npj | materials degradation

Article

Published in partnership with CSCP and USTB

https://doi.org/10.1038/s41529-024-00529-8

A data-driven strategy for phase field nucleation modeling

Yang Hu[1,2], Kai Wang[1,3] ✉ & Robert Spatschek[1,2,4]

We propose a data-driven strategy for parameter selection in phase field nucleation models using machine learning and apply it to oxide nucleation in Fe-Cr alloys. A grand potential-based phase field model, incorporating Langevin noise, is employed to simulate oxide nucleation and benchmarked against the Johnson-Mehl-Avrami-Kolmogorov model. Three independent parameters in the phase field simulations (Langevin noise strength, numerical grid discretization and critical nucleation radius) are identified as essential for accurately modeling the nucleation behavior. These parameters serve as input features for machine learning classification and regression models. The classification model categorizes nucleation behavior into three nucleation density regimes, preventing invalid nucleation attempts in simulations, while the regression model estimates the appropriate Langevin noise strength, significantly reducing the need for time-consuming trial-and-error simulations. This data-driven approach improves the efficiency of parameter selection in phase field models and provides a generalizable method for simulating nucleation-driven microstructural evolution processes in various materials.

For many years, the phase field (PF) method is widely utilized for simulating microstructure evolution processes[1,2], with applications to various fields like solidification[3], solid-state transformations[4], grain growth and coarsening[5], additive manufacturing[6], fracture[7–9], biological systems[10] and many other use cases. The fundamental concept of the methodology is based on the use of "phase fields" (order parameters), which allow to distinguish between the different phases or grains, and the use of partial differential equations for the evolution of the microstructure as expressed through the dynamics of these order parameters. These evolution equations replace a tedious tracking of the interface locations, which is required for complementary sharp interface modeling, and therefore largely simplify the algorithmic effort for these moving boundary problems. A characteristic feature is the use of diffusive interfaces, and their thickness is usually chosen significantly larger than physical interface thicknesses, in order to avoid too fine discretizations of the computational domain. However, this increase of the interface thickness requires careful extrapolation of results towards the sharp interface limit to ensure reliable results. This step can be supported by superior model formulations involving thin interface approaches[3,11] and nondiagonal phase field models[12,13].

Although these approaches nowadays allow to perform large scale quantitative simulations of technically relevant multi-components alloy systems in the deterministic growth regime, the formulation of nucleation as a random event is still a challenge. For example, oxidation can significantly limit materials service through degradation processes. Although the phase field method is employed to tailor the materials design in various ways, reproducing properly the initial nucleation of oxides is still difficult for this approach. A central reason for this difficulty is the mentioned increase of the interface thickness, as it naturally implies that nuclei have to appear on scales much larger than the numerical interface thickness, far beyond their expected physical size, as otherwise they would shrink and disappear immediately for insufficient supersaturations and undercooling. A natural choice of a Langevin noise strength in the evolution equations to enable thermal fluctuations is therefore not sufficient for obtaining a proper nucleation behavior, therefore requiring model-specific and technical code modifications beyond the established deterministic phase field evolution equations.

Generally, there are two different approaches to modeling the nucleation behavior in phase field simulations. The straightforward method is in the spirit of the above discussion that one can explicitly include a Langevin noise term into the governing phase field equations. Another method has been proposed by Gránásy et al.[14], in which the critical fluctuations are calculated as a function of local composition, i.e. chemical driving force. Through statistically assigning the critical fluctuations to different areas of the simulation domain and adding a Langevin noise with small amplitude, a potential nucleus growth beyond critical fluctuations is determined. Although the second method can significantly reduce the

[1]Institute of Energy Materials and Devices IMD-1, Forschungszentrum Jülich GmbH, Jülich, 52428, Germany. [2]Georesources and Materials Engineering, RWTH Aachen University, Aachen, 52062, Germany. [3]Mechanics of Functional Materials Division, Institute of Materials Science, Technical University of Darmstadt, Darmstadt, 64287, Germany. [4]JARA Energy, Jülich, 52428, Germany. ✉e-mail: kai.wang@tu-darmstadt.de

computational effort, the nucleation formation in the phase field simulations remains sensitive to the underlying grid spacing of the simulation domain, the local driving force and the amplitude of the fluctuations. Improper parameters may lead to the decay of the fluctuations or numerical instabilities of the simulations. Using a trial and error method can help to identify proper parameters for nucleation in the phase field simulations, to be in agreement with experimental observations. However, existing phase field models are still inherently computationally expensive and require high-performance computing resources.

A particular application of phase field nucleation models is related to the formation of oxides. Zaeeme and Kadiri developed a multi-phase field model with simple quadratic free energy density for non-selective oxidation to study oxidation kinetics and stress generation in the Zr-ZrO_2 system[15]. Sherman et al. formulated a phase field model to study the equilibrium state of an oxide where the film thickness is below the Debye length[16]. Kim et al. examined the kinetics of diffusion-controlled oxide growth by using Sherman's model[17]. Wang et al. investigated the impact of alloy composition, oxide nuclei size, shape and distribution on the internal to external oxide growth transition[18]. However, these investigations only consider the oxidation processes in sandwich structures or with pre-existing oxide nuclei. These artificially predefined morphologies may lead to inconsistent grain numbers and shapes, as well as improper grain boundary distributions compared to experimental observations.

Contrary, machine learning generated significant progress for materials modeling, e.g. for material design, materials properties prediction, microstructure investigations, and improving and accelerating ab initio simulations[19,20]. In particular, in the field of corrosion, data-driven approaches successfully predict electrochemical corrosion[21], stress corrosion cracking[22], the corrosion rate of alloys[23] and the corrosion resistance for multi-principal element alloys[24]. Therefore, the question arises whether machine learning (ML) techniques can be suitable for the required parameter optimization in phase field nucleation models. With the development in recent decades, more and more powerful machine learning models have emerged, among them several deep learning algorithms, namely, convolutional neural networks (CNN), recurrent neural networks (RNN) and other artificial neural networks (ANN)[25]. A machine learning model is considered as a black box that can analyze complicated relationships by training a mathematical model according to a given data set. The construction of relationships between the input parameters and output results without consideration of underlying physical models can be powerful for situations, where appropriate physical descriptions are difficult to develop or the computing time for an existing model becomes too long. As a result of such an approach, the computing speed can be accelerated tremendously through the use of a trained model. For example, Z. Shen et al. developed a phase field model for the breakdown process of polymer-based dielectrics by incorporating the electrical, thermal, and mechanical effects and training a machine learning model on the basis of high-throughput simulation results to produce an analytical expression for the breakdown strength as a function of different material parameters[26]. D. Zapiain et al. built a model for learning the microstructural evolution of targeted systems directly by combining phase field and history-dependent machine learning techniques to accelerate the calculations[27].

In the present work, we aim to demonstrate such a fruitful combination of phase field and machine learning models for an efficient formulation of oxidation during the service of ferritic steel-based interconnects for solid oxide fuel and electrolyzer applications. Solid oxide fuel cells (SOFCs) are an attractive approach for efficiently producing energy in an environmentally friendly manner[28–31]. Generally, SOFCs are operated as a serial "stack" with common interconnects between the cells, which link the anodes and cathodes of adjacent units. The interconnects not only establish the electrical connection between the electrodes of neighboring cells but also block the contact of the oxidizing and reducing atmospheres. Therefore, interconnects are of crucial importance for the performance of SOFCs, and their properties significantly influence the overall performance of SOFCs[32]. Different metallic alloys are broadly selected as interconnects due to their high mechanical strength, sufficient thermal conductivity, excellent electronic conductivity and low price[33]. At the high working temperature of SOFCs, ferritic stainless steels with formation of protective chromia scales are promising candidates, such as AISI430[34,35], ZMG332[36] and Crofer 22 APU[37,38]. The formation of continuous Cr_2O_3 protective layers with 20–25% Cr content increases the oxidation resistance of the interconnects. The oxidation patterns are complex and depend on temperature, Cr concentration, atmospheric conditions and composition[39–41]. Already for a Fe-Cr binary alloy, the oxide layers typically consist of Fe-rich oxides, Fe-Cr mixed oxide and Cr-rich oxides from the exposed interface to the matrix. Two severe challenges arise when ferritic stainless steels are chosen as interconnects. First, the overall electrical resistance during service time increases. Second, the Cr_2O_3 layer can react with dry or wet air, which leads to the evaporation of gaseous hexavalent Cr species, such as $CrO_3(g)$ and $CrO_2(OH)_2$. As a result, cathode materials like $La_xSr_{1-x}Co_yFe_{1-y}O_{3-\delta}$ (LSCF) can subsequently react with the gaseous Cr species, and a formation of a Sr- and Cr-containing oxides is observed in the cathodes, leading to Cr-related degradation[42]. Therefore, a deep understanding of the Cr-oxide formation processes in ferritic steels is important to optimize the performance and longevity of SOFCs.

Aiming to develop a machine learning model that accurately predicts nucleation parameters, thereby reducing computational costs and improving model fidelity, we propose a data-driven strategy for the phase field oxide nucleation modeling in the present work. The integration of phase field modeling of the Cr_2O_3 nucleation process during the initial oxidation process in the interconnect with machine learning predictions of nucleation parameters for the desired Cr_2O_3 oxide nucleation density is explored. To this end, a grand potential-based phase field nucleation model is developed in section "Phase field nucleation model". The phase field nucleation model is benchmarked against classical models in section "Phase field model validation". In view of the fact that the nucleation behavior in phase field simulations depends on the grid spacing, the chemical driving force as well as the amplitude of the fluctuation, the influence of these parameters on the nucleation density in the phase field simulations is investigated in section "Phase field simulation analysis". Then, based on the phase field results, the predictions of the nucleation density via a developed ANN model are demonstrated. First, a classification of the different nucleation density regimes is done in section "Phase field simulation analysis", followed by a quantitative regression in section "Regression model". As a result, we obtain machine learning-based predictions for the choice of numerical parameters in the phase field model to get the desired nucleation behavior.

Results and discussion

Phase field nucleation model

As mentioned in the introduction, there are different approaches to model nucleation in phase field simulations, with different advantages and disadvantages. In the present work, we adapt the approach proposed by Ode and Ohnuma[43]. A central observation is, that the direct addition of an additive white noise term to the phase field evolution equation (14) at each grid point would typically require unreasonably high noise strengths, as otherwise, the fluctuations would be insufficient to trigger the formation of a nucleus. This is a result of the fact that nuclei typically appear on scales lower than those considered in mesoscale phase field simulations, and therefore the interface thickness overrides the physical critical radius for nucleation, hence affecting artificially the nucleation energy barrier. However, overcoming this effect by increasing the noise strength tends to destabilize the numerics and is therefore not a favorable approach.

The starting point of our considerations is to use a grand potential-based phase field model without consideration of nucleation, as detailed in the methods section. This phase field model allows for more accurate representation of the free energy landscape, which is crucial for predicting microstructure evolution[44]. We add an uncorrelated, white Gaussian noise

to this basic phase field model, such that the evolution equation becomes as extension of Eq. (14) (see methods section)

$$\tau\dot{\eta}_{\alpha i} = \kappa\nabla^2\eta_{\alpha i} - M_0\left(\eta_{\alpha i}^3 - \eta_{\alpha i} + 2\eta_{\alpha i}\sum_{\beta=1}^{N}\sum_{j=1,\alpha i\neq\beta j}^{p_\beta}\gamma_{\alpha i\beta j}\eta_{\beta j}^2\right) - \sum_{\beta=1}^{N}\frac{\partial h_\beta}{\partial\eta_{\alpha i}}\omega_\beta + \tau\zeta, \tag{1}$$

where ζ is a stochastic scalar field, which obeys

$$\langle\zeta(\vec{x},t)\zeta(\vec{x'},t')\rangle = A\delta(\vec{x}-\vec{x}')\delta(t-t'), \tag{2}$$

where δ_{ij} is the Kronecker delta. For the definition of all parameters, we refer to the methods section. As will be discussed below, the selection of the proper noise strength is essential, and a suitable approach for it is presented in this work.

The second element is that supercritical fluctuations, which may lead to the formation of a nucleus, are spatially extended. Therefore, if at a certain location, the phase field (including the thermal fluctuations) exceeds a critical value, the phase field is elevated above the nucleation threshold inside a spherical environment with a randomly selected radius. In this way, the nucleus, which would physically appear on a scale below the grid resolution, is magnified to a scale which is spatially resolved in the phase field simulations. Its further growth or shrinkage is then described by the evolution equation (1).

It is obvious that the behavior of this nucleation model depends on physical aspects and numerical details of the implementation, which are analyzed in the following. The goal is to identify the most relevant dependencies, in order to find a strategy for selecting the noise strength such that the desired nucleation density is obtained. We note that due to the aforementioned nucleus enhancement, the noise strength is becoming a numerical parameter and is therefore not determined by the fluctuation-dissipation theorem.

Physically, we expect the nucleation to depend on the bulk driving force, which is in the current context the grand potential difference $\Delta\omega$ between the mother and oxide phase. In analogy to the work in ref. 43, the first derivative of the well potential term and chemical free energy are considered, while the interface gradient term is omitted. Hence, for an assumed grain j of phase β growing from grain i of the α matrix, the simplified expression of the critical fluctuations is $\frac{\partial\omega_{\mathrm{mw}}}{\partial\eta_{\alpha i}} + \frac{\partial\omega_{\mathrm{chem}}}{\partial\eta_{\alpha i}} = 0$. After expansion, we have

$$M_0(\eta_{\alpha i}^3 - \eta_{\alpha i} + 2\eta_{\alpha i}\gamma_{\alpha i\beta j}\eta_{\beta j}^2) + \frac{\partial h_\alpha}{\partial\eta_{\alpha i}}(\omega_\alpha - \omega_\beta) = 0. \tag{3}$$

With $\eta_{\beta j} = 1 - \eta_{\alpha i}$ and dimensionless $\gamma_{\alpha i\beta j} = 1.5$, $\eta_{\mathrm{crit}}(\Delta\omega)$ can be solved numerically, where $\Delta\omega = \omega_\alpha - \omega_\beta$ is the grand potential difference. This critical phase field value is used explicitly in the procedure explained above, as the order parameter inside the artificially stabilized nucleus is set to this value. From this physical argument, we can therefore expect the nucleation process to be dependent on the dimensionless ratio A/η_{crit}.

The second dependence, which is based on physical arguments, is that the nucleation density will depend on the local value of the critical nucleus. Whereas in the above consideration, the gradient square term from the phase field model has been ignored, it is directly linked to the formation of a critical nucleus as a balance between the above bulk energy gain and an increase of interfacial energy. Therefore, the characteristic length scale for the critical nucleus is given by $R_{\mathrm{crit}} = \sigma/\Delta\omega$ with the interfacial energy σ. It is calculated via the integration of ω_{grad} and ω_{mw}, i.e. $\sigma = \int_{-\infty}^{\infty}(\omega_{\mathrm{grad}} + \omega_{\mathrm{mw}})\,dx$. A larger chemical driving force or smaller interface energy requires a smaller nucleation radius, which directly leads to larger nucleation density. Therefore, we expect the choice of the Langevin noise to depend on the dimensionless ratio R_{crit}/W with the phase field interface thickness W, and the latter is normalized to one in the following. Finally, as elaborated in ref. 43, the phase field nucleation behavior depends on the chosen grid spacing as a numerical parameter. Therefore, the third expected relevant dependency for the selection of the Langevin noise strength is the ratio $\Delta x/W$.

Altogether, we therefore expect that the noise strength for determining the desired nucleation is density will depend in a nontrivial way on the three parameters R_{crit}/W, $\Delta x/W$ and A/η_{crit}. A direct way to determine the noise strength A would be by performing many phase field simulations and then iteratively determining the suitable choice for this parameter. However, such an iterative approach is tedious and time-consuming, and therefore we propose here a strategy based on selected small-scale phase field simulations in combination with machine learning techniques, for accelerated determination of the noise strength. We illustrate the approach in detail in the following sections for the given phase field model. We note that the selection of suitable parameters may additionally depend on other internal model parameters, which are used e.g. for the detection of nuclei within a certain environment and the rate of nucleus magnification. However, these parameters are kept constant throughout the simulations and are therefore not screened. On a more general level, we emphasize that the overall approach will also be beneficial for other phase field applications and different nucleation models, which suffer from the combined dependence on physical and numerical influencing factors.

Phase field model validation

For the application of the proposed phase field nucleation model, it is essential that it leads to physically reasonable predictions. Therefore, we start with a comparison of phase field nucleation simulations with the established and frequently used Johnson-Mehl-Avrami-Kolmogorov (JMAK) model[45–49]. The JMAK theory describes the time evolution of the volume fraction of a new phase that forms during a phase transformation, and it provides a framework for analyzing the rate of the transformation. Typically, JMAK theory can be expressed as

$$Y(t)/Y_{eq} = 1 - \exp(-kt^n), \tag{4}$$

where $Y(t)$ is the fraction of the new phase at time t. Y_{eq} is the equilibrium fraction of the new phase, k is a constant relating to the nucleation and growth rates, n is the so called Avrami exponent, which is determined by the mechanisms of nucleation and growth. The expression for n is given by

$$n = qd + B, \tag{5}$$

where typically q equals 1 for interface-controlled growth while $q \simeq 0.5$ for diffusion-controlled growth patterns. d is the spatial dimensions, and B is related to the nucleation generation mode. For preexisting nuclei, $B = 0$, while $B = 1$ for continuous nucleation with a constant rate. Generally, the Avrami exponent n can conveniently be extracted through plotting $\ln\{-\ln[1 - Y(t)/Y_{eq}]\}$ versus $\ln(t)$[50].

Figure 1 compares two-dimensional phase field simulations with the JMAK theory for a critical nucleation radius of $R_{\mathrm{crit}}/W = 0.1$. In these simulations, the grid spacing is chosen as $\Delta x = \Delta y = 0.8W$, and the amplitude of the fluctuation is A/η_{crit}=0.12 and 0.06 to generate different nucleation densities. In Fig. 1a, five independent two-dimensional phase field simulations with continuous nucleation mode using different random number initializations are presented as the cross points when A/η_{crit}=0.12, all leading to essentially the same results. The corresponding average nucleation density after convergence is, which is defined as the number of nuclei per unit area of the two-dimensional system size. We note that nucleation occurs continuously during the simulation, but as the supersaturation decreases due to the formation and growth of oxide nuclei, it converges to a density, which is optimized in the following investigations. Snapshots of one of the phase field simulations are displayed in the embedded figures. In Fig. 1b, the average nucleation density is $1.13 \times 10^{-4}W^{-2}$, which is much lower than in Fig. 1a, leading to coarser structures.

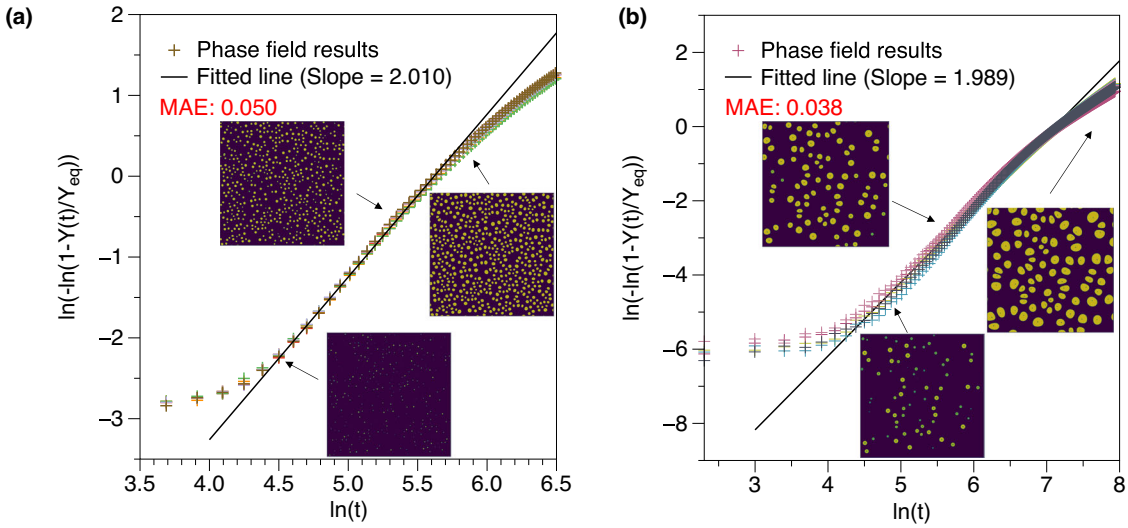


Fig. 1 | Comparison of the phase field results with the Johnson-Mehl-Avrami-Kolmogorov (JMAK) theory in two-dimensions for two different nucleation densities. a $1.05 \times 10^{-3} W^{-2}$ and (**b**) $1.13 \times 10^{-4} W^{-2}$ for continuous nucleation. The agreement between phase field results and the JMAK theory supports the phase field nucleation concept. To verify the robustness of the validation, the error analysis of phase field results in (**a**) and (**b**) are measured via mean absolute error (MAE).

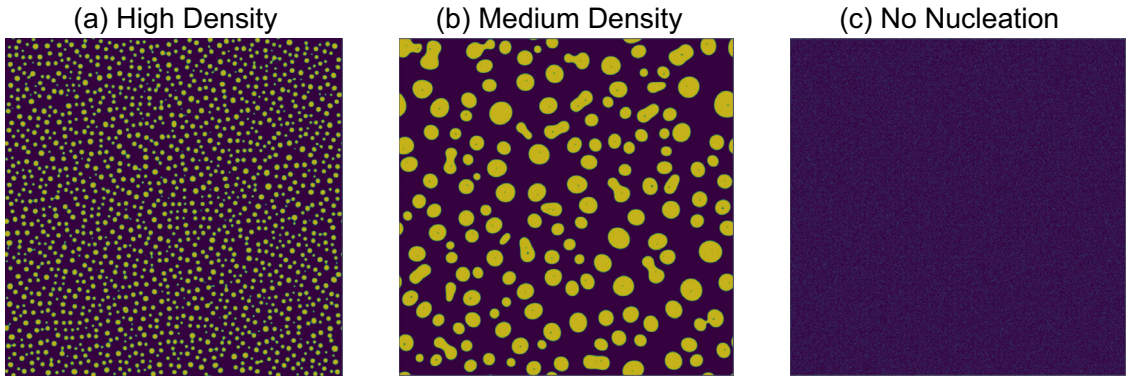


Fig. 2 | Snapshots of microstructures for different nucleation densities. The snapshots are labeled as (**a**) high density (using $R_{crit}/W = 0.1$, $\Delta x/W = 0.8$, $A/\eta_{crit} = 0.2$), (**b**) medium density ($R_{crit}/W = 0.1$, $\Delta x/W = 0.8$, $A/\eta_{crit} = 0.12$) and (**c**) no nucleation ($R_{crit}/W = 0.12$, $\Delta x/W = 0.4$, $A/\eta_{crit} = 0.07$). High, medium and no densities refer to nucleation densities are in the range of $[4.47 \times 10^{-4} W^{-2}, \infty]$, $[4.47 \times 10^{-4} W^{-2}, 0]$ and $[-\infty, 0]$, respectively. At here, density $4.47 \times 10^{-4} W^{-2}$ corresponding to 300 nuclei in the selected simulation domain. The higher nucleation density indicates the stronger nucleation driving force and thus the smaller average size of oxide grains.

Next, the phase field results are fitted against the JMAK expression (shown as solid lines) in the intermediate time regime. We find that the corresponding slopes of the fitted lines are 2.01 and 1.989 for the two panels, which are in excellent agreement with the expected Avrami exponent $n = 2$. Furthermore, the error analysis of the phase field results in Fig. 1a and b are also implemented via mean absolute errors (MAE) according to the fitted slopes of five independent phase field simulations. The small MAEs indicates the the robustness of the validation. Therefore, we can conclude that the nucleation generation model in the present work agrees with the JMAK theory and is independent of the nucleation sites and densities.

Phase field simulation analysis

In order to construct a database of nucleation density for training the machine learning models, 547 phase field simulations are carried out for different dimensionless nucleation parameters. In this section, the central results of these simulations are summarized. In Fig. 2a–c, selected phase field results representing "high density", "medium density" and "no nucleation" nucleation are presented. When the nucleation density is larger than $4.47 \times 10^{-4} W^{-2}$, corresponding to 300 nuclei in the chosen size of the simulation domain, the phase field results are marked as "high density". When the nucleation density is in between 0 and $4.47 \times 10^{-4} W^{-2}$, the label "medium density" is assigned. For the phase field results without nucleation, the "no nucleation" label is used. The choice of these dividing criteria has no direct physical meaning and therefore the actual choice of the limiting values is not critical. However, it assists in selecting the desired nucleation density in the machine learning regression model.

For the nucleation example labeled with "high density" in Fig. 2a, the dimensionless nucleation parameters R_{crit}/W, $\Delta x/W$ and A/η_{crit} are chosen as 0.1, 0.8 and 0.2, respectively. Here, the nucleation density is about $1.6 \times 10^{-3} W^{-2}$ and the dimensionless average radius of the grains is about $5.5W$ after 10^6 simulation time steps. A decrease of the temperature during operation can lead to large chemical driving forces, and more suitable nucleation sites are activated by the larger fluctuations in the matrix phase. In turn, when the chemical driving force and the fluctuations decrease, the number of suitable nucleation sites drops significantly and the average grain radius increases due to more pronounced grain growth. For the medium density case in Fig. 2b, the nucleation density and dimensionless average grain radius are approximately $1.79 \times 10^{-4} W^{-2}$ and $30W$, respectively. Finally, in Fig. 2c, when the chemical driving force or the fluctuations are not large enough,

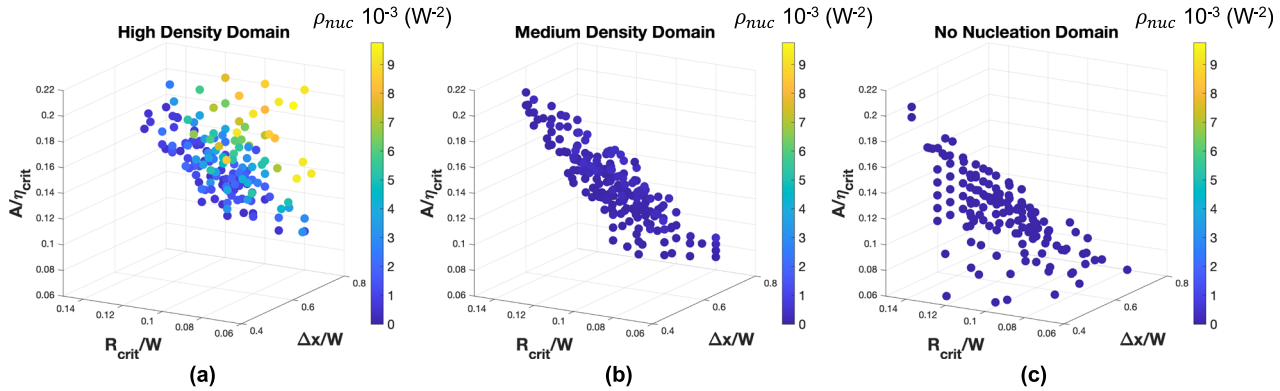


Fig. 3 | Representation of the phase field results as the dataset for the machine learning method. From (**a**) to (**c**), the data are classified according to the dimensionless nucleation density, i.e. more than $4.47 \times 10^{-4} W^{-2}$, up to $4.47 \times 10^{-4} W^{-2}$ and no nucleation. The influence of the key parameters, R_{crit}/W, $\Delta x/W$ and A/η_{crit}, are explicitly demonstrated. Increasing $\Delta x/W$, A/η_{crit}, and decreasing R_{crit}/W leads to a higher nucleation density.

nucleation cannot occur and only subcritical white noise fluctuations are visible inside the simulation domain. Figure 3 includes all the phase field results. As mentioned above, these results serve as input for training the machine learning models. The dimensionless nucleation parameters R_{crit}/W, $\Delta x/W$ and A/η_{crit} are in the range 0.05 to 0.15, 0.4 to 0.8 and 0.06 to 0.22, respectively, and are shown in Fig. 3 for the three categories with high and medium density, as well as no nucleation. Firstly, the sensitivity of the phase field results to the chosen parameters, i.e. R_{crit}, $\Delta x/W$ and A/η_{crit}, are summarized. The parameter R_{crit}/W is crucial in determining the critical radius for nucleation. Variability in this parameter directly impacts the nucleation rate, where a smaller R_{crit}/W promotes more frequent nucleation events, potentially leading to a higher nucleation density and finer microstructure. Conversely, a larger value can suppress nucleation, resulting in lower nucleation density and coarser structures. $\Delta x/W$ is a numerical parameter, which represents the grid resolution. It is introduced by the numerical method, i.e. finite difference method. It contributes the Laplacian term in the phase field equations. A larger $\Delta x/W$ can leads to higher nucleation density. Additionaly, A/η_{crit} affects the amplitude of the noise term in the Langevin equation, influencing the stochasticity of the nucleation process. Higher values of A/η_{crit} increase the influence of thermal fluctuations, potentially leading to more random nucleation events and a less predictable microstructure. Lower values significantly reduce this stochastic effect, making the nucleation fomation unsuccessfully. Typically, the medium density category is the most relevant range for the phase field simulations, as it corresponds to phase transformation under the condition of quasi-equilibrium or with moderate chemical driving force, while the high nucleation density region corresponds to transformations with large undercooling or supersaturation. From the phase field results, we find that the medium density parameter case only occupies the smallest parameter region. Obviously, this region depends on the choice of the category criteria, as discussed above. Consequently, the manual determination of suitable nucleation parameters in the phase field simulations demands large computational resources due to the high sensitivity, especially when the desired value needs to be located in the medium density domain.

Data driven strategy

In this paper, we propose a data-driven strategy for phase field modeling, where suitable machine learning techniques based on the initial phase field results are developed, which can optimize the phase field nucleation parameters, such that the intended nucleation density is achieved. To this end, a deep learning ANN is employed, as detailed in the methods section. Figure 4 illustrates our data-driven strategy for linking the phase field and the machine learning modeling. For this purpose, the nucleation density ρ_{nuc} is obtained from the phase field simulations, together with the previously identified relevant and dimensionless features amplitude A/η_{crit}, grid spacing $\Delta x/W$ and critical nucleation radius R_{crit}/W. This four-dimensional data set is taken as input data for the further data driven regression analysis, as detailed in the methods section. For the classification, the nucleation density is translated to the three different regimes according to the description above. Here, a non-linear relation between the nucleation density and three dimensionless nucleation features is expected. Both a classification and a regression multi-layer ANN model are trained in the present work, in the spirit of a fast but coarse analysis using a classification model first and a more refined regression model for tuning the continuous model parameters afterwards. In detail, the purpose of the classification model is to assign categories of "no nucleation", "medium density" and "high density" of nuclei to different spatial regions in the simulation domain, as discussed in the previous section "Phase field simulation analysis". This information can be used to avoid time-consuming nucleation attempts in regions of the phase field simulation domain where it is not expected to play a role. The regression model predicts the nucleation density with regard to specific nucleation parameters. Consequently, the combination of these two models can significantly decrease the computation efforts of the phase field simulations. For instance, the chemical driving and the surface energy are essentially constant during the typical operation of an interconnect material for a specific temperature. Hence, the nucleation parameter R_{crit}/W is constant. To reproduce experimental results, suitable amplitudes of the fluctuations have to be determined through the regression model, depending on the chosen grid spacing.

Classification model

As a first step, a classification model is developed to evaluate the nucleation behavior. Here, we use the nucleation density as the criterion to divide the parameter space into three different domains, as mentioned before, i.e., no nucleation, medium density and high density. According to this classification model, unnecessary computations in the phase field simulation could be avoided, such as the possibility of nucleation events in the "no nucleation region" could be skipped and the check loop for the appearance of new nuclei in the "medium density region" could be done less frequently, or larger time steps could be used.

Technically, the prediction dataset in our classification model consists of 47 data points and the trained model accuracy reaches 94.54 %. Figure 5 shows the distribution of the three different nucleation regimes. The classification algorithm calculates actually the possibility of each category and the final category is determined according to the highest possibility. The yellow, green and blue regimes in the plot reflect the high, medium and no nucleation density regimes, respectively. Although the domain boundaries between these regimes are not straight, it can be useful to use approximating dividing planes instead, as indicated in the figure. Such separators are easy to parameterize and to use directly in PF simulations to discriminate locally between the different nucleation regimes. Obviously, stronger thermal noise

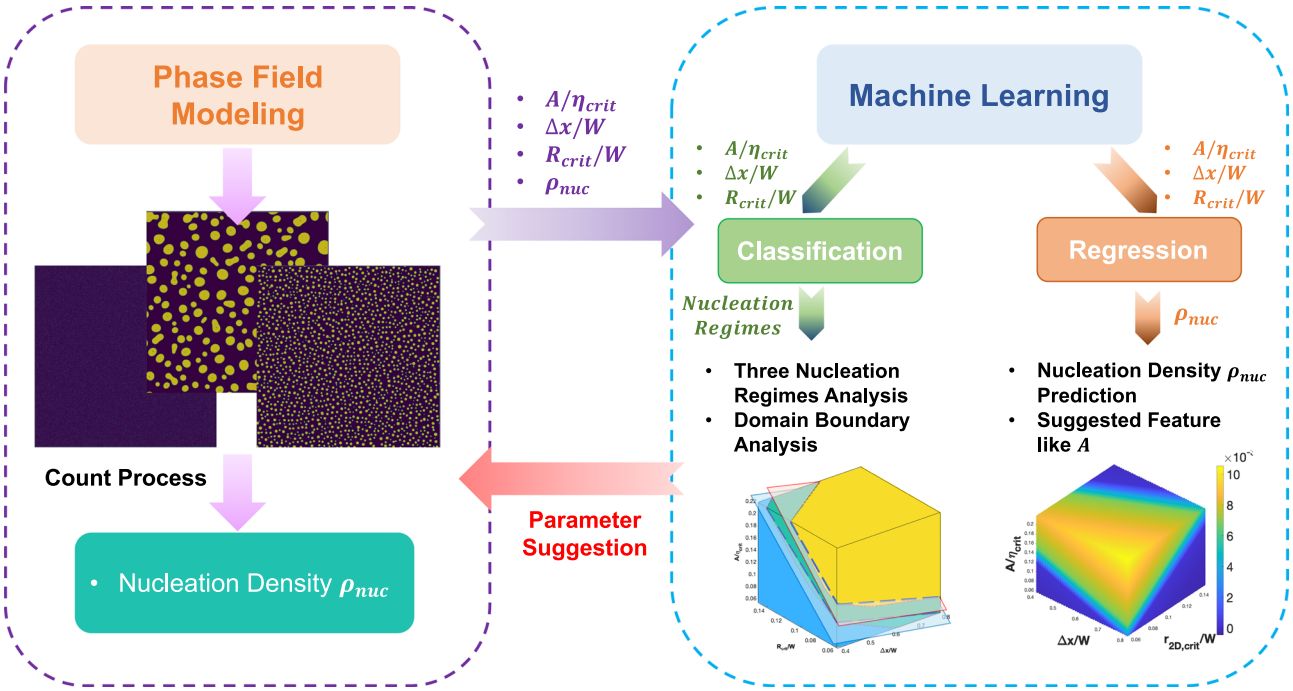
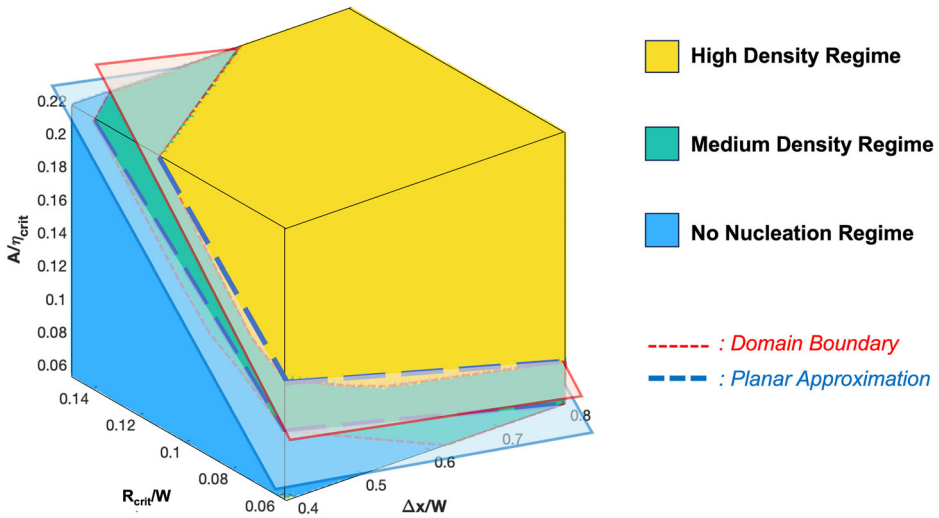


Fig. 4 | Data-driven strategy for linking phase field and machine learning models. Classification and regression machine learning models are developed based on three dimensionless features amplitude A/η_{crit}, grid spacing $\Delta x/W$ and critical nucleation radius R_{crit}/W and the nucleation density ρ_{nuc} from phase field modeling.

Fig. 5 | 3D plot of the predicted nucleation regime based on the machine learning classification model. Three different colors yellow, green and blue represent high, medium and no nucleation regimes, respectively. The surrounding red dashed line marks the border of the domain boundaries of the three different categories, whereas the blue dashed line indicates the intersection of the planar approximation. These planar approximations of the regimes can be used as simplified boundaries in phase field simulations e.g. to suppress unnecessary nucleation attempts in the "no nucleation regime" for saving computer time.



and lower critical radii favor high nucleation densities. The dependence on the grid spacing as a numerical parameter is less pronounced, exhibiting easier nucleation for coarser grids, similar to the discussion in the Ode et al. research[43]. Overall, the machine learning classification model shows good consistency with the phase field model, see Fig. 3.

Regression model

Beyond the classification, also a regression model is developed to predict the nucleation density directly. As mentioned before, the goal is to predict efficiently suitable parameters for the desired nucleation behavior in the PF simulations. Typically, a single PF simulation with the system dimensions given in the methods section requires single node computation times of several minutes. In contrast, the resulting ML predictions are about 10^6 times faster. Figure 6 (a) shows the predicted nucleation density versus the phase field results. The mean absolute error (MAE), root mean squared error (RMSE) and R-square are 0.0002 W^{-2}, 0.0002 W^{-2} and 0.9863, respectively. It is worthwhile to mention that the MAE and RMSE are identical, which indicates that the error distribution of the prediction is uniform. Furthermore, the R-square here is close to one, expressing that the machine learning predictions match the phase field values very well. Close to the origin, some deviations can be found, which is due to the fact that the ML model can predict unphysical negative nucleation densities, which do not play a role. Thus, the machine learning model possesses comparable accuracy and shows a good agreement with the phase field model.

To investigate the impact of three identified independent features, A/η_{crit}, R_{crit}/W and $\Delta x/W$, on the nucleation density, Fig. 7 (a) shows a 3D heat map of the nucleation density based on the machine learning regression model. Overall, it shows a similar tendency as the previous classification model. Furthermore, to clarify the interaction of the different features'

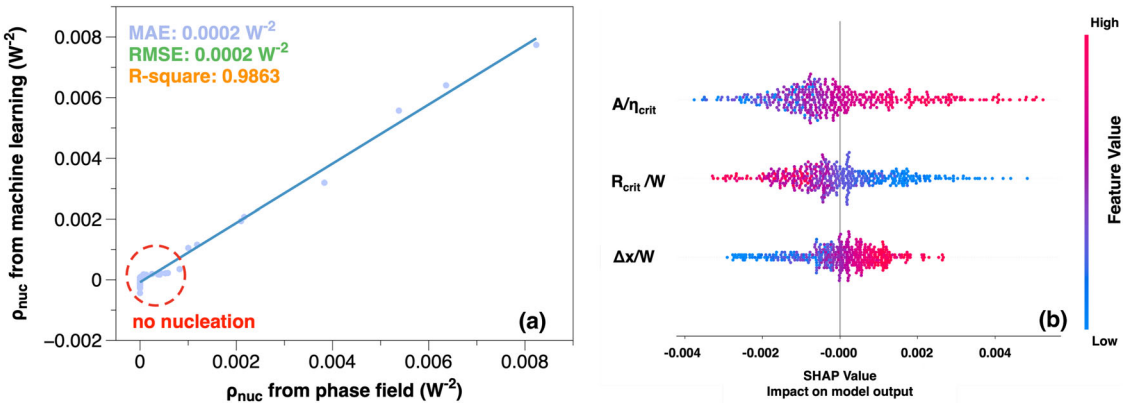


Fig. 6 | Evaluation of machine learning regression model. a Comparison of the machine learning predicted nucleation density in the prediction data set to the phase field results, (**b**) SHAP value analysis for the nucleation density regression model.

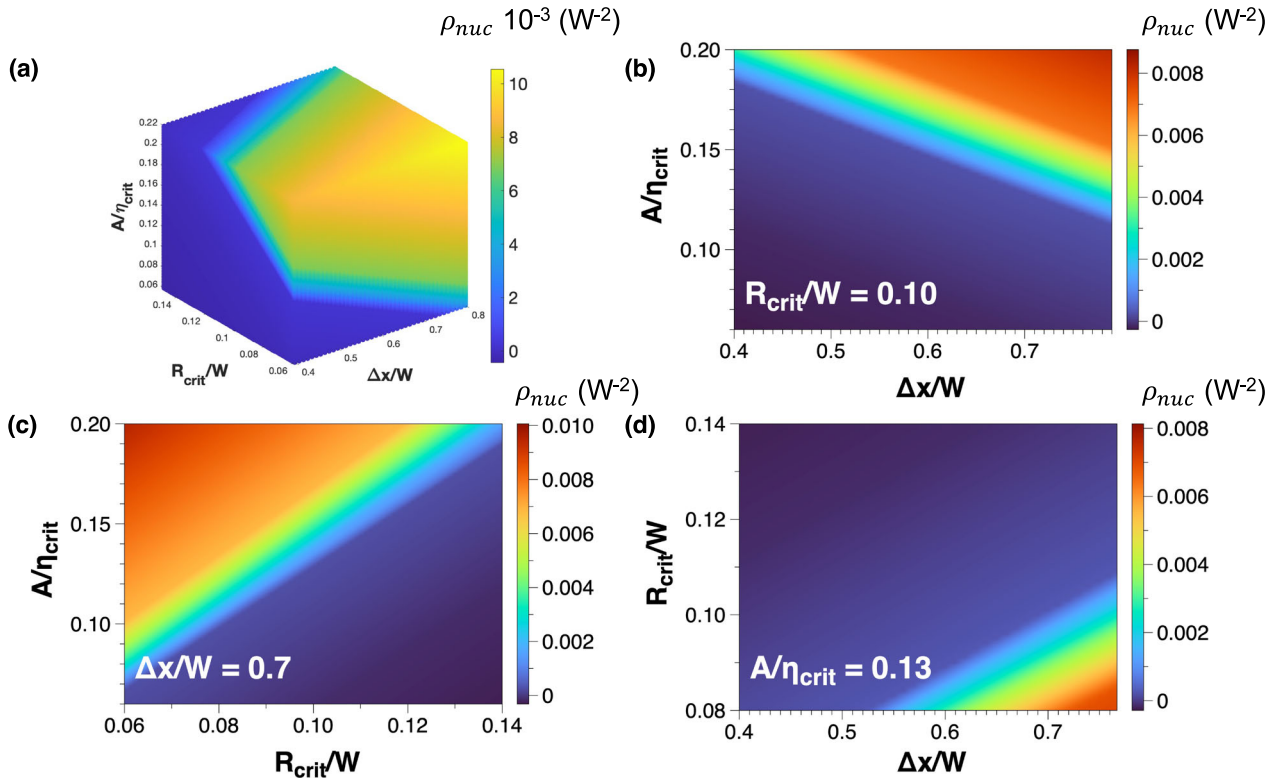


Fig. 7 | Analysis of machine learning regression model results. a 3D heatmap of the predicted nucleation density ρ_{nuc} based on the machine learning regression model, (**b**) 2D plot at $R_{crit}/W = 0.10$, (**c**) 2D plot at $\Delta x/W = 0.7$, (**d**) 2D plot at $A/\eta_{crit} = 0.13$. The color bar in each graph reflects the values of nucleation density. Whereas the influence on the nucleation density is weak in wide parameter ranges, strong gradients appear in an intermediate regime, emphasizing the importance of accurate parameter selection for obtaining the desired nucleation density.

impact on the nucleation behavior, three cross-sectional views are shown in Fig. 7, where one of the variables is set constant and the other two features vary in a certain range. Specifically, R_{crit}/W, $\Delta x/W$ and A/η_{crit} are set as 0.1, 0.71 and 0.13, respectively, see Fig. 7b–d. We observe a monotonic dependence of the nucleation density on the control parameters, exhibiting a rather steep increase in the medium nucleation density, following the rationale discussed in the preceding section. An additional SHAP value analysis[51] is employed, which can be used to explain the contribution of individual features to a model's predictions, see Fig. 6b. The color bar indicates the magnitude of the feature values, and a positive/negative SHAP value means that the related features have a positive/negative influence on the final prediction. Here, a higher amplitude A/η_{crit}, a lower critical nucleus radius R_{crit}/W and a higher grid spacing $\Delta x/W$ may generally lead to a higher nucleation density. The SHAP value calculation supports the previous analysis and is consistent with the previous phase field simulation in Fig. 3.

At this point we can return to our intended goal of the coupled phase field and machine learning simulations, to identify efficiently suitable numerical parameters for the nucleation model. In a phase field model, the grid spacing Δx is selected by the chosen discretization and the (local) value of the critical radius R_{crit} is related to the supersaturation or undercooling, hence, it is a known quantity, as discussed in section "Data driven strategy". The goal is then to determine a suitable noise strength A, in order to obtain the desired nucleation density. As suggested by Fig. 7, the functional dependence between these parameters is monotonic, hence a numerical inversion is straightforward, as demonstrated explicitly in Fig. 8. Interestingly, the functional form looks like a

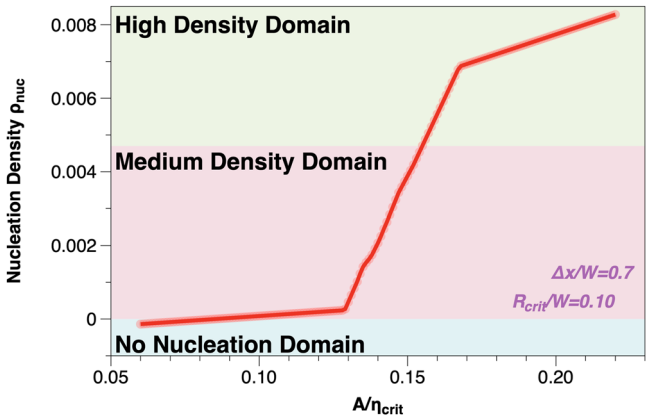


Fig. 8 | Relationship between the nucleation density ρ_{nuc} and the noise of fluctuation A/η_{crit} for fixed values $\Delta x/W = 0.7$ and $R_{\mathrm{crit}}/W = 0.10$.

piecewise definition, and the detailed shape may depend on the training data set, the chosen network topology and the transfer function. Nevertheless, the results show clearly, that (i) a proper noise strength can be selected uniquely, and (ii) that the dependence of the nucleation density on the numerical model parameters can be rather sensitive, and therefore the developed coupled phase field and machine learning model is beneficial for identifying the required parameters efficiently, which can save high computational efforts by avoiding trial-and-error approaches.

As a result, we have therefore developed an efficient scheme for identifying the required noise strength for the selected phase field model. We note that this parameter selection strategy is not restricted to the current application of oxide nucleation, but could also be transferred to other applications, like the determination of optimized kinetic parameters depending on the chosen numerical discretization, complementing techniques like the thin interface approach in situations, where exact relationships are not available.

Discussion

Tuning numerical parameters for phase field simulations in order to reproduce a desired nucleation behavior can be a tedious and computationally expensive task. In this paper, we have proposed a data-driven strategy for phase field nucleation modeling to overcome this challenge. Therefore, a combination of a grand canonical phase field model together with machine learning approaches for the classification and regression of the nucleation behavior has been investigated, which allows to predict a suitable noise strength for obtaining a desired nucleation density. The approach is demonstrated for Cr_2O_3 oxide nucleation in ferritic interconnector steels. According to the analysis of the results the following main conclusions can be drawn:

(i) The discussed phase field nucleation model is in agreement with expectations from a classical JMAK description, which demonstrates the suitability of the approach. We note that the choice of the underlying PF model is not critical, and similar parameter optimization strategies can also be used for other models.

(ii) We have identified that the most relevant features for quantifying the nucleation behavior are the Langevin noise strength, the critical radius for oxide nucleation and the numerical grid spacing of the phase field simulations. Such a feature identification from physical arguments is useful for efficient and accurate machine learning models. As the nucleation density depends smoothly on the phase field parameters, the used number of data points leads already to an accurate description, and additional training data is expected to lead only to minor improvements of the predictions.

(iii) The machine learning approach allows to classify the expected nucleation behavior, depending on the local values of the aforementioned parameters. This enables accelerated simulations by neglecting nucleation attempts in regions where such events are unlikely. Also, it is conceivable that a coarser time discretization can be accepted, provided that the ML approach predicts unfavorable nucleation conditions.

(iv) The selection of a suitable Langevin noise strength in order to reproduce the desired nucleation behavior using machine learning can help to avoid additional phase field screening simulations. Instead, the pretrained machine learning model allows us to skip this step with high accuracy and low computational effort.

Finally, we mention that such a data-driven strategy can be also of use for other applications related to phase-field simulations, as selected training results, which are obtained from small-scale sample simulations, can be used efficiently to optimize large-scale simulations. In particular, the developed model can be used for further investigations of oxide scale formation and growth for SOFC applications and degradation prevention.

Methods

Grand potential-based phase field model

Based on the work by Plapp[52] and Aagesen[53], a grand potential-based phase field model is formulated for N different phases and K components. For each phase, enumerated by α, there are p_α different grain orientations, hence the phase field order parameter for phase α is denoted as $\vec{\eta}_\alpha = (\eta_{\alpha 1}, \eta_{\alpha 2}, \ldots, \eta_{\alpha p_\alpha})$.

The number density ρ_I of component I is defined as $\rho_I = c_I/V_a$, where c_I is the local atomic fraction of solvent I, and V_a is the average atomic volume.

The grand potential functional is written as

$$\Omega = \int_V \left(\omega_{\mathrm{grad}} + \omega_{\mathrm{mw}} + \omega_{\mathrm{chem}}\right) dV. \tag{6}$$

The first term ω_{grad} contains the gradient energy, which is given by

$$\omega_{\mathrm{grad}} = \frac{\kappa}{2} \sum_{\alpha=1}^{N} \sum_{j=1}^{p_\alpha} (\nabla \eta_{\alpha j})^2, \tag{7}$$

with κ being the (isotropic) gradient energy coefficient. The second term reflects the multi-well potential density and has the form

$$\omega_{\mathrm{mw}} = M_0 \sum_{\alpha=1}^{N} \sum_{i=1}^{p_\alpha} \left(\frac{\eta_{\alpha i}^4}{4} - \frac{\eta_{\alpha i}^2}{2}\right) + \sum_{\alpha=1}^{N} \sum_{i=1}^{p_\alpha} \left(\sum_{\beta=1}^{N} \sum_{j=1, \alpha i \neq \beta j}^{p_\beta} \frac{\gamma_{\alpha i \beta j}}{2} \eta_{\alpha i}^2 \eta_{\beta j}^2 \right) + \frac{1}{4}, \tag{8}$$

where M_0 is a constant of dimension energy per unit volume. $\gamma_{\alpha i \beta j}$ is related to the interface profile between the grain i of phase α and grain j of phase β. At here, $\gamma_{\alpha i \beta j} = \gamma_{\beta j \alpha i} = 1.5$ to generate symmetrical phase field profiles[54]. The last term, ω_{chem}, is the contribution of chemical grand potential density, which can be expressed as

$$\omega_{\mathrm{chem}} = \sum_{\alpha=1}^{N} h_\alpha \omega_\alpha, \tag{9}$$

with h_α being an interpolation function, which indicates the phase fraction and is formulated as

$$h_\alpha = \frac{\sum_i^{p_\alpha} \eta_{\alpha i}^2}{\sum_\beta^N \sum_j^{p_\beta} \eta_{\beta j}^2}. \tag{10}$$

In addition, ω_α is the grand potential density of phase α, which can be written as

$$\omega_\alpha = f_\alpha - \rho_A \mu_A - \rho_B \mu_B - \ldots - \rho_{K-1} \mu_{K-1}, \tag{11}$$

where f_α is the chemical free energy density of phase α, and μ_A is the chemical potential of component A in the α phase, relative to the solvent with index K. In this work, parabolic free energy functions are utilized for constructing the dimensionless expressions f_α, i.e.

$$f_\alpha = \sum_{I=1}^{K-1} \epsilon_I^\alpha (c_I - c_I^\alpha)^2/2 + f_\alpha^0. \tag{12}$$

The parabolic coefficient ϵ_I^α determines the curvature of the parabola (dimension: energy density) of component I in phase α. The dimensionless chemical potential of component I in phase α can be expressed as $\mu_I = \partial f_\alpha/\partial \rho_I = V_a \epsilon_I^\alpha (c_I - c_I^\alpha)$. In the spirit of the grand canonical phase field model we assume in the following that they are locally balanced for each component, that we can deal with a single chemical potential μ_I at each point in space. The grand potential density of phase α can be expressed for the given choice of free energy densities as

$$\omega_\alpha = f_\alpha^0 - \sum_{I=1}^{K-1} \mu_I^2/(2V_\alpha^2 \epsilon_I^\alpha) - \mu_I c_I^\alpha / V_\alpha. \tag{13}$$

According to the Allen-Cahn equation, the time evolution of order parameter $\eta_{\alpha i}$ becomes

$$\begin{aligned} \tau \dot{\eta}_{\alpha i} = \kappa \nabla^2 \eta_{\alpha i} - M_0 \left(\eta_{\alpha i}^3 - \eta_{\alpha i} + 2\eta_{\alpha i} \sum_{\beta=1}^{N} \sum_{j=1,\alpha i \neq \beta j}^{p_\beta} \gamma_{\alpha i \beta j} \eta_{\beta j}^2 \right) \\ - \sum_{\beta=1}^{N} \frac{\partial h_\beta}{\partial \eta_{\alpha i}} \omega_\beta. \end{aligned} \tag{14}$$

Then, for each solute species A the diffusion equation reads

$$\frac{\partial \rho_A}{\partial t} = \nabla \cdot \sum_{I=1}^{K-1} M_{AI} \nabla \mu_I, \tag{15}$$

where M_{AI} is the mobility coefficient. Using the chain rule, we have

$$\frac{\partial \rho_A}{\partial t} = \sum_{I=1}^{K-1} \frac{\partial \rho_A}{\partial \mu_I} \frac{\partial \mu_I}{\partial t} + \sum_{\beta=1}^{N} \sum_{i=1}^{p_\beta} \frac{\partial \rho_A}{\partial \eta_{\beta i}} \frac{\partial \eta_{\beta i}}{\partial t} \tag{16}$$

Substituting Eq. (15) in Eq. (16), we obtain

$$\sum_{I=1}^{K-1} \chi_{AI} \frac{\partial \mu_I}{\partial t} = \nabla \cdot \sum_{I=1}^{K-1} M_{AI} \nabla \mu_I - \sum_{\beta=1}^{N} \sum_{i=1}^{p_\beta} \frac{\partial \rho_A}{\partial \eta_{\beta i}} \frac{\partial \eta_{\beta i}}{\partial t}, \tag{17}$$

with χ_{AI} being the susceptibility, known as the thermodynamic factor. It is given by

$$\chi_{AI} = \frac{\partial \rho_A}{\partial \mu_I} = \frac{\partial}{\partial \mu_I} \sum_{\beta=1}^{N} h_\beta \rho_A^\beta = \sum_{\beta=1}^{N} h_\beta \chi_{AI}^\beta. \tag{18}$$

The explicit form of the chemical potential evolution equations provided by Eq. (17) can be obtained from a matrix inversion,

$$\begin{bmatrix} \dot{\mu}_A \\ \vdots \\ \dot{\mu}_{K-1} \end{bmatrix} = \begin{bmatrix} \chi_{AA} & \cdots & \chi_{AK-1} \\ \vdots & \ddots & \vdots \\ \chi_{K-1A} & \cdots & \chi_{K-1K-1} \end{bmatrix}^{-1} \begin{bmatrix} Q_A \\ \vdots \\ Q_{K-1} \end{bmatrix}, \tag{19}$$

with

$$Q_{X,X=A,B,\ldots,K-1} = \nabla \cdot \sum_{I=1}^{K-1} M_{XI} \nabla \mu_X - \sum_{\beta=1}^{N} \sum_{i=1}^{p_\beta} \frac{\partial \rho_X}{\partial \eta_{\beta i}} \frac{\partial \eta_{\beta i}}{\partial t}. \tag{20}$$

In the present work, the nucleation of β grains from an α matrix is modeled via PF simulations by using the aforementioned approaches. The total size of the two-dimensional PF simulation domain is $1024\,\Delta x \times 1024\,\Delta y$, using $\Delta x/W = \Delta y/W = 0.4$. The thermodynamic descriptions of Cr_2O_3 and the ferritic matrix are taken from the Fe-Cr-O ternary database by Taylor et al.[55]. These free energy densities are fitted by a parabolic function near the local-equilibrium states for different temperatures. As Cr_2O_3 is stoichiometric, the curvature of the parabola of the Cr_2O_3 free energy density is 1000 times larger than the ferritic matrix phase. Consequently, the concentrations of Cr and O in Cr_2O_3 grains are nearly independent of the temperature. These treatments not only reproduce all the necessary thermodynamic properties but also save computational efforts. Then, the dimensionless capillary length $d_0 = M_0 W/6$ is set as 0.1 in the PF simulations, which is sufficient for the production of the scale of the curvature of Cr_2O_3 grains, and the dimensionless parameter κ in Eq. (7) is $\kappa = 3d_0 W/4$. In addition, the diffusivity in this work is normalized by using $\tilde{D}_i = D_i/D_{Cr}$ (i=Cr, O), with D_{Cr} being the diffusivity of Cr in the ferritic matrix phase. The dimensionless diffusivities of Cr and O in the ferritic matrix phase are set as 1 and 10^4 according to the diffusivity ratio in ref. 56, while the diffusion of Cr and O in Cr_2O_3 grains are negligible. In order to enhance the computational efficiency, periodic boundary conditions are prescribed at all the boundaries and GPU acceleration is applied in the simulations[57].

Artificial neural networks

The ANN algorithm is inspired by the biological neural network of animal brains and consists of multi-layer structures. A precious and reliable model could be obtained by utilizing suitable hyperparameters like efficient activation functions and appropriate layers/nodes[58]. Generally, the multi-layer structure involves input layer, hidden layer and output layer. The different layers play significant roles in the neural network and could be considered as a filter for data. Deep learning is built on different, densely connected neural layers and implements a form of progressive data distillation[59].

In order to carry out the machine learning method, the open-source package Tensorflow is used in the present work[60]. A total number of 547 data sets obtained from the phase field simulations are randomly split into the training (85% of the entire data) and testing dataset (15%). All the data are normalized by using the MinMaxScaler algorithm from the open-source package scikit-learn[61]. The classification model in this work consists of four different hidden layers, where each dense layer has 8 nodes, and the output layer has the aforementioned three different categories. As an optimizer, the root mean square propagation (RMSprop) is used. The metric and loss function for the classification model are sparse categorical accuracy and sparse categorical cross entropy, respectively. The regression model also contains four different hidden layers, each dense layer possesses 8 nodes. The final output layer delivers the desired nucleation density. The popular optimizer Adaptive Moment Estimation (Adam) is employed for the model training. Here, we use MSE as the metric and loss function. The evaluation of final model is based on the MAE, MSE and R^2 calculation. Additionally, a dropout layer is introduced between each hidden layer with a dropout rate of 0.5 and an early stopping with a patience of 10 is ustilized to avoid overfitting. For the further data prediction, the parameter ranges are A/η_{crit}: [0.06, 0.22], $\Delta x/W$: [0.4, 0.8] and R_{crit}/W: [0.06, 0.15], respectively.

Data availability

The data and codes used in this study can be shared upon reasonable request.

Received: 10 June 2024; Accepted: 14 October 2024;
Published online: 30 October 2024

References

1. Steinbach, I. Phase-field model for microstructure evolution at the mesoscopic scale. *Annu. Rev. Mater. Res.* **43**, 89–107 (2013).

2. Chen, Long-Qing Phase-field models for microstructure evolution. *Annu. Rev. Mater. Res.* **32**, 113–140 (2002).
3. Echebarria, B., Folch, R., Karma, A. & Plapp, M. Quantitative phase-field model of alloy solidification. *Phys. Rev. E* **70**, 061604 (2004).
4. Spatschek, R., Müller-Gugenberger, C., Brener, E. & Nestler, B. Phase field modeling of fracture and stress-induced phase transitions. *Phys. Rev. E* **75**, 066111 (2007).
5. Moelans, N., Wendler, F. & Nestler, B. Comparative study of two phase-field models for grain growth. *Comput. Mater. Sci.* **46**, 479–490 (2009).
6. Ji, K., Clarke, A. J., McKeown, J. T. & Karma, A. Microstructure development during rapid alloy solidification. *MRS Bulletin* **49**, 556–567 (2024).
7. Karma, A., Kessler, D. A. & Levine, H. Phase-field model of mode iii dynamic fracture. *Phys. Rev. Lett.* **87**, 045501 (2001).
8. Brener, E., Spatschek, R. & Karma, A. Phase field modeling of crack propagation. *Philos. Mag.* **91**, 75–95 (2011).
9. Spatschek, R., Hartmann, M., Brener, E., Müller-Krumbhaar, H. & Kassner, K. Phase field modeling of fast crack propagation. *Phys. Rev. Lett.* **96**, 015502 (2006).
10. Castro, M., Travasso, RuiD. M. & Oliveira Joana, C. R. E. The phase-field model in tumor growth. *Philos. Mag.* **91**, 183–206 (2011).
11. Karma, A. & Rappel, Wouter-Jan Quantitative phase-field modeling of dendritic growth in two and three dimensions. *Phys. Rev. E* **57**, 4323 (1998).
12. Wang, K., Boussinot, G., Brener, E. A. & Spatschek, R. Quantitative nondiagonal phase field modeling of eutectic and eutectoid transformations. *Phys. Rev. B* **103**, 184111 (2021).
13. Wang, K., Boussinot, G., Hüter, C., Brener, E. A. & Spatschek, R. Modeling of dendritic growth using a quantitative nondiagonal phase field model. *Phys. Rev. Mater.* **4**, 033802 (2020).
14. Gránásy, L. ászló, Börzsönyi, Tamás & Pusztai, Tamás Nucleation and bulk crystallization in binary phase field theory. *Phys. Rev. Lett.* **88**, 206105 (2002).
15. Zaeem, MohsenAsle & El Kadiri, H. An elastic phase field model for thermal oxidation of metals: Application to zirconia. *Comput. Mater. Sci.* **89**, 122–129 (2014).
16. Sherman, Q. C. & Voorhees, P. W. Phase-field model of oxidation: Equilibrium. *Phys. Rev. E* **95**, 032801 (2017).
17. Kim, K., Sherman, Q. C., Aagesen, L. K. & Voorhees, P. W. Phase-field model of oxidation: Kinetics. *Phys. Rev. E* **101**, 022802 (2020).
18. Wang, R. et al. A phase-field study on internal to external oxidation transition in high-temperature structural alloys. *JOM* **74**, 1435–1443 (2022).
19. Choudhary, K. et al. Recent advances and applications of deep learning methods in materials science. *npj Computational Mater.* **8**, 59 (2022).
20. Morgan, D. & Jacobs, R. Opportunities and challenges for machine learning in materials science. *Annu. Rev. Mater. Res.* **50**, 71–103 (2020).
21. Coelho, LeonardoBertolucci et al. Reviewing machine learning of corrosion prediction in a data-oriented perspective. *npj Mater. Degrad.* **6**, 8 (2022).
22. Alamri, A. H. Application of machine learning to stress corrosion cracking risk assessment. *Egypt. J. Pet.* **31**, 11–21 (2022).
23. Diao, Y., Yan, L. & Gao, K. Improvement of the machine learning-based corrosion rate prediction model through the optimization of input features. *Mater. Des.* **198**, 109326 (2021).
24. Roy, A. et al. Machine-learning-guided descriptor selection for predicting corrosion resistance in multi-principal element alloys. *npj Mater. Degrad.* **6**, 9 (2022).
25. Géron, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. "O'Reilly Media, Inc.", (2022).
26. Shen, Zhong-Hui et al. Phase-field modeling and machine learning of electric-thermal-mechanical breakdown of polymer-based dielectrics. *Nat. Commun.* **10**, 1843 (2019).
27. Montes de Oca Zapiain, D., Stewart, J. A. & Dingreville, R. émi Accelerating phase-field-based microstructure evolution predictions via surrogate models trained by machine learning methods. *npj Computational Mater.* **7**, 3 (2021).
28. Chehrmonavari, H. et al. Hybridizing solid oxide fuel cells with internal combustion engines for power and propulsion systems: A review. *Renew. Sustain. Energy Rev.* **171**, 112982 (2023).
29. Golkhatmi, SanazZarabi, Asghar, MuhammadImran & Lund, P. D. A review on solid oxide fuel cell durability: Latest progress, mechanisms, and study tools. *Renew. Sustain. Energy Rev.* **161**, 112339 (2022).
30. Xu, Q. et al. A comprehensive review of solid oxide fuel cells operating on various promising alternative fuels. *Energy Convers. Manag.* **253**, 115175 (2022).
31. Hanif, MuhammadBilal et al. Recent advancements, doping strategies and the future perspective of perovskite-based solid oxide fuel cells for energy conversion. *Chem. Eng. J.* **428**, 132603 (2022).
32. Hu, Ying-Zhen, Yao, Shu-Wei, Li, Cheng-Xin, Li, Chang-Jiu & Zhang, Shan-Lin Influence of pre-reduction on microstructure homogeneity and electrical properties of aps mn1.5co1.5o4 coatings for sofc interconnects. *Int. J. Hydrog. Energy* **42**, 27241–27253 (2017).
33. Han, M., Peng, S., Wang, Z., Yang, Z. & Chen, X. Properties of fe–cr based alloys as interconnects in a solid oxide fuel cell. *J. Power Sources* **164**, 278–283 (2007).
34. Ebrahimifar, H. & Zandrahimi, M. Oxidation and electrical behavior of aisi 430 coated with cobalt spinels for sofc interconnect applications. *Surf. Coat. Technol.* **206**, 75–81 (2011).
35. Thublaor, T. & Chandra-ambhorn, S. High temperature oxidation and chromium volatilisation of aisi 430 stainless steel coated by mn-co and mn-co-cu oxides for sofc interconnect application. *Corros. Sci.* **174**, 108802 (2020).
36. Fujita, K., Ogasawara, K., Matsuzaki, Y. & Sakurai, T. Prevention of sofc cathode degradation in contact with cr-containing alloy. *J. Power Sources* **131**, 261–269 (2004).
37. Manjunath, N., Santhy, K. & Rajasekaran, B. Thermal expansion of crofer 22 apu steel used for sofc interconnect using in-situ high temperature x-ray diffraction. *Materials Today: Proceedings*, (2023).
38. Topcu, A., Öztürk, B. ülent & Cora, ÖmerNecati Performance evaluation of machined and powder metallurgically fabricated crofer®22 apu interconnects for sofc applications. *Int. J. Hydrog. Energy* **47**, 3437–3448 (2022).
39. Dheeradhada, V. S., Cao, H. & Alinger, M. J. Oxidation of ferritic stainless steel interconnects: Thermodynamic and kinetic assessment. *J. Power Sources* **196**, 1975–1982 (2011).
40. Williams, C. A., Unifantowicz, P., Baluc, N., Smith, GeorgeD. W. & Marquis, E. A. The formation and evolution of oxide particles in oxide-dispersion-strengthened ferritic steels during processing. *Acta Materialia* **61**, 2219–2235 (2013).
41. Safikhani, A., Esmailian, M., Tinatiseresht, T. & Darband, GhasemBarati High temperature cyclic oxidation behavior of ferritic stainless steel with addition of alloying elements nb and ti for use in sofcs interconnect. *Int. J. Hydrog. Energy* **41**, 6045–6052 (2016).
42. Yin, X., Spatschek, R., Menzler, N. H. & Hüter, C. A pragmatic transfer learning approach for oxygen vacancy formation energies in oxidic ceramics. *Materials* **15**, 2879 (2022).
43. Ode, M. & Ohnuma, I. A thermal fluctuation-based nucleation method for phase-field models. *Computational Mater. Sci.* **194**, 110448 (2021).
44. Choudhury, A., Kellner, M. & Nestler, B. A method for coupling the phase-field model based on a grand-potential formalism to thermodynamic databases. *Curr. Opin. Solid State Mater. Sci.* **19**, 287–300 (2015).

45. William, J. & Mehl, R. Reaction kinetics in processes of nucleation and growth. *Trans. Metall. Soc. AIME* **135**, 416–442 (1939).
46. Avrami, M. Kinetics of phase change. i general theory. *J. Chem. Phys.* **7**, 1103–1112 (1939).
47. Avrami, M. Kinetics of phase change. ii transformation-time relations for random distribution of nuclei. *J. Chem. Phys.* **8**, 212–224 (1940).
48. Avrami, M. Granulation, phase change, and microstructure kinetics of phase change. iii. *J. Chem. Phys.* **9**, 177–184 (1941).
49. Kolmogorov, AndreiNikolaevich On the statistical theory of the crystallization of metals. *Bull. Acad. Sci. URSS (Cl. Sci. Math. Nat.)* **3**, 335 (1937).
50. Wu, W. et al. Phase field benchmark problems for nucleation. *Computational Mater. Sci.* **193**, 110371 (2021).
51. Lundberg, S. M. & Lee, Su-In. A unified approach to interpreting model predictions. In Guyon, I. et al. editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., (2017).
52. Plapp, M. Unified derivation of phase-field models for alloy solidification from a grand-potential functional. *Phys. Rev. E* **84**, 031601 (2011).
53. Aagesen, L. K., Gao, Y., Schwen, D. & Ahmed, K. Grand-potential-based phase-field model for multiple phases, grains, and chemical components. *Phys. Rev. E* **98**, 023309 (2018).
54. Moelans, N., Blanpain, B. & Wollants, P. Quantitative analysis of grain boundary properties in a generalized phase field model for grain growth in anisotropic systems. *Phys. Rev. B* **78**, 024113 (2008).
55. Taylor, J. R. & Dinsdale, A. T. A thermodynamic assessment of the cr-fe-o system. *Int. J. Mater. Res.* **84**, 335–345 (1993).
56. Auinger, M., Naraparaju, R., Christ, H. J. & Rohwerder, M. Modelling high temperature oxidation in iron–chromium systems: Combined kinetic and thermodynamic calculation of the long-term behaviour and experimental verification. *Oxid. Met.* **76**, 247–258 (2011).
57. Thörnig, P. & von St. Vieth, B. JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre. *J. large-scale Res. facilities* **7**, A182 (2021).
58. Arriagada, J., Olausson, P. & Selimovic, A. Artificial neural network simulator for sofc performance prediction. *J. Power Sources* **112**, 54–60 (2002).
59. Chollet, F. *Deep learning with Python*. Simon and Schuster, (2021).
60. Abadi, M. et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, (2016).
61. Pedregosa, F. et al. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Acknowledgements

The work was funded by the BMBF projects WirLebenSOFC (03SF0622B) and SOC-Degradation 2.0 (03SF0621A). The authors gratefully acknowledge computing time on the supercomputer JURECA[57] at Forschungszentrum Jülich under grant no. CJIEK2C. Open access was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the framework of the Collaborative Research Centre 1548 (CRC 1548, project number 463184206, subproject B03).

Author contributions

Y.H: conceptualization, methodology, machine learning, visualization, analysis and writing; R.S: methodology, analysis and writing; K. W: conceptualization, methodology, phase field, visualization, analysis and writing.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Kai Wang.