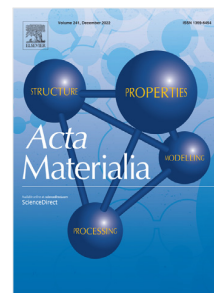


Journal Pre-proof

Towards inverse microstructure-centered materials design using generative phase-field modeling and deep variational autoencoders

Vahid Attari, Danial Khatamsaz, Douglas Allaire, Raymundo Arroyave



PII: S1359-6454(23)00534-7

DOI: <https://doi.org/10.1016/j.actamat.2023.119204>

Reference: AM 119204

To appear in: *Acta Materialia*

Received date: 27 March 2023

Revised date: 25 July 2023

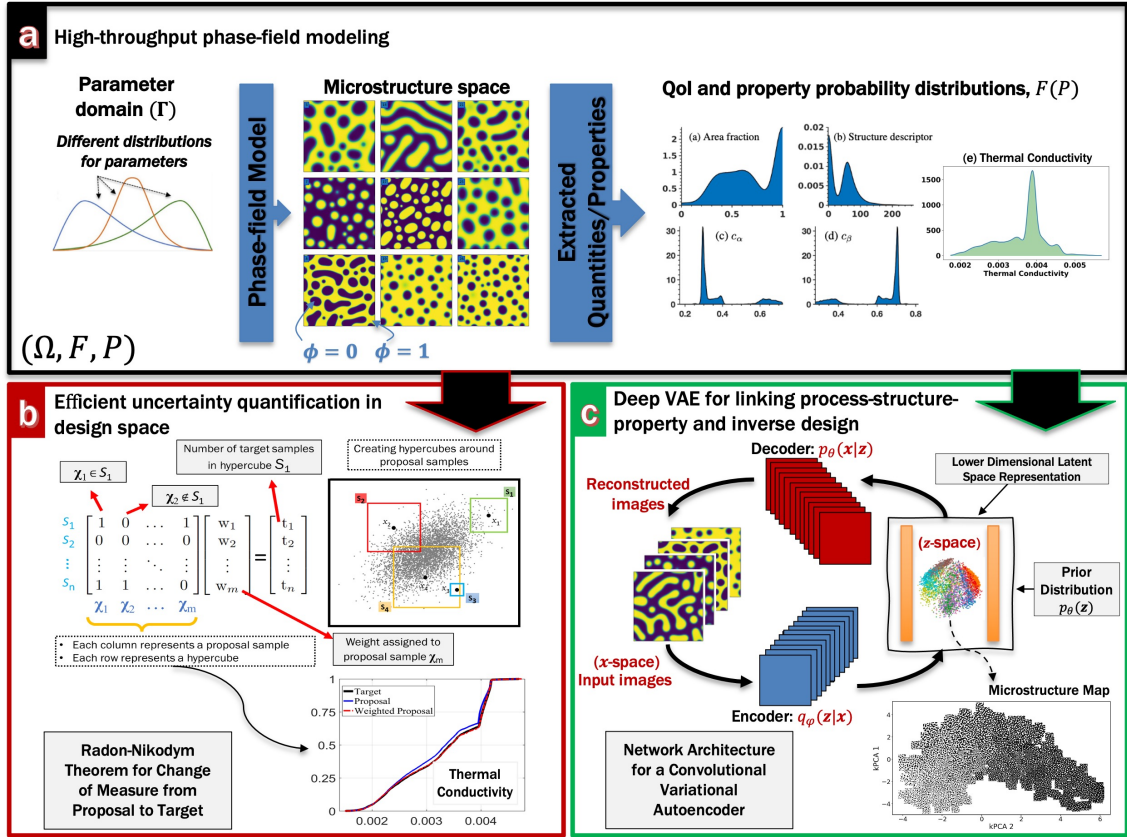
Accepted date: 31 July 2023

Please cite this article as: V. Attari, D. Khatamsaz, D. Allaire et al., Towards inverse microstructure-centered materials design using generative phase-field modeling and deep variational autoencoders, *Acta Materialia* (2023), doi: <https://doi.org/10.1016/j.actamat.2023.119204>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Acta Materialia Inc. Published by Elsevier Ltd. All rights reserved.

Graphical Abstract



Towards Inverse Microstructure-Centered Materials Design using Generative Phase-field Modeling and Deep Variational AutoEncoders

Vahid Attari^{a**}, Danial Khatamsaz^a, Douglas Allaire^b, Raymundo Arroyave^{a,b}

^aMaterials Science and Engineering Department, Texas A&M University, College Station, TX, USA 77840

^bMechanical Engineering Department, Texas A&M University, College Station, TX, USA 77840

Abstract

The field of Integrated Computational Materials Engineering (ICME) combines a broad range of methods to study materials' responses over a spectrum of length scales. A relatively unexplored aspect of microstructure-sensitive materials design is uncertainty propagation and quantification (UP/UQ) of materials' microstructure, as well as establishing process-structure-property (PSP) relationships for inverse material design. In this study, an efficient UP technique built on the idea of changing probability measures and a deep generative unsupervised representative machine learning method for microstructure-based design of thermal conductivity of materials is proposed. Probability measures are used to represent microstructure space, and Wasserstein metrics are used to test the efficiency of the UP method. By using deep Variational AutoEncoder (VAE), we identify the correlations between the material/process parameters and the thermal conductivity of heterogeneous dual-phase microstructures. Through high-throughput screening, UP, and the deep-generative VAE method, PSP relationships that are too complex can be revealed by exploiting the materials' design space with an emphasis on microstructures. As a last point, we demonstrate generative machine learning serves as a useful tool for inverse microstructure-centered materials design, and we demonstrate this by examining the inverse design of thermal conductivity in nano-structured materials. The results reveal the effects of morphology, volume fraction, characteristic length scale, and the individual thermal diffusivity of phases on the thermal conductivity of dual-phase alloys. Our findings emphasize the advantages of high-throughput phase-field modeling and generative deep learning for linking PSP and inverse microstructure-centered materials design.

Keywords: computational materials design, process-structure-property linkage, microstructure, generative modeling, phase-field modeling, uncertainty quantification, variational autoencoder

1 Introduction

The material's **structure** is key to connect processing conditions to resultant properties and performance, forming Process-Structure (PS) and Structure-Property (SP) relationships. Quasi-random micro-/nano-structures, central to the Integrated Computational Materials Engineering (ICME) paradigm, frequently appear in high-dimensional probabilistic materials science problems. Accelerating knowledge acquisition necessitates microstructure quantification, which relies on physical models and data science methods, essential for investigating and applying PSP relationships [1]. Materials informatics employs data, machine learning, and optimization techniques to systematically discern PSP relationships, guiding the predictive design of materials and processes to achieve desired properties [2]. Leveraging advancements in phase-field modeling [3], uncertainty quantification [4], and materials informatics methods [5], generative modeling [6] can build intelligent frameworks predicting emergent material properties, such as electrical, thermal, mechanical, and magnetic characteristics, highlighting microstructure within the Integrated Computational Materials Engineering (ICME) context.

While precise mesoscale simulation tools are currently available, *in silico* microstructure-centered materials design is still frequently pursued without considering optimal search strategies. Additionally, when search strategies are deployed, they are paired with a quantity of microstructures, rather than the entire microstructure with all its complexity. By combining existing modeling tools and knowledge of the materials, current methods often develop a forward computational relationship with some microstructural metric (e.g., volume fraction [7], grain size distributions [8]) or its statistics (e.g., 2-point correlations [9]). For instance, Molkeri et al. [7] proposed a Bayesian framework for microstructure-aware materials design to identify the best combination of chemistry and processing parameters for maximizing a targeted mechanical property of a model dual-phase steel. According to

*Corresponding author

Email addresses: Email: attari.v@tamu.edu (Vahid Attari^a), Email (Vahid Attari^a)

¹Address: Materials Science and Engineering Department, Texas A&M University, College Station, TX, USA 77840

their analysis, deliberate incorporation of microstructure information into materials design methodology dramatically improves materials optimization. Despite this, they did not incorporate the microstructure seamlessly into the design process, instead, used a subsequent quantity of the microstructure, the volume fraction of the martensite phase (a potential attribute of the phase diagram). Similarly, in another work, Saunders et al. [10] have developed a Gaussian process-based method for establishing PSP in additive manufacturing processes that indirectly establishes the connection with microstructure through microstructure features statistics, rather than the microstructure itself. Although microstructural information has traditionally been captured by metrics such as grain size distributions and phase volume fractions, many properties of interest are not sufficiently accounted for by these approaches due to the inherent complexity of the microstructure. Identifying a design workflow that links a desired property to the underlying process/material parameters without reducing the microstructure to some subsequent quantity is a challenging task. Also, it is necessary to quantify uncertainty in the PSP relationship for reliable designs [11].

The stochastic nature inherent in the random heterogeneous structure of materials can introduce considerable uncertainty, thereby compromising precision. It is thus important for any ICME framework to properly quantify the uncertainty emanating from these variations [12]. This uncertainty, particularly at the mesoscale, originates from complex and unpredictable short- and long-range interactions, manifesting as spatio-temporal fluctuations in field-induced gradients (e.g., chemomechanical interactions). These fluctuations generate hard-to-quantify uncertainties across all relevant scales (i.e., atomistic, molecular, and meso-scale). Material reliability and performance can be compromised by this uncertainty, requiring meticulous and robust design considerations. In general, uncertainty in modeling material structure originates from five sources [4]: (i) model selection uncertainty, (ii) epistemic uncertainty in model variables due to limited computed or measured data or large ranges in prior knowledge, (iii) stochasticity in phase arrangements at any material point, (iv) randomness in phase morphology owing to significant spatial variation, and (v) variation of material properties due to the heterogeneous nature and scarcity of lower-scale characteristics, including defects, interfaces, and bulk regions, which influence the macroscopic average property from a mesoscopic perspective. Uncertainty propagation, on the other hand, facilitates the estimation of potential variability in material property or performance arising from these uncertain sources. Monte Carlo (MC) methods are generally recognized as the gold standard for Uncertainty Propagation (UP) [13, 14]. However, despite their ubiquity and simplicity, MC simulations can be computationally demanding due to the necessity of propagating a substantial number of samples through a computational model. To mitigate the computational burden of MC simulation methods, more efficient approaches such as importance sampling [15] and adaptive sampling [16] have been proposed. However, a high-throughput (HTP) exploration becomes inefficient when a model is either (1) expensive or (2) not fully parametrically sampled. Alternative strategies for executing Uncertainty Propagation (UP) include local expansion-based methods [17], which may not perform well in the presence of significant input variability, functional expansion-based methods [18], and numerical integration-based methods [19]. An efficient approach to tackle UP problems suggests changing the probability measure to a desired input distribution (target) from a previously evaluated dataset regardless of its distribution (proposal) [20–22].

This study aims to develop a data-driven general-purpose framework for materials design placing microstructure at the core for establishing forward PSP relations and furthermore inverse design of materials. For predictive modeling in unexplored domains, a wide array of unsupervised representation learning methods [23] has been extensively utilized across various fields. These methods include deep belief networks [24], deep Boltzmann machines [25], deep diffusion models [26], and deep Variational Autoencoders (VAEs) [27]. The deep VAEs have gained popularity for their ability to generate novel data, including audio, text, and images, through density estimation and subsequent sampling. We generate the microstructure dataset used to demonstrate this framework through high-throughput phase-field modeling. Subsequently, we compute the desired microstructural property using previously established constitutive relations. At its core, our framework harnesses microstructure data alongside composition and processing conditions drawn from known probability distributions. Consequently, bottom-up material design entails the use of a deep VAE to establish PSP relations, generate novel microstructures, and conduct inverse design through the capture of spatial relations between pixels in existing microstructure data. Additionally, the design would take into account the uncertainty by the Radon-Nikodym (R-N) theorem [28, 29] that transforms a proposal measure (i.e., known probability measure) to a target measure to save resources by reducing any experimentation costs as the input distributions change due to induced uncertainties.

Recently, there have been several applications of VAEs reported in the literature, spanning various fields. These include the design of drug-like molecules and novel chemical structures [30], the exploration of optimal microstructure spaces for mechanical properties [31], and the design of phononic bandgaps in metamaterials [32]. Additional applications feature the development of ferritic-martensitic steels [33], establishing links between interface energy and the morphology of microstructures undergoing heterogeneous nucleation and growth [34], and database development [35]. While data methods have progressed significantly, standard methods for microstructure-centered materials design have yet to be developed. We anticipate that our developed data-driven framework can contribute to the development of approaches to rational, uncertainty-aware microstructure-based material design by streamlining the creation of models utilizing microstructure data, such as those available in the Open Phase-field Microstructure Database (OPMD)[36]. The rest of this article is arranged as follows: section 2 discusses our proposed methods;

section 3 presents the results and their discussion; and section 4 provides a brief conclusion. Additional details related to some of the results are provided in the supplementary document.

2 Methods

In this study, we present a cost-minimal microstructure-centered materials design framework for linking process-structure-property, and inverse materials design incorporating efficient uncertainty propagation built upon the idea of the change of probability measure. This framework consists of (i) a microstructure evolution model based on phase-field theory, in which the thermal conductivity of the microstructures is subsequently computed, (ii) an efficient UP framework using R-N theorem, and (iii) a deep generative VAE model for PSP linking and inverse design. A schematic of the proposed framework is shown in Fig. 1 where box (a) illustrates the HTP process of calculating tens of thousands of time-series of microstructures from probability distributions of input parameters and further calculation of thermal conductivity and other low-dimensional representations of the microstructure. Box (b) and (c) each schematically show the efficient UP method and VAE method developed in this study, respectively.

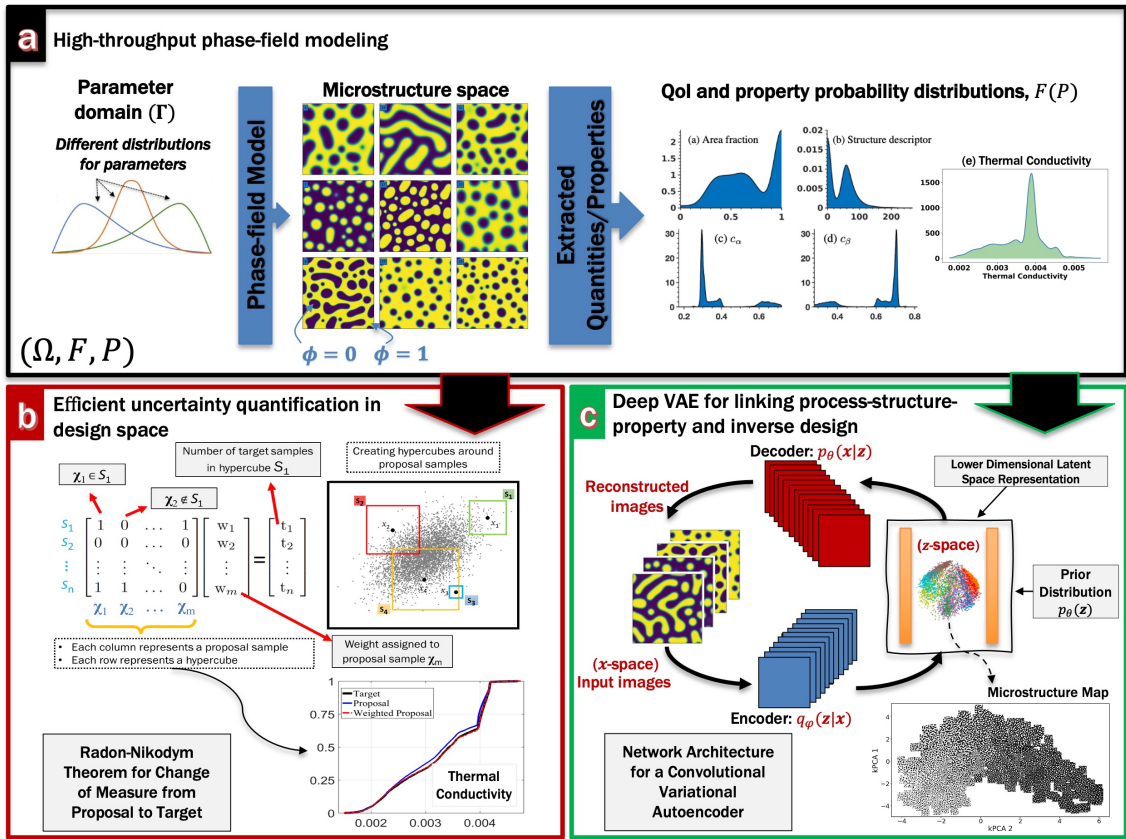


Figure 1: A schematic for efficient UP within a microstructure-sensitive materials design scenario. (a) Given the probability distributions of the material/process parameters, a feasible microstructure space and associated properties is determined by a high-throughput phase-field model using 10,000 simulation instances. (b) The process of efficient propagation of uncertainty by generating hypercubes centered around input samples and assigning weights to input samples. Several example hypercubes of different sizes along with the linear system of equations that needs to be solved, Calculated cumulative density functions of thermal conductivity for proposal, target, and weighted proposal estimations are shown. (c) Schematic of a Variational Autoencoder with microstructure images as input. After receiving input, the encoder produces a compressed representation, which is then fed into the decoder. Then, an image reconstruction is produced by the decoder. VAE involves learning stochastic mappings between microstructure spaces, whose empirical distributions are typically complicated.

2.1 Terminologies for Uncertainty in Microstructure Modeling

Random heterogeneous materials feature intricate and diverse microstructures. These microstructures can be characterized using correlation functions (e.g., n-point probability, surface correlation, nearest-neighbor functions), fundamental to determining the materials' effective properties [37]. A material's microstructure can be defined probabilistically by an indicator function (e.g., a

phase-field variable, $\phi(r)$), where the likelihood $\mathcal{P}\phi(r) = 1$ signals the presence of a phase at position r . Treating the microstructure as a random process, a probability triple $(\Omega, \mathcal{F}, \mathcal{P})$ provides a mathematical model for the material's microstructure. Here, Ω represents all possible microstructures, $\mathcal{F}$ signifies feasible, calculable microstructures, and $\mathcal{P}$ serves as a probability measure assigning each microstructure in Ω a likelihood between 0 and 1. For each microstructure $F \in \mathcal{F}$, its probability $\mathcal{P}(F)$ is given by the expected value under $\mathcal{P}$ of the indicator of the microstructure ($\mathcal{P}(F) = E_{\mathcal{P}}(I_F)$). We view $\mathcal{P}(F)$ as a microstructure representation, encapsulating physical quantities of interest (e.g., volume fraction), microstructural correlation functions (e.g., n-point probability functions), or desired properties (e.g., thermal conductivity).

For convenience, our exposition is restricted to boundary value problems. Let $\mathcal{D} \in \mathbb{R}^d, d = 1, 2, 3$, be a fixed physical domain with boundary $\partial\mathcal{D}$, and $r = \{r_1, \dots, r_d\}$ be the coordinates. Let us consider a PDE,

$$\begin{aligned} \mathcal{L}(r, \xi, u; y) &= 0 & \text{in } \mathcal{D}; \\ B(r, u; y) &= 0 & \text{on } \partial\mathcal{D}; \end{aligned} \quad (1)$$

where $\mathcal{L}$ is a differential operator and B is a boundary operator. $x \in \mathcal{D}$ is a spatial variable, and $y \in \Gamma$ is a vector of random variables in parameter domain Γ . Γ is often considered to be an N-dimensional hyper-cube. ξ denotes a (Gaussian) space-time white noise in $\mathcal{D}$. The solution of the stochastic computation to this problem set is:

$$u = u(t, r; y) \quad (2)$$

where $t \in [0, T]$ is a temporal variable in a temporal interval. The solution is a function of time, spatial variable, and random parameters y . When the PDE is a random differential equation it can contain random inhomogeneous coefficients, initial conditions, boundary conditions, or force terms. The random or stochastic Cahn-Hilliard PDE is given as:

$$\partial_t u = \Delta^2 u + P(u) + \xi, \quad (3)$$

where Δ is the Laplacian and $P(u)$ is a potential function that can take logarithmic, polynomial, or other similar forms. In practice, we are interested in a set of quantities,

$$g = (g_1, \dots, g_K) \in \mathbb{R} \quad (4)$$

called (QoIs) that are functions of the solution u , in addition to the solution itself. QoI is a model output, which may be the PDE solution itself, but is more often a consequence of the solution and could contain statistical information about the output of interest. The outputs of interest in the phase-field model may be interpreted as a function of the PDE solution or statistical information of such a solution. UP enables determining statistical information about the outputs of the phase-field model given samples generated using statistical information about the inputs. In the next section, we describe the model, the uncertain inputs, and the QoIs that we are interested in following the ML method for connecting PSP and inverse microstructure design for a desired property.

2.2 Microstructure Model

To demonstrate our proposed framework, we first generate an extensive microstructure dataset. Leveraging high-throughput phase-field simulations, we generate time-series of synthetic microstructures, aiding in understanding the impact of propagated uncertainties on the microstructure landscape of a hypothetical medium during isothermal thermal annealing. A phase-field model, based on the Cahn-Hilliard model [38], is employed to generate microstructures while accounting for bulk and curvature effects. This model addresses the process of spontaneous phase separation in an isotropic alloy with constant molar volume. The associated boundary value problem is presented as follows:

$$\begin{aligned} \frac{\partial c}{\partial t} &= \nabla \cdot \left\{ M \nabla \left(\frac{\partial f_{\text{bulk}}}{\partial c} - \kappa \nabla^2 c \right) \right\} & \begin{aligned} 0 &< x, y < L_x, L_y \\ 0 &< t < t^* \end{aligned} \\ \text{BC:} & \quad c(0, y, t) = c(L_x, y, t) & c(x, 0, t) = c(x, L_y, t) \\ \text{IC:} & \quad c(x, y, 0) = c^* + A\zeta \end{aligned} \quad (5)$$

where c is concentration, M is the inherently positive effective atomic mobility of the species, f is alloy free energy, and κ is the gradient energy coefficient. The lengths of the simulation domain are set to $L_x = L_y = 200$ with grid size of 256×256 and t^* is the final model run time. BC and IC denote the used boundary and initial conditions, respectively. c^* is the initial average value of the order parameter that is perturbed by a constant noise magnitude A , and ζ is a Gaussian random number with the interval of $[-1, +1]$. Here, we consider two free energy function forms for the bulk free energy:

$$f(c) = W(c - c_\alpha)^2(c - c_\beta)^2 \quad (6)$$

$$f(\phi) = a(T) + \frac{a_2(T)}{2}\phi^2 + \frac{a_4(T)}{4}\phi^4 \quad (7)$$

where the first one (hereafter called model I) approximates a symmetric well for a two-phase alloy with barrier height of transformation of W . $c_\alpha = 0.35$ and $c_\beta = 0.69$ are the equilibrium composition of α and β phases, respectively, and are fixed to reduce the model dimension to four. The latter free energy function (hereafter called model II) is also a symmetric well in the form of a forth-order Landau polynomial where the material-dependent coefficients a_2 and a_4 not only control the barrier height of transformation, but also the corresponding equilibrium values of field-order parameter (ϕ). Below the critical temperature where the material becomes unstable and decomposition occurs, $a_2(T) < 0$ and $a_4(T) > 0$.

By using the Fourier-spectral method for the spatial variables, and, treating the linear fourth-order operators implicitly and the nonlinear terms explicitly, the first-order semi-implicit Fourier-spectral scheme [39, 40] is used to solve the boundary value problem. The resulting numerical scheme is:

$$(1 + M\Delta t k^4)\tilde{c}^{n+1}(\mathbf{k}) = \tilde{c}^n(\mathbf{k}) - \Delta t, k^2\{\tilde{\mu}(c^n)\}_{\mathbf{k}} \quad (8)$$

where $\mathbf{k} = (k_1, k_2)$ is a frequency vector in the Fourier space, $k = \sqrt{k_1^2 + k_2^2}$ is the magnitude of this vector and $\tilde{c}(\mathbf{k}, t)$ represents the Fourier transform of $c(\mathbf{r}, t)$.

In our study, all model parameters are continuous data, albeit some are constrained within certain bounds. For instance, the composition ranges between 0 and 1. Yet, the phase diagram introduces additional constraints on the composition data, further narrowing the feasible range. Consequently, potential statistical distributions can include normal, lognormal, F, chi-square, exponential, t-student, Weibull, and non-normal distributions. While we do not enforce a specific distribution, we ensure that the distribution type and range align with the physical context. For instance, we employ truncated normal and uniform distributions to represent the alloy composition of model I in both proposal and target spaces. This distribution type is then considered when performing the random sampling procedure.

The HTP process of sampling from parameter posterior distributions and microstructure data generation is briefly shown in Fig. 1(a). To generate the proposal (i.e., known probability measure, μ) and target microstructure data, we sampled 10,000 different combinations of $[c^*, W, \kappa_c, M]$ and $[\phi^*, a_2, a_4, \kappa_\phi, M]$ parameters out of the distributions shown in Fig. S12 and Fig. S13, respectively, and ran forward the phase-field model.

Moreover, we conducted a global sensitivity analysis using the Sobol method on time series of microstructure data to evaluate the impact of model inputs on its outputs and their potential interactions. The detailed results are documented in the supplementary material in Tables S1, S2, and S3. In essence, the composition significantly influences the model output, specifically thermal conductivity, during the initial stages of microstructure evolution. However, as the evolution progresses, although the composition effect remains substantial, the interaction between the composition and the gradient energy coefficient becomes increasingly crucial.

Nonetheless, the study incorporates two types of free energy models and two spaces (target and proposal), resulting in a total of 40,000 phase-field runs. From each phase-field run, we acquired a series of 50 micrographs, yielding 40,000 images for each combination and an overall total of 200,000 microstructure images. Fig. 3b presents examples of microstructures in the proposal probability space of model I.

The computation time for solving the Cahn-Hilliard equation is approximately 40 seconds per simulation. Each simulation utilized a single CPU and 7 GBs of random access memory. By implementing batch job submission routines, we distributed the 40,000 simulations across 500 CPUs. This approach significantly reduced the overall simulation time to roughly 2 hours, in contrast to the 444 hours it would have required for sequential execution.

2.3 Material Property: Thermal Conductivity

The framework is motivated by the prediction of the thermal conductivity of heterogeneous structures. Over the past few decades, researchers have conducted a significant amount of research on the thermal conductivity (λ) of nanostructured materials, such as thin films [41], nanowires [42], and bulk nanostructured materials [43] and the results often highlight the influence of quantum confinement [44]. The thermal conductivity of amorphous solids and polycrystalline cubic materials is in general isotropic and scalar. Typically, only one type of energy carrier dominates λ . In a multiple carrier system, thermal conductivity is calculated by taking into account a sum of contributions from the lattice (λ_L), electrons (λ_e), photons (λ_r), i.e., $\lambda = \lambda_L + \lambda_e + \lambda_r$. Among the prominent systems in which multiple carriers are critical include thermoelectric materials (λ_L and λ_e) and thermal barrier coatings at high temperature (λ_L and λ_r) [45]. Based on Carnot efficiency ($\eta_{\text{Carnot}} = \frac{T^{\text{hot}} - T^{\text{cold}}}{T^{\text{hot}}}$) and generalized Zener criterion ($Ze = (\sqrt{1 + ZT} - 1)/(\sqrt{1 + ZT} + 1)$), the maximum power-generation efficiency of isotropic thermoelectric materials is defined

as [46, 47]:

$$\eta_{local}^{Max} = \eta_{Carnot} Z e^{\left[\frac{\sqrt{1+ZT} - 1}{\sqrt{1+ZT} + \left(\frac{T_{cold}}{T_{hot}} \right)} \right]} \quad (9)$$

where Z is the thermoelectric figure of merit of isotropic material ($Z = \frac{S^2 \sigma}{\lambda}$) derived using macroscopic heat balance in thermoelectric legs. A composite material's ZT factor cannot exceed the ZT of its individual phases, according to Bergman and Levy [48]. Nonetheless, Joule heating and Peltier effects invoked by defects and interfaces have been left out in this assessment [44]. In this context, it has been demonstrated by Bian et al. [49] that heterogeneity is important in maximizing thermoelectric response. Likewise, Snyder et al. [50] proposed a compatibility factor for designing functionally graded thermoelectric materials. Through this, the grand problem of thermoelectric energy conversion is shifted to optimizing three macroscopically measurable transport parameters (σ , S , and λ) as a group. To build upon the framework, we, however, concentrate on the design of effective thermal conductivity in composite dual-phase materials. Provided a diverse set of dual-phase microstructures, the effective thermal conductivity of these microstructures can potentially be determined using generalized Ohm's and Fourier's constitutive relations [47]).

2.4 Calculation of Isotropic Thermal Conductivity

The Fourier heat conductivity equation with heterogeneous coefficient is solved for the steady-state case to simulate heat flow through the computed microstructures. This boundary value problem is:

$$\begin{aligned} \nabla \alpha \cdot \nabla T &= 0 & 0 < x, y < L_x, L_y \\ IC : T(x, y, t = 0) &= 50 \\ BC : q \cdot n &= 0.05 \end{aligned} \quad (10)$$

where T is temperature, α is space-dependent thermal diffusivity that is given by $\alpha = \phi \alpha_{p_1} + (1 - \phi) \alpha_{p_2} + A_{GB} \phi^2 (1 - \phi^2)$ where ϕ is the order parameter and A_{GB} is an amplitude factor that controls the grain boundary conductivity. To reduce the complexity, A_{GB} is set to zero in this paper that ensures a linear change in thermal conductivity in the grain boundary zone. q is the local heat flux density. The constant heat-flux boundary conditions on the microstructure in the direction of an orthogonal heat flow from bottom-left corner to top-right corner is applied. The thermal flow is driven by constant temperature difference boundary conditions across the structure in the primary flow direction, and a constant temperature is used as initial condition. The equation was solved by successive-over relaxation method where convergence were achieved by satisfying an L_2 norm of less than 1×10^{-7} to guarantee that the computed thermal conductivity is independent of the relaxation time. The computational grids coincides with the grid points used for Cahn-Hilliard equation. For a classical, isotropic material, the defining equation for heat conduction is Fourier's law [45],

$$q = -\lambda \nabla T \quad (11)$$

where λ represents the material's scalar thermal conductivity expressed in the International System of Units (SI) as $\text{Wm}^{-1} \text{K}^{-1}$ and ΔT is the temperature gradient vector.

The computation time required to calculate the effective thermal conductivity, satisfying the defined tolerance, is approximately 250 seconds, utilizing 28 CPUs. We executed a total of 160,000 calculations across three microstructural stages (early, middle, and final) to determine their effective thermal conductivities. By distributing these computations across 500 CPUs, we substantially reduced the total computational time to roughly 14 hours.

2.5 Uncertainty Propagation

Undertaking uncertainty propagation (UP) with computationally intensive models, such as atomistic Density Functional Theory (DFT) simulations or meso-scale multi-physics phase-field models, can present a challenge. Propagating an ample number of samples from a desired distribution through such models is often impractical due to computational cost. Furthermore, generating samples from a specific distribution can also be computationally demanding. A potential solution lies in utilizing samples from different distributions that have already been propagated through the same model, or devising a distribution that is less computationally taxing for sample extraction. We then apply the concept of change of probability measure to these samples.

In instances where both the target and proposal distributions are known, the Radon-Nikodym (R-N) theorem provides a valid solution for a change of measure by calculating the R-N derivatives (the ratio of the target to proposal densities) and assigning weights to the samples. However, in our scenario, we presume that both the target and proposal distributions are unknown. Consequently, direct application of the R-N theorem becomes infeasible. [51] In the sample-based approach, our objective is to determine the importance weights for each proposal sample. These weights enable us to precisely approximate the target sample density within any subspace of the feature (measure) space by resampling the proposal samples, weighted accordingly. Indeed,

this concept mirrors that used in calculating probability density functions to determine the probability of encountering samples at a specific location or within a defined range in the distribution's support.

To use the notion of change of measure, we assume that there are two sets of samples available: one set is called the target, which we are interested to quantify the uncertainty of a quantity of interest after propagating this set of samples to the model, and one is known as the proposal, which we aim to represent the target samples with it via change of probability measure. The change of probability measure refers to assigning importance weights (equivalent to proposing a new density function) to samples drawn from a proposal distribution, thus, the distribution of samples after resampling according to the assigned weights represents a target distribution. The importance weights are the ratio of probability densities at different locations in the space.

An illustrative example of efficient uncertainty propagation (UP) through the change of measure is depicted in Fig. 1(b), which demonstrates a 2-D distribution of proposal samples. In this approach, we segment the feature space into subspaces by generating differently sized hypercubes centered around the available samples, as exemplified by the four squares shown. A linear system of equations is then constructed, with each equation linked to a specific hypercube to correlate the target density with individual proposal samples. By resolving this system, we can estimate the importance weights between the proposal and target probability measures. Although hypercube sizes can vary, we adopt a strategic method to maintain process tractability: initially, a hypercube encompassing the entire feature space is created, which is then successively divided by two in all dimensions to produce smaller hypercubes. This halving process repeats until the hypercubes reach a pre-determined size threshold.

After defining the hypercubes for each sample, the empirical measure of a given hypercube S_i is calculated as:

$$\mu_i = \frac{1}{n} \sum_{j=1}^n \mathbb{I}_{S_i}(x_j) \quad (12)$$

where μ_i defines the proposal probability measure in hypercube S_i and $\mathbb{I}_{S_i}(x^j)$ is 1 if $x^j \in S_i$, else its value is 0. The goal here is to weigh the empirical proposal measure such that

$$\mu_i = \sum_{j=1}^n w_j \mathbb{I}_{S_i}(x_j) \rightarrow \nu_i \quad (13)$$

where ν_i is the target probability measure. To find the importance weights, a system of equations in the form of $\mathbf{A}\mathbf{w}=\mathbf{B}$ is formed and solved for the importance weights. First, using values P_{ij} for each point and each hypercube as

$$P_{ij}(\mu) = \mathbb{I}_{S_i}(x_j) \quad (14)$$

matrix $\mathbf{A}$ is arranged as

$$\mathbf{A} = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{k1} & P_{k2} & \dots & P_{kn} \end{bmatrix} \quad (15)$$

where n is the number of proposal samples and k is the total number of generated hypercubes. By creating the importance weights vector as $\mathbf{w} = [w_1, w_2, \dots, w_n]$, the weighted empirical proposal measure for each hypercube is $\mathbf{A}_i \mathbf{w}$. Then, matrix $\mathbf{B}$ is a k by 1 column vector that each entry is associated to a hypercube and is the ratio of target samples in that hypercube to the total number of target samples. Basically, this is an estimate of the probability of having target samples in each hypercube. Note that if the target distribution is known, each entry can be calculated by integrating the probability density function over the corresponding hypercube and this approach is still valid. Finally, to solve the system of equation, $\mathbf{A}\mathbf{w}=\mathbf{B}$, any least squares technique can be employed satisfying two conditions: weights must be bounded to $[0,1]$ and they must sum up to unity.

The merit of the sample-based change of measure lies in its convenience and cost-efficiency. With this technique, it is not necessary to conduct new simulations or experiments when the distribution of input samples changes. Instead, one can utilize the results from previously propagated samples, reweighting them to achieve the target distribution of a quantity of interest. This approach is applicable even without knowledge of the actual sample distribution, as it merely relies on the existence of proposal and target samples within the input space.

2.6 Deep Generative Machine Learning

Variational autoencoders (VAEs) are a type of unsupervised learning algorithm that can be used to generate new data by means of a latent representation. VAEs consist of an encoder, a decoder, and a loss objective (ELBO: Evidence Lower Bound). The encoder takes input data and maps it to a latent space. The decoder takes the latent space and maps it back to the original data space.

The loss objective is used to measure the difference between the original data and the freshly generated data typically derived through Jensen's inequality [27]. Consider N images $\{x^{(n)}\}_{n=1}^N$, with $x^{(n)} \in \mathcal{R}^{N_x \times N_y \times N_c}$; N_x and N_y represent the number of pixels in each spatial dimension, and N_c denoting the number of color bands in the image ($N_c = 1$ for gray-scale images and $N_c = 3$ for RGB images). The learning process is conducted by minimizing the variational ELBO. Figure 2 illustrates an schematic with the underlying network architectures for encoder and decoder networks used in this study. The detailed information regarding each network can be found in Fig. S20.

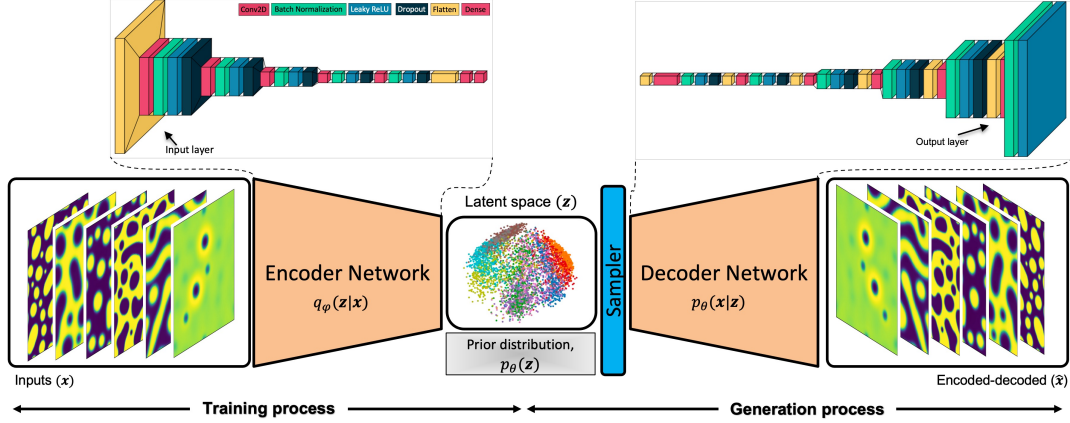


Figure 2: VAE model and the architecture of the encoder (inference) and decoder (generative) networks.

In the latent space of the VAE, a Gaussian distribution is utilized where a sampling layer is performed as per the formula $z = \mu + \epsilon e^{\sigma/2}$. In this formula, ϵ is a random tensor corresponding to the size of the latent space, holding values from a normal distribution with zero mean and unit variance. The terms μ and σ originate from the last convolutional layer, which are then flattened and connected to two dense layers. For visualizing this in 2D, the latent space is further reduced using kernel principal component analysis (kPCA) with a linear basis function kernel. The architecture of the encoder model is mirrored in the decoder, ensuring optimal tuning between the encoder and decoder networks.

The optimization objective of the VAE is the evidence lower bound that is for any choice of inference model $q_\phi(z|x)$ is given as:

$$\mathcal{L}_{\theta,\phi}(x) = \log p_\theta(x) - D_{KL}(q_\phi(z|x) || p_\theta(z|x)) \quad (16)$$

where the loss function is composed of a reconstruction term that makes the encoding-decoding scheme efficient and a regularization term that makes the latent space regular. The VAE loss is provided by the sum of the Kullback–Leibler divergence and the image reconstruction binary cross entropy.

Consequently, a complex microstructure space characterized by a complicated empirical distribution is stochastically mapped to a simpler, compressed latent space. The encoder model, represented as $q_\phi(z|x)$, provides an approximation to the true but computationally infeasible posterior of the generative model, denoted by $p_\theta(z|x)$. This generative model learns a prior distribution, a joint distribution expressed as $p_\theta(x, z)$, employing the latent space and a stochastic decoder along with a prior distribution over the latent space.

3 Results and Discussion

To facilitate the discussion, we first delve into the characterization of synthetic microstructures that are generated with the HTP phase-field method. We then discuss the microstructure-thermal conductivity connection for the calculated set of microstructures with an emphasis on distinct types of microstructures. Next, we explain the efficient UP results and how to link PSP and inverse microstructure-centered materials design using the deep VAE method for designing for thermal conductivity of multi-phase composite materials.

3.1 Microstructure representation

Microstructure-based design faces two challenges: presenting high-dimensional microstructure images concisely, and determining the effective macroscopic properties linked to heterogeneous microstructures. A microstructure image of a multi-phase

material that contains bulk and interfacial features is a valuable source of information regarding the material's properties. Pixels in these images represent degrees of freedom, resulting in a remarkably complex information space.

Over the past decade, numerous studies have attempted to represent images as low-dimensional objects by reducing their complexity. An overview is provided here [52]. Besides conventional physical descriptors (e.g., area/volume fraction of phases), other features have been proposed to aggregate local information into global descriptors. A number of these are based on Haralick texture statistics [53], as well as local binary patterns, contrast features, visual bag of words [54], auto- and cross-correlation functions (e.g., two-point correlations). An example application of two-point correlations can be found in [55] and also articles published by a portion of the authors in this paper [56, 57].

In this study, we begin by computing a time-series of microstructural evolution to generate a 2-D microstructure dataset. Subsequently, we calculate the effective thermal conductivity at three selected time points during the evolution of the microstructures - the early, middle, and final stages. Besides determining the effective thermal conductivity of the microstructure, we also compute four physical microstructure descriptors: the area fractions for the 2D microstructure images (A_f), the characteristic length scales derived from Fourier transforms of the microstructure, and the bulk compositions of the yellow and blue phases, represented by c_{min} and c_{max} respectively. These basic microstructure descriptors are computed for assessing potential correlations between effective thermal conductivity and these descriptors. Figure 3(a) shows the probability density functions used to represent model I's input parameters. To reflect the physical limitations in the parameter range, these distributions are chosen to be truncated Gaussian. By way of example, the free energy model limits alloy composition within the 0.35-0.69 range. Furthermore, phase transformation must have a positive gradient energy coefficient, mobility, and barrier height. These limitations are imposed by setting the lower and upper bounds of parameter distributions accordingly.

Figure 3(b) illustrates a 15×15 microstructure palette representing the morphological variability of the computed microstructures. Figure 3(c) shows empirical cumulative density functions (eCDF) for the four different physical descriptors extracted. In time, the eCDFs change to reflect changes in the microstructure ensembles.

The area fraction of phases (A_f) is determined by applying a global image threshold using Otsu's method [58] to grayscale images. This method is particularly effective for semi- and fully-decomposed microstructures with phase area fractions greater than 0.2. For the estimation of characteristic length scale, we compute the circularly averaged Fast Fourier Transform (FFT) of the phase-field order parameter. As established in our previous work [11], this yields a unique curve for each microstructure at a fixed time, with the curve and its peak progressively shifting towards the right as the annealing time increases. To illustrate this, we present a graph of the empirical cumulative distribution functions (eCDFs) of effective thermal conductivity at early, middle, and late stages of evolution in Fig. 3(d).

3.2 Microstructure/Property Connection: Thermal Conductivity

This section examines the role of microstructure characteristics on effective thermal conductivity using the constitutive boundary value problem described in section 2.4. We assume that close to equilibrium, transport is well understood through Fourier's law, and we expect the thermodynamic fields to be smooth and thus to have good homogenization properties [59]. Precisely determining λ is challenging due to measurement difficulties and subtle microstructure variations. However, our focus is to employ thermal transport as a test problem, acknowledging that thermal conductivity is influenced by the phases' nature, volume, shape, size, and topology.

We initially examine two microstructures with distinct morphologies - circular and bicontinuous - with fixed phase area fractions (0.23 and 0.5). To understand the individual phase's thermal diffusivity role, we vary the precipitate thermal diffusivity to matrix phase ratio from 0.01, 0.1, 1, 10, to 100 and repeat thermal conductivity calculations. Figure 4(a and b) presents steady-state heat conduction calculations for these cases and 5 ratio calculations each. Figure 4(c) offers a comparison of these microstructures' thermal conductivity as a function of the thermal diffusivity ratio between phases, morphologies, and particle area fractions.

Both particle area fractions and phase thermal diffusivity significantly influence the effective thermal conductivity. For instance, in a microstructure with highly diffusive precipitates ($\alpha_{p_1}/\alpha_{p_2} > 1$), a circular morphology results in lower thermal conductivity than a bicontinuous one. In contrast, for a highly diffusive matrix phase ($\alpha_{p_1}/\alpha_{p_2} < 1$), the bicontinuous morphology exhibits reduced thermal conductivity. Here, $p_1 : \phi(r) = 1$ and $p_2 : \phi(r) = 1$ represent the two phases, with phase p_1 having a higher volume fraction. Also, the ratios of 0.1 and 10 essentially interchange the thermal diffusivity values of the phases. Results indicate that a material with circular precipitate morphology and a high thermal diffusivity ratio above 100 can decrease overall λ , which may be suitable for power generation or refrigeration. Additionally, there is approximately an order of magnitude difference in λ between circular and bicontinuous morphologies with the same thermal diffusivity ratio.

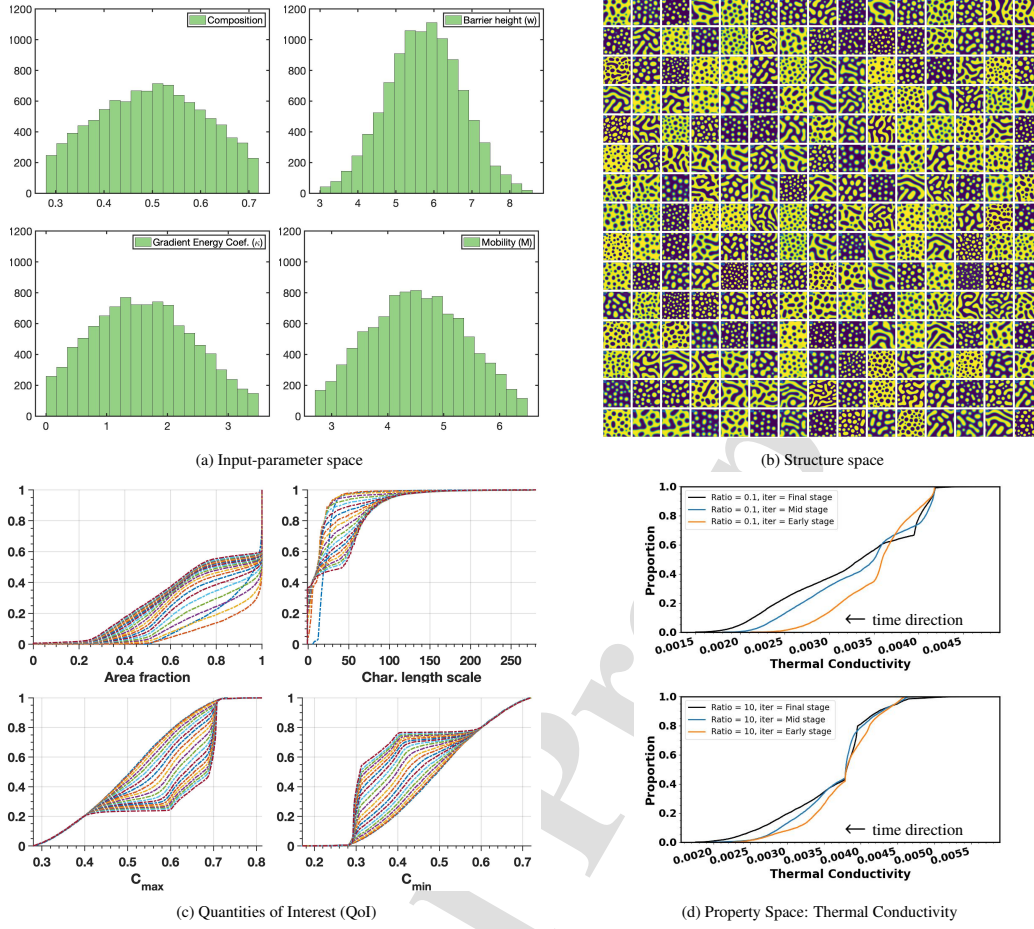


Figure 3: Proposal probability space for model I. (a) Proposed distributions for input parameters of the phase-field model, free energy model I (i.e., $[c^*, \rho_s, \kappa_c, M]$). (b) Typical microstructures obtained from solving Cahn-Hilliard equation using a symmetric double-well potential. (c) Transient eCDF plots of four QoI (i.e., Area fraction of phase α , radially averaged characteristic length scale, Max. and Min. concentrations in the domain (i.e., c_α, c_β) extracted from the 10,000 microstructures. Different eCDF lines show evolution in probability measure as a function of time for different sets of microstructure images synthetically annealed the same amount of time. (d) Transient eCDF plots of thermal conductivity for three sets of microstructures (each set contains 10,000 microstructures) for two cases of thermal conductivity ratios, $\alpha_{p1}/\alpha_{p2} = 0.1$ and $\alpha_{p1}/\alpha_{p2} = 10$. Set one corresponds to time 95.2 (early stage), set two corresponds to time 232.2 (mid stage), and set three corresponds to time 600 (final stage).

To delve deeper into the influence of phase fraction and characteristic length scale on λ , we ran heat conductivity calculations on 6,040 diverse dual-phase microstructures selected from 10,000 proposal samples. We obtained A_f and characteristic length scale from each microstructure's final simulation time-step. Figures 5(a-d) present 2D joint plots illustrating the impact of thermal diffusivity ratios ($\alpha_{p1}/\alpha_{p2} = 0.1$ and $\alpha_{p1}/\alpha_{p2} = 10$) on thermal conductivity, which ranges approximately between 0.002 and 0.006. In both cases, the lowest thermal conductivity is typically found in circular or semi-circular morphologies when the matrix phase has lower thermal diffusivity. The thermal conductivity is inversely related to the area fraction A_f for $\alpha_{p1}/\alpha_{p2} = 0.1$, and directly proportional for $\alpha_{p1}/\alpha_{p2} = 10$, implying an inversion in the pairplot distributions for different ratios.

Determining an immediate correlation in the characteristic length scale and thermal conductivity pair plots is difficult. Traditional pair plots do not provide direct visual insight into microstructures and often require further labeling, impeding the establishment of PSP relationships. Due to the limitations of classical pair plots, it is difficult to understand the effects of alloy composition or model parameters on thermal conductivity and we will conduct deep VAE to establish PSP in section 3.4. Our next objective is to create sequentially optimal sampling policies for the non-intrusive and efficient UP from model inputs to model outputs for ICME models.

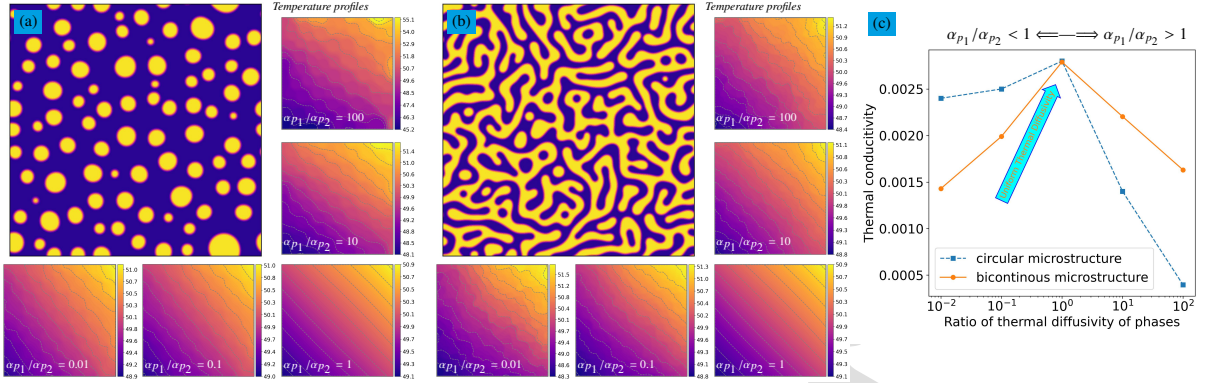


Figure 4: Example calculation of effective thermal conductivity for the microstructure cases shown in (a) circular and (b) bicontinuous. We have assumed five hypothetical materials in which the ratio of thermal diffusivity between the two phases are 0.01, 0.1, 1, 10 and 100. The heat flows from bottom-left corner to top-right corner, and the calculated steady-state temperature profiles for each ratio are shown next to each microstructure. (c) The change in effective thermal conductivity as a function of the ratio of thermal diffusivity of phases for the two microstructure cases and five hypothetical materials.

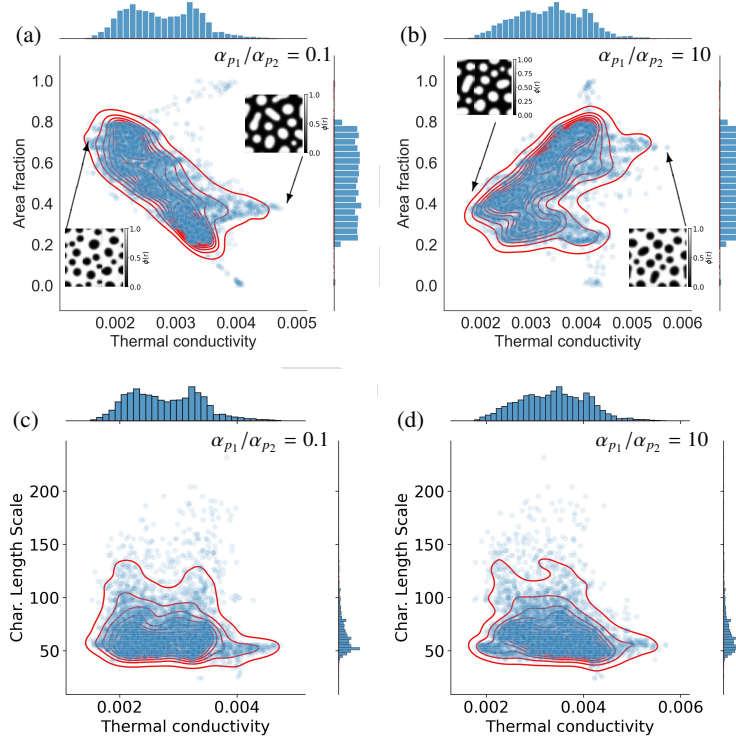


Figure 5: (a) and (b) 2D kernel density estimates for thermal conductivity and area fraction with data points overlaid over this density for cases (a) $\alpha_{p1}/\alpha_{p2} = 0.1$ and (b) $\alpha_{p1}/\alpha_{p2} = 10$. (c) and (d) 2D kernel density estimates for thermal conductivity and characteristic length scale with data points overlaid over this density for cases (c) $\alpha_{p1}/\alpha_{p2} = 0.1$ and (d) $\alpha_{p1}/\alpha_{p2} = 10$. In the plots, the time of evolution is fixed at 600.

3.3 Efficient Uncertainty Quantification via Change of Measure

Next, we implement a general sample-based technique [21] for change of probability measure, regardless of the distributions from which samples were generated from. The motivation here is that computational complexity and cost are serious bottlenecks for the design process and if input space (e.g., distribution range or type) changes, we can therefore transform the results to a new subspace.

Figure 6 presents the empirical cumulative distribution functions (eCDFs) for four quantities of interest (QoI) at early and final simulation stages for free energy models I and II. The overlay of target samples' eCDFs confirms the accurate representation of their distribution by our weighted proposal samples. For detailed insights into the evolution of the eCDF shapes for each QoI over time, see Figures S16 to S19 in the Supplementary Document.

Fig. 7 shows the eCDFs of the thermal conductivity that we obtained in the case of free energy models I. The goal here is to use the sample-based technique to estimate importance weights to determine the uncertainty in QoI or property space by weighting the proposal samples accordingly, regardless of the type of initial prior distributions. Note that a reasonable choice to generate proposal samples when there is little knowledge about the target sample distribution is the uniform distribution. However, this is not a requirement.

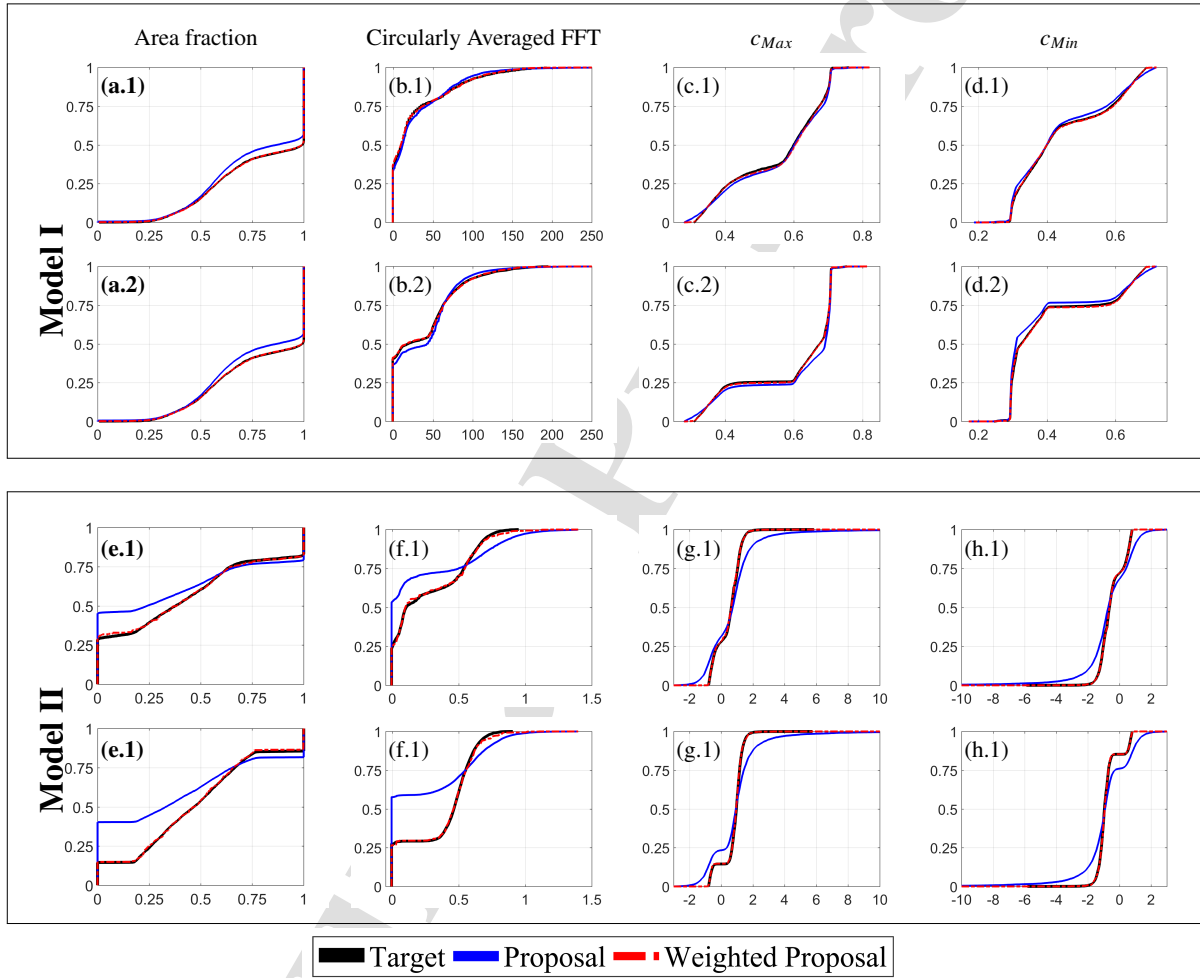


Figure 6: eCDFs corresponding to all QoIs at middle and final time increments for models I and II. Model I: (a.1) and (a.2) correspond to the first QoI (Area fraction) at middle and final states. (b.1) and (b.2) correspond to the second QoI (circularly averaged FFT of microstructure) at middle and final states. (c.1) and (c.2) correspond to the third QoI (c_{Max}) at middle and final states. (d.1) and (d.2) correspond to the fourth QoI (c_{Min}) at middle and final states. The same order applies for model II.

Here, we used 5,000 target and 10,000 proposal samples. These values might vary for different problems based on feature space dimensionality and input distributions. To validate the sufficiency of our proposal samples, we ran 100 simulations, each time randomly selecting samples. Starting from 500 samples and increasing in steps of 500, we calculated importance weights for each QoI. Instead of using eCDFs, we used the Wasserstein metric as a measure of similarity between the weighted proposal and target distributions. Lower metric values indicate more similarity, with zero indicating identical distributions. For more on the Wasserstein metric, see Ref. [60].

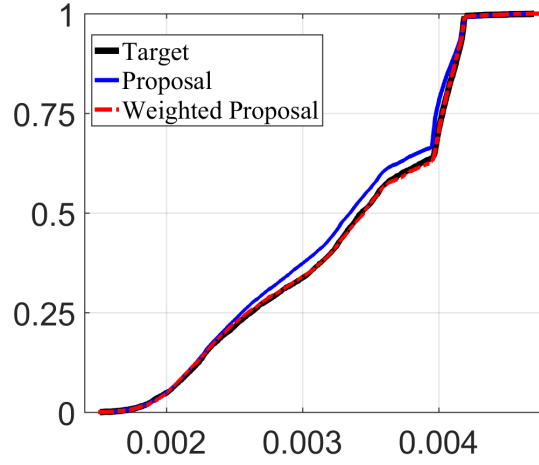


Figure 7: eCDFs corresponding to thermal conductivity at final time increments for model I.

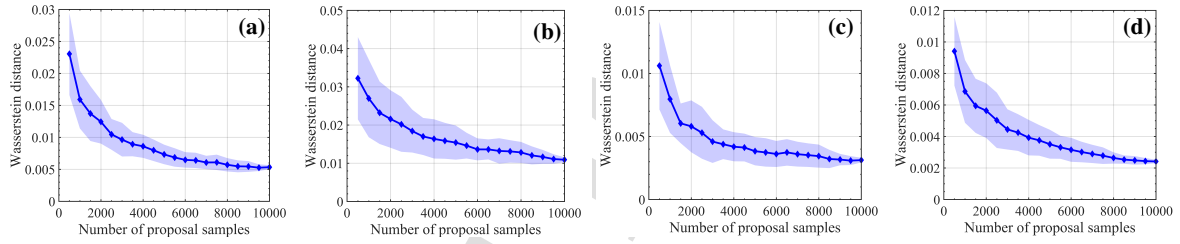


Figure 8: Mean and confidence interval of Wasserstein distances in model I between target and weighted proposal distributions of different QoI as a function of number of proposal samples resulted from 100 replication of simulations. (a) First QoI. (b) Second QoI. (c) Third QoI. (d) Fourth QoI. As the number of available proposal samples increases, target and weighted proposal distributions get more similar.

In Fig. 8, Wasserstein distances between target and weighted proposal for all four QoI are depicted versus number of proposal samples. Increasing the number of proposal samples results in better convergence to target distribution (as Wasserstein distance is decreasing). However, at some point, the improvement saturates, as the sampled points are already sufficient to produce the target distribution accurately.

In this section, we used advanced UP and high-throughput phase-field modeling to explore the microstructure design space in a forward manner. Our results suggest that weighed sampling approaches can yield very accurate descriptions of propagated uncertainty in complex models at a fraction of the expense necessary if one were to use brute force approaches.

In the past few years, advances in machine learning have culminated in sophisticated generative models that can create realistic novel instances of the system under investigation by learning from distributions over real examples. the focus of this effort was on *forward* propagation of information. An important aspect that remains to be explore in this work is how to exploit the available information to carry out *inverse* design.

Here, traditional MC-based UP generated close to 40,000 time-series of microstructure data (total of 40,000×40 images). This data provides sufficient information to train generative models (based in this case in VAEs). Provided the training sample is sufficient, generative models can greatly advance microstructure-centered materials design workflows, enabling the materials development in a fraction of the time needed in traditional methods. Our next section will explore the use of VAEs to link PSPs and design inverse microstructures for thermal conductivity. Yet, the framework is also useful for designing mechanical, electrical, or chemical properties.

3.4 PSP Relations and Inverse Microstructure-Centered Materials Design

Figure 9(a) displays the microstructure distribution in the latent space of the deep VAE model. Each microstructure image is positioned in the kernel Principal Component Analysis (kPCA) representation, capturing the first two components. The la-

latent space provides a qualitative environment to analyze correlations between model input variables and associated properties. By employing a basis transformation through a dense layer, the kPCA space combines the VAE images and input parameters. This representation reveals clear separations between microstructure classes, indicating significant correlations among the microstructures, the material parameters, and the resulting thermal conductivity. Microstructures in the kPCA space exhibit a clear correlation with the physical phase diagram: yellow precipitation is located on the left, dark precipitation on the right, and bicontinuous structures in the middle. Increasing $kPCA_1$ corresponds to a coarsening effect.

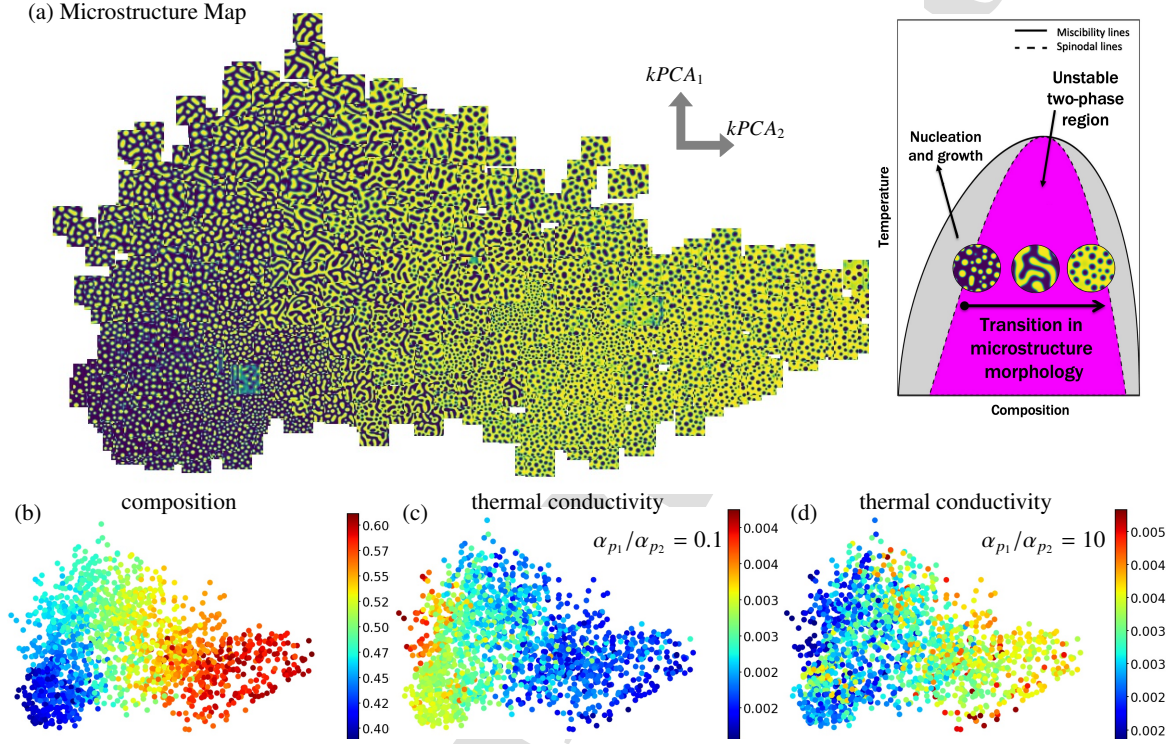


Figure 9: The visualization of composition-microstructure-thermal conductivity relations. (a) Microstructure map in the latent space of the VAE model showing the transition from yellow circular to bicontinuous and then dark circular microstructures with a yellow matrix. In the phase diagram, we see the same morphological transition moving from left to right in the unstable two-phase region. (b) The change in alloy composition as a function of the microstructure morphology. (c) and (d) The change in thermal conductivity as a function of the microstructure morphology for the cases of (c) $\alpha_{p1}/\alpha_{p2} = 0.1$ and (d) $\alpha_{p1}/\alpha_{p2} = 10$.

Figures 9(b-d) show the color-coded latent space, revealing the relationship between alloy composition and thermal conductivity for ratios of 0.1 and 10. The results indicate an inversion in thermal conductivity between the two cases, with outliers observed for the ratio 10. Specifically, when the matrix phase has lower thermal diffusivity ($\alpha_{p1}/\alpha_{p2} = 0.1$), increasing alloy composition leads to a decrease in thermal conductivity. The results reveal an inverted relationship for higher thermal diffusivity in the matrix phase ($\alpha_{p1}/\alpha_{p2} = 10$). Microstructures with finer particle morphology and lower matrix phase diffusivity exhibit lower thermal conductivity. The thermal conductivity of the alloy is influenced by alloy composition, microstructure morphology, volume fraction, characteristic length scale, and the thermal diffusivity of individual phases. Engineering the phase interface offers potential control over the alloy's thermal conductivity.

While the inference model and latent space characterization allowed us to establish PSP relationships, the generative model enables inverse design. With the use of the VAE decoder model, the material design process can now be inverted; instead of attempting to find the source of improved performance from existing data, the model can be used to generate new microstructures with the respective processing/material parameters that potentially have the desired performance. Figure 10(a) compares microstructures generated by the VAE decoder model with those from the original dataset, showing their close resemblance. The predicted estimates versus true target values in Fig. 10(b) and Fig. 10(c) demonstrate high accuracy with low residual MSE. The phase-field and VAE reconstructions in Fig. 10(d) indicate successful interpolation by the generative model. However, in extrapolation, the VAE may struggle to recognize the boundary between unstable dual-phase and nucleation/growth regions, suggesting the need for training with a wider range of microstructures to fully capture the underlying physics.

Figure 11(a) presents a composition-thermal conductivity diagram designed using the VAE model. The diagram allows for selecting alloy compositions with desired microstructures and thermal conductivity while fixing the other model parameters. The trend of alloy thermal conductivity as a function of the ratio of thermal diffusivity of the phases is depicted in Fig. 11(b). The design approach takes into account the influence of relative thermal diffusivity of phases on alloy microstructures and thermal conductivity. By understanding the relationship between alloy composition and thermal conductivity, the design identifies optimal alloy compositions with desired microstructures and thermal conductivity. This optimized design has practical applications in fabricating materials with enhanced thermal properties for specific uses. It is especially valuable for thermoelectric materials, where thermal conductivity plays a crucial role in performance. Through the optimization of alloy composition and microstructures, the design can improve power output and efficiency in thermoelectric materials.

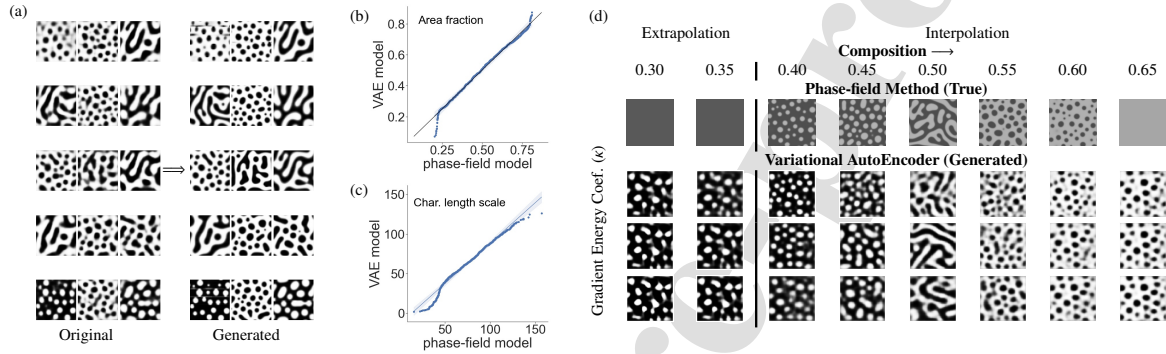


Figure 10: VAE model performance. (a) Analyzing the difference between actual and simulated microstructures as calculated by the phase-field method (original) and the VAE method (generated). (b) and (c) True (phase-field) vs. predicted (VAE) estimates of microstructures for extracted QoI (b) phase area-fraction and (c) characteristic length scale. The VAE exhibits excellent interpolation performance, while extrapolation cases show slight deviations (the tails). (d) Detailed view of the phase-field and VAE microstructures for different compositions and λ . When the composition is below 0.4, the alloy does not show miscibility and phase-field modeling does not predict morphology. VAE, however, extrapolates a circular morphology that is not physical.

4 Summary and Conclusion

In summary, we introduced a general-purpose machine learning framework for inverse microstructure-centered materials design for predicting the effective thermal conductivity of multi-phase materials. We also introduced the use of R-N derivatives to perform uncertainty propagation effectively when the design and in turn the range and type of data change. We built the framework on microstructure data generated with high-throughput phase-field modeling. The created synthetic microstructure data demonstrate spontaneously decomposing alloys that are curated on the Open Phase-field Microstructure Database, OPMD [36] (<http://microstructures.net>) and Zenodo platform [61] (<https://doi.org/10.5281/zenodo.7702179>) for community access. Overall, the methods proposed in this study allow generation of new microstructures with desired properties of the material as a function of material composition and other variables.

Our work uses a deep VAE to link alloy composition and model parameters to microstructures and consequently thermal conductivity. In the VAE model, the microstructure information (i.e. image) plays a fundamental role, and both material parameters and properties can be easily linked after successful training, leading to a process-structure-property relationship. Moreover, the VAE can utilize the learned representations to generate new microstructures as well as the creation of embeddings for microstructural parameters, allowing prediction of thermal conductivity. Additionally, deep VAE eliminates the need for reduction of microstructures to physical or non-physical descriptors.

In addition, we explored the microstructure design space using advanced UP in a forward manner. We found that weighted sampling approaches can yield accurate descriptions of propagated uncertainty in complex models. We sampled from desired probability distributions and as the number of available proposal samples increased, the Wasserstein distance between target and weighted proposal distributions got smaller, but at some point the improvement saturates. By implementing the R-N theorem here, we are able to change a measure from any set of evaluated data to a target measure without having to run new expensive experiments.

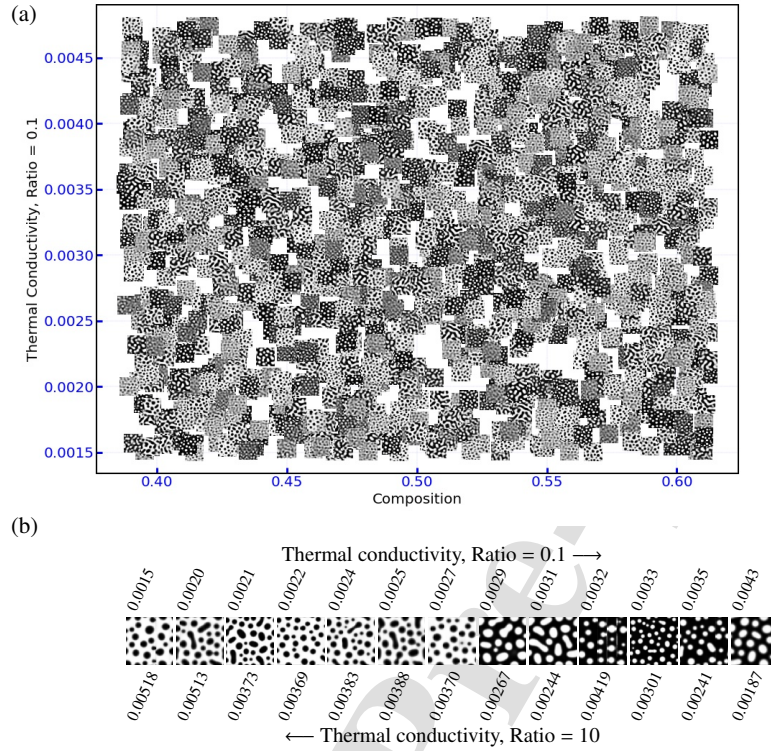


Figure 11: (a) A microstructure map predicted by the VAE model in the space of composition and thermal conductivity. We assume that the other parameters (i.e., barrier height, mobility, and gradient energy coef.) are fixed for this visualization. (b) The general trend in thermal conductivity as a function of ratios of thermal diffusivity of the phases (i.e., 0.1 and 10).

We found that circular morphologies with lower thermal diffusivity in the matrix phase had lower thermal conductivity. The change in thermal conductivity with respect to microstructure and corresponding material properties (e.g., alloy composition) are mapped in the latent space of the VAE allowing to establish PSP linkages. Additionally, the generator model is used to inversely design microstructures for a desired thermal conductivity. Thermoelectric designs should focus on creating a microstructure with as much of a difference in thermal diffusivity between the matrix and precipitate phases as possible, while also emphasizing a circular morphology to maximize the reduction of thermal conductivity. To further optimize thermoelectric performance, it is critical to consider the microstructure, its morphology, and the respective impacts on electrical and Seebeck coefficient.

While the current work represents a modest demonstration of the importance of uncertainty-aware analysis of PSP relationships, it points toward important future research directions. It would be interesting, for example, to examine how this framework can be incorporated into more general frameworks aiming at microstructure-sensitive design and optimization. Other potential avenues for research are to investigate how the present VAE model can be further improved. Several directions can be explored for improving this approach to microstructure design. There are many ways to extend the VAE models. You can change the prior, the posterior/variational posterior, regularize the posterior, and change the architecture. Changing the architecture includes changing the layers to RNNs/LSTMs/CNN layers, and using other divergence measures instead of KL divergence. Many of these variations often include convolutional layers, even if not explicitly stated. Additionally, the effect of local grain boundaries on the effective thermal conductivity of a material will be considered in future studies.

4.1 Data Handling

The dataset called OPMD version 1.0, contains 40,000×50 or 200,000 computationally generated microstructure images that are obtained from 40,000 distinct phase-field simulations where 50 time sequences is saved from each simulation. The dataset is divided into two equal-sized sets called proposal samples and target samples. All images are 1167×875 px² that composition

order parameter to a Red-Green-Blue (RGB) color scheme. The extraction of QoI is performed using the original raw data. The images were further resized to 64×64 before being fed into the VAE model. The microstructure dataset features a diverse set of chemical spinodal morphologies.

5 Data availability

A Data DOI is available for the image data that the VAE model in this manuscript relies upon: <https://doi.org/10.5281/zenodo.7702179> [62]. Alternatively, you can access the entire raw and image data at the Open Phase-field Microstructure Database (OPMD) website via the link <https://microstructures.net> [36].

6 Acknowledgments

The authors acknowledge the support of the National Science Foundation through Grant No. CDSE-2001333, as well as ARPA-E through contract DE-AR0001427. V.A. and R.A. thank the support by the National Science Foundation under NSF Grants No. DMR-1905325 and ARO through Grant No. W911NF-22-2-0117. V.A. also thanks Texas A&M University Data Science Institute (TAMIDS) for postdoctoral grant support that partially enabled this work. The authors would like to acknowledge the Terra supercomputing facility of the Texas A&M University for providing the computing resources to carry out the research reported in this article.

7 Authors' contributions

V.A. planned, developed, and implemented the high-throughput phase-field and machine learning models reported herein, processed the results, and drafted the manuscript. D.K. developed the efficient UP method and assisted in writing the paper. D.A. and R.A. supervised the project.

8 Authors' details

¹ Department of Materials Science and Engineering, Texas A&M University, 77840 College Station, TX, USA.

² Mechanical Engineering Department, Texas A&M University, 77840, College Station, TX, USA

9 Competing interests

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.

10 Additional information

Supplementary information: The online version contains supplementary material.
Correspondence and requests for materials should be addressed to Vahid Attari.

References

- [1] D. L. McDowell, R. A. LeSar, The need for microstructure informatics in process–structure–property relations, *MRS Bulletin* 41 (8) (2016) 587–593.
- [2] J. Hill, G. Mulholland, K. Persson, R. Seshadri, C. Wolverton, B. Meredig, Materials science with large-scale data and informatics: unlocking new opportunities, *Mrs Bulletin* 41 (5) (2016) 399–409.
- [3] D. Tourret, H. Liu, J. LLorca, Phase-field modeling of microstructure evolution: Recent applications, perspectives and challenges, *Progress in Materials Science* 123 (2022) 100810.
- [4] P. Honarmandi, R. Arróyave, Uncertainty quantification and propagation in computational materials science and simulation-assisted materials design, *Integrating Materials and Manufacturing Innovation* 9 (1) (2020) 103–143.
- [5] J. M. Rickman, T. Lookman, S. V. Kalinin, Materials informatics: From the atomic-level to the continuum, *Acta Materialia* 168 (2019) 473–510.
- [6] R. Salakhutdinov, Learning deep generative models, *Annual Review of Statistics and Its Application* 2 (2015) 361–385.
- [7] A. Molkeri, D. Khatamsaz, R. Couperthwaite, J. James, R. Arróyave, D. Allaire, A. Srivastava, On the importance of microstructure information in materials design: Psp vs pp, *Acta Materialia* 223 (2022) 117471.
- [8] D. G. Backman, D. Y. Wei, D. D. Whitis, M. B. Buczek, P. M. Finnigan, D. Gao, ICME at GE: accelerating the insertion of new materials and processes, *JOM* 58 (2006) 36–41.
- [9] N. H. Paulson, M. W. Priddy, D. L. McDowell, S. R. Kalidindi, Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics, *Acta Materialia* 129 (2017) 428–438.
- [10] R. N. Saunders, K. Teffer, A. Elwany, J. G. Michopoulos, D. Lagoudas, Metal AM process-structure-property relational linkages using gaussian process surrogates, *Additive Manufacturing* (2023) 103398.
- [11] P. Honarmandi, V. Attari, R. Arroyave, Accelerated materials design using batch bayesian optimization: A case study for solving the inverse problem from materials microstructure to process specification, *Computational Materials Science* 210 (2022) 111417.
- [12] J. H. Panchal, S. R. Kalidindi, D. L. McDowell, Key computational modeling issues in integrated computational materials engineering, *Computer-Aided Design* 45 (1) (2013) 4–25.
- [13] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, *Bayesian data analysis* chapman & hall, CRC Texts in Statistical Science (2004).
- [14] L. P. Swiler, Bayesian methods in engineering design problems., Tech. rep., Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA ... (2006).
- [15] R. Melchers, Importance sampling in structural systems, *Structural safety* 6 (1) (1989) 3–10.
- [16] C. G. Bucher, Adaptive sampling—an iterative fast monte carlo procedure, *Structural safety* 5 (2) (1988) 119–126.
- [17] P. Thoft-Cristensen, M. J. Baker, *Structural reliability theory and its applications*, Springer Science & Business Media, 2012.
- [18] D. Xiu, G. E. Karniadakis, Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos, *Computer methods in applied mechanics and engineering* 191 (43) (2002) 4927–4948.
- [19] D. H. Evans, An application of numerical integration techniques to statistical tolerancing, *Technometrics* 9 (3) (1967) 441–456.
- [20] S. Amaral, D. Allaire, K. Willcox, Optimal L_2 L_2 -norm empirical importance weights for the change of probability measure, *Statistics and Computing* 27 (3) (2017) 625–643.
- [21] M. Sanghvi, P. Honarmandi, V. Attari, T. Duong, R. Arroyave, D. L. Allaire, Uncertainty propagation via probability measure optimized importance weights with application to parametric materials models, in: *AIAA Scitech 2019 forum*, 2019, p. 0967.
- [22] T. Kloek, H. K. Van Dijk, Bayesian estimates of equation system parameters: an application of integration by monte carlo, *Econometrica: Journal of the Econometric Society* (1978) 1–19.
- [23] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, *arXiv preprint arXiv:1511.06434* (2015).
- [24] G. E. Hinton, Deep belief networks, *Scholarpedia* 4 (5) (2009) 5947.
- [25] R. Salakhutdinov, H. Larochelle, Efficient learning of deep boltzmann machines, in: *Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings*, 2010, pp. 693–700.
- [26] P. Dhariwal, A. Nichol, Diffusion models beat gans on image synthesis, *Advances in Neural Information Processing Systems* 34 (2021) 8780–8794.
- [27] D. P. Kingma, M. Welling, et al., An introduction to variational autoencoders, *Foundations and Trends® in Machine Learning* 12 (4) (2019) 307–392.
- [28] R. D. Bourgin, *Geometric aspects of convex sets with the Radon-Nikodym property*, Vol. 993, Springer, 2006.
- [29] L. A. Shepp, Radon-nikodym derivatives of gaussian measures, *The Annals of Mathematical Statistics* (1966) 321–354.

- [30] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules, *ACS central science* 4 (2) (2018) 268–276.
- [31] Y. Kim, H. K. Park, J. Jung, P. Asghari-Rad, S. Lee, J. Y. Kim, H. G. Jung, H. S. Kim, Exploration of optimal microstructure and mechanical properties in continuous microstructure space using a variational autoencoder, *Materials & Design* 202 (2021) 109544.
- [32] Z. Wang, W. Xian, M. R. Baccouche, H. Lanzerath, Y. Li, H. Xu, Design of phononic bandgap metamaterials based on gaussian mixture beta variational autoencoder and iterative model updating, *Journal of Mechanical Design* 144 (4) (2022) 041705.
- [33] Z. Pei, K. A. Rozman, Ö. N. Doğan, Y. Wen, N. Gao, E. A. Holm, J. A. Hawk, D. E. Alman, M. C. Gao, Machine-learning microstructure for inverse material design, *Advanced Science* 8 (23) (2021) 2101207.
- [34] V. Attari, R. Arroyave, Machine learning-assisted high-throughput exploration of interface energy space in multi-phase-field model with calphad potential, *Materials Theory* 6 (1) (2022) 1–20.
- [35] S. Sundar, V. Sundararaghavan, Database development and exploration of process–microstructure relationships using variational autoencoders, *Materials Today Communications* 25 (2020) 101201.
- [36] V. Attari, D. Saucedo, Open Phase-field Microstructure Database (OPMD), 2019 (2019).
- [37] S. Torquato, H. Haslach Jr, Random heterogeneous materials: microstructure and macroscopic properties, *Appl. Mech. Rev.* 55 (4) (2002) B62–B63.
- [38] J. W. Cahn, Phase separation by spinodal decomposition in isotropic systems, *The Journal of Chemical Physics* 42 (1) (1965) 93–99.
- [39] V. Attari, P. Honarmandi, T. Duong, D. J. Saucedo, D. Allaire, R. Arroyave, Uncertainty propagation in a multiscale CALPHAD-reinforced elastochemical phase-field model, *Acta Materialia* 183 (2020) 452–470.
- [40] L. Q. Chen, J. Shen, Applications of semi-implicit fourier-spectral method to phase field equations, *Computer Physics Communications* 108 (2-3) (1998) 147–158.
- [41] D. G. Cahill, H. E. Fischer, T. Klitsner, E. Swartz, R. Pohl, Thermal conductivity of thin films: Measurements and understanding, *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 7 (3) (1989) 1259–1266.
- [42] X. Lü, W. Shen, J. Chu, Size effect on the thermal conductivity of nanowires, *Journal of applied physics* 91 (3) (2002) 1542–1552.
- [43] P. Vaqueiro, A. V. Powell, Recent developments in nanostructured materials for high-performance thermoelectrics, *Journal of Materials Chemistry* 20 (43) (2010) 9577–9584.
- [44] A. Shakouri, Recent developments in semiconductor thermoelectric physics and materials, *Annual review of materials research* 41 (2011) 399–431.
- [45] E. S. Toberer, L. L. Baranowski, C. Dames, Advances in thermal conductivity, *Annual Review of Materials Research* 42 (2012) 179–209.
- [46] J. He, T. M. Tritt, Advances in thermoelectric materials research: Looking back and moving forward, *Science* 357 (6358) (2017) eaak9997.
- [47] D. M. Rowe, *Thermoelectrics handbook: macro to nano*, CRC press, 2018.
- [48] D. J. Bergman, O. Levy, Thermoelectric properties of a composite medium, *Journal of Applied Physics* 70 (11) (1991) 6821–6833.
- [49] Z. Bian, A. Shakouri, Beating the maximum cooling limit with graded thermoelectric materials, *Applied Physics Letters* 89 (21) (2006) 212101.
- [50] G. J. Snyder, T. S. Ursell, Thermoelectric efficiency and compatibility, *Physical review letters* 91 (14) (2003) 148301.
- [51] P. Billingsley, *Probability and measure*, John Wiley & Sons, 2008.
- [52] M. Larmuseau, M. Sluydts, K. Theuwissen, L. Duprez, T. Dhaene, S. Cottenier, Compact representations of microstructure images using triplet networks, *npj Computational Materials* 6 (1) (2020) 156.
- [53] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, *IEEE Transactions on Systems, Man, and Cybernetics SMC-3* (6) (1973) 610–621. doi : 10.1109/TSMC.1973.4309314.
- [54] B. L. DeCost, E. A. Holm, A computer vision approach for automated analysis and classification of microstructural image data, *Computational materials science* 110 (2015) 126–133.
- [55] D. T. Fullwood, S. R. Niezgoda, S. R. Kalidindi, Microstructure reconstructions from 2-point statistics using phase-recovery algorithms, *Acta Materialia* 56 (5) (2008) 942–948.
- [56] C. Kunselman, V. Attari, L. McClenny, U. Braga-Neto, R. Arroyave, Semi-supervised learning approaches to class assignment in ambiguous microstructures, *Acta Materialia* 188 (2020) 49–62.
- [57] C. Kunselman, S. Sheikh, M. Mikkelsen, V. Attari, R. Arróyave, Microstructure classification in the unsupervised context, *Acta Materialia* 223 (2022) 117434.
- [58] N. Otsu, A threshold selection method from gray-level histograms, *IEEE transactions on systems, man, and cybernetics* 9 (1) (1979) 62–66.

- [59] T. Bodineau, I. Gallagher, L. Saint-Raymond, A microscopic view of the fourier law, *Comptes Rendus Physique* 20 (5) (2019) 402–418.
- [60] C. Villani, The wasserstein distances, in: *Optimal transport*, Springer, 2009, pp. 93–111.
- [61] K. Nowak, L. H. Nielsen, A. T. Ioannidis Pantopikos, Zenodo, a free and open platform for preserving and sharing research output. (May 2016). doi:10.5281/zenodo.51902.
- [62] V. Attari, Opmd microstructure image data version 1.0 (Mar. 2023). doi:10.5281/zenodo.7702179.
URL <https://doi.org/10.5281/zenodo.7702179>

Declaration of interests

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: