several important sources of complexity. First, it contains three phases. In theory, the incorporation of additional local states is simple: expand the number of input and output channels in the SBG network. Practically, this extension requires larger networks and complicates the conditioning process by increasing uncertainty in the local state. Second, the reference has several length scales associated with its salient features: the layered dark blue and yellow lath features are characterized by an average ($\pm 1\sigma$) chord length (in the x -direction) of 2.0 ± 1.5 and 3.1 ± 2.8 pixels, respectively. In contrast, the teal matrix phase has characteristic x -chord lengths of 48 ± 52 pixels. Third, a diverse set of local patterns are found within the yellow and dark blue composite phases; for example, there are disordered regions, regions with thin layering, and regions of thick layering. Mathematically, this third characteristic is especially challenging because it means the neighborhood distribution is multimodal — therefore, the SBG must learn and be able to synthesize from each mode.

5.3.1. Dataset, training, and architecture

Using the original 512×400 microstructure, we constructed a patch dataset composed of 8000 individual 64×64 patches. Mirroring the previous case study, the patch size had to representatively describe all salient features in each pixel's neighborhood. As noted, this includes the local featuring of the yellow and dark blue phases as well as the larger phase boundaries between this composite region and the teal background phase.

Because of the larger patch size, we employed a different hyperparameterization of the U-net to achieve a larger receptive field. A network with a latent dimension of 100 and 3 down-sampling layers (approximately $\approx 1,000,000$ trainable parameters) achieved the best results. Additionally, this SBG model takes a 3 channel microstructure as input; each phase is explicitly included. This is in stark contrast with the 2-phase case, where the second phase is defined implicitly via point-wise conservation of the volume fraction. During experimentation, we found that explicitly including each phase improved the framework's ability to control their volume fractions without requiring a larger network. For an in-depth discussion of the training results and visualization of the patches, please see Appendix D.

5.3.2. Qualitative analysis

As before we generated 200 synthetic microstructures using the GRF, and then postprocessed each with the denoiser. For the SBG, we initialized conditional sampling at a reduced initial noise level²⁴ of $\sigma = 0.75$, sampling took 16.3 ± 0.6 seconds per microstructure.

Fig. 9c,e,g depict representative samples from just the Gaussian Random Field model. Although the visual difference is immediately stark, the GRF actually successfully approximates several of the reference's salient features. In particular, it successfully synthesizes the distinct length scales. The yellow and dark blue phases are clearly intermixed with fine individual features, and the remaining teal phase forms larger, continuous regions. This demonstrates that even with an extremely challenging reference, the GRF model can successfully support its role in the proposed hybrid framework: maintaining long range context.

Perhaps unsurprisingly, the GRF struggles locally. In addition to significant noise, the unimodal GRF replaces the lath heterogeneities in the yellow-and-dark-blue regions with amorphous, pseudo-connected features [12].

²⁴ In this case, because of the extremely poor performance of the GRF model locally, we forgo segmenting the initial GRF output before denoising.

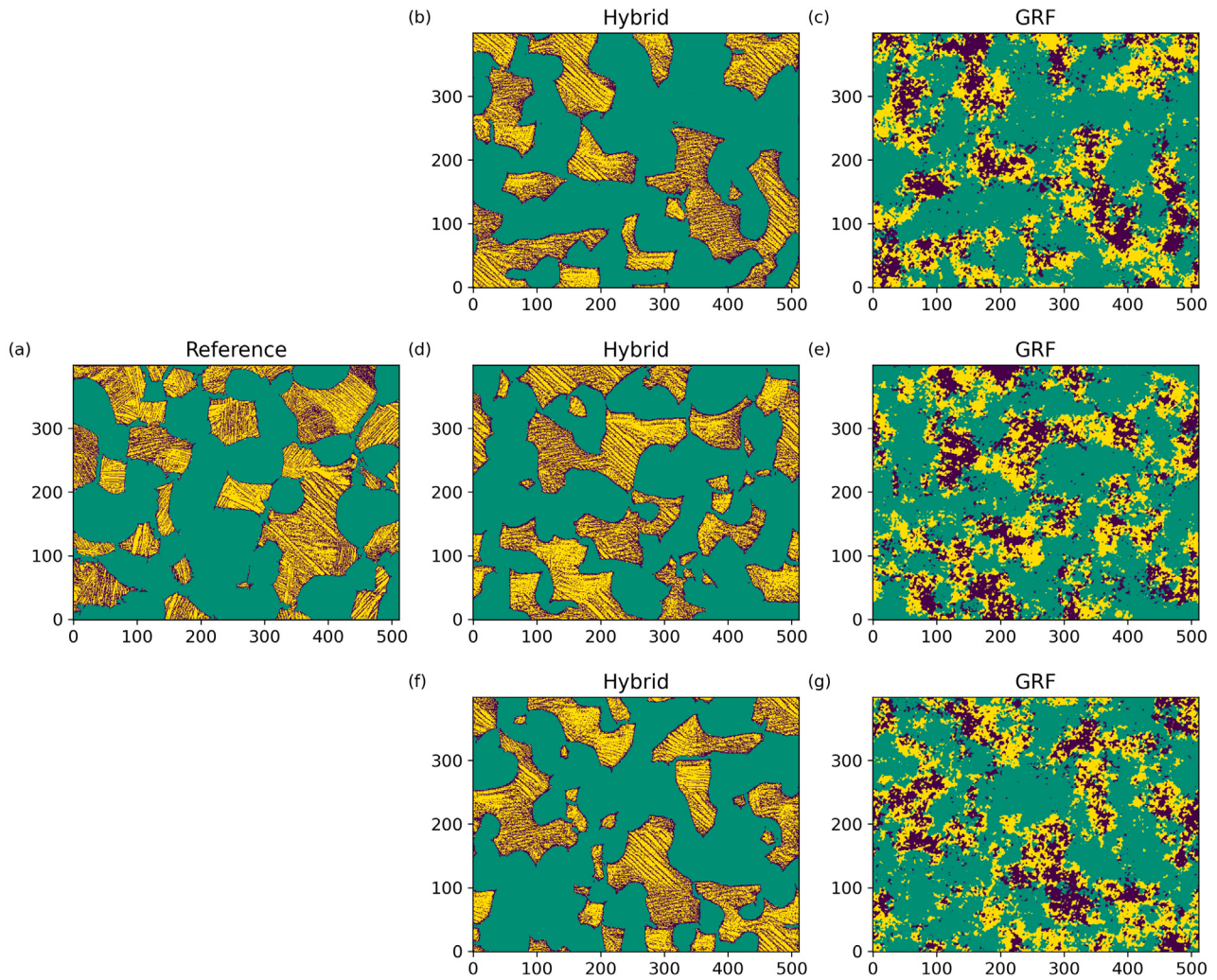


Fig. 9. Images contrasting (a) the original reference microstructure, (b,d,f) samples from the proposed hybrid model, and (c,e,g) corresponding samples from the Gaussian Random Field model created from the output of the first layer of the hybrid model.

The hybrid model produces samples with much more realistic local features while preserving global patterns present in the corresponding GRF samples. For example, regions where the GRF predicts high densities of yellow and dark blue phases retain this density in the hybrid samples. Likewise, large teal regions are retained (and the noise is largely removed). The local disorder is replaced by neighborhoods similar to those observed in the reference microstructure: small, disordered regions, thick layered regions, thin layered regions, and clear phase boundaries between the teal and composite regions. The achieved heterogeneity demonstrates the value of the SBG model. It is able to successfully learn the multimodality in the local feature distribution and consistently generate neighborhoods from each mode to produce diverse large microstructures.

Even with significantly improved visual similarity, some subtle differences remain. First, as in the first case study, we observe that the synthetic microstructures are less noisy than the reference. This is especially noticeable on the cusps of the teal phase boundary. Small noisy segments (such as the bottom-center of the reference image) are largely absent. Because this noise is largely an artifact of segmentation error, its notable absence is evidence that our model is not overfitting the training data. Second, the synthetic neighborhoods are slightly less visually diverse than the reference image. While the prominent patterns are retained (for example, thick laths), some of the small ones are lost. Importantly, this is not an artifact of insufficient model capacity — increasing the latent dimension did not alter this trend (and

eventually training became unstable). Instead, we hypothesize that this occurs because the original reference is quite noisy. The greater homogeneity could reflect the model's inability to discern reasonable patterns in a subset of training patches, leading it to effectively ignore them. This could likely be overcome by expanding the training dataset with a second image or using a less noisy reference. This issue aside, the hybrid model produces much more realistic local features while retaining the GRF's global coherence.

5.3.3. Statistical analysis

These observations are supported quantitatively by comparing the microstructure statistics. Fig. 10 depicts 1D sections from each autocorrelation (Fig. 10a-c) as well as the crosscorrelations with the yellow phase (Fig. 10d-f). The 2-point statistics of samples from the hybrid model are in excellent agreement with the reference. This is especially true in the ensemble average — the dark cyan line is almost indistinguishable from the black line in many regions. Furthermore, compared to the GRF the hybrid model actually *improves* the agreement near the origin. This effect is even more pronounced than in Case Study 1. For example, in the yellow-dark blue cross-correlation (Fig. 10f), the hybrid model is able to recreate the sharp transition from no correlation at the origin to strong correlation near $x = \pm 2$. In fact, in the central peak, the hybrid ensemble is in such strong agreement that every sample's statistics are almost indistinguishable from the reference's.

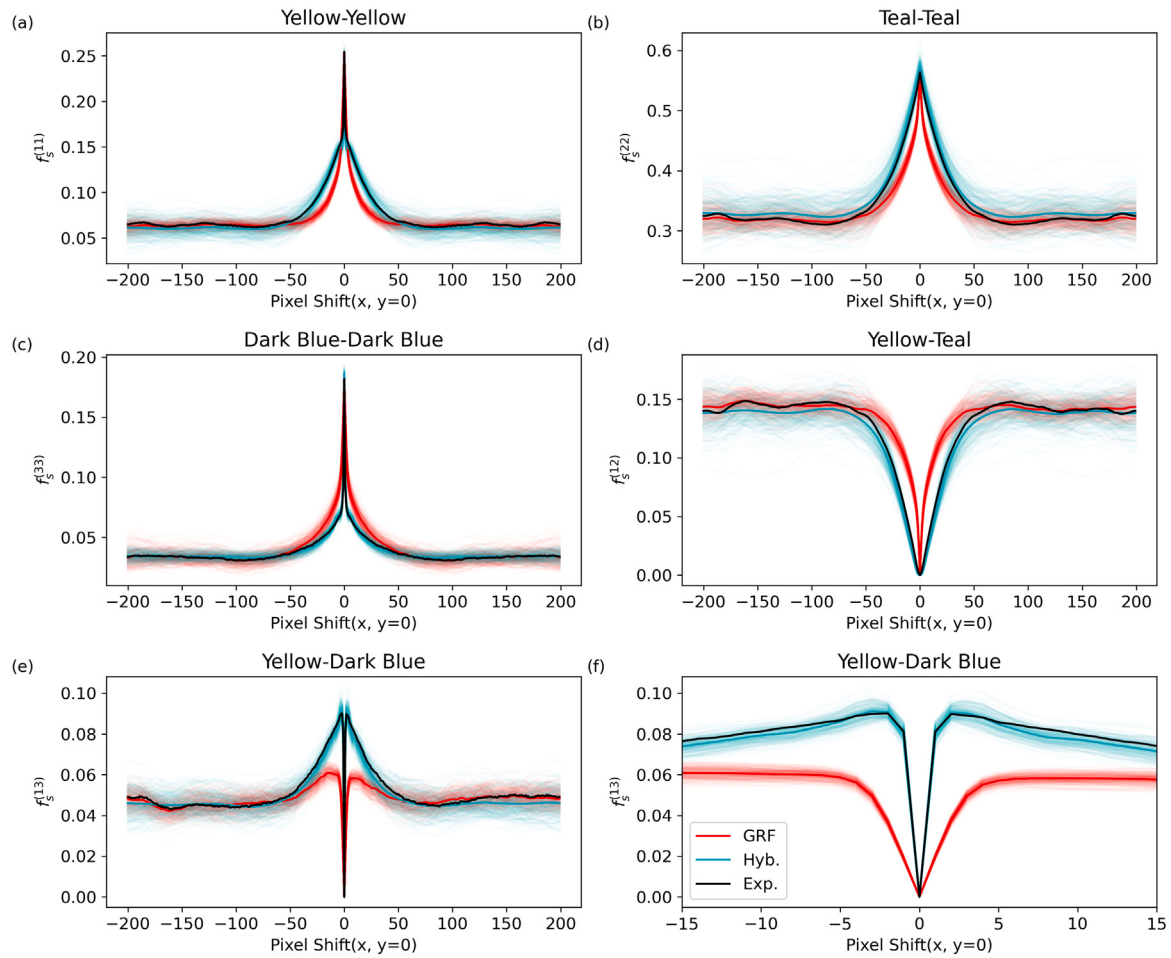


Fig. 10. Plots contrasting 1-D segments from the reference's 2-point statistics (black) with the 2-point statistics of an ensemble of samples drawn from the GRF (red) and the hybrid model (cyan). Ensemble averages are depicted in bold. Plots (a), (b), and (c) contain the yellow, teal, and dark blue autocorrelations, respectively, while (d) and (e) show the yellow-teal and yellow-dark blue crosscorrelations. Plot (f) highlights the central feature in the yellow-dark blue crosscorrelation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

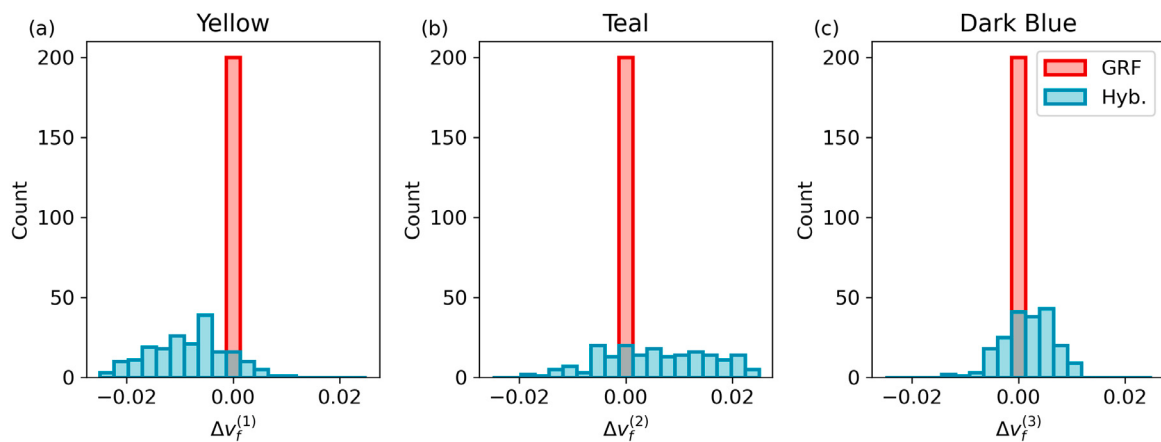


Fig. 11. Histograms summarizing the distribution of volume fractions for each phase estimated using an ensemble of samples from the GRF (red) and the hybrid model (cyan). The width of the GRF distribution is artificially added for visualization. By construction, the samples from the GRF have no variance in their volume fractions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

While the samples from the hybrid model match many salient features of the reference's 2-point statistics, the same subtle deviations appear here as in the first case study. In contrast to the GRF ensemble, the hybrid ensemble's autocorrelation displays greater variance at the

peak (for 0 shift) and for very large shifts. This occurs due to greater uncertainty in the 1-point statistics, Fig. 11, which propagates upward through statistical coupling. Unfortunately, controlling the volume fraction is more challenging for the 3-phase case. The error in volume

Table 1
X-CLD statistics for each set of structures (mean $\pm$ 1 standard deviation).

Model	Yellow	Dark blue	Teal
GRF	5.7 $\pm$ 5.3	5.3 $\pm$ 5.9	13 $\pm$ 20
Hybrid	3.0 $\pm$ 2.8	2.1 $\pm$ 1.6	53 $\pm$ 46
Reference	3.1 $\pm$ 2.8	2.0 $\pm$ 1.5	48 $\pm$ 52

fraction for each phase is -0.008 ± 0.008 , 0.007 ± 0.01 , and 0.002 ± 0.005 for the yellow, teal, and dark blue phases, respectively.²⁵ We hypothesize that volume fraction errors occur because the denoising process is defined over the space of continuous numbers instead of the natural domain of N-phase microstructure functions (the set 0 and 1). As the noise level is decreased during generation (and, as a result, the changes to the microstructure at each Langevin step become smaller) its capacity to transition a pixel from one phase to another decreases significantly.²⁶ Such artifacts could likely be mitigated by utilizing a sampling process specialized to N-phase generation (e.g., an adaptation of [73]). However, in the interest of keeping the basic framework agnostic to the choice of the local state, we do not pursue this in this work. Even so, this uncertainty is sufficiently small that the framework is able to acceptably match both the 2-point statistics and salient higher-order statistics.

We next consider the higher-order statistics. The mean ($\pm 1\sigma$) chord length for each phase across reference, GRF-, and hybrid-generation structures are presented in Table 1. These results confirm the previous visual analysis: for the finer phases (yellow and dark blue) the GRF's tendency to produce overly-smooth, stretched features results in overestimated average chord lengths. We note that the apparent mismatch in the teal average chord length for the GRF model is likely a numerical artifact caused by the presence of noise, as in Fig. 9c,e,g. The denoiser corrects these trends almost perfectly. Additionally, for the larger teal phase, the hybrid model produces a wide range of large individual features. Noting the significant standard deviation, many of these features far exceed the patch size, providing quantitative support for the GRF's value in the proposed framework. Given the multi-scale nature of this reference, the capacity to generate features at multiple length scales is an exciting strength of this framework and is a direct result of the assumed local-global decomposition.

Finally, we compare the same subset of the 3-point statistics used in Case Study 1; this time statistics are computed on each phase individually. The two distributions are indistinguishable after the first two principal components, Fig. 12. As in the 2-phase example, a clear separation exists between samples from the GRF and hybrid model for all three phases. The reference's 3-point statistics lie clearly within the hybrid distribution. Again, we emphasize that this agreement is achieved without explicitly incorporating the higher-order statistics into the training process for the SBG.

Overall, this case study demonstrates that the flexibility of the combined GRF and SBG models allow it to handle generation problems involving higher numbers of phases, complex local morphologies, and multi-scale spatial patterns. The assumed local-global decomposition as well as the careful construction of the secondary layer facilitate strong agreement in both lower- and higher-order statistics while also producing visually realistic microstructures. This case study also highlighted two characteristic challenges when extending to larger numbers of phases. First, the extension required a significant increase in network size (from 100,000 parameters in the first case study to 1,000,000) in

order to capture the complexity of the local neighborhood distribution. Second, increasing the number of states was accompanied by a slight increase in the variance of the synthetic microstructure's microstructure statistics (especially the volume fraction). However, we emphasize that the hybrid model still achieves excellent agreement in second- and third-order statistics. Future work will test whether similar uncertainty is observed for increasing N and for continuous local states.

6. Conclusions

In this work, we present a theoretical and computational framework for conditional approximations to the generating processes of material microstructures (i.e., their stochastic microstructure functions). In addition, we present the necessary algorithms to implement the proposed framework as well as document possible challenges that arise when training a hybrid model. Our hybrid model displays several important strengths compared with previous generative models. First, we are able to efficiently and systematically generate microstructures with realistic local features while retaining explicit control over their 1- and 2-point statistics. Second, the hierarchical framework allows us to extrapolate, with limited error, and generate microstructures outside of the model's training data. Third, the necessary learning models can be trained using just a single training image, opening up a number of data-scarce applications. Fourth, the generated microstructures are periodic even if the reference is not. Therefore, the synthetically generated microstructures can be utilized in efficient numerical simulations with periodic boundary conditions by future efforts. Finally, generation is computationally efficient. Although slightly slower than our Gaussian Random Field framework [12], the two-layer method outpaces other approaches by more than an order of magnitude [14,66,76].

The primary theoretical contribution of this paper is the proposition of a novel probabilistic decomposition approximating a stochastic microstructure function. Specifically, we propose a two layer semi-directed Probabilistic Graphical Model.

$$p(\mathbf{m}_1, \dots, \mathbf{m}_S; \boldsymbol{\mu}, \mathbf{f}_r) = \mathcal{N}(\hat{\mathbf{m}}_1, \dots, \hat{\mathbf{m}}_S; \boldsymbol{\mu}, \mathbf{f}_r) \prod_{i=1}^K p^{\text{cond}}(N_i | \hat{N}_i, N_i^c; \Phi^{(3, \dots)})$$

We argue that this two layer construction is a direct mathematical realization of the global-local assumption presented in our previous work [12]. An N-output Gaussian Random Field acts as a fully-correlated first layer, introducing global correlations contained in the 1- and 2-point statistics. The second layer, which assumes conditional dependence only within compact spatial regions, incorporates higher-order statistical information to produce realistic local neighborhoods. We argue that this second layer is best approximated using a Score-Based Denoising Deep Generative learning model. In addition to promising greater stability than popular alternatives (i.e., distribution-to-distribution GANs [81,82]), these models can be directly conditioned using Bayes Rule. We use this technique to softly constrain sampling from the neighborhood distribution in order to maintain the 1- and 2-point statistics output by the Gaussian Random Field. In total, we argue that sampling can be performed using the following expression.

$$\nabla_{\mathbf{m}} \log p(\mathbf{m} | \mathbf{v}_f; \theta, \sigma, \sigma_{\mu}) = -\frac{1}{S\sigma_{\mu}^2} \left(\frac{\mathbf{m}^T \mathbf{1}}{S} - \mathbf{v}_f \right) \mathbf{1} + \frac{s_{\theta}(\mathbf{m})}{\sigma}$$

To support our analysis, we presented two case studies demonstrating the practical implementation of the proposed framework for generating 2-phase and 3-phase microstructures. In both case studies, we compared against experimental micrographs. In addition, we explored the stability of the model by reapplying it – without re-training – to approximate stochastic microstructure functions with different lower-order statistics but identical local features. In doing so, we demonstrated a gentle form of extrapolation to stochastic microstructure functions outside of the training set. This ability stems from the bi-level directed nature of our framework. The learned second

²⁵ The reference volume fraction of each phase is 0.254, 0.564, and 0.189.

²⁶ In the 2-phase case, this problem is minimized because the transition is clear: white to black. However, in the higher phase case, the ambiguity of which phase to transition to seems to incorporate uncertainty into the conditioning process. This could potentially pose an issue for high-entropy systems with a large number of phases.

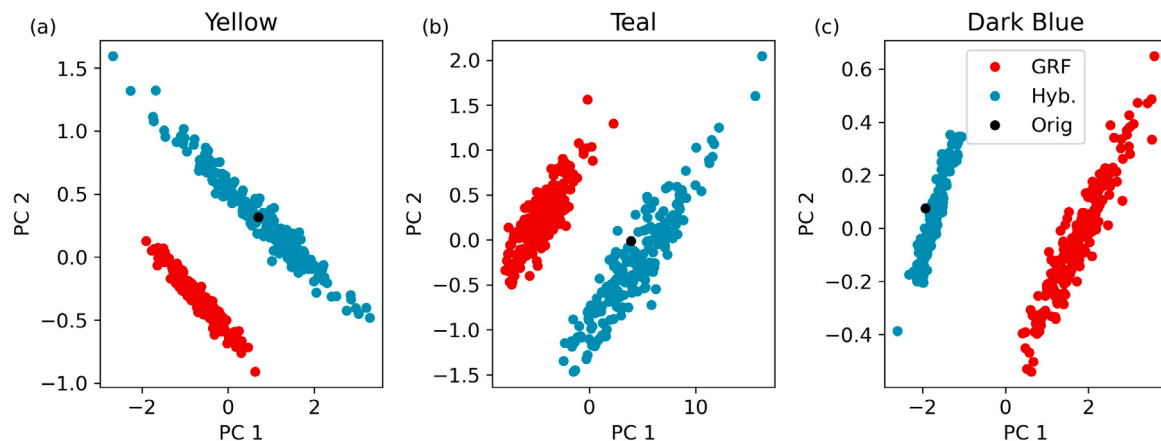


Fig. 12. Principal Component projections of a subset of the 3-point statistics computed for each phase: yellow, teal, and dark blue for (a), (b) and (c), respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

layer is shielded by the more robust (but less flexible) GRF layer, allowing the entire framework to extrapolate without significantly reducing generative performance.

This study lays the groundwork for a number of future explorations. First, from a modeling perspective, incorporating higher numbers of local states proves challenging. In this paper, we explored both 2-phase and 3-phase generation; however, even this simple transition led to a significant increase in the required capacity of the score-based denoising model. To facilitate extension to increasingly complex local states, more research is needed to identify methods of training these models without incurring parameter explosion (e.g., the curse of dimensionality). Second, the SBG model's generalization (i.e., ability to learn distributions over arbitrary local states) leaves space for improvement via application specific specialization of the local model. For example, for N-phase generation, models specialized for categorical random variables [73] could address the documented scatter in generated volume fractions. Third, we restricted the focus to 2D generation because of the greater accessibility of 2D data. Although the 3D case is theoretically equivalent, we noted that its implementation could be practically difficult primarily because of data scarcity. For 3D generation, recent research has proposed techniques for training 3D generative models on orthogonal 2D data slices in order to minimize these challenges [98]. Incorporation of these ideas into this framework would simplify implementation in 3D. Fourth, in this paper we ignore any material context, focusing instead on statistically-conditioned generation. This decision stemmed from the fundamental relationship between microstructure statistics and relevant material properties [1,3–6,57] and produced a generalized conditional generation framework. However, we acknowledge that better generation quality can likely be achieved on a system-to-system basis by incorporating material-specific symmetries and physics. For example, the titanium structures considered in this paper have highly-constrained lath arrangements. Further research into systematically incorporating “physics-driven information” into the generative process is an interesting – yet challenging – open question. Fifth, from a practical use perspective, continued stability analysis during extrapolation is necessary. Ideally one would have a model which can accept changing 2-point statistics without impacting the higher-order statistics.

Code availability

The implementation of the hybrid model, the trained diffusion models, and the training framework described in this paper are freely available under an open-source license at <https://github.com/arobertson38/LocalGlobalDecompositions>.

CRediT authorship contribution statement

Andreas E. Robertson: Conceptualization, Methodology, Software, Data curation, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Conlain Kelly:** Conceptualization, Methodology, Writing – review & editing. **Michael Buzzy:** Software, Writing – review & editing. **Surya R. Kalidindi:** Supervision, Funding acquisition, Writing – original draft, Writing – review & editing.

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

A.E. Robertson would like to acknowledge Dr. Laurent Capolungo and Dr. Aritra Chakraborty for their valuable discussions as well as Ravyn Malatesta for her support throughout the process. In addition, A.E. Robertson thanks the Jack Kent Cooke Foundation for their continued support. A.E. Robertson, C. Kelly, and S.R. Kalidindi acknowledge support from Vannevar Bush Faculty Fellowship program with the grant ONR N00014-18-1-2879. M. Buzzy acknowledges support from the National Science Foundation under the grant NSF DMREF 2119640. C. Kelly acknowledges support from the National Science Foundation under NSF 2027105 and NSF GRFP DGE-1650044. This research was supported in part through research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, Georgia, USA.

Appendix A. Derivation of posterior score function

This appendix contains the derivation of the volume-fraction-conditioned posterior score function which enforces a desired global volume fraction. We begin by expanding the abstract definition of the posterior distribution using Bayes rule. For simplicity, we consider microstructures with scalar local states; $\mathbf{m}$ is a vectorized representation of the entire spatial microstructure.

$$\nabla_{\mathbf{m}} \log p(\mathbf{m}|v_f) = \nabla_{\mathbf{m}} \log p(v_f|\mathbf{m}) + \nabla_{\mathbf{m}} \log p(\mathbf{m}) - \nabla_{\mathbf{m}} \log Q \quad (\text{A.1})$$

Here, Q is a normalization term which is assumed to remain constant across all volume fractions and microstructures. Because the gradient of a constant is zero, this term can be ignored. The second

term, $\nabla_{\mathbf{m}} \log p(\mathbf{m})$, is the score function of the prior. This term is replaced by its learned representation. Finally, the distribution of the likelihood is estimated using a Gaussian centered at the volume fraction of the current microstructure.

$$\nabla_{\mathbf{m}} \log p(\mathbf{m}|v_f) = \nabla_{\mathbf{m}} \log \mathcal{N}\left(v_f; \frac{\mathbf{1}^T \mathbf{m}}{S}, \sigma_{\mu}^2\right) + \frac{s_{\theta}(\mathbf{m})}{\sigma} \quad (\text{A.2})$$

Here, $\mathbf{1}$ is the vector of all ones. Therefore, $\frac{\mathbf{1}^T \mathbf{m}}{S}$ is the volume fraction of the current microstructure.

$$\nabla_{\mathbf{m}} \log \mathcal{N}\left(v_f; \frac{\mathbf{1}^T \mathbf{m}}{S}, \sigma_{\mu}^2\right) = \nabla_{\mathbf{m}} \log \exp\left[-0.5 \frac{\left(v_f - \frac{\mathbf{1}^T \mathbf{m}}{S}\right)^2}{\sigma_{\mu}^2}\right] \quad (\text{A.3})$$

$$= \frac{1}{S\sigma_{\mu}^2} \left(v_f - \frac{\mathbf{1}^T \mathbf{m}}{S}\right) \mathbf{1} \quad (\text{A.4})$$

Altogether, the augmented score function is given as:

$$\nabla_{\mathbf{m}} \log p(\mathbf{m}|v_f) = -\frac{1}{S\sigma_{\mu}^2} \left(\frac{\mathbf{1}^T \mathbf{m}}{S} - v_f\right) \mathbf{1} + \frac{s_{\theta}(\mathbf{m})}{\sigma} \quad (\text{A.5})$$

Because the score of the likelihood is prescribed analytically, in theory, it does not require an annealing schedule for stable sampling. However, in practice, we found a mild annealing schedule improved numerical stability. With the same variance exploding schedule archetype, we achieved the best results with $\sigma_{\mu,i} = (1.526e - 7)S$ and $\sigma_{\mu,f} = (1.526e - 9)S$, using S to account for differences in the total number of pixels as the domain size changes.

For vector valued local states, conditioning is achieved by applying the likelihood expression to each dimension individually.

Appendix B. Ablation study: Volume fraction conditioning

In Section 3, we propose and derive an analytic transformation to the learned score function to condition the denoising model's sampling process on a desired 1-point statistic (see Eq. (12) and Appendix A). Without this transformation, the denoising process synthesizes microstructures over a wide distribution of 1-point statistics. Such spread is observed both when sampling the SBG alone and in the proposed hybrid framework (i.e., when conditioning the SBG on the output of the GRF as described in Section 3.2). In the context of the generation problem described in Case Study 1, we compare the volume fractions of 200 synthetic microstructures generated using four different generating strategies to demonstrate the importance of the proposed custom sampling rule. We use the same domain size and sampling hyperparameters utilized in Case Study 1 unless explicitly noted.

1. **SBG**: Samples are drawn using the standard SBG score function and utilizing the entire denoising process (i.e., starting at $\sigma = 20.0$).
2. **SBG + V_f** : Samples are drawn using the custom SBG sampling rule, Eq. (12), and the entire denoising process. We slightly adjusted the volume fraction noise schedule to $\sigma_{\mu,i} = 0.02$, $\sigma_{\mu,f} = 0.0003$ to stably generate microstructures.
3. **SBG + GRF**: Samples are drawn using the standard SBG score function and GRF conditioning (i.e., starting the denoising process at $\sigma = 0.75$ and initializing with a GRF sample.).
4. **Hyb: SBG + GRF + V_f** : *The Proposed Framework*. Samples are drawn using the custom SBG sampling rule and GRF conditioning.

Fig. B.1 depicts the distributions of the volume fractions of the synthetic microstructures generated using each strategy. Case Study 1's reference (i.e., target) volume fraction, 0.379, is demarcated with a black line. Contrasting Strategies 1 and 2 as well as Strategies 3 and 4, the impact of the custom sampling rule's inclusion is clear. In both cases, samples generated with the custom sampling rule display much closer agreement with the target volume fraction. The difference is

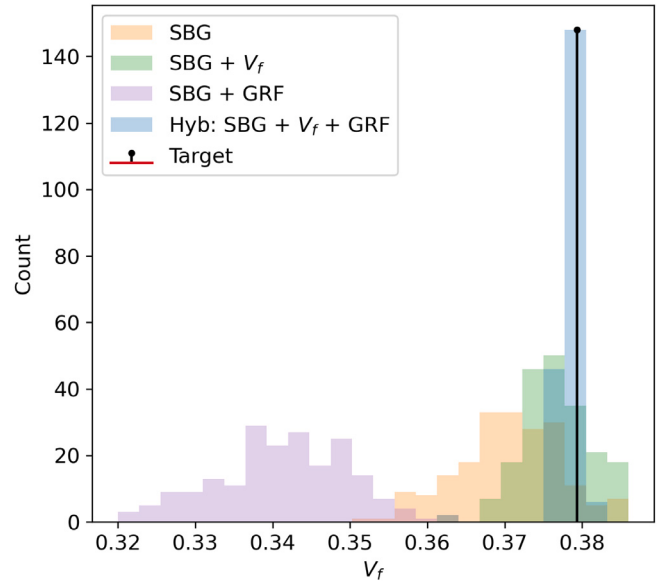


Fig. B.1. Histogram contrasting the distributions of volume fractions of synthetic microstructures generated using four different generation strategies. The target volume fraction (0.379) corresponds to the reference microstructure in Case Study 1.

particularly pronounced for the hybrid frameworks (Strategies 3 and 4). In this case, the correction adjusts the volume fraction from 0.341 ± 0.009 (mean $\pm 1\sigma$) to 0.379 ± 0.001 , respectively (compared to 0.370 ± 0.007 and 0.377 ± 0.004 for Strategies 1 and 2, respectively). Interestingly, comparing Strategies 2 and 4, the GRF and the volume fraction conditioning together seem to further improve the generating strategy's capacity to match the 1-point statistics, 0.377 ± 0.004 to 0.379 ± 0.001 . This deviation occurs because, without the GRF, generative instability forces us to use a more diffuse volume fraction likelihood (i.e., to utilize a higher final volume fraction noise level). We hypothesize that the GRF helps to numerically stabilize generation because it initializes sampling with a microstructure with the correct 1-point statistics.

Appendix C. Case study 1: Training

The following section outlines important steps in designing and training the SBG model used in Case Study 1. We emphasize that many of these observations – regarding the design of these models to ensure good statistical performance during application – are quite general and apply to Case Study 2 as well.

C.1. Generating the dataset

For this case study, we used a patch size of 40×40 pixels (the original SEM image is 256×256 pixels). From the Chord Length Distribution, Fig. 6a,b, the largest prevalent individual feature size is 8 pixels.²⁷ Therefore, a 40×40 patch size is sufficiently large to contain most individual features as well as their first and segments of their second nearest neighbors. We cut a training dataset of 4000 patches from the original image (e.g., Fig. C.1 - bottom row.). The patch centers were selected using a Latin hyper cube sampling [158] of the reference domain to encourage representative coverage of the reference microstructure.

²⁷ We used the largest prevalent size instead of the average because of the sharp cutoff in feature size observable both visually and in the chord length distributions.

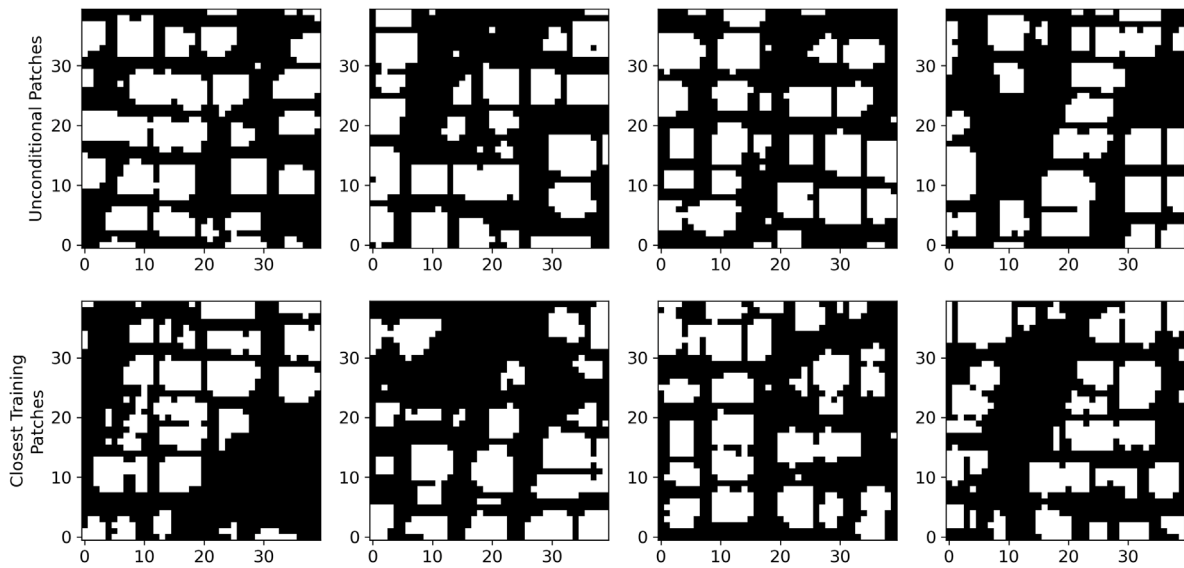


Fig. C.1. Images contrasting randomly selected unconditional samples from the trained SBG model (row 1) with their closest image in the training dataset (row 2). Closeness is measured using the L_2 -norm.

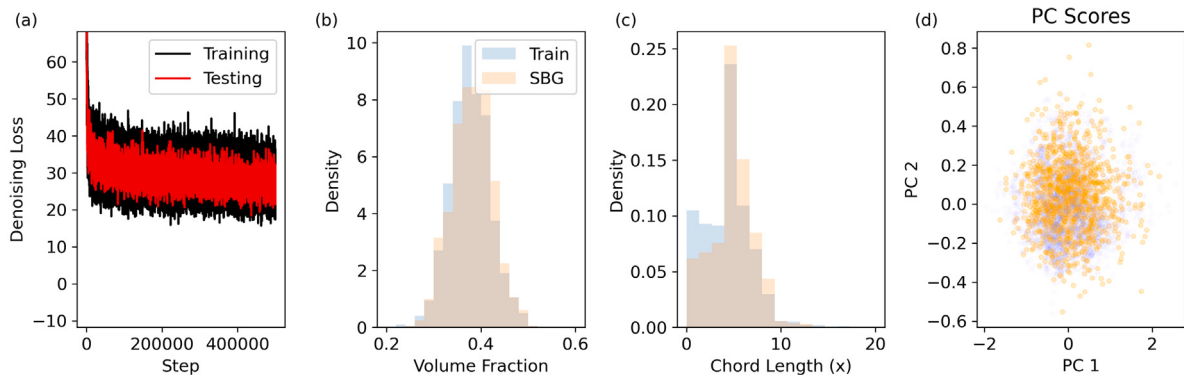


Fig. C.2. Images summarizing the training process for the final score based denoising model used in Case Study 1. (a) Plot depicting the denoising score matching loss computed during training on batches from the train and testing sets. 1000 sample patches were drawn from the trained SBG model and compared against the original training data to produce images (b), (c), and (d). (b) White phase volume fractions. (c) White phase chord length distribution. (d) PC projection of the nonperiodic patch 2-point statistics.

C.2. Training

Fig. C.2 summarizes the results of the training process of the final model. As shown in Fig. C.2a, the model shows similar performance on the training and test data (we used a 85 – 15% training–testing split), demonstrating the stability of the model. Fig. C.2b–d summarize the statistical agreement between the training data and synthetic patches generated *unconditionally* by the trained model. As demonstrated by the strong agreement in the statistics, the SBG model successfully learns to generate realistic patches. Importantly, the statistical agreement holds for all three orders of statistics. When deployed, this agreement means that the model should be able to successfully correct the higher-order statistics without excessively perturbing the lower-order ones. The patches in Fig. C.1 further affirm this conclusion. Fig. C.1 contrast four randomly selected synthetic patches (the first row) with the closest, in the L_2 -norm, patches to each from the training dataset (the second row). The synthetic patches share many qualitative characteristics with the original training data: geometric individual features, fine separation between the features, and semi-regular spatial arrangement between the features and their nearest neighbors. Moreover, the network does not seem to be simply recreating the training dataset.

During architecture and hyperparameter tuning, we observed that the statistical generation quality – referring to the agreement between

the microstructure statistics of the *unconditionally* generated patches and the microstructure statistics of the experimental patches (i.e., the training dataset) – decreased rapidly with growing architecture size (i.e., the number of trainable parameters). This behavior contrasted sharply with the visual generation quality – referring to the qualitative visual agreement between the two datasets – which, generally, improved as the architecture size increased.²⁸ The proclivity of neural networks to overfit is well documented in the deep learning community [159]. However, we believe that it remains an especially important observation in this context because our solution (using smaller, lightweight architectures) goes against the general trend of extremely large architectures in diffusion based distributional learning [86,87, 114,115,127]. We hypothesize two justifications for this divergence. First, the hybrid two-level construction (i.e., the GRF construction followed by the denoising step) significantly simplifies the complexity of the distribution that the denoising model is tasked to learn, allowing

²⁸ In addition to statistical errors, large architectures often also introduced training instability. When present, repeatedly restarting the training process would lead to a fraction of the models generating white noise instead of discernible patches. With sufficiently lightweight architectures, we never observe this instability. However, there are still minor fluctuations in generated microstructure statistics between training runs.

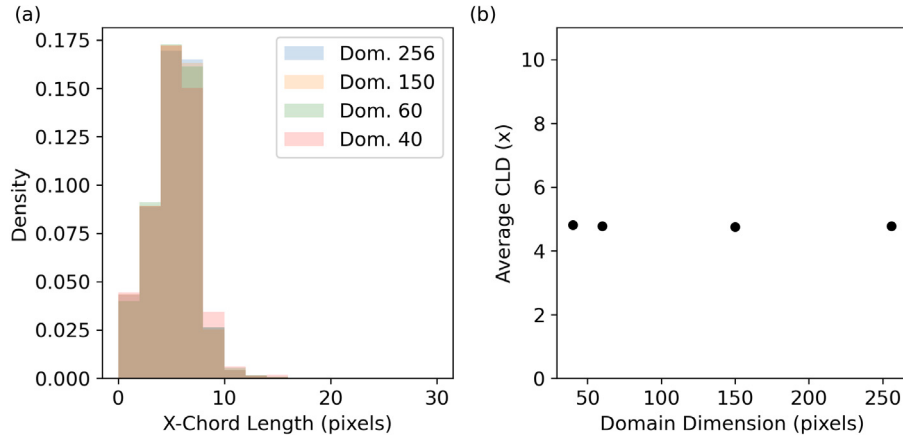


Fig. C.3. Images summarizing the statistical stability of the generated images as the size of the spatial domain is varied (40, 60, 150, 256 pixels). (a) An overlay of the Chord Length Distributions (computed in the x -direction) at each scale. (b) The average of the Chord Length Distribution (x -direction) at each scale.

us to use a less complex model. Second, limited dataset sizes leave us much more susceptible to training instability if we were to use a large architecture. Researchers have developed an extensive suite of regularization techniques, such as dropout²⁹ [160] and weight decay [161], to combat overfitting and instability. However, we argue that, for this application, systematically designing a lightweight model is a better solution. Specifically, smaller architectures result in decreased computational cost of the forward pass. Although the difference might be imperceptible for a single-pass model, the composite cost over the many forward passes required in denoising is noticeable. As an extreme example, generation time using Song et al.'s original NCSN architecture [109] (approximately 8 million parameters) took us several minutes per microstructure, whereas generation time with the augmented U-net takes 1.9 ± 0.6 seconds per sample on an Nvidia Tesla V100 GPU.

In addition to the measures described above, we also tested the statistical stability of the trained model as the size of the spatial dimension is changed. Because the model is trained at a constant patch size (40×40 for Case Study 1) and then evaluated at a different, larger spatial dimension during application (here, 256×256), such stability is vital to the performance of the entire generative framework. Fig. C.3 summarizes the results of the stability test. At each spatial dimension (40, 60, 150, and 256), 40 samples were drawn to compute statistics. Fig. C.3a,b demonstrate that the chord length distributions of the generated features remain acceptably stable as the spatial dimension is increased. This is clearly emphasized by the stability of the mean of the CLD. Altogether, this indicates that the denoising model will retain its documented performance from training when applied to a larger domain during application. During experimentation, we observed an important, negative relationship between the visual receptive field of our model and the statistical stability. Receptive fields significantly larger than the patch size caused instability, likely due to changes in the boundary information when we transition to larger spatial domains.

C.3. Selecting the reduced initial noise level

Before evaluating the conditional hybrid model, the final necessary step is to identify the adjusted starting noise level. As outlined in Section 3, sampling is conditioned by the output of the GRF via a shortened noise schedule (by starting at a reduced initial noise level, σ) and by replacing the white noise initial iterate with the output of the GRF. This hyperparameter must be tuned. As a guiding principle, it should be sufficiently high to enable pixels to transition from one

state to another (i.e., on the order of $\sigma \approx 0.5$), but not too high to wash out information from the initial GRF state. We swept a large range of values and recorded the error in our aforementioned statistical indicators. As seen in Fig. C.4, a balance is achieved when the change induced by the SBG transformation is sufficiently small to maintain the important features of the GRF (thereby matching the desired 2-point statistics) and sufficiently large to perturb the local features (thereby matching higher-order statistics such as chord length distributions). $\sigma = 0.75$ balances these competing goals. During our experiments, we observed that reduced initial noise levels near this one worked satisfactorily in many situations. However, performance could almost always be improved by application-specific tuning.

Appendix D. Case study 2: Training

For Case Study 2, during training, we used an initial noise level of 50.0 and a final noise level of 0.01. Mirroring observations in the literature [87], we employed a longer training regime – 750,000 iterations – compared to 500,000 in the 2-phase microstructures case. Otherwise, the methodology and training hyperparameters were the same as reported in Section 4.4.

Fig. D.1 summarizes the training process. Again, we compare microstructure statistics (this time for all three phases), rather than the loss curve, to measure quality-of-fit. Volume fractions (Fig. D.1a-c) and chord length distributions (Fig. D.1d-f) are computed for each phase individually. The principal component projection is computed in aggregate [29]. Although for some statistics there are larger inconsistencies than in Case Study 1 (this occurs because this neighborhood distribution is much more complex, Section 5.3), altogether, the performance is similar to the two-phase case and acceptable. Importantly, the performance for the higher-order statistics is strong. The difference between the neighborhood 2-point statistics distributions is almost imperceptible. Further, the chord length distributions, especially for the yellow and dark blue phases display excellent agreement. Its worth noting that the features for the teal phase are often larger than the patch size, so some error there is acceptable since such features will be predominantly dictated by the GRF. Fig. D.2 contrasts randomly selected patches generated using the final model with the closest patch to each from the training data. The patches contain both salient features of the finer yellow and dark blue phases as well as representative segments from the teal phase boundary. Again we note that the model does not simply memorize the training data.

²⁹ Dropout is used extensively in the DDPM framework [86].

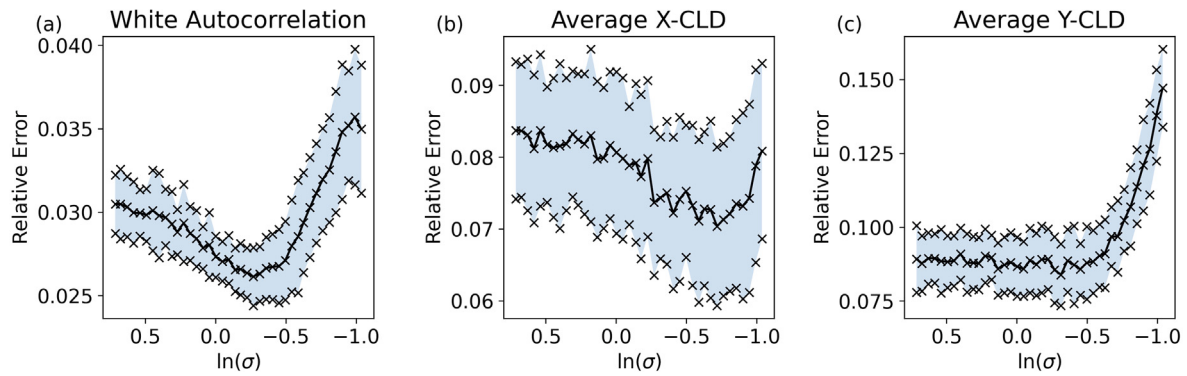


Fig. C.4. Images summarizing the experiments to determine the initial index to use for conditional generation. All plots have the noise index as their x-axis. Plots contrast the relative error of a set of 80 samples against the reference for several microstructure statistics. The mean and one standard deviation are depicted. (a) 2-point statistics. (b) Average Chord Length (x). (c) Average Chord Length (y).

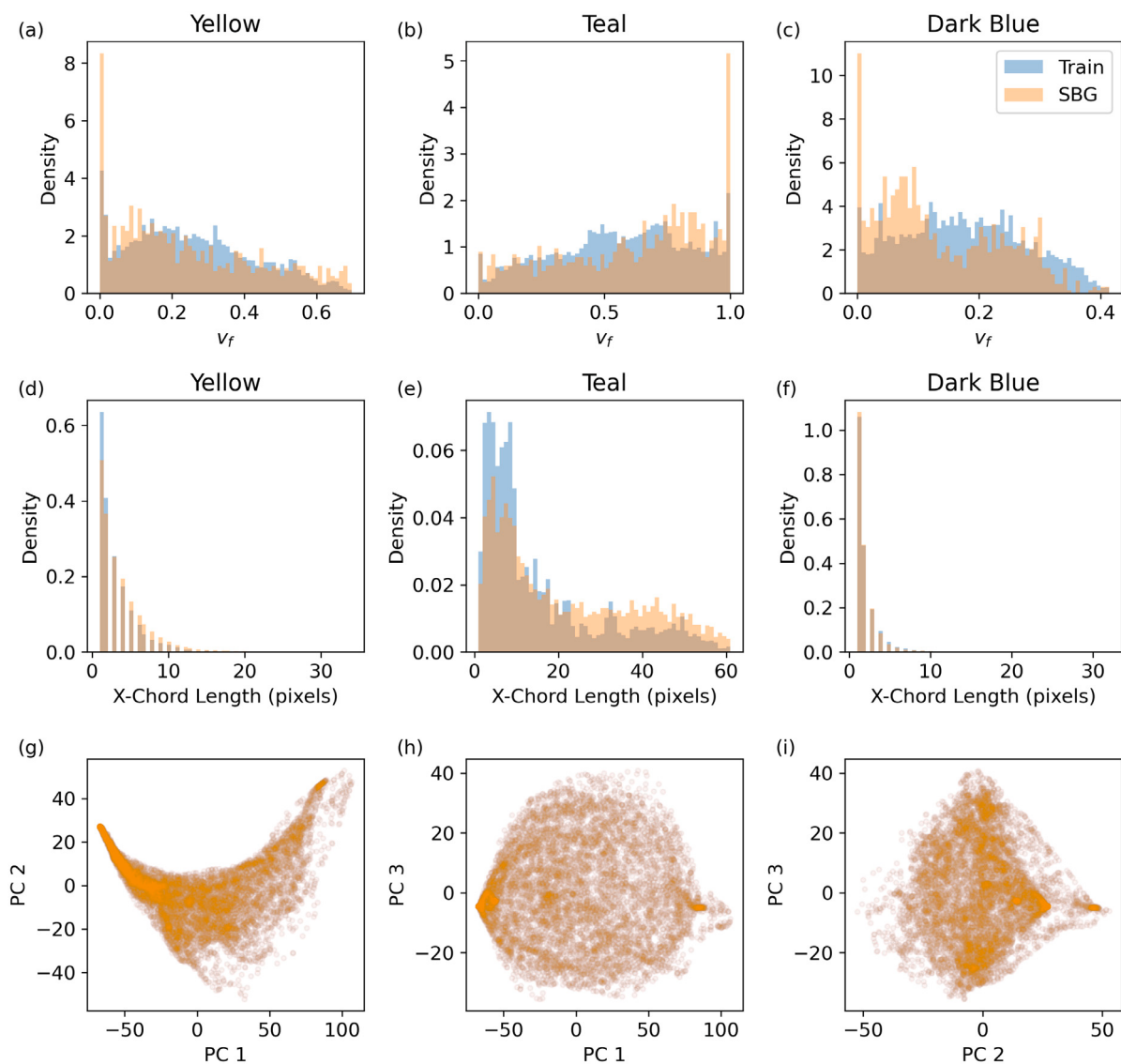


Fig. D.1. Visual summary of the training process metrics and output. The statistics of the training data are depicted in blue, the statistics of the synthetic patches are in orange. Distributions contrasting the volume fractions (a-c) and X chord length (d-f) between the training data (blue) and synthetic data from the unconditional SBG model (orange) for each phase. (h-j) 2-dimensional Principal Components of local patch 2-point statistics. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

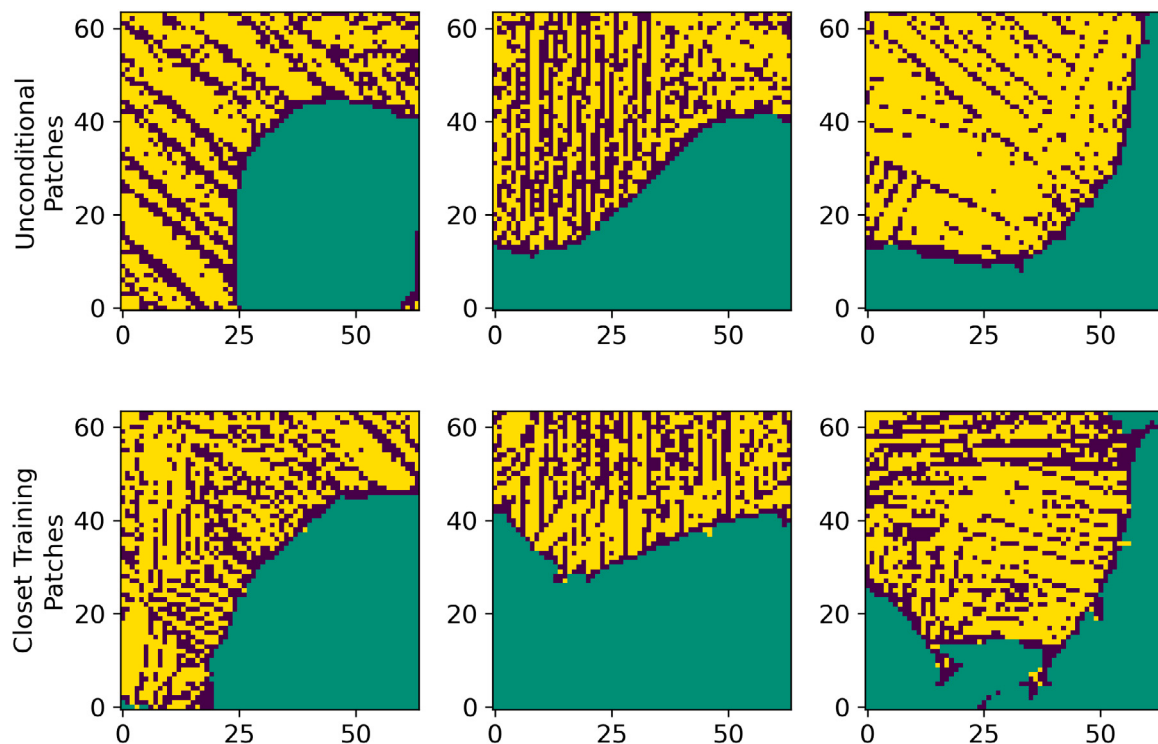


Fig. D.2. Images contrasting randomly selected unconditional samples from the trained SBG model (row 1) with their closest image in the training dataset (row 2). Closeness is measured using the L_2 -norm.

References

- [1] E. Kroner, Bounds for effective elastic moduli of disordered materials, *J. Mech. Phys. Solids* 25 (1977) 137–155.
- [2] E. Kroner, *Statistical Continuum Mechanics*, Springer, New York, NY, 1972.
- [3] W. Brown Jr., Solid mixture permittivities, *J. Chem. Phys.* 23 (1955) 1514–1517.
- [4] D. Fullwood, B. Adams, S. Kalidindi, A strong contrast homogenization formulation for multi-phase anisotropic materials, *J. Mech. Phys. Solids* 56 (2008) 2287–2297.
- [5] S. Torquato, Effective stiffness tensor of composite media: 1. Exact series expansions, *J. Mech. Phys. Solids* 45 (1997) 1421–1448.
- [6] S. Torquato, Effective stiffness tensor of composite media: 2. Applications to isotropic dispersions, *J. Mech. Phys. Solids* 46 (1998) 1411–1440.
- [7] N. Paulson, M. Priddy, D. McDowell, S. Kalidindi, Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics, *Acta Mater.* 129 (2017) 428, <http://dx.doi.org/10.1016/j.actamat.2017.03.009>.
- [8] M. Latypov, L. Toth, S. Kalidindi, Materials knowledge system for nonlinear composites, *Comput. Methods Appl. Mech. Engrg.* 346 (2019) 180, <http://dx.doi.org/10.1016/j.cma.2018.11.034>.
- [9] S. Niezgoda, Y. Yabansu, S. Kalidindi, Understanding and visualizing microstructure and microstructure variance as a stochastic process, *Acta Mater.* 59 (2011) 6387–6400, <http://dx.doi.org/10.1016/j.actamat.2011.06.051>.
- [10] S. Niezgoda, D. Fullwood, S. Kalidindi, Delineation of the space of 2-point correlations in a composite material system, *Acta Mater.* 56 (2008) 5285–5292, <http://dx.doi.org/10.1016/j.actamat.2008.07.005>.
- [11] Materials genome initiative for global competitiveness, 2011.
- [12] A. Robertson, S. Kalidindi, Efficient generation of 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>.
- [13] L. Mosser, O. Dubrule, M. Blunt, Stochastic reconstruction of oolitic limestone by generative adversarial networks, *Transp. Porous Media* 125 (2018) 81–103, <http://dx.doi.org/10.1007/s11242-018-1039-9>.
- [14] D. Fullwood, S. Niezgoda, S. Kalidindi, Microstructure reconstruction from 2-point statistics using phase recovery algorithms, *Acta Mater.* 56 (2008) 942–948, <http://dx.doi.org/10.1016/j.actamat.2007.10.044>.
- [15] S. Cheng, Y. Jiao, Y. Ren, Data-driven learning of 3-point correlation functions as microstructure representations, *Acta Mater.* 229 (2022) 117800, <http://dx.doi.org/10.1016/j.actamat.2022.117800>.
- [16] Y. Jiao, F. Stillinger, S. Torquato, Modeling heterogeneous materials via two-point correlation functions: basic principles, *Phys. Rev. E* 76 (2007) 031110, <http://dx.doi.org/10.1103/PhysRevE.76.031110>.
- [17] Y. Gao, Y. Jiao, Y. Liu, Ultra-efficient reconstruction of 3D microstructure and distribution of properties of random heterogeneous materials containing multiple phases, *Acta Mater.* 204 (2021) 116526, <http://dx.doi.org/10.1016/j.actamat.2020.116526>.
- [18] Y. Gao, Y. Liu, Reliability-based topology optimization with stochastic heterogeneous microstructure properties, *Mater. Des.* (2021) <http://dx.doi.org/10.1016/j.matdes.2021.109713>.
- [19] A. Marshall, S. Kalidindi, Autonomous development of a machine-learning model for the plastic response of two-phase composites from micromechanical finite element models, *JOM* 73 (2021) 2085–2095, <http://dx.doi.org/10.1007/s11837-021-04696-w>.
- [20] S. Kalidindi, M. Binci, D. Fullwood, B. Adams, Elastic Properties closures using second-order homogenization theories: Case studies in composites of two isotropic constituents, *Acta Mater.* 54 (2006) 3117–3126, <http://dx.doi.org/10.1016/j.actamat.2006.03.005>.
- [21] M. Hasan, Y. Mao, F. Tavazza, A. Choudhary, A. Agrawal, P. Acar, Data-driven multi-scale modeling and optimization for elastic properties of cubic microstructures, *Integr. Mater. Manuf. Innov.* (2022) <http://dx.doi.org/10.1007/s40192-022-00258-3>.
- [22] M. Diehl, M. Groeber, C. Haase, D. Molodov, F. Roters, D. Raabe, Identifying structure-property relationships through dream.3D representative volume elements and DAMASK crystal plasticity simulations: an integrated computational materials engineering approach, *JOM* 69 (2017) 848–855, <http://dx.doi.org/10.1007/s11837-017-2303-0>.
- [23] M. Abram, K. Burghardt, G.V. Steeg, A. Galstyan, R. Dingreville, Inferring topological transitions in pattern forming processes with self supervised learning, *NPJ: Comput. Mater.* 8 (2022) <http://dx.doi.org/10.1038/s41524-022-00889-2>.
- [24] S. Hashemi, S.R. Kalidindi, Gaussian Process Autoregression models for the evolution of polycrystalline microstructures subjected to arbitrary stretching tensors, *Int. J. Plast.* 162 (2023) 103532, <http://dx.doi.org/10.1016/j.ijplas.2023.103532>.
- [25] S. Torquato, *Random Heterogeneous Materials*, Springer, New York, NY, 2002.
- [26] B. Adams, S. Kalidindi, D. Fullwood, *Microstructure Sensitive Design for Performance Optimization*, Butterworth-Heinemann, Waltham, MA, 2013.
- [27] D. Fullwood, S. Niezgoda, B. Adams, S. Kalidindi, Microstructure sensitive design for performance optimization, *Prog. Mater. Sci.* 55 (2010) 477–562, <http://dx.doi.org/10.1016/j.pmatsci.2009.08.002>.
- [28] M.F. Horstemeyer, Multiscale modeling: a review, in: *Practical Aspects of Computational Chemistry: Methods, Concepts and Applications*, Springer Netherlands, Dordrecht, 2010, pp. 87–135, http://dx.doi.org/10.1007/978-90-481-2687-3_4.
- [29] A. Robertson, S. Kalidindi, Digital representation and quantification of discrete dislocation structures, *JOM* 73 (2021) 2143–2158, <http://dx.doi.org/10.1007/s11837-021-04669-z>.

- [30] T. Fast, O. Wodo, B. Ganapathysubramanian, S. Kalidindi, Microstructure taxonomy based on spatial correlations: application to microstructure coarsening, *Acta Mater.* 108 (2016) 176, <http://dx.doi.org/10.1016/j.actamat.2016.01.046>.
- [31] M. Yuan, S. Paradiso, B. Meredig, S. Niezgoda, Machine learning-based reduced order polycrystalline modelling for ICME applications, *Integr. Mater. Manuf. Innov.* 7 (2018) 214–230, <http://dx.doi.org/10.1007/s40192-018-0123-x>.
- [32] A. Mann, S. Kalidindi, Development of a robust CNN model for capturing microstructure-property linkages and building property closures supporting material design, *Front. Mater.* 9 (2022) 851085, <http://dx.doi.org/10.3389/fmats.2022.851085>.
- [33] D.M. de Oca Zapiain, J. Stewart, R. Dingreville, Accelerating phase field based microstructure evolution predictions via surrogate models trained by machine learning methods, *NPJ Comput. Mater.* 3 (2021) 1–11, <http://dx.doi.org/10.1038/s41524-020-00471-8>.
- [34] S. Hashemi, S. Kalidindi, A machine learning framework for the temporal evolution of microstructure during static recrystallization of polycrystalline materials simulated by cellular automaton, *Comput. Mater. Sci.* 188 (2021) 110132, <http://dx.doi.org/10.1016/j.commatsci.2020.110132>.
- [35] G. Harrington, C. Kelly, V. Attari, R. Arroyave, S. Kalidindi, Application of a chained-ANN for learning the process-structure mapping in $Mg_{2.5}Si_xSn_{1-x}$ spinodal decomposition, *Integr. Mater. Manuf. Innov.* 11 (2022) 433–449, <http://dx.doi.org/10.1007/s40192-022-00274-3>.
- [36] J. Rossin, P. Leser, K. Pusch, C. Frey, S. Vogel, A. Saville, C. Torbet, A. Clarke, S. Daly, T. Pollock, Single crystal elastic constants of additively manufactured components determined by resonant ultrasound spectroscopy, *Mater. Charact.* 192 (2022) 112244, <http://dx.doi.org/10.1016/j.matchar.2022.112244>.
- [37] A. Castillo, A. Venkatraman, S. Kalidindi, Mechanical responses of primary- α Ti grains in polycrystalline samples: part II – bayesian estimation of the crystal-level elastic-plastic mechanical properties from spherical-indentation measurements, *Integr. Mater. Manuf. Innov.* 10 (2021) 99–114, <http://dx.doi.org/10.1007/s40192-021-00204-9>.
- [38] R. Hill, A self-consistent mechanics of composite materials, *J. Mech. Phys. Solids* 13 (4) (1965) 213–222, [http://dx.doi.org/10.1016/0022-5096\(65\)90010-4](http://dx.doi.org/10.1016/0022-5096(65)90010-4).
- [39] G.J. Dvorak, A.M. Wafa, Y.A. Bahei-El-Din, Implementation of the transformation field analysis for inelastic composite materials, *Comput. Mech.* 14 (3) (1994) 201–228, <http://dx.doi.org/10.1007/BF00370073>.
- [40] H. Moulinec, P. Suquet, A numerical method for computing the overall response of nonlinear composites with complex microstructure, *Comput. Methods Appl. Mech. Engrg.* 157 (1) (1998) 69–94, [http://dx.doi.org/10.1016/S0045-7825\(97\)00218-1](http://dx.doi.org/10.1016/S0045-7825(97)00218-1).
- [41] Z. Liu, M.A. Bessa, W.K. Liu, Self-consistent clustering analysis: An efficient multi-scale scheme for inelastic heterogeneous materials, 306 (2016) 319–341, <http://dx.doi.org/10.1016/j.cma.2016.04.004>.
- [42] R.A. Lebensohn, A.D. Rollett, Spectral methods for full-field micromechanical modelling of polycrystalline materials, *Comput. Mater. Sci.* 173 (2020) 109336, <http://dx.doi.org/10.1016/j.commatsci.2019.109336>.
- [43] C. Kelly, S. Kalidindi, Recurrent localization networks applied to the lippmannschwinger equation, *Comput. Mater. Sci.* 192 (2021) 110356, <http://dx.doi.org/10.1016/j.commatsci.2021.110356>.
- [44] P. Acar, V. Sundararaghavan, Stochastic design optimization of microstructural features using linear programming for robust design, *AIAA J.* 57 (2019) 448–455.
- [45] F. Roters, P. Eisenlohr, L. Hantcherli, D. Tjahjanto, T. Bieler, D. Raabe, Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite element modeling: Theory, experiments, applications, *Acta Mater.* 58 (2010) 1152.
- [46] M. Groeber, S. Ghosh, M. Uchic, D. Dimiduk, A framework for automated analysis and simulation of 3D polycrystalline microstructures. Part 2: Synthetic Microstructure Generation, *Acta Mater.* 56 (2008) 1274–1287, <http://dx.doi.org/10.1016/j.actamat.2007.11.040>.
- [47] M. Groeber, M. Jackson, Dream3D: a digital representation environment for the analysis of microstructure in 3D, *Integr. Mater. Manuf. Innov.* 3 (2014) 56–72, <http://dx.doi.org/10.1186/2193-9772-3-5>.
- [48] M. Prasad, N. Vajragupta, A. Hartmaier, Kanapy: A Python package for generating complex synthetic polycrystalline microstructures, *J. Open Source Softw.* 4 (2019) 1732, <http://dx.doi.org/10.21105/joss.01732>.
- [49] T. Fast, S.R. Kalidindi, Formulation and calibration of higher-order elastic localization relationships using the MKS approach, *Acta Mater.* 59 (2011) 4595–4605, <http://dx.doi.org/10.1016/j.actamat.2011.04.005>.
- [50] K. Frydrych, K. Karimi, M. Pecelerowicz, R. Alvarez, F.D. Gutierrez, F. Rovaris, S. Papanikolaou, Materials informatics for mechanical deformation: a review of applications and challenges, *Materials* 14 (2021) 5764, <http://dx.doi.org/10.3390/ma14195764>.
- [51] F. Bock, R. Aydin, C. Cyron, N. Huber, S. Kalidindi, B. Klusemann, A review of the application of machine learning and data mining approaches in continuum materials mechanics, *Front. Mater.* 6 (2019) 110, <http://dx.doi.org/10.3389/fmats.2019.00110>.
- [52] B.L. DeCost, M. Hecht, T. Francis, B.A. Webler, Y.N. Picard, E. Holm, UHSCDB: ultra high carbon steel micrograph database, *Integr. Mater. Manuf. Innov.* 6 (2017) 197–205, <http://dx.doi.org/10.1007/s40192-017-0097-0>.
- [53] S. Goswami, M. Yin, Y. Yu, G. Karniadakis, A physics-informed variational DeepONet for predicting crack path in quasi-brittle materials, *Comput. Methods Appl. Mech. Engrg.* 391 (2022) 11457, <http://dx.doi.org/10.1016/j.cma.2022.114587>.
- [54] G. Landi, N. Niezgoda, S. Kalidindi, Multi-scale modeling of elastic properties of three-dimensional voxel-based microstructure datasets using novel DFT-based knowledge systems, *Acta Mater.* 58 (2010) 2716–2725, <http://dx.doi.org/10.1016/j.actamat.2010.01.007>.
- [55] O. Zerhouni, S. Brisard, K. Danas, Quantifying the effects of two-point correlations on the effective elasticity of specific classes of random porous materials with and without connectivity, *Internat. J. Engrg. Sci.* (2021) <http://dx.doi.org/10.1016/j.ijengsci.2021.103520>.
- [56] N. Paulson, M. Priddy, D. McDowell, S. Kalidindi, Reduced-order microstructure-sensitive protocols to rank-order the transition fatigue resistance of polycrystalline microstructures, *Int. J. Fatigue* 119 (2019) 1, <http://dx.doi.org/10.1016/j.ijfatigue.2018.09.011>.
- [57] M. Safdari, M. Baniassadi, H. Garmestani, M. Al-Haik, A modified strong-constraint expansion for estimating the effective thermal conductivity of multiphase heterogeneous materials, *J. Appl. Phys.* 112 (2012) 114318.
- [58] A. Papoulis, S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, Chennai, 2002.
- [59] D. Koller, N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, Cambridge, MA, 2011.
- [60] P. Du, A. Zebrowski, J. Zola, B. Ganapathysubramanian, O. Wodo, Microstructure design using graphs, *Comput. Mater.* 4 (2018) <http://dx.doi.org/10.1038/s41524-018-0108-5>.
- [61] Y.C. Yabansu, A. Isakov, A. Kapustina, S. Rajagopalan, S. Kalidindi, Application of Gaussian Process Regression models for capturing the evolution of microstructure statistics in aging of nickel-based superalloys, *Acta Mater.* 178 (2019).
- [62] N. Millan-Espitia, S. Kalidindi, Study of bimodal α - β Ti alloy microstructure using multi-resolution spherical indentation stress-strain protocols, *J. Compos. Sci.* 6 (2022) 1–13, <http://dx.doi.org/10.3390/jcs6060162>.
- [63] Y. Jiao, F. Stillinger, S. Torquato, Modeling heterogeneous materials via two-point correlations. II. Algorithmic details and applications, *Phys. Rev. E* 77 (2008) 031135, <http://dx.doi.org/10.1103/PhysRevE.77.031135>.
- [64] M. Baniassadi, H. Garmestani, D. Li, S. Ahzi, M. Khaleel, X. Sun, Three-phase solid oxide fuel cell anode microstructure realization using two-point correlation functions, *Acta Mater.* 59 (2011) 30–43, <http://dx.doi.org/10.1016/j.actamat.2010.08.012>.
- [65] A. Hasanabadi, M. Baniassadi, K. Abrinia, M. Safdari, H. Garmestani, 3D microstructural reconstruction of heterogeneous materials from 2D cross sections: A modified phase-recovery algorithm, *Comput. Mater. Sci.* 111 (2016) 107–115, <http://dx.doi.org/10.1016/j.commatsci.2015.09.015>.
- [66] P. Seibert, M. Ambati, A. Rablöff, M. Kastner, Reconstructing Random Heterogeneous Media through differentiable optimization, *Comput. Mater. Sci.* 196 (2021) 110455, <http://dx.doi.org/10.1016/j.commatsci.2021.110455>.
- [67] P. Seibert, A. Rablöff, M. Ambati, M. Kastner, Descriptor-based reconstruction of three-dimensional microstructures through gradient-based optimization, *Acta Mater.* 227 (2022) 117667, <http://dx.doi.org/10.1016/j.actamat.2022.117667>.
- [68] Y. Jiao, F. Stillinger, S. Torquato, A superior descriptor of random textures and its predictive capacity, *Proc. Natl. Acad. Sci. USA* 106 (2009) 17634–17639, <http://dx.doi.org/10.1073/pnas.0905919106>.
- [69] A. Roberts, M. Knackstedt, Structure-Property correlations in model composite materials, *Phys. Rev. E* 54 (1996) <http://dx.doi.org/10.1103/PhysRevE.54.2313>.
- [70] Y. Liu, J. Li, B. Yu, Advances in gaussian random field generation: a review, *Comput. Geosci.* 23 (2019) 1011–1047, <http://dx.doi.org/10.1007/s10596-019-09867-y>.
- [71] C. Dietrich, G. Newsam, A fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix, *SIAM J. Sci. Comput.* 18 (1997) 1088–1107, <http://dx.doi.org/10.1137/S1064827592240555>.
- [72] A. Cecen, B. Yucl, S. Kalidindi, A generalized and modular framework for Digital Generation of Composite Microstructures, *J. Compos. Sci.* 5 (2021) 1–20, <http://dx.doi.org/10.3390/jcs5080211>.
- [73] R. Bostanabad, A. Bui, W. Xie, D. Apley, W. Chen, Stochastic microstructure characterization and reconstruction via supervised learning, *Acta Mater.* 103 (2016) 89–102, <http://dx.doi.org/10.1016/j.actamat.2015.09.044>.
- [74] X. Liu, V. Shapiro, Random heterogeneous materials via texture synthesis, *Comput. Mater. Sci.* 99 (2015) 177–189, <http://dx.doi.org/10.1016/j.commatsci.2014.12.017>.
- [75] I. Javaheri, V. Sundararaghavan, Polycrystalline microstructure reconstruction using markov random fields and histogram matching, *Comput. Aided Des.* 120 (2020) 102806, <http://dx.doi.org/10.1016/j.cad.2019.102806>.
- [76] A. Senthilnathan, P. Acar, M.D. Graef, Markov Random Field based microstructure reconstruction using the principal image moments, *Mater. Charact.* 178 (2021) 111281, <http://dx.doi.org/10.1016/j.matchar.2021.111281>.
- [77] D. Turner, S.R. Kalidindi, Statistical construction of 3D Microstructures from 2-D exemplars collected on oblique sections, *Acta Mater.* 102 (2016) 136–148, <http://dx.doi.org/10.1016/j.actamat.2015.09.011>.

- [78] S. Kamrava, P. Tahmasebi, M. Sahimi, Enhancing images of shale formations by a hybrid stochastic and deep learning algorithm, *Neural Netw.* 118 (2019) 310–320, <http://dx.doi.org/10.1016/j.neunet.2019.07.009>.
- [79] T. Bai, P. Tahmasebi, Hybrid geological modeling: combining machine learning and multi-point statistics, *Comput. Geosci.* 142 (2020) 104519, <http://dx.doi.org/10.1016/j.cageo.2020.104519>.
- [80] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, *NeurIPS* (2014).
- [81] J. Zhu, T. Park, P. Isola, A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *ICCV* (2014) 1–10.
- [82] J.-W. Lee, N. Goo, W. Park, M. Pyo, K.-S. Sohn, Virtual microstructure design for steels using generative adversarial networks, *Eng. Rep.* 3 (2021) 1–14, <http://dx.doi.org/10.1002/eng2.12274>.
- [83] P. Isola, J. Zhu, T. Zhou, A. Efros, Image-to-image translation with conditional adversarial networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [84] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, B. Catanzaro, High-resolution image synthesis and semantic manipulation with conditional GANs, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [85] S. Radev, U. Mertens, A. Voss, L. Ardizzone, U. Kothe, BayesFlow: learning complex stochastic models with invertible neural networks, *IEEE Trans. Neural Netw. Learn. Syst.* 33 (2022) 1452–1466, <http://dx.doi.org/10.1109/TNNLS.2020.3042395>.
- [86] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, *NeurIPS* (2020).
- [87] Y. Song, J. Sohl-Dickstein, D.P. Kigima, A. Kumar, S. Ermon, B. Poole, Score-based generative modeling through stochastic differential equations, *Int. Congr. Learn. Represent.* (2021) 1–36.
- [88] S. Bond-Taylor, A. Leach, Y. Long, C. Willcocks, Deep generative modelling: a comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models, *IEEE Trans. Pattern Anal. Mach. Intell.* (2022) 1–20.
- [89] L. Gatys, A. Ecker, M. Bethge, Image style transfer using convolutional neural networks, *CVPR* (2016).
- [90] A. Bhaduri, A. Gupta, A. Olivier, L. Graham-Brady, An Efficient Optimization based microstructure reconstruction approach with multiple loss functions, *Comput. Mater. Sci.* 199 (2021) 110709, <http://dx.doi.org/10.1016/j.commatsci.2021.110709>.
- [91] D. Jangid, N. Brodnik, A. Khan, M. Goebel, M. Echlin, T. Pollock, S. Daly, B. Manjunath, 3D grain shape generation in polycrystals using generative adversarial networks, *Integr. Mater. Manuf. Innov.* 11 (2022) 71–84, <http://dx.doi.org/10.1007/s40192-021-00244-1>.
- [92] S. Chun, S. Roy, Y. Nguyen, J. Choi, H. Udaykumar, S. Baek, Deep learning for synthetic microstructure generation in a materials-by-design framework for heterogeneous energetic materials, *Sci. Rep.* 10 (2020) 13307, <http://dx.doi.org/10.1038/s41598-020-70149-0>.
- [93] T. Hsu, W.K. Epting, H. Kim, H.W. Abernathy, G.A. Hackett, A.D. Rollett, P.A. Salvador, E.A. Holm, Microstructure generation via generative adversarial network for heterogeneous, topologically complex 3D materials, *JOM* 73 (2021) 90–102, <http://dx.doi.org/10.1007/s11837-020-04484-y>.
- [94] R. Narikawa, Y. Fukatsu, Z.-L. Wang, T. Ogawa, Y. Adachi, Y. Tanaka, S. Ishikawa, Generative adversarial networks-based synthetic microstructures for data-driven materials design, *Adv. Theory Simul.* 2 (2022) 2100470, <http://dx.doi.org/10.1002/adts.202100470>.
- [95] O. Ogoke, K. Johnson, M. Glinsky, C. Laursen, S. Kramer, A. Farimani, Deep-learned generators of porosity distributions produced during additive manufacturing, 2022, *ArXiv: Submitted to Additive Manufacturing*.
- [96] J. Tang, X. Geng, D. Li, Y. Shi, J. Tong, H. Xiao, F. Peng, Machine Learned-based microstructure prediction during laser sintering of alumina, *Sci. Rep.* 11 (2021) 10724, <http://dx.doi.org/10.1038/s41598-021-89816-x>.
- [97] D. Fokina, E. Muravleva, G. Ovchinnikov, I. Oseledets, Microstructure synthesis using style-based generative adversarial networks, *Phys. Rev. E* 101 (2020) 043308, <http://dx.doi.org/10.1103/PhysRevE.101.043308>.
- [98] 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>.
- [99] S. Noguchi, J. Inoue, Stochastic Characterization and reconstruction of material microstructures for establishment of process-structure-property linkage using the deep generative model, *Phys. Rev. E* 104 (2021) 025302, <http://dx.doi.org/10.1103/PhysRevE.104.025302>.
- [100] C. Dureth, P. Seibert, D. Rucker, S. Handford, M. Kastner, M. Gude, Conditional diffusion-based microstructure reconstruction, 2022, *ArXiv*.
- [101] K. Lee, G. Yun, Microstructure reconstruction using diffusion-based generative models, 2022, *ArXiv*.
- [102] A. Iyer, B. Dey, A. Dasgupta, W. Chen, A conditional generative model for predicting material microstructures from processing methods, 2019, *ArXiv*.
- [103] J. Tang, X. Geng, D. Li, Y. Shi, J. Tong, H. Xiao, F. Peng, Machine-learning-based microstructure prediction during laser sintering of alumina, *Sci. Rep.* 11 (2021) 10724, <http://dx.doi.org/10.1038/s41598-021-89816-x>.
- [104] C. Luo, Understanding diffusion models: a unified perspective, 2022, <http://dx.doi.org/10.48550/arXiv.2208.11970>, *ArXiv*.
- [105] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, Y. Shao, W. Zhang, B. Cui, M. Yang, Diffusion models: a comprehensive survey of methods and applications, 2022, *ArXiv*.
- [106] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, *Int. Conf. Mach. Learn.* (2015) 2256–2265.
- [107] A. Nichol, P. Dhariwal, Improved denoising diffusion probabilistic models, *Int. Conf. Mach. Learn.* (2021).
- [108] P. Vincent, A connection between score matching and denoising autoencoders, *Neural Comput.* 23 (2011) 1661–1674, http://dx.doi.org/10.1162/NECO_a.00142.
- [109] Y. Song, S. Ermon, Generative modeling by estimating gradients of the data distribution, in: *NeurIPS* 2019, 2019.
- [110] A. Vahdat, K. Kreis, J. Kautz, Score-based generative modeling in latent space, in: *NeurIPS* 2021, 2021, pp. 1–21.
- [111] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J. Zhu, S. Ermon, SDEdit: guided image synthesis and editing with stochastic differential equations, 2022, *ArXiv*.
- [112] M. Zhao, F. Bao, C. Li, J. Zhu, EGSD: unpaired image-to-image translation via energy-guided stochastic differential equations, 2022, pp. 1–21, *ArXiv*.
- [113] P. Dhariwal, A. Nichol, Diffusion models beat GANs on image synthesis, in: *NeurIPS*, 2021.
- [114] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen, Hierarchical text-conditional image generation with CLIP latents, 2022, <http://dx.doi.org/10.48550/arXiv.2204.06125>, *ArXiv*.
- [115] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, D. Fleet, Video diffusion models, 2022, <http://dx.doi.org/10.48550/arXiv.2204.03458>, *ArXiv*.
- [116] W. Harvey, S. Naderiparizi, V. Masrani, C. Wilbach, F. Wood, Flexible diffusion modeling of long videos, 2022, <http://dx.doi.org/10.48550/arXiv.2205.11495>, *ArXiv*.
- [117] E. Hoogeboom, V. Satorras, C. Vignac, M. Welling, Equivariant diffusion for molecule generation in 3D, *Int. Conf. Mach. Learn.* (2022) <http://dx.doi.org/10.48550/arXiv.2205.15019>.
- [118] N. Anand, T. Achim, Protein structure and sequence generation with equivariant denoising diffusion probabilistic models, 2022, <http://dx.doi.org/10.48550/arXiv.2205.15019>, *ArXiv*.
- [119] N. Chen, Y. Zhang, H. Zen, R. Weiss, M. Norouzi, W. Chan, WaveGrad: estimating gradients for waveform generation, 2020, <http://dx.doi.org/10.48550/arXiv.2009.00713>, *ArXiv*.
- [120] Y. Song, S. Ermon, Improved techniques for training score-based generative models, in: *NeurIPS*, 2020.
- [121] Y. Song, D. Kingma, How to train your energy based models, 1, 2021, pp. 1–22, *ArXiv*.
- [122] A. Jolicœur-Martineau, K. Li, R. Piche-Taillefer, T. Kachman, I. Mitliagkas, Gotta go fast when generating data with score-based models, 2021, pp. 1–24, *ArXiv*.
- [123] J. Song, C. Meng, S. Ermon, Denoising diffusion implicit models, 2021, <http://dx.doi.org/10.48550/arXiv.2010.02502>, *ArXiv*.
- [124] F. Schaefer, H. Zheng, A. Anandkumar, Implicit competitive regularization in GANs, in: *Proceedings of the 37th International Conference on Machine Learning, PMLR*, 2020, pp. 8533–8544, URL <https://proceedings.mlr.press/v119/schaefer20a.html>.
- [125] I. Kobyzev, S. Prince, M. Brubaker, Normalizing flows: an introduction and review of current methods, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (2021) 3964–3979, <http://dx.doi.org/10.1109/TPAMI.2020.2992934>.
- [126] Y. Song, S. Garg, J. Shi, S. Ermon, Sliced score matching: a scalable approach to density and score estimation, *Proc. Mach. Learn. Res.* (2020).
- [127] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, M. Norouzi, Palette: image-to-image diffusion model, in: *ACM SIGGRAPH Conference Proceedings*, 2022, <http://dx.doi.org/10.1145/3528233.3530757>.
- [128] J. Ho, T. Salimans, Classifier-free diffusion generation, 2022, <http://dx.doi.org/10.48550/arXiv.2207.12598>, *ArXiv*.
- [129] J. Choi, S. Kim, Y. Jeong, Y. Gwon, S. Yoon, ILVR: conditioning method for denoising diffusion probabilistic models, 2021, *ArXiv*.
- [130] P. Embrechts, F. Lindskog, A. McNeil, Modelling dependence with copulas and applications to risk management, 2001, URL https://mx.nthu.edu.tw/~jtyang/Teaching/Risk_management/Papers/Correlations/Modelling%20Dependence%20with%20Copulas%20and%20Applications%20to%20Risk%20Management.pdf.
- [131] V. Monga, Y. Li, Y.C. Eldar, Algorithm unrolling: interpretable, efficient deep learning for signal and image processing, *IEEE Signal Process. Mag.* 38 (2) (2021) 18–44, <http://dx.doi.org/10.1109/SPM.2020.3016905>.
- [132] S. Pang, G. Barbastathis, Machine learning regularized solution of the lippmannschwinger equation, 2020, <http://dx.doi.org/10.48550/arXiv.2010.15117>, URL <http://arxiv.org/abs/2010.15117>, *arXiv:2010.15117 [physics]*.
- [133] J. Adler, O. Öktem, Learned primal-dual reconstruction, *IEEE Trans. Med. Imaging* 37 (6) (2018) 1322–1332, <http://dx.doi.org/10.1109/TMI.2018.2799231>, *arXiv:1707.06474 [cs, math]*.
- [134] S. Bai, J. Kolter, V. Koltun, Deep equilibrium models, *Adv. Neural Inf. Process. Syst.* 32 (2019).

- [135] P. Putzky, M. Welling, Recurrent inference machines for solving inverse problems, 2017, <http://dx.doi.org/10.48550/arXiv.1706.04008>, ArXiv.
- [136] H. You, Q. Zhang, C. Ross, C. Lee, Y. Yu, Learning Deep Implicit Fourier Neural Operators (IFNOs) with applications to heterogeneous material modeling, *Comput. Methods Appl. Mech. Engrg.* 398 (2022) 115296, <http://dx.doi.org/10.1016/j.cma.2022.115296>.
- [137] G. Adomian, *Solving Frontier Problems of Physics: The Decomposition Method*, Springer Netherlands, Dordrecht, 1994, <http://dx.doi.org/10.1007/978-94-015-8289-6>, URL <http://link.springer.com/10.1007/978-94-015-8289-6>.
- [138] K.E. Atkinson, *The Numerical Solution of Integral Equations of the Second Kind*, in: Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 1997, <http://dx.doi.org/10.1017/CBO9780511626340>, URL <https://www.cambridge.org/core/books/numerical-solution-of-integral-equations-of-the-second-kind/91E77DDA93701D3B024830429043A037>.
- [139] R. Lebensohn, A. Rollett, Spectral Methods for full-field micromechanical modelling of polycrystalline materials, *Comput. Mater. Sci.* 173 (2020) 109336, <http://dx.doi.org/10.1016/j.commatsci.2019.109336>.
- [140] C. Shorten, T. Khoshgoftaar, A survey of image data augmentation for deep learning, *J. Big Data* (2019).
- [141] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: *NeurIPS*, 2017.
- [142] N. Brodnik, C. Muir, N. Tulshibagwale, J. Rossin, M. Echlin, C. Hamel, S. Kramer, T. Pollock, J. Kiser, C. Smith, S. Daly, Perspective: Machine learning in experimental solid mechanics, *J. Mech. Phys. Solids* 173 (2023) 105231, <http://dx.doi.org/10.1016/j.jmps.2023.105231>.
- [143] A. Gupta, A. Bhaduri, L. Graham-Brady, Accelerated multiscale mechanics modeling in a deep learning framework, 2023, <http://dx.doi.org/10.48550/arXiv.2212.14601>, ArXiv.
- [144] Z. Yang, C. Yu, M. Buehler, Deep Learning Model to predict complex stress and strain fields in hierarchical composites, *Sci. Adv.* 7 (2021) <http://dx.doi.org/10.1126/sciadv.abd7416>.
- [145] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015*, 9351, 2015, http://dx.doi.org/10.1007/978-3-319-24574-4_28.
- [146] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, 2016.
- [147] W. Luo, Y. Li, R. Urtasun, R. Zemel, Understanding the effective receptive field in deep convolutional neural networks, in: *Advances in Neural Information Processing Systems*, Vol. 29, 2016.
- [148] P. Ramachandran, B. Zoph, Q. Le, Searching for activation functions, 2017, <http://dx.doi.org/10.48550/arXiv.1710.05941>, ArXiv.
- [149] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2018.
- [150] M. Guo, T. Xu, J. Liu, Z. Liu, P. Jiang, T. Mu, S. Zhang, R. Martin, M. Cheng, S. Hu, Attention mechanisms in computer vision: a survey, *Comput. Vis. Media* 8 (2022) 331–368, <http://dx.doi.org/10.1007/s41095-022-0271-y>.
- [151] A. Cecen, T. Fast, S. Kalidindi, Versatile algorithms for the computation of 2-point spatial correlations in quantifying material structure, *Integr. Mater. Manuf. Innov.* 5 (2016) 1–15, <http://dx.doi.org/10.1186/s40192-015-0044-x>.
- [152] J. Berryman, Relationship between specific surface area and spatial correlation functions for anisotropic porous media, *J. Math. Phys.* 28 (1987) 244–245.
- [153] S. Blair, P. Berge, J. Berryman, Using two-point correlation functions to characterize microgeometry and estimate permeabilities of sandstone and porous glass, *J. Geophys. Res.* 101 (1996) 20359–20375, <http://dx.doi.org/10.1029/96JB00879>.
- [154] R. Dimitrakopoulos, H. Mustapha, E. Gloaguen, Higher-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-gaussian and non-linear phenomena, *Math. Geosci.* 42 (2010) 65–99, <http://dx.doi.org/10.1007/s11004-009-9258-9>.
- [155] J. Shlens, A tutorial of principal component analysis, 2020, URL <https://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition.jp.pdf>. (Accessed 28 November 2020).
- [156] A. Generale, S. 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>.
- [157] S. Torquato, J. Beasley, Y. Chiew, Two-point cluster function for continuum percolation, *J. Chem. Phys.* 88 (1988) 6540, <http://dx.doi.org/10.1063/1.454440>.
- [158] J. Helton, F. Davis, Latin hypercube sampling and propagation of uncertainty in analyses of complex systems, *Reliab. Eng. Syst. Saf.* (2003) 23–69, [http://dx.doi.org/10.1016/S0951-8320\(03\)00058-9](http://dx.doi.org/10.1016/S0951-8320(03)00058-9).
- [159] C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals, Understanding deep learning requires rethinking generalization, 2017, URL <http://arxiv.org/abs/1611.03530>, arXiv:1611.03530 [cs].
- [160] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* 15 (2014) 1929–1958.
- [161] A. Krogh, J. Hertz, A simple weight-decay can improve generalization, *Adv. Neural Inf. Process. Syst.* 4 (1991).