ORIGINAL ARTICLE   **Open Access**

Machine learning-assisted high-throughput exploration of interface energy space in multi-phase-field model with CALPHAD potential

Vahid Attari[1][*] and Raymundo Arroyave[1,2]

*Correspondence:
attari.v@tamu.edu
[1]Materials Science and Engineering
Department, Texas A&M University,
College Station 77840, TX, USA
Full list of author information is
available at the end of the article

Abstract

Computational methods are increasingly being incorporated into the exploitation of microstructure–property relationships for microstructure-sensitive design of materials. In the present work, we propose non-intrusive materials informatics methods for the high-throughput exploration and analysis of a synthetic microstructure space using a machine learning-reinforced multi-phase-field modeling scheme. We specifically study the interface energy space as one of the most uncertain inputs in phase-field modeling and its impact on the shape and contact angle of a growing phase during heterogeneous solidification of secondary phase between solid and liquid phases. We evaluate and discuss methods for the study of sensitivity and propagation of uncertainty in these input parameters as reflected on the shape of the Cu_6Sn_5 intermetallic during growth over the Cu substrate inside the liquid Sn solder due to uncertain interface energies. The sensitivity results rank σ_{SI}, σ_{IL}, and σ_{IL}, respectively, as the most influential parameters on the shape of the intermetallic. Furthermore, we use variational autoencoder, a deep generative neural network method, and label spreading, a semi-supervised machine learning method for establishing correlations between inputs of outputs of the computational model. We clustered the microstructures into three categories ("wetting", "dewetting", and "invariant") using the label spreading method and compared it with the trend observed in the Young-Laplace equation. On the other hand, a structure map in the interface energy space is developed that shows σ_{SI} and σ_{SL} alter the shape of the intermetallic synchronously where an increase in the latter and decrease in the former changes the shape from dewetting structures to wetting structures. The study shows that the machine learning-reinforced phase-field method is a convenient approach to analyze microstructure design space in the framework of the ICME.

Keywords: Microstructure-sensitive materials design, Interface energy, Phase-field modeling, Uncertainty propagation, Global sensitivity analysis, Semi-supervised machine learning, Variational Autoencoder

Introduction

Computational materials science and engineering utilizes a diverse set of techniques across different length-scales to study the response of materials in different scales. These techniques often rely on solving partial differential equations (PDEs) where multiples of these PDEs are integrated to create multi-ensemble complex nonlinear spaces. Many of the emerging techniques do not have analytical solutions, and the reliability of the results may become a challenge due to the insufficiency of the numerical solvers. In materials science and engineering, the broad field of **microstructure-sensitive materials** design (Fullwood et al. 2010; McDowell 2018) and chemistry-sensitive materials design are examples of PDE-constrained optimal control problems with uncertainty in which interactions are complex and nonlocal. The curse of dimensionality in the microstructure-sensitive materials design problems calls for a more objective function measuring tool to predict the design space reasonably and accurately, avoiding full-factorial designs (Forrester et al. 2008) in order to establish accurate and quantitative Process-Structure-Property (PSP) links (Khatamsaz et al. 2021).

To successfully design the microstructure of materials for a desired property under the framework of Integrated Computational Materials Engineering (ICME), high through-put screening using the multi-scale, multi-ensemble phase-field simulations constitutes a plausible pathway (Chen 2002). In recent work (Attari et al. 2020), we have developed and demonstrated a microstructure design framework involving uncertainty quantification (UQ) and propagation (UP). The model and parameter uncertainty in phase-field models originate from the choice of the empirical potential/free energy functional. For instance, Cahn and Hilliard (1958) initially used a logarithmic potential for the free energy of unstable material. Very close to the critical point, one can approximate the logarithm-containing potential using an easier to handle polynomial expression. Far from the critical point below or when the resulting gradients in the temperature field are considerable, a non-smooth double-obstacle potential is introduced (Bosch 2016). Another source of parameter uncertainty is due to the selection of temperature-independent interface mobility, interface energies, and location-independent diffusion coefficient (Wang and McDowell 2020). UQ/UP in the microstructure modeling paves the road for high-throughput exploitation of the design space to reveal PSP relationships that are too complicated/complex. However, high-throughput screening requires high-throughput post-processing and analysis.

The primary objective of the current work is to show the myriad impact of phase-field based high-throughput screening of the material parameter space in combination with machine learning (ML) for high-throughput analysis of the results space to establish relations between inputs and the outputs. This is the key for establishing accurate and quantitative links between the PSP chain in materials. Specifically, we are interested in the interface energy space in heterogeneous phase growth as it is often uncertain or unknown.

In this work, we explore the effect of interface energy on heterogeneous phase growth over a substrate using a multi-phase-field model with CALPHAD potential that is reinforced with sensitivity analysis, semi-supervised machine learning method (Kunselman et al. 2020), and Variational Auto-Encoder (VAE) (Kingma and Welling 2013; Pu et al. 2016; Banko et al. 2020). Traditional supervised classifiers (Gola et al. 2018; Prakash et al. 2011) often suffer from nonphysical representation of the images and their performance

is limited by the assumptions it 'is told' to make. On the other hand, popular pre-trained convolutional neural network (CNN) layers (Choudhury et al. 2016; Niezgoda et al. 2008; DeCost et al. 2017) require large sets of labeled data to efficiently classify the images. Instead, we map structure diversity (phase-field model outputs) in the 3-D interface energy space (inputs) using a semi-supervised machine learning model that uses small amount of initially labeled data and VAE as a relatively new deep generative recognition model that does not require large datasets. Deep generative models have been successfully used in generating realistic-looking images, voices, or movies and gained attention in public domain (Kingma and Welling 2019; Ruthotto and Haber 2021).

This paper is structured as follows: The strategy for the propagation of uncertainty across the PDE-constrained PFM-model is explained in "Pragmatic modeling with PDEs" section. The details of the multi-phase-field method, and its uncertain inputs, and a desirable quantity of interest are described in "Computational model" section. In "Machine learning models" section, the high-throughput materials informatics procedure, including sensitivity analysis, sampling schemes, sem-supervised machine learning algorithm, and VAE, are described. In "Results" section, we discuss the results by motivating the present work through a summary of the treatment of heterogeneous phase growth using phase-field theory. Next, we discuss the uncertainty propagation strategy and the resulting sensitivity due to variation in inputs of the phase-field model. We also discuss the establishment of PSP linkage in inputs and outputs of the phase-field model by reinforcing semi-supervised machine learning and VAE as a generative deep recognition model. Furthermore, we present a summary of our findings and draw our conclusions in "Conclusion" section.

Pragmatic modeling with PDEs

The uncertainties in the model components such as material properties, coefficients, domain geometry, boundary, and initial data demand to take into account the noises in the inputs of the realistic simulations of complex systems governed by nonlinear PDEs. In practice, only statistical properties of the uncertain inputs are known, resulting in models involving PDEs having random inputs that are usually exactly or approximately represented in terms of a finite number of random parameters or variables. The strategy for high-throughput microstructure-sensitive exploration of process-structure-property chain in materials engineering is schematically represented in Fig. 1. At the most general level, we consider a material model governed by PDEs having random inputs, and use UP to determine information about the uncertainties in the outputs of the PDE, given information about the uncertainties in its inputs. Ultimately, a data-driven framework for establishing the links between the inputs and outputs is created to accelerate the process of microstructure-sensitive materials design and study.

Governing equations

For notational convenience, the exposition is restricted to boundary value problems. The framework is nevertheless applicable to general time dependent problems. Let $\mathcal{D} \in \mathbb{R}^d, d = 1, 2, 3$, be a fixed physical domain with boundary $\partial \mathcal{D}$, and $x = \{x_1, ..., x_d\}$ be the coordinates. Let us consider a PDE,
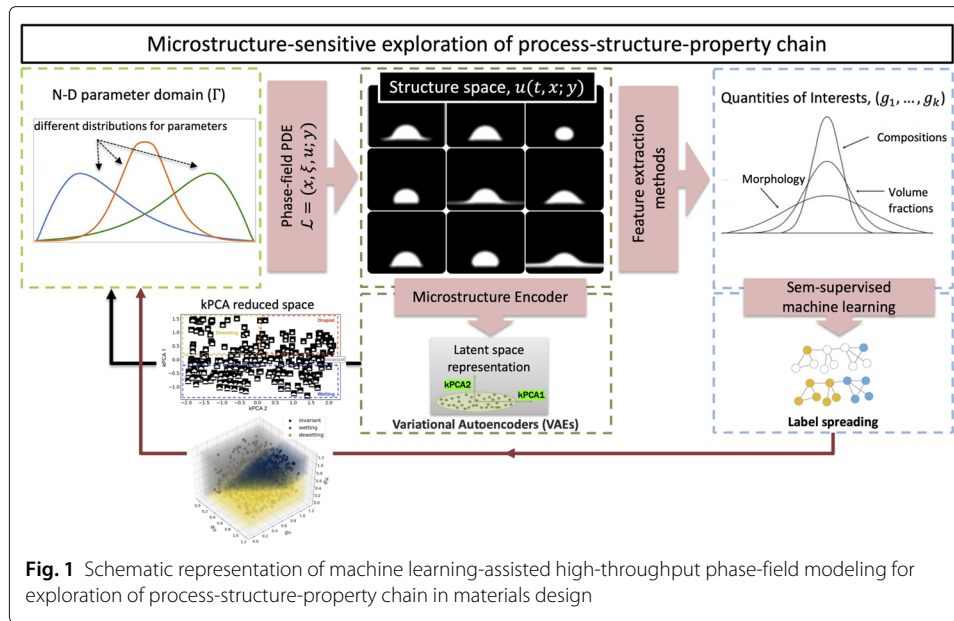
Fig. 1 Schematic representation of machine learning-assisted high-throughput phase-field modeling for exploration of process-structure-property chain in materials design

$$\mathcal{L}(x, u; y) = 0 \quad in \quad \mathcal{D};$$
$$B(x, u; y) = 0 \quad on \quad \partial\mathcal{D};$$
$$\tag{1}$$

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. The solution of the stochastic computation to this problem set is:

$$u = u(t, x; y) \tag{2}$$

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 or boundary conditions or random force terms. The random PDE reads as

$$\mathcal{L}(x, \xi, u; y) = 0 \quad in \quad \mathcal{D}; \tag{3}$$

where ξ denotes a (Gaussian) space-time white noise in $\mathcal{D}$. The random or stochastic Cahn-Hilliard PDE is given as:

$$\partial_t u = \Delta^2 u + P(u) + \xi, \tag{4}$$

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,

$$g = (g_1, \cdots, g_K) \in \mathbb{R} \tag{5}$$

called *Quantities of Interests* (QoIs), that are functions of the solution u, in addition to the solution itself. The QoIs are model outputs that may be the solution of the PDE itself but more often are functionals of that solution and contain the statistical information about the output of interest. QoIs often have to be represented using high-dimensional set of parameters. These outputs of interest in the phase-field model may be interpreted as a function of the PDE solution or statistical information of such solution. UQ/UP enables

determining statistical information about outputs of the phase-field model given statistical information about the inputs. In the next section, we describe the model, the uncertain inputs and the QoI that we are interested in.

Computational model

The applied multi-phase-field method

In this study, a multi-phase-field formalism is used to study the evolution of microstructure at isobaric and isothermal state. A set of non-conserved ($0 \leq \phi_i \leq 1$) and conserved variables ($0 \leq c_i \leq 1$) describe the components of the microstructure, where the non-conserved variables define the spatial fraction of available phases over the domain (Ω) and the conserved variables define the phase compositions. The total free energy of chemically heterogeneous material involves interfacial and bulk interactions:

$$\mathcal{F}^{tot} = \int_\Omega \left[f^{int} + f^{bulk} \right] d\Omega \tag{6}$$

where the two contributing factors are respectively formulated as:

$$\begin{cases} f^{int} = \sum_i \sum_{j>i} \frac{4\sigma_{ij}}{\eta_{ij}} \left[-\frac{\eta_{ij}^2}{\pi^2} \nabla\phi_i.\nabla\phi_j + |\phi_i\phi_j| \right] \\ f^{bulk} = \sum_i \phi_i f_i^0(c_i) \end{cases} \tag{7}$$

with σ_{ij} as interface energy coefficient, η_{ij} as the interface width, and $|\phi_i\phi_j|$ as the double obstacle potential. f_i^0 is the free energy of the homogeneous phase i and c_i is it's molar concentration. In this study, all concentrations are based on molar concentration of Sn in each phase. Using the model of the total free energy as a function of the field variables (ϕ_i) and (c), the following forms of the kinetic equations (phase-field and diffusion) are postulated as the governing equations:

$$\frac{\partial \phi_i}{\partial t} = -\sum_{i \neq j} \frac{M_{ij}}{N_p} \left[\frac{\delta \mathcal{F}^{tot}}{\delta \phi_i} - \frac{\delta \mathcal{F}^{tot}}{\delta \phi_j} \right] \tag{8}$$

$$\frac{\partial c}{\partial t} = \nabla. \left[D(\vec{\phi}) \sum_{i=1}^N \phi_i \nabla c_i \right] \tag{9}$$

where M_{ij} is the interface mobility, and N_p is the number of coexisting phases at the neighboring grid points. The phase-field equation works only in the interface where ϕ_i changes between 0 and 1. Interdiffusivity (D) is defined as a function of the phase-field order parameter to easily take into account the diffusivity in various features of the microstructure (i.e., interfaces, GBs, and bulk phases). The grid points in the solution domain may comprise a mixture of different phases. The coexistence of the phases in the interface is stated by the condition of equality of chemical potentials and the mass conservation as:

$$f_{c_1}^1[c_1(x,t)] = f_{c_2}^2[c_2(x,t)] = ... = f_{c_N}^N[c_N(x,t)] \tag{10}$$

$$c(x,t) = \sum_{i=1}^N \phi_i c_i \tag{11}$$

where $f_{c_N}^N$ stands as the derivative of the free energy with respect to the composition of the phase N.

Model parameters and uncertain inputs

In this study, the free energy is modeled using logarithmic potentials in CALPHAD formalism (Shim et al. 1996). In the CALPHAD method, the constant terms in the free energy model are optimized against experimental equilibrium data, though, uncertainty data does not exist. Both diffusion and interfacial mobility are temperature- and location-variant in this study and the parameters are reported in (Attari et al. 2018). On the other hand, surface energy perhaps is the most uncertain and less revealed parameter. It is defined as the surface excess free energy per unit area of a particular crystal facet (Vitos et al. 1998). Experimentally, it can be obtained by measuring surface tension, and computationally, it can be obtained from first-principles (De Waele et al. 2016) or by semi-empirical methods (Vitos et al. 1998). It varies by temperature and pressure and exact determination is hard. In equilibrium, the relevant heterogeneous interface free energies are related to each other by the Young-Laplace (YL) equation (Liu and Cao 2016):
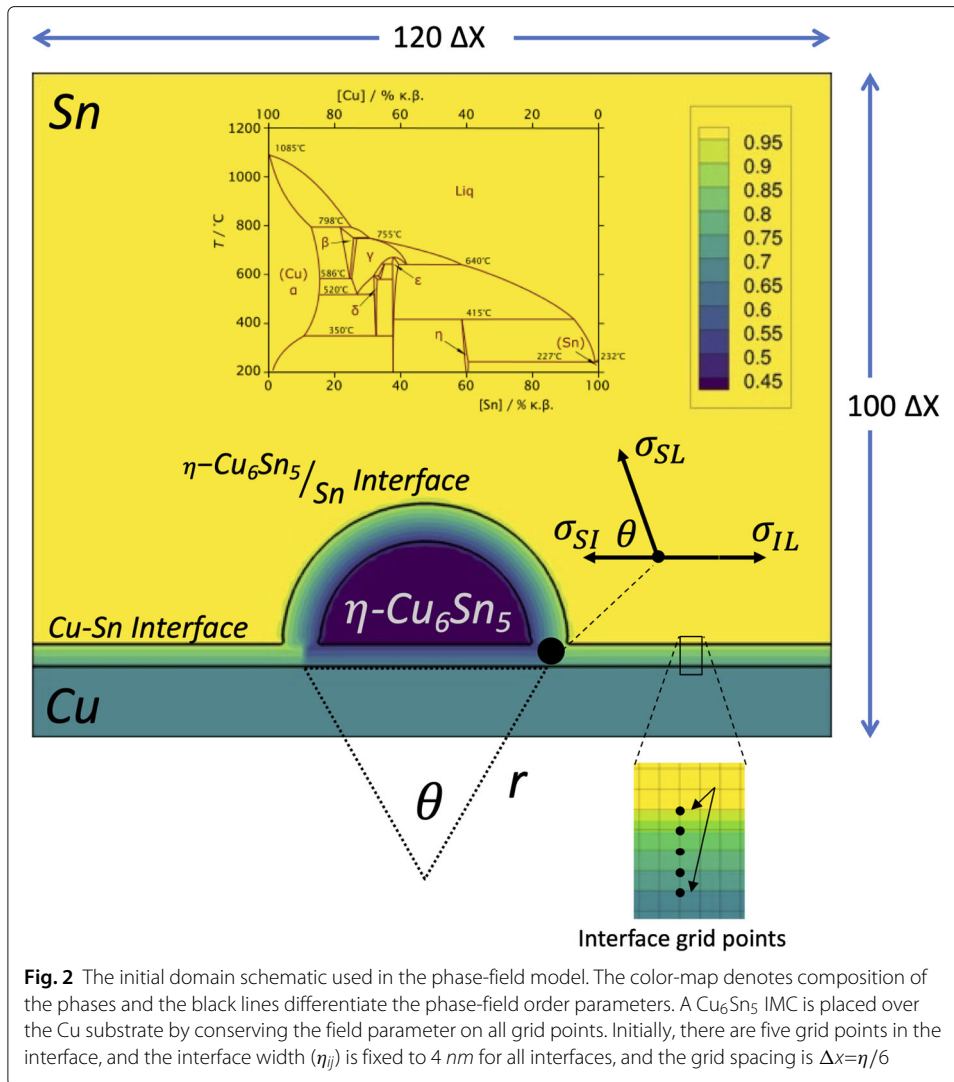
$$\sigma_{SL} = \sigma_{SI} + \sigma_{IL} Cos(\theta) \tag{12}$$

where σ_{SL}, σ_{SI}, and σ_{IL} stand for the solid-liquid, solid-IMC (Intermetallic), and IMC-liquid interfacial free energies, respectively, whereas θ is the contact angle. On the other hand, the YL equation fails when there is a contact angle hysteresis since it dictates only one equilibrium for a homogeneous phase on a perfectly flat surface (Good 1979). A contact angle hysteresis can be defined for real materials based on the advancing contact angle and the receding contact angle. High-throughput screening of the interface energy space using the multi-phase-field modeling reinforced with high-throughput machine learning analysis to exploit the existing patterns can help gain further insight into the physics of different IMC shapes and wetting characteristics.

To prevent unattended dynamics of particle pushing, we initiate the particle in a 2D simulation cell in the form of a half-circular particle with a fixed shape and size. The simulation cell size is 120×100, and the initial state of the microstructural domain is shown in Fig. 2. The color map denotes the composition of the phases, and the black lines differentiate the phase-field order parameters. The initial composition of each phase is set to it's equilibrium value. As the close-up in the lower right corner of the figure shows, five grid points exist in the interface at the simulation start. The phase diagram of the Cu/Sn reacting system denoting the important IMCs is also represented in this figure. In the simulations, L_x and L_y are 79.3 and 66 nm, as $\Delta x = 0.66$ nm whereas the time step has been $\Delta t = 0.4 \times 10^{-8} s$. The domain is assumed to be at the constant temperature of 250°C, and we don't consider any heat release during crystallization. We solve the governing Eqs. (8) and (9) using an explicit finite difference solver discretized in space and time. The boundary conditions are periodic at the right and left side of this domain and Neumann with zero flux at the top and bottom.

Quantity of interest (Fourier descriptor)

Since we are interested in the shape of the IMC and wetting characteristics, we use Fourier descriptor (FD) analysis (Persoon and Fu 1977; Bowman et al. 2001) to obtain a compact

Fig. 2 The initial domain schematic used in the phase-field model. The color-map denotes composition of the phases and the black lines differentiate the phase-field order parameters. A Cu_6Sn_5 IMC is placed over the Cu substrate by conserving the field parameter on all grid points. Initially, there are five grid points in the interface, and the interface width (η_{ij}) is fixed to 4 *nm* for all interfaces, and the grid spacing is $\Delta x = \eta/6$

representation of IMC shape, universal enough to store sufficient information about the shape. This approach uses border of 2D particles plotted in polar coordinates r and θ where the origin is the particle centroid. The boundary of the particle is circumnavigated in the complex plane at constant speed. The step size is chosen such that one circumnavigation takes time 2θ and the number of steps is $2k$. The concise Cartesian coordinates (x_k, y_k) of the IMC edge is extracted using the phase-field order parameter (ϕ^{IMC}) and by looking for local maxima of the gradients in the coordinates. The coordinates (x_k, y_k) of the edge is represented as a complex vector, and the discrete inverse Fourier transform of this complex vector taken as:

$$x_m + iy_m = \sum_{k=-N/2+1}^{N/2} (x_k + iy_k) exp\left(-\frac{2\pi i}{N}k\mu\right) \tag{13}$$

where N is the total number of descriptors; k is the descriptor number; N is the total number of points describing the particle; m is the index number of a point on the particle and i denotes an imaginary number.
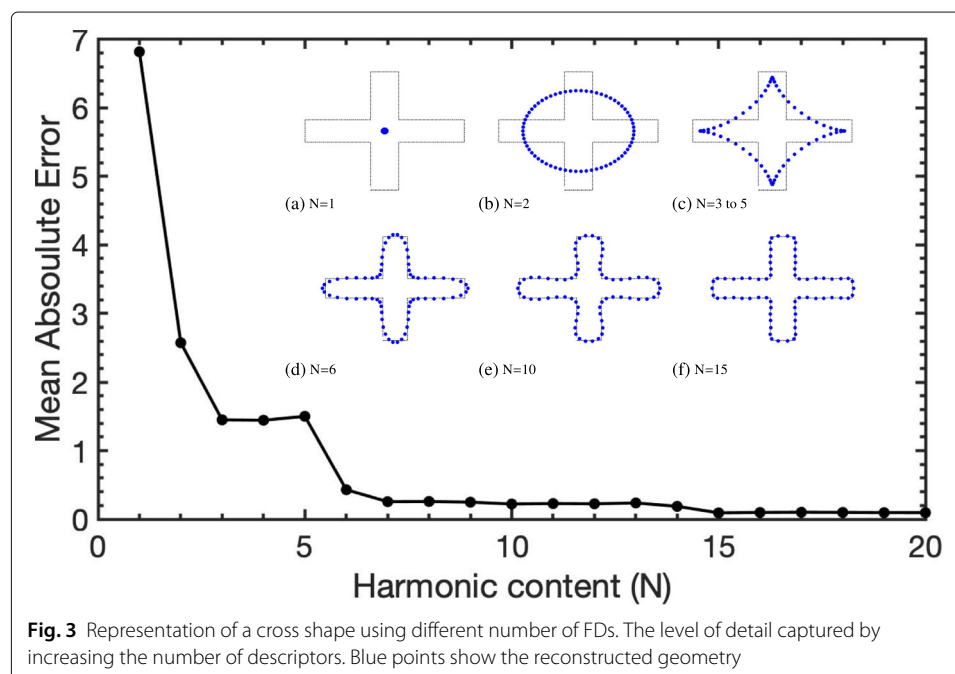
We normalize all the FDs obtained by the largest descriptor to obtain k invariant results with respect to position, size, and orientation. Hence, the shape will be represented by a series of independent coefficients in FD analysis in an effective and efficient way. The output from the FFT is a vector of n complex numbers, $f(k)$, and their magnitude $c(k) = |f(k)|$ allows to achieve rotation invariance, as the phase information is ignored. Figure 3(a–f) show reconstruction of a cross shape using different number of FDs. As the number of FDs increases, the level of captured detail during reconstruction process also increases. Figure 3 shows the mean standard error (MAE) of the position of reconstructed points with the original shape. It is also shown that to accurately represent corners, we need to include a greater number of descriptors.

Machine learning models

The sobol' sensitivity analysis

Assuming a non-linear and non-monotonic model function of form of $Y = f(x_n)$, where the input $x = (x_1, ..., x_n)$ is an n-dimensional vector of input variables and Y is a scalar quantity of interest, sensitivity analysis studies relationships between model inputs and outputs. Sensitivity analysis are often categorized based on if i) the approach is deterministic versus stochastic or if the method is ii) global or local. Local sensitivity analysis evaluates changes in the model outputs for variations in a single parameter input. In contrast, in global sensitivity analysis, the simultaneous variation of all model parameters allows to evaluate the relative contributions of each individual parameter, and the interactions in between to the model output variance. Given that the model inputs can span a wide range, global sensitivity analysis can determine which reactions and processes contribute most to the overall response system.

Among many global sensitivity analysis methods, variance decomposition-based Sobol' sensitivity analysis is so far one of the most powerful techniques (Zhang et al. 2015) and

Fig. 3 Representation of a cross shape using different number of FDs. The level of detail captured by increasing the number of descriptors. Blue points show the reconstructed geometry

we use this technique to evaluate the global sensitivity in the interface energies of the multi-phase-field model. An important distinction between Sobol' and classical sensitivity is that the Sobol' sensitivity analysis detects interactions of input variables through the second and higher order terms, whilst classical sensitivity methods give only derivatives with respect to single variables. The influences of higher order interactions (e.g., the influence of doubles) on the monitored output are significant in systems comprising more members.

The sensitivity analysis of the IMC shape to three interface energy parameters was evaluated using Sobol' sensitivity analysis in this study. The Sobol' first-order sensitivity index is written in the form:

$$S_i = \frac{V\left(E(Y|X_i)\right)}{V(Y)} \tag{14}$$

where S_i measures the main effect of X_i on the model output Y. V is variance and E is expectation symbol. Interactions between the model input parameters (interface energy parameters ($\gamma_{\alpha\beta}$) determine important features of the multi-phase-field model, and are more difficult to detect than first-order effects. The Sobol' second-order sensitivity index given as:

$$S_{ij} = \frac{V\left(E(Y|X_i, X_j)\right)}{V(Y)} - S_i - S_j \quad (\text{for} \quad i \neq j) \tag{15}$$

measures the interaction effect between pairs (X_i, X_j). $V(E(Y|X_i, X_j))$ measures the joint effect of the pair (X_i, X_j) on output Y. In other words, S_{ij} is the variation in Y due to X_i and X_j that cannot be explained by the sum of individual effects of X_i and X_j. An analogous formula for higher-order Sobol' sensitivity terms is:

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \sum_i \sum_{j>i} \sum_{k>j} S_{ijk} + \dots + S_{123\dots M} = 1 \tag{16}$$

The sum of all sensitivity indices must be equal to 1. The number of terms in Eq. (16) increases exponentially with the number of input variables M. The total sensitivity index S_{T_i} is defined as the sum of all sensitivity indices:

$$S_{T_i} = \frac{E\left(V(Y|X_{\sim i})\right)}{V(Y)} = 1 - \frac{E\left(V(Y|X_{\sim j})\right)}{V(Y)} \tag{17}$$

Where $X_{\sim i}$ are all input variables that do not include the index i.

Sampling

In this study, we use Saltelli sampler as implemented in SALib python package (Usher et al.) instead of conventional Monte Carlo sampling. While the latter is in general robust, it demands extensive computational resources. Saltelli (Saltelli 2002; Saltelli et al. 2010) proposed two methods for overcoming the curse of dimensionality in the full-factorial design of experiments. In the first case, one may analyze at a reduced computational cost all effects of the first and total order, plus all those of order $k-2$, at the cost of $n(k+2)$ simulations. The second case demands an excessively higher number of function evaluations ($n(2k+2)$) than the first case. However, it estimates the index of the first and total orders

plus all indices of order 2 and $k-2$. Although the second case is computationally more expensive, it provides more insight into pair interactions. We used the second method to generate four different sample sets of the three interface energy parameters with 8, 120, 520, and 1024 samples, respectively. Table 1 summarizes the Saltelli sampling parameters.

Learning the microstructures with local and global consistency

The forward propagation of Sobol samples resulted in 1672 simulations (all 4 cases) with transient data for each that will be used for further analysis of the effect of model input parameters on the particle shape. A semi-supervised learning method called label spreading (Zhou et al. 2003) is used first to train the data by only labeling a small portion. The semi-supervised estimators can use this additional unlabeled data to better capture the shape of the underlying data distribution and generalize better to new samples (Kunselman et al. 2020). Since labeling often requires human labor, whereas unlabeled data is far easier to obtain, semi-supervised learning is beneficial in many real-world problems, and it has recently attracted a considerable amount of research.

Given a microstructure-set defined by scalar quantity of interest $QoI = \mathcal{X} = \{x_1, ..., x_l, x_{l+1}, ..., x_n\} \subset \mathbb{R}^m$ and a label-set $\mathcal{L} = \{1, ..., c\}$, the first l microstructures have labels $\{y_1, ..., y_l\} \in \mathcal{L}$ and the remaining microstructures are unlabeled. The goal is to predict the labels of these unlabeled microstructures with the assumption of the consistency of the scalar QoI extracted from the microstructure data, which means: (1) nearby points are likely to have the same label, and (2) points on the same cluster are likely to have the same label.

The LS technique considers a finite weighted graph $G = (V, E, \omega)$, consisting of a set of nodes, $V = \{1, 2, ..., n\}$ based on a data set $QoI = \mathcal{X} = \{x_i, i = 1, ..., n\}$, a set of edges $E \subseteq V \times V$, and edge-weight function $\omega(i, j) > 0$. The weight function $\omega(i, j)$ is interpreted as a similarity measure between the vertices x_i and x_j, and it is defined on G using a distance metric ρ:

$$\omega(i, j) = h\left(\frac{\rho(x_i, x_j)^2}{\mu \sigma^2}\right), \tag{18}$$

for some function h with exponential decay at infinity (A common choice is $h(x) = e^{-x}$). μ and σ are hyper-parameters and σ is learned by the mean distance to K-Nearest Neighborhoods. The initial labels are represented by membership vectors in an $n \times c$ matrix Y, where $Y_{i,l} = 1$ if node i has initial label l and $Y_{i,l} = 0$ otherwise. Given an initial guess $F^{(0)} \in \mathbb{R}^{n \times c}$, the label spreading algorithm iteratively computes:

$$F^{(r+1)} = \beta S F^{(r)} + \gamma Y \quad r = 0, 1, 2, ... \tag{19}$$

Table 1 Saltelli sampler parameters

	n (Evaluation arg.)	k (# of inputs)	$n(2k+2)$ (Sample-set size)
case 1	1	3	8
case 2	15		120
case 3	65		520
case 4	128		1024

**Table 2** The algorithm for the label spreading semi-supervised method (Zhou et al. 2003)

1. Form the affinity matrix ω defined by $\omega_{ij} = exp(-|x_i - x_j|^2/2\sigma^2)$ if $i \neq j$ and $\omega_{ii} = 0$.

2. Construct the matrix $S = D^{-1/2}\omega D^{-1/2}$ in which D is a diagonal matrix with its (i, i)-element equal to the sum of the ith row of ω.

3. Iterate $F(r + 1) = \beta SF(r) + (1 - \beta)Y$ until convergence where β is a parameter in (0, 1).

4. Let F^* denote the limit of the sequence $\{F(r)\}$. Label each point x_i as a label $y_i = argmax_{j \leq c}\tilde{F}_{ij}^*$.

with $\beta, \gamma \geqslant 0$ and $\beta + \gamma = 1$. β is clamping factor and it's value is in (0, 1) that specifies the relative amount that an instance should adopt the information from its neighbors as opposed to its initial label. $\beta = 0$ means keeping the initial label information; $\beta = 1$ means replacing all initial information. The iterates converge to the solution of the linear system $(1 - \beta S)F^* = \gamma Y$, but in practice, a few iterations of Eq. (19) with the initial point $F^{(0)} = Y$ suffices. This yields an approximate solution $\tilde{F}^*$. The prediction on an unlabeled node j is then $argmax_l \tilde{F}_{j,l}^*$. The algorithm is given in Table 2.

Variational auto-Encoder

The conventional Convolutional Neural Network (CNN) methods for image recognition perform extensive image labeling in a supervised way and require a large amount of data to determine desired classes, e.g., cat and non-cat classes. In this study, to determine the correlations between inputs and outputs of the phase-field model, we use the VAE approach, a class of deep generative methods that overcome this limitation by using a dimensionally much smaller latent space than the input data space. The VAE consists of a neural network image encoder and image decoder, and a latent space. In this approach, we convert the input data into an encoded vector where each dimension represents some learned attribute about the data. The encoder network often results in a single value for each encoding dimension, and the decoder network subsequently uses these values to reconstruct the original input.

Figure 4 represents a general schematic of the structure of the VAE where a five-layer 2D convolutional neural network is used as an encoder, and the output of the last layer is flattened. The phase-field synthetic images were used as an input for the encoder, and we trained the generator for the microstructure images on hand. Still, we didn't use the decoder to reconstruct new images. The input images only illustrate IMC phase-field order parameter that reflects the IMC shape, and the composition data abandoned. The images were cropped and binarized before the training step. We are interested in learning the IMC structure space to establish correlations between the inputs and outputs of
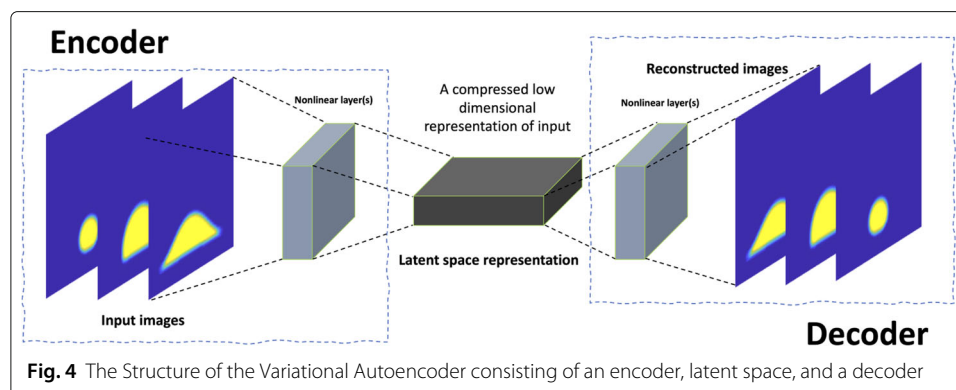
Fig. 4 The Structure of the Variational Autoencoder consisting of an encoder, latent space, and a decoder

the model in kernel Principal Component Analysis (kPCA) space (Schölkopf et al. 1997). The decoder section of the VAE is omitted in this study and will be addressed in a future article.

Results

In this section, we first motivate the work by discussing the computational treatment of heterogeneous phase growth in the Cu/Sn reacting system in "Treatment of heterogeneous phase growth using phase-field method" section. Next, in "Response variability" section, we explain the response variability in interface energy space using the multi-phase-field model and YL equation and then the sensitivity results of FD analysis for characterization of the IMC shape. We then explain the results of semi-supervised learning of the microstructures with respect to contact angle classes in interfacial energy space in "Effect of interface energies on contact angle" section. The learning of microstructural data using VAE and correlations between inputs and outputs of multi-phase-field is explained in "Visualization of interface energy–microstructure relations using VAE" section.

Treatment of heterogeneous phase growth using phase-field method

Heterogeneous phase nucleation and growth is the case which a secondary solid phase nucleates between a solid and liquid medium. It has been studied using single-, two-, and multi-phase-field methods in a number of articles (Granasy et al. 2019). Traditionally, Cahn has treated this by considering a foreign wall in the single phase-field formulation (Cahn 1977). Granasy et al. (2007) has proposed two distinct models where the model (I) assumes presence of a flat wall does not perturb the structure of the equilibrium between solid and liquid phase. In model (II), the barrier height of nucleation is overcome employing surface spinodal in which the undercooling is a function of the phase-field value. Similarly, heterogeneous nucleation is realized in binary materials, too. A thorough review of this literature is beyond the scope of the paper, and we limit ourselves to just these references to provide some illustrative examples (Granasy et al. 2019). In the framework of multi-phase-field modeling, a second solid phase is realized instead of the wall. In this method, a supercritical seed of the heterogeneous phase is either place by hand in the domain or activated through induced fluctuations (Azizi et al. 2021; Attari and Arroyave 2016). In multi-phase-field formalism, the user often doesn't have to calibrate the combination of interface energies to approximate shape growth for the material. This task can be pretty complicated as heterogeneous liquid-state reactions are not only influenced by chemical driving forces but also by the interfacial and elastic energy contributions. Following the guidelines in "Pragmatic modeling with PDEs" section, we use an isotropic multi-phase-field model with adjustable interfacial energies as uncertain inputs to reasonably and accurately map the inputs to the structure of IMCs for proper understanding of the formation of Cu_6Sn_5 IMCs in Cu/Sn reacting systems. The initial morphology and size of the Cu_6Sn_5 intermetallic in experiments (although Cu_3Sn is likely present but not resolvable in SEM) is often hard to determine and is dependent on many factors including the cooling rate in solder materials (Deng et al. 2003). The initial size of the IMC in the simulation domain is taken to be sufficiently larger than the critical size that can be calculated using the classical nucleation theory. Herein, we investigate a fundamental role of
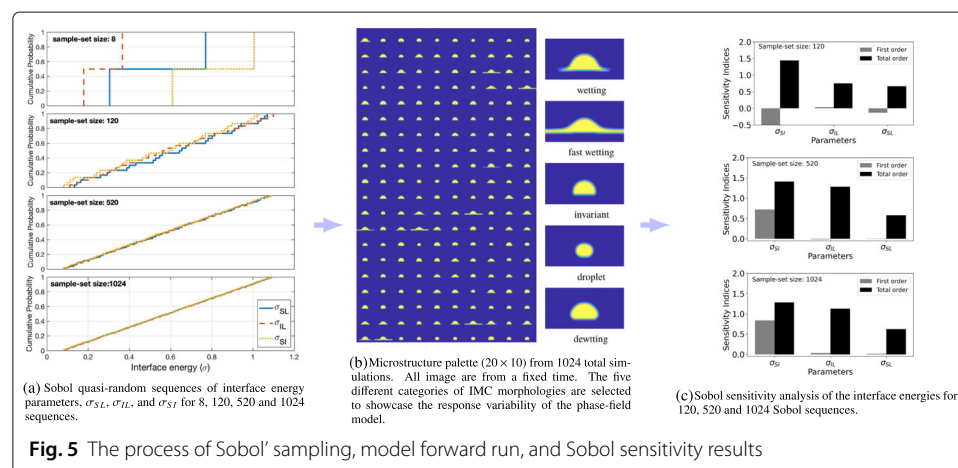
the surface energy on heterogeneous growth of an intermetallic during the solidification process using the phase-field method.

Response variability

Figure (5) shows the process of propagation of uncertainties in input space (interface energy parameters) using the phase-field model and the analysis of response variability in the microstructure due to change in interface energy parameters. We sampled four sequences of sample sets from a sufficiently large interfacial energy space range and modeled the transition of the initial seed of IMC into various morphologies. The cumulative probability of quasi-random samples from the interface energy space is shown in Fig. (5a). By feeding these samples to the phase-field model, we obtain extensive microstructure data that changes in time. A number of these microstructures extracted from a fixed time are shown in Fig. (5b). The microstructure pallet shows only 200 microstructure cases out of the total of 1672 simulations, and we specifically focus on the five different variants shown next to this pallet. These are the most common cases that we observed by briefly checking the images.

In the **first and second cases**, the IMC tends the wet the substrate. Though, the wetting behavior is different in each case. The IMC tends to wet the substrate slightly in the former, but it does not spread over it. In the latter, the IMC behaves like a liquid and spreads over the substrate very fast. In the **third case**, the shape of the IMC remains invariant to the initial condition. In this case, the IMC either grows or shrinks with the initial contact angle ($\theta = 90°$) and looks very similar to the initial condition. In the **fourth case**, evolution starts by dewetting the substrate and proceeds further until the IMC turns to a droplet over the substrate. In some cases, upon good waiting, IMC gets separated from the substrate surface over time. Finally, the IMC dewets the substrate in the **fifth category** and grows with a contact angle higher than 90°.

To rank and understand the role of the three interface energies on these observed IMC shapes, we performed the sensitivity analysis using the method described in "The sobol' sensitivity analysis" section and the results are shown in Fig. 5c. The Sobol's first-order and total sensitivity indices for the three of the largest sample sets (i.e., 120, 520, and 1024) are shown in three distinct plots. As indicated in the plots by gray-colored

(a) Sobol quasi-random sequences of interface energy parameters, σ_{SL}, σ_{IL}, and σ_{SI} for 8, 120, 520 and 1024 sequences.

(b) Microstructure palette (20×10) from 1024 total simulations. All image are from a fixed time. The five different categories of IMC morphologies are selected to showcase the response variability of the phase-field model.

(c) Sobol sensitivity analysis of the interface energies for 120, 520 and 1024 Sobol sequences.

Fig. 5 The process of Sobol' sampling, model forward run, and Sobol sensitivity results

bars, all or some of the first-order sensitivity indices are negative for 120 and 520 samplings, and it displays all positive for 1024 sampling. These trials show that at least 1024 PFM simulations are required to realize the parametric effect of these input free energies on the output of the phase-field model (i.e., Fourier Descriptor). The sensitivity analysis infers that the solid-IMC interface energy σ_{SI} is the most effective interface energy on the obtained FDs, and the second foremost important parameter is IMC-liquid interface energy σ_{IL}. Overall, the differences in sensitivity of σ_{SI} and σ_{IL} are not noteworthy, and it is hard to conclude if one of these parameters alters the shape more than the other. The results suggest that the three interface energies affect the shape considerably altogether, maybe not equally.

Effect of interface energies on contact angle

In equilibrium, the relevant heterogeneous interface free energies are related to each other by the YL equation, i.e., $\sigma_{SL} = \sigma_{SI} + \sigma_{IL}Cos(\theta)$ with a hysteresis involved for all materials. We have identified the three contact angle regions in the interface energy by using YL equations. We first inserted the sampled points in YL to obtain the contact angle values. Not all of these points are feasible in the YL equation, and some of these points got rejected. Figure 6(a) illustrates 766 points out of 1024 inserted samples into YL equation in 3d interface energy space where the colors indicate the calculated contact angles using the YL equation. This plot differentiates three regions that the IMC forms a contact angle below 85 (green region), between 85 and 95 (blue region), and above 95 degrees (yellow region). Besides the color, the size of the spheres indicates that the contact angle increases as we move in the direction of σ_{SI}.

Similarly, we used multi-phase-field calculations with semi-supervised machine learning analysis to carry out contact angle and shape analysis. The key to the semi-supervised learning method is the consistency assumption, which essentially requires a classifying function to be sufficiently smooth. The results allow us to identify the same three contact angle regions using the FD feature extracted from the phase-field images. Similar to the results shown in Fig. 6(a) for the YL equation, we have identified the three contact angle
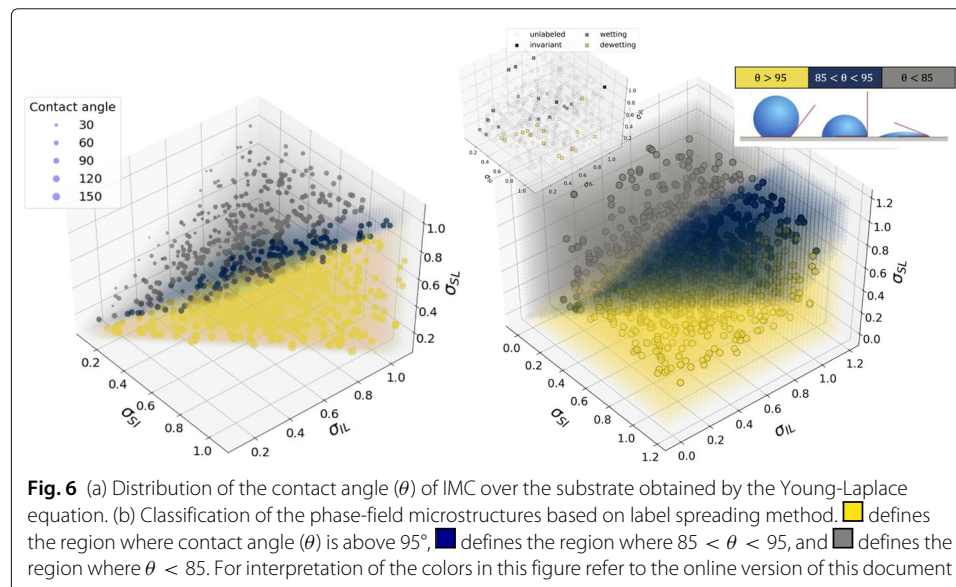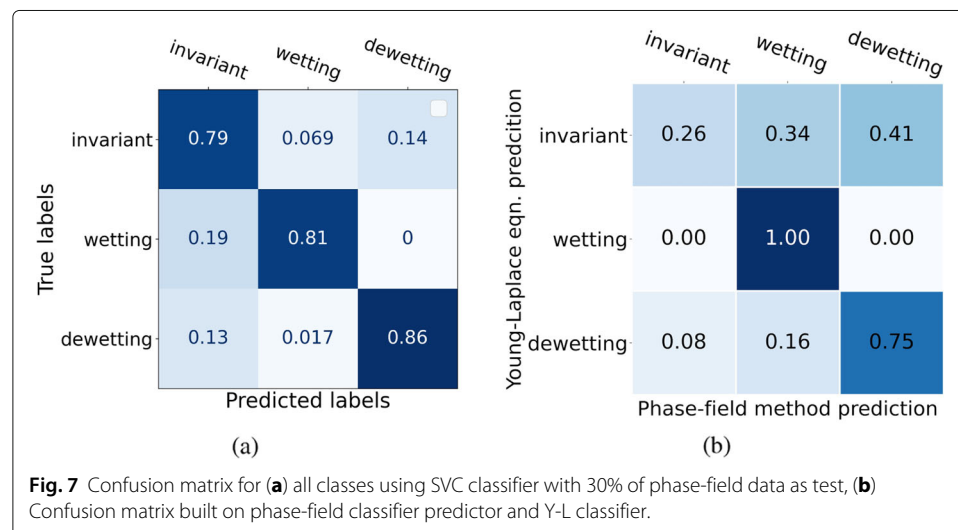

Fig. 6 (a) Distribution of the contact angle (θ) of IMC over the substrate obtained by the Young-Laplace equation. (b) Classification of the phase-field microstructures based on label spreading method. ■ defines the region where contact angle (θ) is above 95°, ■ defines the region where $85 < \theta < 95$, and ■ defines the region where $\theta < 85$. For interpretation of the colors in this figure refer to the online version of this document

regions in the interface energy by the combination of phase-field and semi-supervised machine learning methods. Figure 6(b) shows these three regions that correspond to region i) green: $\theta < 85°$, region ii) dark blue: $85° < \theta < 95°$, and region iii) yellow: $\theta > 95°$. In calculating these regions, we have used LS clustering with KNN kernel and clamping factor (β) of 0.2, meaning that we preserve more data similar to initial labeling.

We only initially labeled 35 samples that are shown with square colored markers in the small scatter plot on the top-left corner of Fig. 6(b). The unlabeled points are shown with white markers with gray edge lines. The decision boundaries are obtained by training the Support Vector Classifier (SVC) with a linear kernel on 70% of the data as training-set and 30% as test-set. In classification problems, good classification accuracy is the primary concern. Therefore, the confusion matrix for the SVC classifier is reported in Fig. 7(a) and true class labels are listed along the x-axis, and the SVC class predictions are listed along the y-axis. Along the first diagonal are the correct classifications, whereas all the other entries show misclassifications. Figure 7(b) shows the confusion matrix for phase-field and Y-L predictors. The prediction accuracy is 85% and 81% for wetting and dewetting class labels, respectively. However, it reduces to only 25% for invariant class labels, and the difference in labeling invariant cases by the semi-supervised and Y-L models is considerable.
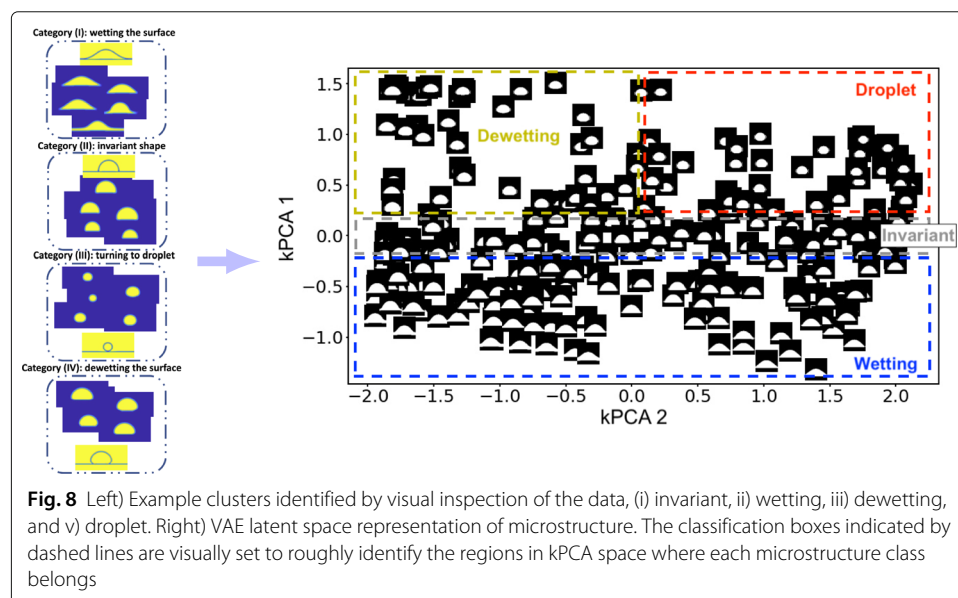
The YL equation dictates a narrow region for invariant shape in the middle area shown by the blue color in Fig. 6(a), where the equilibrium angle remains close to 90°. Moreover, the points form an inverted pyramid region in the 3d interface energy space. These empty regions in the YL equation input space are due to the fact that the arccosine value is out of range, and it does not predict any conceivable contact angle. On the contrary, not all the shapes that are obtained by the phase-field model reached equilibrium. Therefore, the obtained contact angles are non-equilibrium values and are different from the YL equation. Besides, as seen in Fig. 6, there are markers outside the region that is accessible by the YL equation where the phase-field model produces results. This observation is attributed to the fact that many of the contact angles we see in these regions are not equilibrium ones.

Fig. 7 Confusion matrix for (**a**) all classes using SVC classifier with 30% of phase-field data as test, (**b**) Confusion matrix built on phase-field classifier predictor and Y-L classifier.

Visualization of interface energy–microstructure relations using VAE.

To further inspect the dataset for understanding the effect of interface energies on IMC shape, we train a VAE reinforced with a regression model to extract the correlations between inputs (i.e., interface energies) and outputs (i.e., microstructure images) of the phase-field model. The model optimizes simultaneously on microstructure images and phase-field model input parameters and achieves a well-structured and dense representation (latent space embedding). The latent space (reduced image data) is further dimensionally reduced by the kPCA with a linear basis function kernel. This helps to visualize the image data in 2D spaces to further detect correlations between parameters. If the microstructure image data and phase-field input parameters correlate, the images should cluster in the VAE latent space.

Figure 8 shows a collection of the microstructures and representative clusters in the image dataset along with the calculated VAE latent space embedding (right figure). The clusters are the same ones that we explained in "Response variability" section, except that we merge the category one and two into one cluster (i.e., wetting and fast wetting). The figure shows the first two components of the kPCA space representation of the validation set. It provides a qualitative overview space that can further be used to find correlations with input parameters, i.e., interface energies. The kPCA axes have no actual physical meaning. They are obtained using a basis transformation to recognize and join the VAE images and the input parameter space in a dense layer. Each microstructure image is plotted at its position in the dimensionally reduced latent space embedding of the VAE. The images cluster in regions of similar sizes and shapes. The border of these clusters is estimated visually by the colored dashed lines in Fig. 8. The figure shows that the wetting structures (both wetting and fast wetting cases) are located at the bottom of the plot covering negative $kPCA_1$ and entire $kPCA_2$ space. The invariant structure (category II) is located in the central section of this plot around $kPCA_1$ and $kPCA_2$. As we move away from the central section, we can see the other shape categories. The droplet category is located in the top right of the kPCA plot, and the dewetting cluster is located in the top

Fig. 8 Left) Example clusters identified by visual inspection of the data, (i) invariant, ii) wetting, iii) dewetting, and v) droplet. Right) VAE latent space representation of microstructure. The classification boxes indicated by dashed lines are visually set to roughly identify the regions in kPCA space where each microstructure class belongs
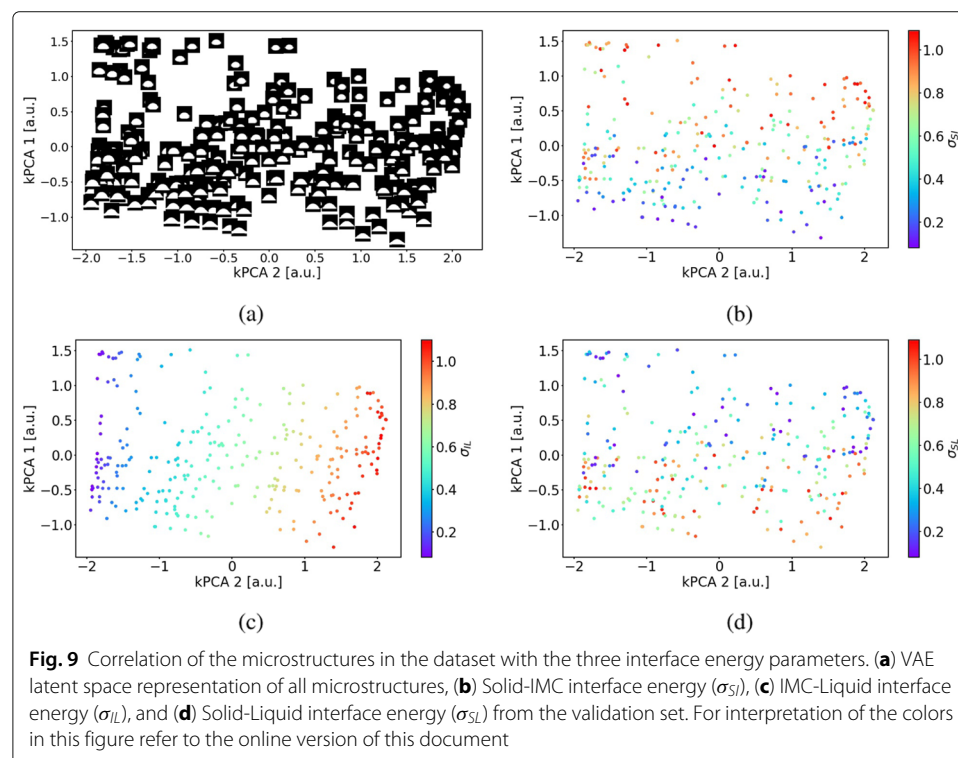
left of the kPCA plot. These clusters infer a good correlation between the inputs and reduced latent space due to the clustering of different categories in kPCA space.

Additionally, we color-coded the latent space with each parameter to find the correlations between the interface energies and microstructures. In contrast, Fig. 9(a-d) shows the position of each sample in the latent space with their respective color-coded input parameters. It visualizes the interplay between parameters and their significance on microstructural features. The results in Fig. 9(b) and Fig. 9(d) indicate that gradual decrease in σ_{SI} together with increase in σ_{SL} leads to wetting morphology where the degree of wetting increases when the difference between σ_{SL} and σ_{SI} is maximum. Moreover, as we move in the wetting class region in the lower section of the kPCA space toward the negative kPCA spaces, the degree of wetting becomes more pronounced. Figure 9(c) shows that σ_{IL} decreases during this movement while the other two parameters are almost constant. Another conclusion can be drawn when σ_{SI} is maximum, σ_{SL} is minimum, and σ_{IL} changes from 0.1 to 1.1. In such conditions, the dewetting structure turns into a droplet structure. The droplet structure becomes vivid when $kPCA_1$ and $kPCA_2$ are maximum or parameter-wise, σ_{SI} and σ_{IL} are maximum, and σ_{SL} is minimum. On the other hand, transition in σ_{IL} space covers all morphologies with a uniform change from the left-side in Fig. 9(c) to the right.

Conclusion

In this work, we proposed a systematic strategy for machine learning-assisted high-throughput study of the microstructure of materials. The ultimate goal is the microstructure-sensitive design of the response of materials using multi-phase-field under the framework of ICME to overcome the curse of dimensionality. We prevailed

Fig. 9 Correlation of the microstructures in the dataset with the three interface energy parameters. (**a**) VAE latent space representation of all microstructures, (**b**) Solid-IMC interface energy (σ_{SI}), (**c**) IMC-Liquid interface energy (σ_{IL}), and (**d**) Solid-Liquid interface energy (σ_{SL}) from the validation set. For interpretation of the colors in this figure refer to the online version of this document

over the large spreads in inputs, the large spreads in outputs, uncertainty in a complex nonlinear model, and complex image featurization tasks through automation of the computation of a large number of simulations and post-processing tasks, employing a semi-supervised machine learning method, and variational autoencoder (VAE), i.e., a deep generative learning method. We layout the challenges and opportunities in optimal design or understating of the microstructure of materials. Microstructure characterization and classification have been identified as essential steps in building processing-structure-property linkages for the ultimate goal of materials by design.

In summary, we sampled four instances of Sobol samples with different sizes from the three interface energies and carried out in total 1672 phase-field simulations. Accordingly, we reduced the dimensionality in the images by using Fourier descriptor analysis. As a result, we were able to perform sensitivity analysis due to variations in inputs and identify and rank the effect of interface energy on the shape of the IMC. Next, we used two distinct machine learning methods to determine the correlations between inputs and outputs of the multi-phase-field method. We first employed the label spreading method (i.e., semi-supervised machine learning method) by labeling only 35 microstructures out of 1672 images into three classes to spread these labels across the other unlabeled data. We compared these results with the Young-Laplace (YL) equation results and denoted the similarities and differences. Overall, the phase-field and YL observations follow a similar trend. However, the main difference is the applicability of the phase-field simulations in the regions that YL equation cannot work. We showed that semi-supervised methods are indeed great tools to train and learn high-performing microstructure classification models.

Furthermore, we used VAE, a deep generative neural network method, to determine the correlations between inputs and outputs of the phase-field method. The VAE method allows establishing direct correlations between reduced microstructure image data and inputs of the physical model. The results show that there is an inverse correlation between σ_{SI} and σ_{SL} and they act simultaneously on the shape of IMC where an increase in the former and decrease in the latter changes the shape from dewetting structures to wetting structures.

We applied automated combinatorial data-driven methods in conjunction with a computationally expensive microstructure model to uncover subtle relationships in microstructural data.

Authors' contributions
VA designed, planned, performed, processed the phase-field simulations, sensitivity, and machine learning analysis and interpreted the results. VA prepared the manuscript, and RA supervised the project. The author(s) read and approved the final manuscript.

Availability of data and materials
Our microstructure dataset contains 1672×25 or 41800 computationally generated images that are obtained from 1672 distinct phase-field simulations where 25 time sequences is extracted from each simulation. For consistency and comparability, we only used 1672 images that are obtained from one of the latest iterations of the phase-field model. All images are 388×343 px^2 that contain only the local characteristics of the IMC phase-field order parameter that basically preservers the shape information. Composition data are stored separately and we didn't use the composition

information in this study. All images were cropped to IMC area and further rescaled to 64×64 px^2 to reduce the cost of the analysis. All of the raw datasets on which the results of the manuscript are based on are publicly available via the following Data DOI: https://doi.org/10.5281/zenodo.5090288 or can be accessed on the Open Phase-field Microstructure Database (OPMD) website accessible via https://microstructures.net link.

Declarations

Author details
[1]Materials Science and Engineering Department, Texas A&M University, College Station 77840, TX, USA. [2]Mechanical Engineering Department, Texas A&M University, College Station 77840, TX, USA.

References

V. Attari, R. Arroyave, Phase field modeling of joint formation during isothermal solidification in 3DIC micro packaging. J. Phase Equilib. Diffus. **37**(4), 469–480 (2016)

H. Azizi, A. Ebrahimi, N. Ofori-Opoku, M. Greenwood, N. Provatas, M. Mohammadi, Characterizing the microstructural effect of build direction during solidification of laser-powder bed fusion of Al-Si alloys in the dilute limit: a phase-field study. Acta Mater. **214**, 116983 (2021)

V. Attari, S. Ghosh, T. Duong, R. Arroyave, On the interfacial phase growth and vacancy evolution during accelerated electromigration in Cu/Sn/Cu microjoints. Acta Mater. **160**, 185–198 (2018)

V. Attari, P. Honarmandi, T. Duong, D. J. Sauceda, D. Allaire, R. Arroyave, Uncertainty propagation in a multiscale CALPHAD-reinforced elastochemical phase-field model. Acta Mater. **183**, 452–470 (2020)

L. Banko, Y. Lysogorskiy, D. Grochla, D. Naujoks, R. Drautz, A. Ludwig, Predicting structure zone diagrams for thin film synthesis by generative machine learning. Commun. Mater. **1**(1), 1–10 (2020)

J. Bosch, *Fast iterative solvers for cahn-hilliard problems. PhD thesis*. (Otto-von-Guericke Universität Magdeburg, 2016)

E. T. Bowman, K. Soga, W. Drummond, Particle shape characterisation using fourier descriptor analysis. Geotechnique. **51**(6), 545–554 (2001)

J. W. Cahn, Critical point wetting. J. Chem. Phys. **66**(8), 3667–3672 (1977)

J. W. Cahn, J. E. Hilliard, Free energy of a nonuniform system. i. interfacial free energy. J. Chem. Phys. **28**(2), 258–267 (1958)

L.-Q. Chen, Phase-field models for microstructure evolution. Annu. Rev. Mater. Res. **32**(1), 113–140 (2002)

A. Choudhury, Y. C. Yabansu, S. R. Kalidindi, A. Dennstedt, Quantification and classification of microstructures in ternary eutectic alloys using 2-point spatial correlations and principal component analyses. Acta Mater. **110**, 131–141 (2016)

B. L. DeCost, T. Francis, E. A. Holm, Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures. Acta Mater. **133**, 30–40 (2017)

X. Deng, G. Piotrowski, J. Williams, N. Chawla, Influence of initial morphology and thickness of Cu_6Sn_5 and Cu_3Sn intermetallics on growth and evolution during thermal aging of Sn-Ag solder/Cu joints. J. Elec. Materi. **32**(12), 1403–1413 (2003)

S. De Waele, K. Lejaeghere, M. Sluydts, S. Cottenier, Error estimates for density-functional theory predictions of surface energy and work function. Phys. Rev. B. **94**(23), 235418 (2016)

A. Forrester, A. Sobester, A. Keane, *Engineering Design Via Surrogate Modelling: a Practical Guide*. (American Institute of Aeronautics and Astronautics, Inc., 2008)

D. T. Fullwood, S. R. Niezgoda, B. L. Adams, S. R. Kalidindi, Microstructure sensitive design for performance optimization. Prog. Mater. Sci. **55**(6), 477–562 (2010)

J. Gola, D. Britz, T. Staudt, M. Winter, A. S. Schneider, M. Ludovici, F. Mücklich, Advanced microstructure classification by data mining methods. Comput. Mater. Sci. **148**, 324–335 (2018)

R. J. Good, in *Surface and colloid science*, Contact angles and the surface free energy of solids (Springer, 1979), pp. 1–29

L. Gránásy, T. Pusztai, D. Saylor, J. A. Warren, Phase field theory of heterogeneous crystal nucleation. Phys. Rev. Lett. **98**(3), 035703 (2007)

L. Granasy, G. I. Tóth, J. A. Warren, F. Podmaniczky, G. Tegze, L. Ratkai, T. Pusztai, Phase-field modeling of crystal nucleation in undercooled liquids–a review. Prog. Mater. Sci. **106**, 100569 (2019)

D. Khatamsaz, A. Molkeri, R. Couperthwaite, J. James, R. Arróyave, D. Allaire, A. Srivastava, Efficiently exploiting process-structure-property relationships in material design by multi-information source fusion. Acta Mater. **206**, 116619 (2021)

D. P. Kingma, M. Welling, Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)

D. P. Kingma, M. Welling, An introduction to variational autoencoders. arXiv preprint arXiv:1906.02691 (2019)

C. Kunselman, V. Attari, L. McClenny, U. Braga-Neto, R. Arroyave, Semi-supervised learning approaches to class assignment in ambiguous microstructures. Acta Mater. **188**, 49–62 (2020)

H. Liu, G. Cao, Effectiveness of the young-laplace equation at nanoscale. Sci Rep. **6**(1), 1–10 (2016)

D. L. McDowell, in *Computational Materials System Design*, Microstructure-sensitive computational structure-property relations in materials design (Springer, 2018), pp. 1–25

S. Niezgoda, D. Fullwood, S. Kalidindi, Delineation of the space of 2-point correlations in a composite material system. Acta Mater. **56**(18), 5285–5292 (2008)

E. Persoon, K.-S. Fu, Shape discrimination using fourier descriptors. IEEE Trans. Syst. Man Cybern. **7**(3), 170–179 (1977)

P. Prakash, V. Mytri, P. Hiremath, Fuzzy rule based classification and quantification of graphite inclusions from microstructure images of cast iron. Microsc. Microanal. **17**(6), 896–902 (2011)

Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, L. Carin, Variational autoencoder for deep learning of images, labels and captions. arXiv preprint arXiv:1609.08976 (2016)

L. Ruthotto, E. Haber, An introduction to deep generative modeling. GAMM-Mitteilungen, 202100008 (2021). Wiley Online Library

A. Saltelli, Making best use of model evaluations to compute sensitivity indices. Comput. Phys. Commun. **145**(2), 280–297 (2002)

A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. Comput. Phys. Commun. **181**(2), 259–270 (2010)

B. Schölkopf, A. Smola, K.-R. Müller, in *International Conference on Artificial Neural Networks*, Kernel principal component analysis (Springer, 1997), pp. 583–588

J.-H. Shim, C.-S. Oh, B.-J. Lee, D. Lee, Thermodynamic assessment of the cu-sn system. Int. J. Mater. Res. **87**(3), 205–212 (1996)

W. Usher, J. Herman, C. Whealton, D. Hadka, xantares, F. Rios, bernardoct, C. Mutel, J. van Engelen, SALib/SALib: Launch!. https://doi.org/10.5281/zenodo.160164

L. Vitos, A. Ruban, H. L. Skriver, J. Kollár, The surface energy of metals. Surf. Sci. **411**(1-2), 186–202 (1998)

Y. Wang, D. L. McDowell, *Uncertainty Quantification in Multiscale Materials Modeling*. (Woodhead Publishing, 2020)

X.-Y. Zhang, M. Trame, L. Lesko, S. Schmidt, Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models. CPT: Pharmacometrics Syst. Pharmacol. **4**(2), 69–79 (2015)

D. Zhou, O. Bousquet, T. Lal, J. Weston, B. Schölkopf, Learning with local and global consistency. Adv. Neural Inf. Process. Syst. **16**, 321–328 (2003)

Publisher's Note