

Learning time-dependent deposition protocols to design thin films via genetic algorithms

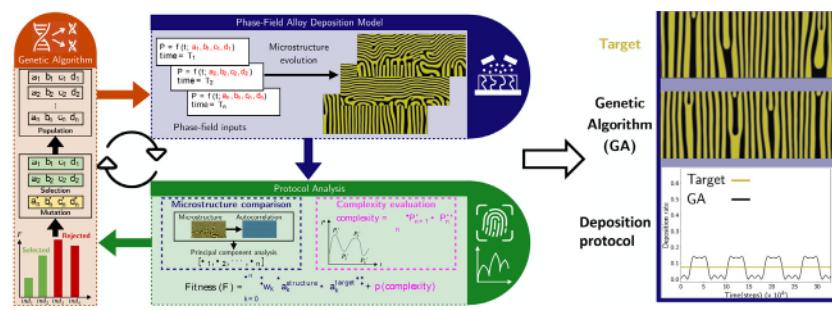
Saaketh Desai, Rémi Dingreville *

Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM 87185, USA

HIGHLIGHTS

- We discover time-dependent physical vapor deposition protocols that can achieve targeted microstructures.
- The design scheme couples a genetic algorithm with phase-field simulations of alloy deposition.
- Time-dependent deposition protocols offer new avenues to fabricate made-to-order thin films.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 15 April 2022

Revised 23 May 2022

Accepted 31 May 2022

Available online 6 June 2022

Keywords:

Physical vapor deposition

Thin films

Genetic algorithm

Phase-field method

ABSTRACT

Designing next generation thin films tailor-made for specific applications relies on the availability of robust process-structure-property relationships. Traditional structure zone diagrams that relate one or two deposition conditions to microstructures are limited to simple mappings, with machine-learning methods only recently attempting to relate multiple processing parameters to the final microstructure. Despite this progress, process-structure relationships are unknown for deposition conditions that vary during thin-film deposition, limiting the range of achievable microstructures and properties. We combine phase-field simulations with a genetic algorithm to identify and design time-dependent deposition protocols that achieve tailor-made microstructures. We simulate the physical vapor deposition of a binary-alloy thin film by employing a phase-field model, where deposition rates and diffusivities of the deposited species vary in time and are controlled via the genetic algorithm. Our genetic-algorithm-guided protocols achieve targeted microstructures with lateral and vertical concentration modulations, as well as more complex, hierarchical microstructures previously not described in classical structure zone diagrams. By elucidating the process-structure mechanisms during physical vapor deposition and using this knowledge to achieve precise thin-film microstructures, our algorithm provides insights to the thin film, physical vapor deposition, and film functionality communities looking for additional avenues to design novel thin-film microstructures.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Abbreviations: PVD, physical vapor deposition; PCA, principal component analysis; LCM, lateral concentration modulation; VCM, vertical concentration modulation; RCM, random concentration modulation; HCM, hierarchical concentration modulation; NPCM, nanoprecipitate concentration modulation; MEMPHIS, Mesoscale Multi-Physics Phase Field Simulator; VAE, variational autoencoder; GA, genetic algorithm.

* Corresponding author.

E-mail address: rdingre@sandia.gov (R. Dingreville).

<https://doi.org/10.1016/j.matdes.2022.110815>

0264-1275/© 2022 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Thin films are widely used in microelectronics and sensing applications [1–3], photovoltaic cells [4–6], non-volatile storage [7], and rechargeable batteries [8,9]. The ability to spatially pattern and tailor the microstructural evolution of materials during the

fabrication process, e.g. by forming concentration modulations within a binary solid solution, is key to achieve high-precision, high-reliability thin-film components [10–13]. Physical vapor deposition (PVD) is one of the fabrication techniques commonly used to grow such modulated microstructures, as shown in recent experiments for Cu-Mo [14] and Cu-Ta [15] systems. Designing thin films, tailor-made for specific applications, relies on thorough knowledge of the relationship between deposition conditions, microstructure, and properties. Traditional process–structure mappings take the form of structure zone diagrams that relate key deposition parameters such as species deposition rate, substrate temperature, and the pressure in the deposition chamber to the thin-film microstructure [16–23]. Established structure zone diagrams require an extensive experimental database of processing conditions and microstructure characterization. They are *de facto* limited to one or two deposition parameters and simplified visualizations, restricting the types of microstructures explored. These structure zone diagrams also fail to explain the underlying mechanisms that result in specific microstructures, hindering complete understanding of the microstructure formation process and associated properties. As an alternative, computational methods such as phase-field simulations [24] can prove useful in rapidly developing structure zone diagrams by simulating the evolution of microstructure formation during the growth of thin films under varying processing conditions [25–29]. More recently, machine-learning methods are being used to relate multiple processing conditions to thin-film microstructures [30–32] to create high-dimensional structure zone diagrams. Such approaches are beginning to be combined with other advanced techniques such as active-learning methods and high-throughput experiments to automate and accelerate the discovery of targeted thin-film materials [33,34].

Despite these advances, one limitation of current process–structure mappings is that they do not yet consider processing conditions that vary *during* the deposition. These mappings are thus potentially missing out on families of deposition protocols that could efficiently and accurately lead to the design of thin-film microstructures, as well as achieve complex, hierarchical microstructures with features across multiple length scales. Varying process conditions such as background gas pressure, background gas flow rate, bias voltage, or substrate temperature in time would change the plasma conditions and the associated growth mechanisms during the deposition, potentially resulting in hierarchical microstructures. Hierarchical microstructures deposited by PVD, as shown in [14,35], have been demonstrated to have improved mechanical properties such as plastic deformability. Researchers have recently demonstrated the opportunities offered by reinforcement learning methods to explore and exploit such time-dependent protocols for chemical vapor deposition [36]. Genetic algorithm approaches such as neuro-evolution have also shown promise in solving state-of-the-art reinforcement learning challenges efficiently [37] and have been recently used to optimize self-assembly protocols [38]. Genetic algorithms in general offer a broadly applicable, simple, gradient-free optimization approach, and have thus been used for materials design problems in conjunction with a computationally cheap, machine-learning based surrogate model. Examples of this approach include nanocluster structure optimization [39], optimization of composite design [40], and designing high glass transition polymers with high bandgap [41]. However, a genetic algorithm or reinforcement learning based optimization has yet to be explored for thin-film design.

In this work, we couple a genetic algorithm to an alloy-deposition phase-field model to discover time-dependent deposi-

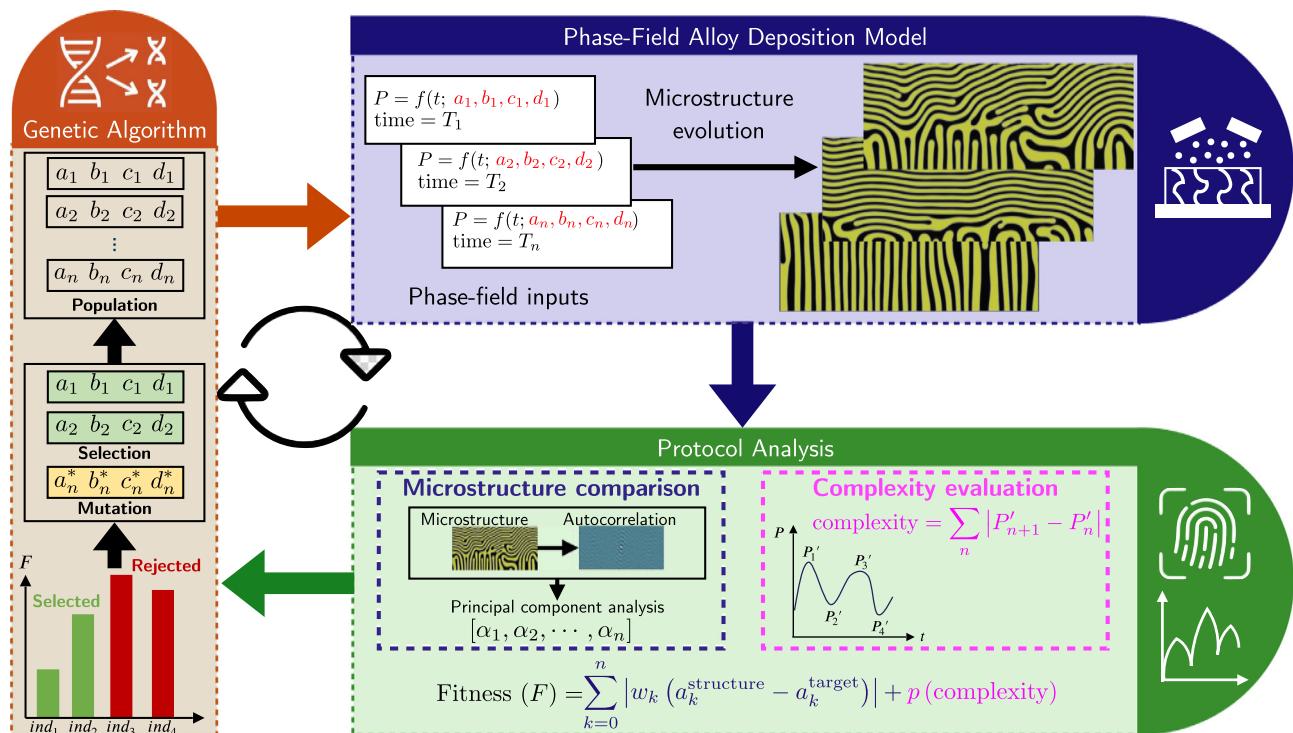


Fig. 1. Coupling a genetic algorithm to a phase-field alloy deposition model. The genetic algorithm optimizes a population of individuals where each individual is defined as a collection of parameters that collectively define a time-dependent protocol. The parameters describing each individual are passed to the alloy-deposition model, which then simulates the microstructure evolution during vapor deposition of the film. Each time-dependent deposition protocol is then assigned a fitness score by computing the accuracy and complexity of the protocol. The fitness score is then used to select certain individuals and discarding the rest, replacing the discarded individuals with mutations of the selected individuals. This selection and mutation step defines the next generation of individuals, and the cycle repeats until the desired microstructure is achieved.

tion protocols that achieve targeted microstructures. Fig. 1 illustrates this workflow. We use an alloy-deposition model that has previously successfully predicted the formation of experimentally observed concentration modulations during PVD [28]. We extend this model to simulate microstructure evolution for time-varying deposition parameters such as deposition rate and species mobility/diffusivity. The time-varying functional forms describing the deposition process could be simple step functions and linear ramps, or more complex time-varying functions represented as a Fourier series. The genetic algorithm chooses and controls the parameters of the functional forms dictating the time-dependent alloy-deposition protocol, and then iterates over multiple generations of deposition protocols to hone in on the deposition conditions necessary for yielding target microstructures. The genetic algorithm uses the fitness of individuals to perform selection and mutation operations such that deposition protocols with best fitness remain after many generations. The fitness of a given time-dependent deposition protocol is evaluated against two criteria: (i) comparison of the resulting microstructure to the target microstructure by using statistical microstructure functions, and (ii) complexity of the deposition protocol using a curvature-inspired metric [42]. The intent of the complexity criterion in our fitness measure is to favor simple protocols over complex ones, as time-varying deposition protocols in an experimental deposition chamber are limited to simple implementations such as pulsed and step function depositions. These two criteria are balanced by a parsimony parameter ‘*p*’ that controls the relative importance of accuracy versus complexity. High-dimensional microstructures are compared in a lower dimension by encoding microstructural features using two-point statistics [43], followed by a dimensionality reduction to a low-dimensional space using Principal Component Analysis (PCA) [44]. This strategy has been used successfully before to develop surrogate models that predict microstructure evolution during phase-field simulations [45], as well as accelerating phase-field simulations using recurrent neural networks [46]. We use a simple evolutionary algorithm with a mutation scheme similar to other genetic-algorithm-based design strategies [38], balancing exploration with exploitation. Details on each step of this workflow are described in Methods.

We find the genetic-algorithm-designed protocols to not only rediscover known structure zone diagrams for simple constant deposition conditions, but also uncover deposition pathways beyond human intuition that generate simple, monomodal concentration modulations that are well known and observed experimentally [47,48], as well as more complex hierarchical concentration modulations that have recently gained attention [14,15]. An analysis of the design space explored by the genetic algorithm indicates that the algorithm learns known correlations between deposition parameters and the resulting microstructure and uses these correlations to navigate the design space and discover optimal protocols.

2. Methods

2.1. Phase field modeling of thin film deposition

The phase-field model used to simulate physical vapor deposition of a binary-alloy thin film [28] uses conserved order parameters, ϕ and c to describe the structural and compositional ordering of the system as the thin film grows. The field variable ϕ ($\phi = \pm 1$) is meant to distinguish between the vapor and solid phases and is used to track the growing thin film. The compositional field variable c ($c = \pm 1$) distinguishes between the two phases in the growing thin film. The free energy functional \mathcal{F} describing the energy

landscape for the binary alloy system is defined in terms of these phase-field variables such that,

$$\mathcal{F} = \int \left\{ f_\phi + \frac{\kappa_\phi}{2} (\nabla \phi)^2 + s(\phi) \left(f_c + \frac{\kappa_c}{2} (\nabla c)^2 \right) \right\} d\Omega, \quad (1)$$

where $f_\phi = \omega_\phi \left(\frac{\phi^4}{4} - \frac{\phi^2}{2} \right)$ and $f_c = \omega_c \left(\frac{c^4}{4} - \frac{c^2}{2} \right)$ are double-well functions in ϕ and c respectively, and ω_ϕ and ω_c are the associated activation barriers. The variables κ_ϕ and κ_c are gradient penalty coefficients. Finally, the function $s(\phi) = \frac{(1+\phi)^2}{4}$ is an interpolation switching function used to consider compositional effects only in the growing thin film.

The composition evolution in the growing film is described via the Cahn-Hilliard equation such that,

$$\frac{\partial c}{\partial t} = \nabla \cdot \left(M_c(\phi, c) \nabla \frac{\partial \mathcal{F}}{\partial t} \right), \quad (2)$$

where M_c is the structurally and compositionally dependent mobility function. The mobility function M_c is constructed to describe both bulk and surface mobilities for each phase, $M_{(A,B)}^{\text{Bulk}}$ and $M_{(A,B)}^{\text{Surf}}$ respectively such that

$$M_c(\phi, c) = M^{\text{Bulk}} + M^{\text{Surf}}, \quad (3)$$

with,

$$M^{\text{Bulk}} = \frac{1}{4} (2 - \phi)(1 + \phi)^2 \left[h(c) M_A^{\text{Bulk}} + (1 - h(c)) M_B^{\text{Bulk}} \right] \quad (4)$$

$$M^{\text{Surf}} = e^{(-(\phi/\sigma_{\text{surf}}))^2} \left[h(c) M_A^{\text{Surf}} + (1 - h(c)) M_B^{\text{Surf}} - M^{\text{Bulk}} \right], \quad (5)$$

$$h(c) = \frac{(2 - c)(1 + c)^2}{4}. \quad (6)$$

A similar equation is solved for the evolution of ϕ , with the addition of a source-term that incorporates the deposition of species A or B.

$$\frac{\partial \phi}{\partial t} = \nabla \cdot \left[M_c(\phi) \nabla \frac{\partial \mathcal{F}}{\partial \phi} \right] + S(\vec{n}(\phi)), \quad (7)$$

$$\frac{\partial \rho}{\partial t} = \nabla \cdot [D_\rho \nabla \rho] - \nabla \cdot [\rho \cdot \mathbf{v}] - S(\vec{n}(\phi)). \quad (8)$$

The vector \mathbf{v} is the velocity field of the incident vapor flux. The term $S(\vec{n}(\phi))$ in Eqs. (7) and (8) couples the thin-film evolution to the incident-vapor flux and acts as the source term for interfacial growth as a function of the normal of the thin-film surface ($\vec{n}(\phi)$) at the expense of the vapor phase. This source term (S) is active only at the solid-vapor interface and is a function of the local density and velocity of the vapor phase (see Ref. [28] for additional details). Both \mathbf{v} and $S(\vec{n}(\phi))$ can be easily related back to experimental deposition conditions. Typical deposition rates in experiments are on the order of nm/s. In an experimental deposition chamber, the deposition rate varies as a function of the background gas pressure, sputtering power, and working distance between the target and the substrate. In our phase-field simulations, the magnitude of the velocity field of the incident vapor flux can be evaluated by calculating the average incoming ion energy for a given discharge voltage, background gas pressure, and working distance for a given PVD chamber geometry. The local vapor density between the target and substrate can be evaluated by calculating the emission rate, flight time, and average incoming ion energy. For instance, we could use the Monte Carlo code SIMTRA [49] to simulate the transport of atoms from the target to the substrate during PVD and calculate the velocity and the density of the gas-phase atoms.

This phase-field model is described in more details in a previous study [28]. This model was implemented, verified, and validated in our in-house Mesoscale Multi-Physics Phase Field Simulator (MEMPHIS), a modular phase-field simulation platform [28,50]. We previously demonstrated the ability of this model to correctly predict the emergence of lateral (LCM), vertical (VCM), random (RCM), and nanoprecipitate (NPCM) concentration modulation microstructures in the film under various constant deposition rates and species mobilities [28,15]. The spatio-temporal evolution of the thin-film growth evolution was solved on a two-dimensional (2D) computational grid using a finite-difference scheme with second-order central difference stencils for all spatial derivatives. Numerical temporal integration of the equations was performed using the explicit Euler method. All 2D simulations were performed on a uniform 512×512 numerical grid, a spatial discretization of $\Delta x = \Delta y = 1$, and a timestep $\Delta t = 10^{-3}$. The spatial discretization was chosen to preserve the conservation properties of the differential equation based on our finite-difference integration scheme. The spatial stability of our implementation has been documented for a representative phase-field model on the PFHUB benchmark problems website (<https://pages.nist.gov/pfhub/simulations/7a.0/>). The temporal discretization was chosen conservatively so that, for the set of parameters used, the incident vapor flux and surface mobility effects can be appropriately captured during the film growth process regardless of the deposition rate. The composition field within the simulation domain was initially randomly populated by sampling a truncated Gaussian distribution between -1 and 1 with a standard deviation of 0.35 . The number of time steps used to complete a given simulation varies depending on the deposition rate used to achieve a nominal film thickness of 150 grid points.

2.2. Simulating physical vapor deposition with time-dependent processing conditions

We included time-dependent deposition rates and diffusivities/mobilities by implementing these quantities as arbitrary functional forms in our phase-field model. These functional forms can be represented as a five term Fourier series expansion, as shown below. A deposition protocol $P(t)$ takes the form:

$$P(t) = \sum_{n=1}^5 a_n \cos\left(\frac{2\pi n t}{T_{\text{period}}}\right) + b_n \sin\left(\frac{2\pi n t}{T_{\text{period}}}\right). \quad (9)$$

The desired functional form and its associated parameters are supplied by the user in the MEMPHIS input file, allowing the user to simulate complex, time-varying deposition conditions. For instance, using a step function deposition protocol requires the user to supply the initial value P_i , the final value P_f , and the time at which the value switches from the initial to the final value t_{switch} , as a fraction of the total simulation time. Similarly, for a linear ramp, the parameters supplied are the initial value P_i , the final value P_f , the time at which to start the ramp $t_{\text{ramp}}^{\text{start}}$, and the time at which the ramp ends $t_{\text{ramp}}^{\text{end}}$. For a Fourier series form, the user supplies the Fourier coefficients a_n, b_n and the time period T_{period} . We additionally bounded the Fourier series to values that are physically meaningful and realistic, based on the alloy deposition model. This protocol is applied to the deposition rate (i.e. magnitude of the velocity field of the incident vapor flux) and/or species mobilities.

2.3. Coupling a genetic algorithm to the alloy deposition model

The genetic algorithm begins by creating an initial population of individuals, each individual deposition protocol is defined by a set of parameters that describe the time-dependent deposition condi-

tions such as deposition rate and species mobilities. Each parameter of the functional form is assigned to a gene in an individual, and each gene is chosen from a uniform distribution. The microstructure from each simulation is then analyzed (see subsection below on definition of fitness score), and the objective function (fitness) of each individual is evaluated. The fittest 25% of the individuals are passed on to the next generation (this is the selection step in the genetic algorithm), while the rest of the population is generated by mutating the fittest 25% of the individuals three times each. This strategy is similar to other design approaches [38] that use genetic algorithms, although other mutation/crossover strategies could be used as well. For each generation, each individual and its fitness are recorded. The algorithm is halted either at the end of ten generations, or when there is no improvement in the best individual's fitness. We use the DEAP Python package to implement the genetic algorithm [51].

2.4. Definition of a fitness score for the genetic algorithm

We defined the fitness F of each individual deposition protocol using two terms: an accuracy term and a complexity term such that,

$$F = \sum_{k=0}^n |w_k (\alpha_k^{\text{structure}} - \alpha_k^{\text{target}})| + p \left(\sum_n |P'_{n+1} - P'_n| \right). \quad (10)$$

The first part of the equation above describes the accuracy, while the second describes the complexity.

To obtain terms in the accuracy criterion of the fitness score, we first compute two-point statistics for each microstructure using the PyMKS package [43], and then reduce this statistical description of the microstructure using Principal Component Analysis (PCA), as implemented in the scikit-learn package [44]. The PCA method reduces the two-point statistic into a low-dimensional space, downgrading our microstructure comparison to comparing the PCA scores (coefficients) of the genetic algorithm structure and the target structure. Specifically, the accuracy term is a weighted sum of the absolute error in each PCA coefficient of the microstructure. The coefficients $\alpha_k^{\text{structure}}$ and α_k^{target} are thus the principal component (PC) scores of the microstructure obtained from a deposition protocol suggested by the genetic algorithm and that of the target microstructure respectively. The coefficient w_k are weights for PC score k . Note that we used weights, w_k , of 1 in this work for simplicity.

We created two PCA microstructure models: one that captures LCM, VCM, RCM, and NPCM structures to help rediscover constant deposition structure zone diagrams (see Section 3.1), and the second solely focused on LCM, VCM, RCM, and hierarchical structures to aid the discovery of time-dependent deposition protocols (see Sections 3.3 and 3.5). For the constant deposition structure zone diagram, the PCA model was constructed using an initial dataset of 71 microstructures obtained using a range of alloy compositions and a range of constant deposition rates. With this dataset, we found a PCA model with 8 components to explain $\sim 98\%$ of the variance in our dataset. For the time-dependent deposition protocols, we constructed a PCA model using an initial dataset of 559 microstructures, obtained using a range of time-dependent protocols such as step functions, linear ramps, and various realizations of a Fourier series expansion, representing arbitrary time-dependent protocols. In this case, we find a 15 component PCA model to describe $\sim 88\%$ of the variance in the dataset. This decrease in accuracy is mostly due to the more complex, hierarchical set of microstructures in our dataset.

The second term in our fitness score is a measure of the complexity of the time-dependent deposition protocol. This term favors simple deposition paths over complex ones and is motivated

by the ease of realizing simple deposition protocols in experiments, as opposed to complex protocols. The complexity of a time-dependent protocol is measured by the sum of the absolute values of the first derivative of the functional form $P(t)$, as used in some machine learning approaches [42]. The coefficients P'_n thus describe the first derivatives of the time-varying functional form, evaluated at each timestep.

The overall objective function, or fitness, of an individual deposition protocol, is a combination of the accuracy and complexity of the individual with the parsimony coefficient 'p' meant to control the relative importance between the two terms. The goal of the genetic algorithm is to minimize this objective function, *i.e.*, a lower objective function is a fitter individual.

3. Results

3.1. Rediscovering structure zone diagrams for constant deposition protocols

We first illustrate the capabilities of the genetic algorithm by using it to (re-) discover known structure zone diagrams for constant deposition conditions [25]. Specifically, we aim at discovering the alloy composition and the deposition rate (constant) needed to achieve specific concentration modulations, namely: LCM, VCM, RCM, and NPCM. Prior work has shown that the origin of various microstructure classes in these systems is a trade-off between deposition of fresh atoms on the surface of the thin film and the phase separation occurring on the surface/bulk of the thin film via spinodal decomposition/ diffusion [28,14]. Low deposition rates allow for a longer time for diffusion to occur on the surface of the film, leading to complete phase separation laterally. This leads to columns of A and B phases forming, which then act as templates for subsequent layers to phase separate, resulting in an LCM microstructure. On the other hand, higher deposition rates do not allow for complete phase separation to occur at the surface. When subsequent layers of materials are now deposited on the surface of the film, phase separation can occur both laterally and vertically, resulting initially in checkerboard microstructures that later coarsen and evolve into a VCM microstructure [25]. Higher deposition rates further restrict surface diffusion, and the microstructure formation is thus dominated by bulk spinodal decomposition, resulting in a RCM microstructure. For non-equiautomic alloy compositions, nanoprecipitate microstructures are formed, with precipitates made of the minority phase.

The genetic algorithm optimizes a population of 40 deposition protocols per generation, with each individual deposition protocol defined via two parameters: a constant deposition rate and an alloy composition. For the initial population, the deposition rate in our simulations (dimensionless) is drawn from a uniform distribution between 0.05 and 1.25. This range corresponds to deposition rates employed in experimental efforts to realize the targeted concentration modulations in thin films [15] and has shown previously to yield similar simulated microstructures for equivalent deposition conditions [28]. The composition, represented as a fraction between 0 and 1, represents the atomic percentage of element B in the model A-B alloy. For each generation, the fittest 25% of the individuals within our population of deposition protocols are retained in the next generation and the remaining 75% of the population is generated by mutating each of the fittest 25% individuals thrice. Each mutated individual deposition protocol is generated by perturbing each parameter by up to 30%. While the genetic algorithm explores the design space to search for parameter combinations that yield specific microstructures, we can classify each microstructure obtained during the optimization into an LCM, VCM, RCM, NPCM, or a hybrid structure. This

classification is performed manually at this point but could be performed with a more robust model such as a convolutional neural network.

We plot the structure zone diagram discovered by the genetic algorithm in Fig. 2. This figure shows the combination of alloy composition and deposition parameters that result in specific microstructures. The deposition/diffusion trade-off is measured as the ratio between a characteristic diffusion time scale and a characteristic deposition time scale, similar to the metric used by Lu and co-worker [25]. Our results clearly demonstrate that the genetic algorithm can yield deposition protocols that distinguish between LCM, VCM, RCM, and NPCM regions, learning that low deposition rates (high characteristic deposition time) result in LCM structures, while high deposition rates (low characteristic deposition time) result in VCM or RCM structures. Equiautomic compositions (50/50) at low deposition rates yield LCM structures. As we get away from the equimolar phase fraction center line along the alloy decomposition ordinate, the genetic algorithm explores non-equiautomic NPCM microstructure regions, as well as more clearly defined hybrid microstructure regions. For instance, our simulations yield LCM structures with some fraction of VCM characteristics (light blue points) at approximately 65/35 concentrations. Beyond this concentration, we observe nanoprecipitates in the microstructures, either in conjunction with LCM characteristics (purple) or with some VCM characteristics (olive), up to 70/30 concentrations. For higher concentrations, the microstructures are predominantly NPCM microstructures (orange/red). Even within the NPCM domains, we find a variation in precipitate sizes and shapes as deposition conditions vary. This variation is not captured in our labelling, nor is it highlighted in this structure zone diagram, but it could be explored in further details for applications that desire such microstructure characteristics. We hypothesize that microstructures with LCM features are formed over a wider compositional region than VCM or RCM, because LCM features are formed under low deposition, high diffusion regimes which enables the templating of LCM structure prior

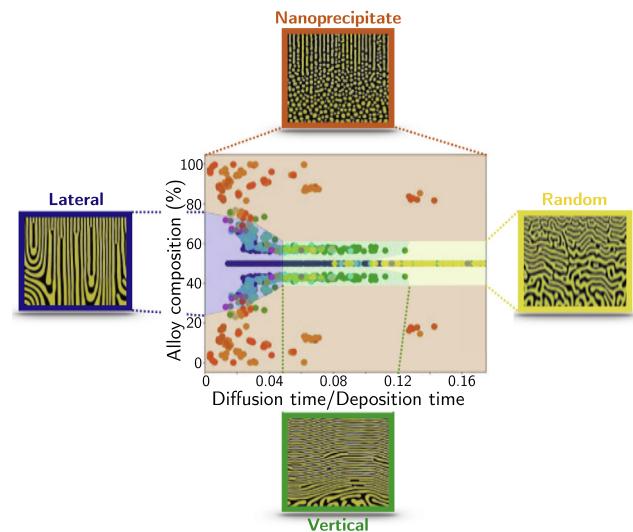


Fig. 2. Structure zone diagram discovered by the genetic algorithm for constant deposition conditions. The deposition/diffusion trade-off fundamental to microstructure formation is defined as a ratio of a characteristic diffusion time and a characteristic deposition time, similar to the metric used in Lu et al. [25]. The genetic algorithm is able to learn known process-structure correlations, such as low deposition rates yielding lateral concentration modulation (LCM) structures, and high deposition rates leading to vertical (VCM) or random concentration (RCM) modulation structures, and non-equiautomic alloy compositions lead to nanoprecipitate concentration modulations (NPCM). Also shown are representative snapshots of LCM, VCM, RCM, and NPCM structures.

the deposition the next atomic layer. In contrast, microstructures with predominantly VCM features are formed under a narrow region of intermediate deposition conditions, beyond which we transition to a RCM microstructure. Overall, this structure zone diagram learnt by the genetic algorithm is qualitatively similar to previous computational and experimental studies [25,28,29], while adding nuances by providing additional details on microstructures formed at non-equatomic compositions.

3.2. Microstructure achieved using simple time-dependent protocols

We now illustrate the type of microstructures achievable using simple, time-dependent deposition rates for equatomic phase fractions. The time-varying protocols tested here are a simple step function and a linear ramp (see Methods for the mathematical description). These protocols not only serve as simple illustrations of the nature of microstructures achievable using time-varying protocols, but also as illustrations of the improvements to our previous alloy-deposition model [28]. The resulting microstructures help further understand the role of competing deposition and diffusion processes in forming the resultant microstructure and will serve as a guide to understand the microstructures generated using the time-dependent protocols explored by the genetic algorithm. As mentioned above, low deposition rates result in LCM microstructures, due to high surface diffusion rates (relative to the deposition rate) which cause fast local spinodal decomposition of the deposited species, see Fig. 3 (top panel). High deposition rates, on the other hand, resulted in VCM microstructures, due to incomplete spinodal decomposition as well as the thermodynamic favorability of forming A-B interfaces beyond a critical film thickness [52], see Fig. 3 (top panel). A step function deposition protocol, switching from a low to a high deposition rate, results in a hierarchical microstructure that consists of both LCM and VCM, appearing in order of the deposition, see Fig. 3 (bottom panel). An inverted deposition protocol results in the concentration modulation being inverted, with LCM domains at the top of the film while the bottom part transforms to a VCM. Longer times at high deposition rates will result in the local transformation of the LCM structure to a VCM structure, again due to the thermodynamic favorability of forming A-B interfaces. The rationale of step function deposition protocols is to trigger specific patterning mechanisms at specific times during the deposition process. These mechanisms include surface-diffusion dominated microstructure formation at low deposition rate or bulk diffusion and phase sepa-

ration dominated mechanisms at higher deposition rates. As shown in Fig. 3 (bottom panel), the result is a multimodal, hierarchical microstructure for which the transition from one concentration modulation to the other corresponds to the change of deposition conditions during the thin film growth. Alternatively, a linear ramp in the deposition rate (low to high) results in a mixed microstructure. The bottom part of the films shows LCM characteristics, as expected for an initially low deposition rate. As the deposition rate linearly increases with time, we observe the central portions of the film beginning to transform to a VCM, while the top layer of the film shows a VCM microstructure.

3.3. Achieving targeted concentration modulations using the genetic algorithm

The previous section illustrated that simple time-dependent deposition protocols lead to complex microstructures beyond the known, monomodal microstructure classes achieved by constant deposition protocols. We now devise time-dependent depositions that achieve precise thin-film microstructures for a given film thickness. In this section, we explore the ability of the genetic algorithm to find deposition protocols that result not only in monomodal microstructures such as LCM and VCM microstructures, but also in more complex, hierarchical concentration modulation (HCM) structures. Our implementation allows deposition protocols to be specified as Fourier series expansion of 5 terms, permitting the exploration of arbitrary time-dependent protocols. Details on the Fourier series are provided in Methods. We are particularly interested in exploring pulsed deposition protocols as they can be easily set up in an actual deposition chamber. Pulsed deposition protocols are completely characterized by three parameters: amplitude, frequency, and width of the pulse. This pulse function can be realized with a Fourier series where all the sine terms have a coefficient of zero, while the cosine terms have a coefficient of $\frac{2A}{\pi} \sin\left(\frac{n\pi T_p}{T_{\text{period}}}\right)$, where A is the amplitude, T_p is the width of the pulse, and T_{period} is the time period of the pulse. An individual deposition protocol is uniquely defined by these three coefficients. For each generation, we use a population size of 40, with each individual consisting of three parameters: A , T_{period} , and duty cycle (ratio of pulse width to time period: T_p/T_{period}). For the initial population (generation 0), the amplitude, A , is drawn from a uniform distribution between 0.05 and 1.00. Note that this amplitude range corresponds to typical deposition rates employed in experimental efforts to realize simple concentration modulations in thin films [15]. The duty cycle is also drawn from a uniform distribution between 0.10 and 0.90. The time period is drawn from a uniform distribution between 1 and 100, and multiplied by a 1000, to realize time periods between 1000 and 1000000 steps. This is performed to have the genetic algorithm choose from time periods that are sufficiently distinct and avoid generating individuals that are too similar. The mutation protocol is the same as mentioned in the section above, where the fittest 25% of the individuals are retained in the next generation and the remainder of the population is generated by mutating each of the fittest 25% individuals thrice. Each mutated individual is generated by perturbing separately the amplitude, time period, and duty cycle by up to 30%.

Fig. 4 shows the ability of the genetic algorithm to generate LCM, VCM, and HCM via time-dependent deposition protocols. In Fig. 4(a) we show the target LCM microstructure and the microstructure generated from the best time-dependent deposition protocol identified by the genetic algorithm. We see that the genetic algorithm is able to define deposition protocols that achieve the target concentration modulation relatively well. The bottom panel of Fig. 4(a) shows the pulsed deposition protocol describing the fittest individual. This pulsed protocol is compared

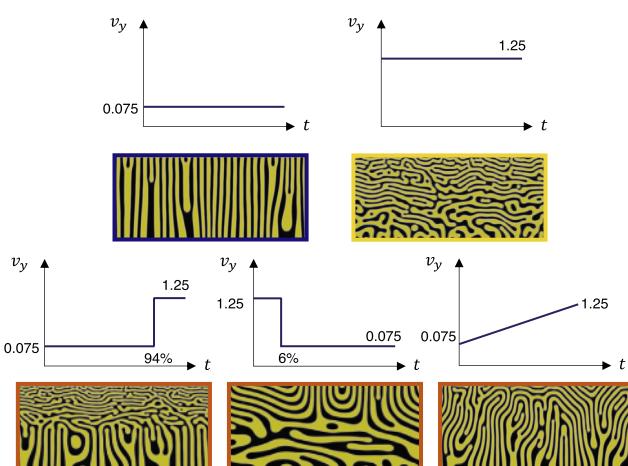


Fig. 3. The effect of time-dependent deposition conditions. (top panel) Microstructures obtained using constant deposition rates (bottom panel) Microstructures obtained from step functions and a linear ramp.

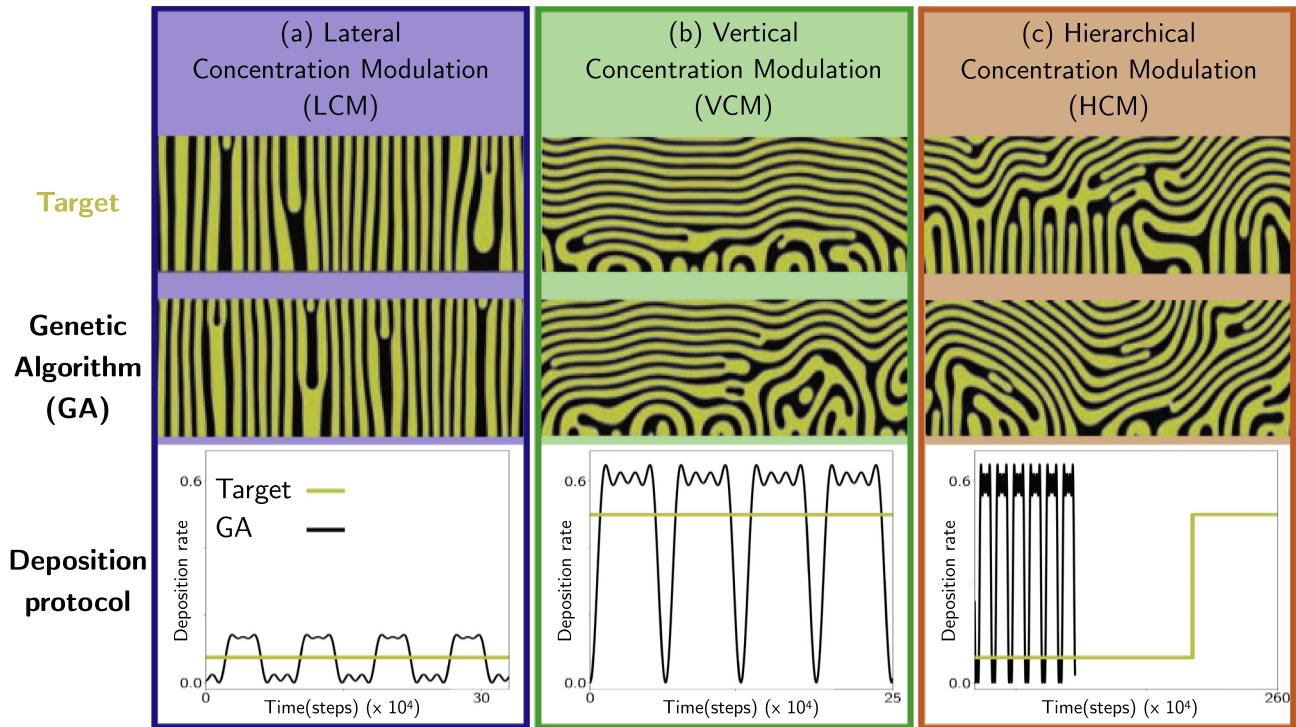


Fig. 4. Discovering time-dependent deposition protocols. (Top panel) Target LCM, VCM, and hierarchical microstructures (a), (b), and (c) respectively. (Middle panel) Genetic algorithm obtained structures for each of the microstructure classes. (Bottom panel) Comparing the genetic algorithm discovered protocols to the protocols used to generate the target microstructure. Duty cycles corresponding to pulses in panels (a), (b), and (c) are $\sim 60\%$, $\sim 31\%$, and $\sim 49\%$ respectively. Note that duty cycle calculations are approximate because a pulse is approximated by a Fourier series expansion, resulting in on-off transitions that are not sharp.

to the constant deposition protocol that was initially used to generate the target microstructure. We observe that the genetic algorithm adopts a pulsed deposition protocol whose amplitude is in roughly the same range as the constant deposition rate condition. Note that the oscillations in the pulse occur because we approximate a pulse function using a Fourier series with limited terms. Fig. 4(a) also shows that a LCM structure can be achieved in multiple ways (constant deposition and pulse deposition), as long as surface diffusion can take place much faster than the rate at which new species are deposited on the growing film. We hypothesize that this pulsed deposition protocol finds a way of combining short bursts of deposition with periods of surface diffusion to achieve the same effect as a constant low deposition rate. However, there is a limit to which we can combine the 'on' and 'off' cycles of the pulse while still achieving the same microstructure. As we will see later, the genetic algorithm finds low amplitude pulses to be closer to the target microstructure overall, while high deposition rates result in either VCM structures or structures transitioning from LCM to VCM, reflecting well-known correlations [28], see Figs. 5(a) and 5(b). The genetic algorithm is also able to achieve a VCM microstructure with high accuracy, as illustrated in Fig. 4(b). The functional form chosen by the genetic algorithm is a pulse with an amplitude close to the constant deposition rate condition yielding the targeted VCM microstructure.

Having seen the ability of the genetic algorithm to find pathways to generate LCM and VCM structures, we next turn our attention to hierarchical microstructures. Fig. 4(c) shows a target hierarchical microstructure generated using a step function deposition path. The target microstructure shows LCM near the bottom of the film due to low deposition rates at the start of the deposition process, followed by a transition to a VCM near the top of the film, with the central region transforming from a LCM to a VCM structure. Yet again, we see that the algorithm finds a deposition protocol to achieve a microstructure with the same characteristics as the

target, although with a smaller fraction of LCM domains. Comparing the processing protocols discovered by the genetic algorithm to the protocol used to generate the target microstructure, we find that the amplitude of the pulse is again higher than the value used during the high deposition rate step of the protocol used (unoptimized and chosen at random) to generate the targeted HCM microstructure, however the total deposition time is much faster.

3.4. Interpreting the decisions made by the genetic algorithm

To understand how the genetic algorithm chose these specific time-dependent protocols, we can look at the design space explored by the algorithm in Fig. 5. In Fig. 5(a), we show the evolution of the pulse amplitude as the number of generations evolves, highlighting the choices of selections and mutations made by the genetic algorithm. The average pulse amplitude in the population at a given generation is denoted by a symbol, and the error bars show the minimum and maximum amplitudes present in each generation. The initial population (generation 0) starts with a wide range of amplitudes, reflecting our choice of a uniform distribution to initialize the genetic algorithm. This uniform distribution can also be considered as an 'uninformed' prior distribution. For a LCM target microstructure, we see that the average pulse amplitude drops a little from generation 0 to the first generation, followed by a rapid drop in the third generation. After that, the average pulse amplitude roughly stays the same for the rest of the optimization, indicating that the genetic algorithm found low amplitude pulses to be suitable to generate LCM structures. Interestingly, we note a sharp decrease in the range of amplitudes retained in each generation after the third generation (generation 3), further indicating that the genetic algorithm not only found low amplitude pulses to be suitable, but also that high amplitude pulses are unsuitable to generate LCM structures. In contrast, when the target is a VCM structure, the genetic algorithm learns that a

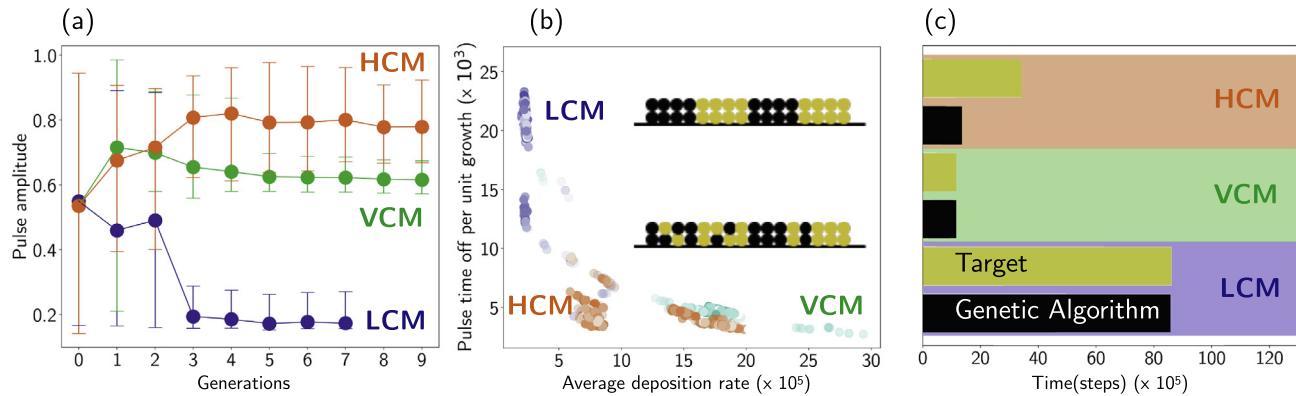


Fig. 5. Evolution of protocol characteristics through generations. (a) Evolution of the amplitude of a pulsed deposition rate protocol while discovering protocols for various microstructures. The genetic algorithm discovers that low amplitudes result in LCM structures, while high amplitudes result in VCM structures. (b) Deposition versus diffusion explored by the genetic algorithm. The algorithm learns that a low deposition rate and a higher time allowed for diffusion results in LCM structures, while high deposition rates result in VCM structures. Hierarchical structures fall in the middle depending on the ratio of LCM and VCM domains in the specific structure. (c) Time taken by the genetic algorithm based protocols to generate the target microstructures. The protocol discovered by the genetic algorithm is more efficient for a hierarchical microstructure, while discovering protocols that take similar times to the constant deposition protocols used to generate the target microstructures.

narrow amplitude range of $\sim 0.55 - 0.65$ is suitable to generate VCM structures as illustrated by the small error bars. While the first generation (generation 1) overshoots the average pulse amplitude, later generations rapidly converge toward the narrow amplitude range corresponding to the VCM regime. Lastly, when the target is a hierarchical structure (HCM), we see that the genetic algorithm favors higher pulse amplitudes on average. However, the large error bars indicate that the algorithm finds individuals with low amplitudes to be also valid and viable deposition candidates to generate HCM structures. This ambiguity is probably due to the composite nature of the HCM structure, requiring periods of low deposition and high deposition, which could be achieved in multiple ways by optimizing amplitude, duty cycle, and time period simultaneously.

Fig. 5(b) shows the deposition/diffusion trade-offs explored by the genetic algorithm by plotting a measure of the time allowed for deposition versus the time allowed for diffusion for each protocol explored during the genetic algorithm's search. The average deposition rate for each protocol is defined as the ratio of the height of the film grown to the time taken for the film to grow to that height, effectively describing the rate at which the film grows. This quantity describes the time spent depositing atoms on the film surface and increases with increasing pulse amplitude and frequency. The quantity 'pulse time off per unit growth' is calculated as the number of steps for which the pulse is 'off' (in the low state) during the deposition, normalized by the film thickness. This effectively describes the time allowed for diffusion at the film surface and decreases as the pulse frequency increases. In Fig. 5(b), points are colored by the target microstructure and shaded by the fitness score, with darker points indicating better protocols to achieve the target microstructure denoted by that color. We see that for a LCM target, the fittest individuals/ best protocols (darkest colors) are deposition protocols for which the pulse is in the 'off' state for longer time (per unit film grown) and has lower deposition rates. This trend again shows that LCM structures are obtained when the deposition protocol is such that it allows enough time for diffusion to occur at the film surface before the next layer of atoms is deposited at the film surface, as seen in Fig. 5(a), as well as in Fig. 2, and prior work [25,28] for constant deposition conditions. Similarly, a VCM target is obtained best with deposition protocols that have high deposition rates and lesser time allowed for diffusion. Hierarchical structures are obtained with a range of deposition rates and diffusion times, likely due to each specific hierarchical structure being composed of different proportions of LCM and VCM domains.

The deposition protocols discovered by the genetic algorithm are also reasonably efficient, as measured by the time taken to achieve a given film thickness, see Fig. 5(c). We see that films with LCM microstructures require a long time to achieve a certain thickness, especially when compared to films with VCM microstructures. This observation is expected and can also be seen in the average deposition rate plotted along the x-axis of Fig. 5(b). We also see that time taken to achieve the target microstructures via the genetic-algorithm-based pulse protocol is similar to the time taken to generate LCM and VCM microstructures using constant deposition protocols. Interestingly, the genetic-algorithm-discovered pulse protocol is $\sim 60\%$ faster in achieving a hierarchical microstructure, highlighting the need to explore various time-dependent processing protocols to discover efficient deposition paths. This is especially important because of the unique combination of mechanical properties displayed by hierarchical microstructures [14,53]. Overall, we find the pulsed deposition protocols discovered by the genetic algorithm to be accurate in achieving LCM, VCM, and HCM microstructures, as well as equally efficient or more efficient than the constant/step deposition protocols used to generate the target microstructures. These results are commensurate with prior work on the fabrication of nanometallic multilayers with different layer thicknesses that have been deposited by controlling the on-time of different sources [54,55] using shutters in the deposition chamber.

3.5. Time-dependent multi-parameter deposition protocols: Varying the deposition rate and the mobility concurrently

The previous subsection showed how we can target various microstructure classes by solely varying the deposition rate as a function of time. We now discuss the extension of this workflow when multiple, critical deposition parameters are varied as a function of time during the deposition process. During physical vapor deposition, critical deposition parameters include not only the deposition rate, but also the substrate temperature, where the substrate surface temperature directly affects the diffusive properties of the atomic species being deposited. We are thus focusing here not only on controlling the deposition rate but also the temperature of the substrate on which a film is grown. In our PVD model, we control the temperature of the substrate indirectly by controlling the bulk and surface mobilities of both species, where the mobility is proportional to the diffusion coefficient. Higher temperatures result in higher atomic mobilities. We assume here that a change of the substrate temperature instantaneously results in a

change in atomic mobilities. This assumption is reasonable if the film being deposited is thin enough ($\sim 100\text{--}250$ nm) so that heat transfer from the substrate to the surface of the film is instantaneous. Such effects of deposition temperature on surface diffusion distance are shown in [56], where increased temperatures result in microstructural features such as hillocks with spacings corresponding to diffusion distances at that temperature. We assume that temperature (and therefore mobility) changes that can be controlled during the deposition follow a step function. In other words, the mobility either increases or decreases sharply once at some time during the deposition, corresponding to the change of the substrate temperature. This simple choice is made to mimic temperature controller limitations while operating a PVD chamber, although our workflow allows for arbitrarily complex functional forms. The choice of step functions corresponds to two distinct periods of constant temperature during deposition, a protocol that is relatively easy to implement and realize in experimental PVD experiments.

The genetic algorithm now must optimize 13 parameters, three parameters describing the pulsed deposition rate as before, as well as 10 other parameters that describe the variation of species mobilities with time, see Methods for a full list of the 13 quantities optimized by the genetic algorithm. To ensure consistent variations in surface and bulk mobilities of both species, we randomly select an increase or decrease in temperature, following which we randomly select initial and final values for each mobility. We show in Fig. 6 protocols in which both the deposition rate and mobilities change as a function of time. These protocols result in (a) LCM, (b) VCM, and (c) HCM microstructures. The sections above showed that favoring diffusion over deposition results in LCM structures. Here, when considering multiple deposition parameters varying in time, we see that pulsed deposition rates with low amplitude, in conjunction with high mobilities result in LCM structures as illustrated in Fig. 6(a). In contrast, low mobilities in conjunction with high amplitude pulsed deposition rates result in

VCM structures, as this combination favors deposition over diffusion, see Fig. 6(b). The time at which the mobility is triggered to switch from low-to-high or high-to-low in panels (a) and (b) corresponds to deposition regimes which favor further surface diffusion in the case of panel (a) and vertical (bulk) diffusion in panel (b). In addition, as noted in panel (b) the genetic algorithm discovers a regime in terms of mobility that nearly suppresses surface diffusion in the later time of the deposition protocol. For a target HCM structure, we observe that we are unable to simultaneously optimize deposition rate and mobilities to obtain a structure close to the target microstructure. This is due to the nature of the protocols explored by the genetic algorithm during the optimization, many of which result in microstructures too close to each other in the low-dimensional representation of the microstructure. The low-dimensional representations of the microstructures cannot be easily distinguished by the genetic algorithm for the optimization but are clearly different when visually inspected. This points to a deficiency in our construction of the accuracy criterion in our fitness score, which in turns points to a weakness in our low dimensional representation of microstructures. Future work is warranted to consider other forms of compact, non-linear representations of microstructures such as autoencoders or variational autoencoders (VAE) for instance. Autoencoders consist of encoder and decoder models (typically neural networks), where the encoder learns a non-linear embedding from microstructural space to a low dimensional latent space and the decoder learns to reconstruct the microstructure back from the low-dimensional latent space such that the microstructure reconstruction error is minimized. In a VAE, the encoder organizes the latent space as a Gaussian distribution based on input variables. Using an autoencoder or a VAE in our approach could help learn non-linear embeddings of our microstructures such that the different microstructures explored here are sufficiently separated in the low-dimensional space, allowing the genetic algorithm to easily optimize processing protocols for a specific microstructure.

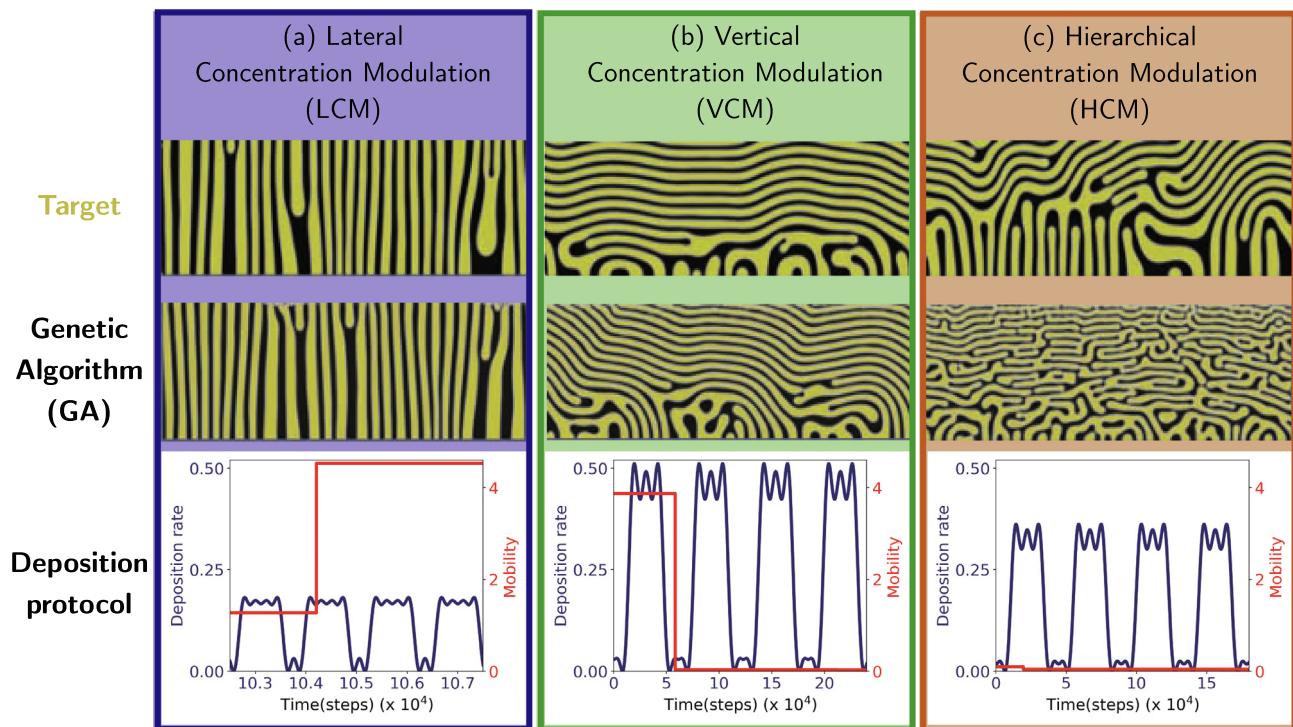


Fig. 6. Discovering time-dependent multi-parameter deposition protocols. Time varying deposition rate and mobilities that result in (a) LCM, (b) VCM, and (c) HCM microstructures.

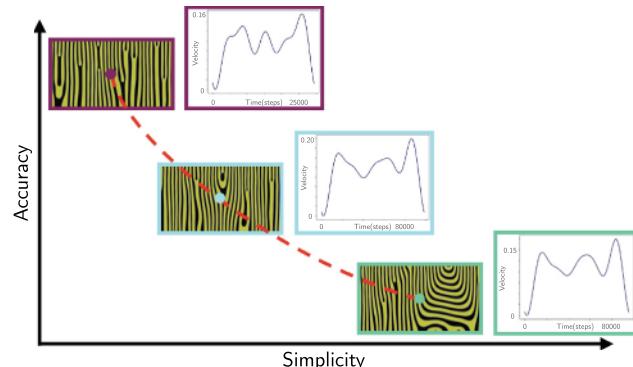


Fig. 7. Balancing simplicity and accuracy of the deposition protocols. A lateral concentration modulation can be achieved by multiple protocols of varying complexity. The axes are schematics of accuracy and simplicity, with a visual inspection clearly showing that the protocol on the top left is more complex than the protocol on the bottom right.

3.6. Balancing accuracy and complexity of deposition protocols

The complexity term in the objective function was introduced to favor simple protocols over complex ones. This bias helps our results to be implemented and reproducible in experiments, where simpler deposition protocols are easier to realize than complex ones. The parsimony coefficient ' p ' balances our ability to reproduce the target microstructure with our bias of favoring simple protocols. Fig. 7 illustrates the effect of changing the parsimony coefficient on the obtained deposition protocols for the deposition rate. A low value for the parsimony coefficient will favor a high accuracy in terms of achieving targeted microstructures at the expense of generating complex protocols. Conversely, a high value for the parsimony coefficient will favor simple protocols but with degrading performance in terms of achieving precise targeted microstructures. We observe in Fig. 7 that an LCM microstructure can be accurately obtained with a deposition protocol of high complexity (top left), while a deposition protocol of lower complexity (bottom right) generates microstructures that are not as accurate. Complexity here can be seen by frequent changes in the protocol (line plot), quantified by the curvature-inspired metric we use, see Methods for more details. The deposition protocol illustrated in the middle of Fig. 7 is neither as complex as the protocol on the top left, nor is it as simple as the protocol on the bottom right. The resulting microstructure from this deposition protocol is almost as accurate as the microstructure on the top left, thus indicating the possibility of a pareto front of optimal protocols that trade accuracy for complexity. This potential Pareto front is sketched schematically with the dashed line. Note that for this demonstration, we attempted to realize the target microstructure using a protocol represented as an arbitrary Fourier series, again showing that multiple protocols can achieve the same microstructure class.

4. Discussion

We have demonstrated above that by coupling a genetic algorithm to a model for the physical vapor deposition of alloys, we can discover and devise novel, time-dependent deposition protocols that achieve targeted thin-film microstructures. We find that the genetic algorithm can discover pulsed deposition protocols that accurately grow known monomodal microstructures such as LCM and VCM microstructures. In fact, the genetic algorithm can rediscover known structure zone diagrams created for constant deposition conditions, while simultaneously providing additional

details for non-equiautomic alloy compositions. However for more complex, hierarchical microstructures, we observed that the achieved microstructures are different from the target, especially for the case of simultaneous optimization of deposition rate and mobilities. This deficiency in accuracy is partially due to our approach in the fitness score to compare microstructures using statistical functions (two-point statistics) and PCA. This scheme struggles to capture complex hierarchical structures, reducing them to the same subspace in reduced dimensions. One way to address this issue would be to consider other techniques to provide a compact description of the microstructure. For instance, one could use pre-trained convolutional neural network (e.g. ResNet-50 v2 [57]) or VAE [58,59] to obtain a faithful, low-dimensional representation of the microstructure. These models have demonstrated that they are able to discern complex shapes and motifs with precision, making them an attractive alternative to the approach used in our current workflow. Another option would be to use other linear dimensionality reductions such as a Karhunen Loeve expansion [60–62], or other non-linear methods such as Kernel PCA [63] to capture microstructures accurately.

Several experimental studies have demonstrated the scientific merit of achieving hierarchical structures [14,15,64] in order to gain emerging functionalities. However, as of now, such structures stem from serendipitous discoveries rather than controlled and predictable fabrication processes. Time-dependent deposition protocols currently rely on sequential use of masking shutter stages to achieve hierarchical microstructures. The present approach offers an alternative to current practices and provide avenues for thin film, physical vapor deposition, and film functionality communities to design tailor-made thin films for specific applications. Our approach could also be viewed as a general optimization approach, similar to other approaches such as active learning, reinforcement learning, or Bayesian optimization in materials design. In this parlance, our approach would be considered 'online learning' since the genetic algorithm launches phase-field simulations in each generation during the optimization. However, the current workflow could also be adapted to an 'offline learning', computationally efficient approach, where a surrogate model such as a variational autoencoder or a polynomial chaos expansion would predict the thin film microstructure [45]. For instance, one could use surrogate models to predict the thin film microstructure under varying deposition conditions and relate changes in deposition rate and mobilities to changes in background gas pressure, sputtering power, working distance, and temperature for instance.

5. Conclusion

In this work, we presented an approach that couples a genetic algorithm to a phase-field alloy deposition model to discover time-dependent deposition protocols to design thin films. The genetic algorithm optimizes a population of deposition protocols where each individual protocol is uniquely defined as a collection of parameters that collectively define a time-dependent protocol. Each time-dependent deposition protocol is evaluated in terms of the accuracy to achieve a targeted microstructure and in terms of the complexity of the protocol. These two criteria are used to select certain deposition protocols. Major take aways are the following:

- Our genetic-algorithm-guided protocols achieve targeted microstructures with known, lateral and vertical concentration modulations, as well as more complex, hierarchical microstructures previously not described in classical structure zone diagrams. Using this approach we can re-create known structure zone diagrams for constant deposition conditions, while adding additional information about non-equiautomic alloy composi-

tions. More importantly, we can also discover novel, time-dependent protocols with time-varying deposition rates and mobilities forming well-known microstructures experimentally observed in these films or new microstructures. The pulsed deposition protocols discovered by the genetic algorithm achieve targeted LCM, VCM, and HCM microstructures with similar efficiency and accuracy than those microstructures obtained with constant/step deposition protocols.

- The genetic algorithm discovers these protocols by learning the trade-off between deposition and diffusion during film growth, *i.e.*, the fundamental physical process governing microstructure formation.
- By building a fitness score that balances accuracy and simplicity of the deposition protocol, we show that one can achieve optimal protocols that trade accuracy for complexity.

To relate our discovered protocols with variables controllable in a deposition chamber such as background gas pressure, sputtering power, or substrate temperature, the methods in this study could be combined other deposition prediction tools such as SIMTRA [49] to relate deposition rate and mobilities to operating conditions of a PVD chamber.

Although the approach presented in this study could be improved both to enhance accuracy and efficiency, this work puts forward novel promising pathways to devise deposition routes to tailor thin-film properties. By elucidating the process-structure mechanisms during physical vapor deposition, and by using this knowledge to achieve precise thin-film microstructures, the present approach offers a path forward to transition from current manufacturing methods to new paradigms for fabricating novel, application-specific thin films.

Data availability

Some of the Python scripts, Jupyter notebooks, and data used to plot figures are provided at <https://github.com/saakethdesai/GA-PVD>. Other raw/processed data, and greater access to the phase field code, are available upon request.

6. CRediT authorship contribution statement

Saaketh Desai: Conceptualization, Methodology, Software, Investigation, Validation, Formal analysis, Data curation, Writing – original draft. **Rémi Dingreville:** Conceptualization, Methodology, Investigation, Writing – original draft, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgments

SD and RD acknowledge funding under the Beyond Fingerprinting Sandia Grand Challenge Laboratory Directed Research and Development (GC LDRD) program. The phase-field framework is supported by the Center for Integrated Nanotechnologies (CINT), an Office of Science user facility operated for the U.S. Department of Energy. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy National Nuclear Security Administration under contract DE-

NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

- [1] Y. Fu, H. Du, W. Huang, S. Zhang, M. Hu, TiNi-based thin films in MEMS applications: a review, *Sens. Actuators, A* 112 (2–3) (2004) 395–408, <https://doi.org/10.1016/j.sna.2004.02.019>.
- [2] Y. Cui, Y. Ke, C. Liu, Z. Chen, N. Wang, L. Zhang, Y. Zhou, S. Wang, Y. Gao, Y. Long, Thermochromic VO₂ for energy-efficient smart windows, *Joule* 2 (9) (2018) 1707–1746, <https://doi.org/10.1016/j.joule.2018.06.018>.
- [3] Gr. Schwartz, B.C.-K. Tee, J. Mei, A.L. Appleton, D.H. Kim, H. Wang, Z. Bao, Flexible polymer transistors with high pressure sensitivity for application in electronic skin and health monitoring, *Nat. Commun.* 4 (1) (2013) 1–8, <https://doi.org/10.1038/ncomms2832>.
- [4] F. Liu, Q. Zeng, J. Li, X. Hao, A. Ho-Baillie, J. Tang, M.A. Green, Emerging inorganic compound thin film photovoltaic materials: progress, challenges and strategies, *Mater. Today* 41 (2020) 120–142, <https://doi.org/10.1016/j.mattod.2020.09.002>.
- [5] M.A. Green, A. Ho-Baillie, H.J. Snaith, The emergence of perovskite solar cells, *Nat. Photonics* 8 (7) (2014) 506–514, <https://doi.org/10.1038/nphoton.2014.134>.
- [6] G. Li, R. Zhu, Y. Yang, Polymer solar cells, *Nat. Photonics* 6 (3) (2012) 153–161, <https://doi.org/10.1038/nphoton.2012.11>.
- [7] J. Ouyang, C.-W. Chu, C.R. Szmanda, L. Ma, Y. Yang, Programmable polymer thin film and non-volatile memory device, *Nat. Mater.* 3 (12) (2004) 918–922, <https://doi.org/10.1038/nmat1269>.
- [8] N.J. Dudney, Solid-state thin-film rechargeable batteries, *Mater. Sci. Eng. B* 116 (3) (2005) 245–249, <https://doi.org/10.1016/j.mseb.2004.05.045>.
- [9] A. Reyes Jimenez, R. Klöpsch, R. Wagner, U.C. Rodehorst, M. Kolek, R. Nolle, M. Winter, T. Placke, A step toward high-energy silicon-based thin film lithium ion batteries, *ACS Nano* 11 (5) (2017) 4731–4744, <https://doi.org/10.1021/acsnano.7b00922>.
- [10] J. Rivnay, L.H. Jimison, J.E. Northrup, M.F. Toney, R. Noriega, S. Lu, T.J. Marks, A. Facchetti, A. Salleo, Large modulation of carrier transport by grain-boundary molecular packing and microstructure in organic thin films, *Nat. Mater.* 8 (12) (2009) 952–958, <https://doi.org/10.1038/nmat2570>.
- [11] S. Chen, A. Walsh, J.-H. Yang, X.-G. Gong, L. Sun, P.-X. Yang, J.-H. Chu, S.-H. Wei, Compositional dependence of structural and electronic properties of Cu₂ZnSn (S, Se)₄ alloys for thin film solar cells, *Phys. Rev. B* 83 (12) (2011) 125201, <https://doi.org/10.1103/PhysRevB.83.125201>.
- [12] A. Sharekno, M.F. Toney, Relationships between lead halide perovskite thin-film fabrication, morphology, and performance in solar cells, *J. Am. Chem. Soc.* 138 (2) (2016) 463–470, <https://doi.org/10.1021/jacs.5b10723>.
- [13] B.K. Derby, T.G. Holesinger, J.A. Valdez, B.P. Überuaga, C.R. Kreller, Microstructural analysis of novel Gd₂Ti₂O₇ thin films processed via sputter deposition, *Mater. Des.* 199 (2021) 109430, <https://doi.org/10.1016/j.matdes.2020.109430>.
- [14] B. Derby, Y. Cui, J.K. Baldwin, R. Arroyave, M.J. Demkowicz, A. Misra, Processing of novel pseudomorphic Cu–Mo hierarchies in thin films, *Mater. Res. Lett.* 7 (1) (2019) 1–11, <https://doi.org/10.1080/21663831.2018.1546237>.
- [15] M. Powers, B. Derby, A. Shaw, E. Raeker, A. Misra, Microstructural characterization of phase-separated co-deposited Cu–Ta immiscible alloy thin films, *J. Mater. Res.* 35 (12) (2020) 1531–1542, <https://doi.org/10.1557/jmr.2020.100>.
- [16] K.-H. Müller, Stress and microstructure of sputter-deposited thin films: Molecular dynamics investigations, *J. Appl. Phys.* 62 (5) (1987) 1796–1799, <https://doi.org/10.1063/1.339559>.
- [17] P.B. Barina, M. Adamik, Fundamental structure forming phenomena of polycrystalline films and the structure zone models, *Thin Solid Films* 317 (1–2) (1998) 27–33, [https://doi.org/10.1016/S0040-6090\(97\)00503-8](https://doi.org/10.1016/S0040-6090(97)00503-8).
- [18] A. Anders, A structure zone diagram including plasma-based deposition and ion etching, *Thin Solid Films* 518 (15) (2010) 4087–4090, <https://doi.org/10.1016/j.tsf.2009.10.145>.
- [19] H. Stein, D. Naujoks, D. Grochla, C. Khare, R. Gutkowski, S. Grützke, W. Schuhmann, A. Ludwig, A structure zone diagram obtained by simultaneous deposition on a novel step heater: A case study for Cu₂O thin films, *Phys. Status Solidi A* 212 (12) (2015) 2798–2804, <https://doi.org/10.1002/pssa.201532384>.
- [20] M.J. Powell, D.B. Potter, R.L. Wilson, J.A. Darr, I.P. Parkin, C.J. Carmalt, Scaling aerosol assisted chemical vapour deposition: Exploring the relationship between growth rate and film properties, *Mater. Des.* 129 (2017) 116–124, <https://doi.org/10.1016/j.matdes.2017.05.017>.
- [21] B. Bouaouina, C. Mastail, A. Besnard, R. Mareus, F. Nita, A. Michel, G. Abadias, Nanocolumnar TiN thin film growth by oblique angle sputter-deposition: Experiments vs. simulations, *Mater. Des.* 160 (2018) 338–349, <https://doi.org/10.1016/j.matdes.2018.09.023>.
- [22] Z. Chen, Y. Zhuo, R. Hu, W. Tu, Y. Pei, B. Fan, C. Wang, G. Wang, Control of morphology and orientation for textured nanocrystalline indium oxide thin film: A growth zone diagram, *Mater. Des.* 131 (2017) 410–418, <https://doi.org/10.1016/j.matdes.2017.06.043>.

- [23] A. Kretschmer, T. Wojcik, R. Schuster, K. Yalamanchili, H. Rudigier, P.H. Mayrhofer, Tuning of structure, grain orientation and mechanical properties in reactively sputtered (Al, Mo, Ta, V, W)N, Mater. Des. 213 (2022) 110346, <https://doi.org/10.1016/j.matdes.2021.110346>.
- [24] L.-Q. Chen, Phase-field models for microstructure evolution, Annu. Rev. Mater. Res. 32 (1) (2002) 113–140, <https://doi.org/10.1146/annurev.matsci.32.112001.132041>.
- [25] Y. Lu, C. Wang, Y. Gao, R. Shi, X. Liu, Y. Wang, Microstructure map for self-organized phase separation during film deposition, Phys. Rev. Lett. 109 (8) (2012) 086101, <https://doi.org/10.1103/PhysRevLett.109.086101>.
- [26] J.A. Stewart, D.E. Spearot, Phase-field models for simulating physical vapor deposition and grain evolution of isotropic single-phase polycrystalline thin films, Comput. Mater. Sci. 123 (2016) 111–120, <https://doi.org/10.1016/j.commatsci.2016.06.021>.
- [27] J.A. Stewart, D.E. Spearot, Phase-field simulations of microstructure evolution during physical vapor deposition of single-phase thin films, J. Mater. Res. 131 (2017) 170–177, <https://doi.org/10.1016/j.commatsci.2017.01.034>.
- [28] J.A. Stewart, R. Dingreville, Microstructure morphology and concentration modulation of nanocomposite thin-films during simulated physical vapor deposition, Acta Mater. 188 (2020) 181–191, <https://doi.org/10.1016/j.actamat.2020.02.011>.
- [29] Y. Lu, B. Derby, H. Sriram, K. Kadirkal, C. Wang, X. Liu, A. Misra, Y. Wang, Microstructure development and morphological transition during deposition of immiscible alloy films, Acta Mater. 220 (2021) 117313, <https://doi.org/10.1016/j.actamat.2021.117313>.
- [30] M. Sivapragash, P. Kumaradhas, B. Stanly Jones Retnam, X. Felix Joseph, U.T.S. Pillai, Taguchi based genetic approach for optimizing the PVD process parameter for coating ZrN on AZ91D magnesium alloy, Mater. Des. 90 (2016) 713–722, <https://doi.org/10.1016/j.matdes.2015.11.027>.
- [31] 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) (2020) 1–10, <https://doi.org/10.1038/s43246-020-0017-2>.
- [32] A. Costine, P. Delsa, T. Li, P. Reinke, P.V. Balachandran, Data-driven assessment of chemical vapor deposition grown MoS₂ monolayer thin films, J. Appl. Phys. 128 (23) (2020) 235303, <https://doi.org/10.1063/5.0017507>.
- [33] B.P. MacLeod, F.G.L. Parlane, T.D. Morrissey, F. Häse, L.M. Roch, K.E. Dettelbach, R. Moreira, L.E. Yunker, M.B. Rooney, J.R. Deeth, et al., Self-driving laboratory for accelerated discovery of thin-film materials, Sci. Adv. 6 (20) (2020) eaaz8867, <https://doi.org/10.1126/sciadv.aaz8867>.
- [34] I. Ohkubo, Z. Hou, J.N. Lee, T. Aizawa, M. Lippmaa, T. Chikyow, K. Tsuda, T. Mori, Realization of closed-loop optimization of epitaxial titanium nitride thin-film growth via machine learning, Mater. Today Phys. 16 (2021) 100296, <https://doi.org/10.1016/j.mtphys.2020.100296>.
- [35] M. Powers, J.A. Stewart, R. Dingreville, B.K. Derby, A. Misra, Compositionally-driven formation mechanism of hierarchical morphologies in Co-deposited immiscible alloy thin films, Nanomaterials 11 (10) (2021) 2635, <https://doi.org/10.3390/nano11102635>.
- [36] P. Rajak, A. Krishnamoorthy, A. Mishra, R. Kalia, A. Nakano, P. Vashishta, Autonomous reinforcement learning agent for chemical vapor deposition synthesis of quantum materials, npj Comput. Mater. 7 (1) (2021) 1–9, <https://doi.org/10.1038/s41524-021-00535-3>.
- [37] F.P. Such, V. Madhavan, E. Conti, J. Lehman, K.O. Stanley, J. Clune, Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning, arXiv preprint arXiv:1712.06567, 2017. <https://doi.org/10.48550/arXiv.1712.06567>.
- [38] S. Whitelam, I. Tamblyn, Learning to grow: Control of material self-assembly using evolutionary reinforcement learning, Phys. Rev. E 101 (5) (2020) 052604, <https://doi.org/10.1103/PhysRevE.101.052604>.
- [39] P.C. Jennings, S. Lysgaard, J.S. Hummelshøj, T. Vegge, T. Bligaard, Genetic algorithms for computational materials discovery accelerated by machine learning, npj Comput. Mater. 5 (1) (2019) 1–6, <https://doi.org/10.1038/s41524-019-0181-4>.
- [40] S.N. Hankins, R.S. Fertig III, Methodology for optimizing composite design via biological pattern generation mechanisms, Mater. Des. 197 (2021) 109208, <https://doi.org/10.1016/j.matdes.2020.109208>.
- [41] C. Kim, R. Batra, L. Chen, H. Tran, R. Ramprasad, Polymer design using genetic algorithm and machine learning, Comput. Mater. Sci. 186 (2021) 110067, <https://doi.org/10.1016/j.commatsci.2020.110067>.
- [42] L. Vanneschi, M. Castelli, S. Silva, Measuring bloat, overfitting and functional complexity in genetic programming, in: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation, 2010, pp. 877–884, <https://doi.org/10.1145/1830483.1830643>.
- [43] D.B. Brough, D. Wheeler, S.R. Kalidindi, Materials knowledge systems in python—a data science framework for accelerated development of hierarchical materials, Integr. Mater. Manuf. Innov. 6 (1) (2017) 36, <https://doi.org/10.1007/s40192-017-0089-0>.
- [44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (85) (2011) 2825–2830. URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [45] E. Herman, J.A. Stewart, R. Dingreville, A data-driven surrogate model to rapidly predict microstructure morphology during physical vapor deposition, Appl. Math. Model. 88 (2020) 589–603, <https://doi.org/10.1016/j.apm.2020.06.046>.
- [46] D. Montes de Oca Zapiain, J.A. Stewart, R. Dingreville, Accelerating phase-field-based microstructure evolution predictions via surrogate models trained by machine learning methods, npj Comput. Mater. 7 (1) (2021) 1–11, <https://doi.org/10.1038/s41524-020-00471-8>.
- [47] C.D. Adams, M. Atzman, Y.T. Cheng, D.J. Srolovitz, Phase separation during co-deposition of Al–Ge thin films, J. Mater. Res. 7 (3) (1992) 653–666, <https://doi.org/10.1557/JMR.1992.0653>.
- [48] B. Derby, Y. Cui, J.K. Baldwin, A. Misra, Effects of substrate temperature and deposition rate on the phase separated morphology of co-sputtered, Cu–Mo thin films, Thin Solid Films 647 (2018) 50–56, <https://doi.org/10.1016/j.tsf.2017.12.013>.
- [49] D. Depla, W.P. Leroy, Magnetron sputter deposition as visualized by Monte Carlo modeling, Thin Solid Films 520 (20) (2012) 6337–6354, <https://doi.org/10.1016/j.tsf.2012.06.032>.
- [50] R. Dingreville, J.A. Stewart, E.Y. Chen, J.M. Monti, Benchmark problems for the mesoscale multiphysics phase field simulator (MEMPHIS), Technical Report SAND2020-12852, Sandia National Laboratories (SNL-NM), Albuquerque, NM, USA, 2020. URL <https://doi.org/10.2172/1615889>.
- [51] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner Gardner, M. Parizeau, C. Gagné, DEAP: Evolutionary algorithms made easy, J. Mach. Learn. Res. 13 (70) (2012) 2171–2175. <http://jmlr.org/papers/v13/fortin12a.htm>.
- [52] C.D. Adams, M. Atzman, Y.-T. Cheng, D.J. Srolovitz, Transition from lateral to transverse phase separation during film co-deposition, Appl. Phys. Lett. 59 (20) (1991) 2535–2537, <https://doi.org/10.1063/1.105944>.
- [53] Y. Cui, B. Derby, N. Li, A. Misra, Design of bicontinuous metallic nanocomposites for high-strength and plasticity, Mater. Des. 166 (2019) 107602, <https://doi.org/10.1016/j.matdes.2019.107602>.
- [54] M.N. Polyakov, T. Chookajorn, M. Mecklenburg, C.A. Schuh, A.M. Hodge, Sputtered Hf–Ti nanostructures: A segregation and high-temperature stability study, Acta Mater. 108 (2016) 8–16, <https://doi.org/10.1016/j.actamat.2016.01.073>.
- [55] J.S. Riano, A.M. Hodge, Exploring the thermal stability of a bimodal nanoscale multilayered system, Scripta Mater. 166 (2019) 19–23, <https://doi.org/10.1016/j.scriptamat.2019.02.043>.
- [56] M. Powers, B. Derby, E. Raeker, N. Champion, A. Misra, Hillcock formation in co-deposited thin films of immiscible metal alloy systems, Thin Solid Films 693 (2020) 137692, <https://doi.org/10.1016/j.tsf.2019.137692>.
- [57] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in: Computer Vision – ECCV 2016, Springer International Publishing, 2016, pp. 630–645, https://doi.org/10.1007/978-3-319-46493-0_38.
- [58] S. Sundar, V. Sundararaghavan, Database development and exploration of process–microstructure relationships using variational autoencoders, Mater. Today Commun. 25 (2020) 101201, <https://doi.org/10.1016/j.mtcomm.2020.101201>.
- [59] Y. Kim, H.K. Park, J. Jung, P. Asghari-Rad, S. Lee, J.Y. Kim, H.G. Jung, Hyoung S. Kim, Exploration of optimal microstructure and mechanical properties in continuous microstructure space using a variational autoencoder, Mater. Des. 202 (2021) 109544, <https://doi.org/10.1016/j.matdes.2021.109544>.
- [60] K. Karhunen, Zur spektraltheorie stochastischer prozesse, Ann. Acad. Sci. Fenn. Math. 37 (1946) 34.
- [61] M. Loeve, Probability Theory: Foundations, Random Sequences, Van Nostrand, New York, NY, 1955.
- [62] R.G. Ghanem, P.D. Spanos, Stochastic finite element method: Response statistics, in: Stochastic Finite Elements: A Spectral Approach, Springer, 1991, pp. 101–119, https://doi.org/10.1007/978-1-4612-3094-6_4.
- [63] B. Schölkopf, A. Smola, K.-R. Müller, Kernel principal component analysis, in: International Conference on Artificial Neural Networks, Springer, 1997, pp. 583–588, <https://doi.org/10.1007/BFb0020217>.
- [64] J. Hui, Q. Hu, H. Zhang, J. Zhao, Y. Luo, Y. Ren, Z. Zhang, H. Wang, High-throughput investigation of structural evolution upon solid-state in Cu–Cr–Co combinatorial multilayer thin-film, Mater. Des. 215 (2022) 110455, <https://doi.org/10.1016/j.matdes.2022.110455>.