Accelerating simulations of strained-film growth by deep learning: Finite element method accuracy over long time scales

Daniele Lanzoni ; Fabrizio Rovaris ; Luis Martín-Encinar ; Andrea Fantasia ;
Roberto Bergamaschini ✉ ; Francesco Montalenti

Check for updates

View Online

Export Citation

Accelerating simulations of strained-film growth by deep learning: Finite element method accuracy over long time scales

View Online   Export Citation   CrossMark

Daniele Lanzoni,[1] [ID]   Fabrizio Rovaris,[1] [ID]   Luis Martín-Encinar,[2] [ID]   Andrea Fantasia,[1] [ID]
Roberto Bergamaschini,[1,a)] [ID]   and Francesco Montalenti[1] [ID]

AFFILIATIONS

[1] L-NESS and Dept. of Materials Science, University of Milano-Bicocca, 20125 Milano, Italy

[2] Dpto. de Electricidad y Electrónica, E. T. S. I. de Telecomunicación, Universidad de Valladolid, 47011 Valladolid, Spain

[a)] Author to whom correspondence should be addressed: roberto.bergamaschini@unimib.it

ABSTRACT

A convolutional neural network is trained on a large dataset of suitably randomized film profiles and corresponding elastic energy densities ρ_ε, computed by the finite element method. The trained model provides quantitative predictions of ρ_ε for arbitrary profiles, surrogating its explicit calculation, and is used for the time integration of partial differential equations describing the evolution of strained films. The close match found between the neural network predictions and the "ground-truth" evolutions obtained by the finite element method calculation of ρ_ε, even after tens-of-thousands of integration time-steps, validates the approach. A substantial computational speed up without significant loss of accuracy is demonstrated, allowing for million-steps simulations of islands growth and coarsening. The intriguing possibility of extending the domain size is also discussed.

I. INTRODUCTION

In the last decade, machine learning (ML) techniques[1–3] have been introduced in almost any field of science, from purely theoretical to applied disciplines and engineering.[4–9] Among the many uses, these methods proved to be an effective way to speed up otherwise computationally heavy calculations. In Condensed Matter Physics and Materials Science, considerable effort has been dedicated both to the acceleration of atomic-scale models[10–12] and, more recently, to the approximation of continuum models based on partial differential equations.[13–18] A few recent studies even proposed novel strategies to bridge the gap between the two paradigms by exploiting ML.[19,20]

Here we show that neural networks (NN)[2] can be conveniently exploited to accelerate continuum simulations of strained-film growth (as encountered, e.g., in lattice-mismatched heteroepitaxy[21]), a key field in Materials Science.[22,23] Indeed, a wide range of behaviors depending on growth parameters can be observed and exploited in various applications. In some cases, controlling

the surface roughness that may arise during the growth is the main issue.[24,25] Other applications are instead based on the formation of three-dimensional (3D) nanostructures, such as islands or low-dimensional objects,[26–29] yielding novel properties of high interest. Understanding the interplay between growth parameters and material properties is thus key and the development of reliable and fast simulation tools to support and anticipate costly experiments would offer great benefits. As atomistic approaches are generally too expensive for simulating the film evolution on realistic spatial and, most critically, temporal scales, continuum modeling becomes fundamental.[30]

Since the seminal work by Mullins,[31] more than 70 years ago, researchers have been working on the development of continuum models dealing with surface dynamics,[22,30] extending the original approach to cope with strain effects,[32–35] multi-component alloys,[36,37] and defects.[38,39] Given a profile h at a time t, its evolution is driven by the (local) surface chemical potential μ. While different regimes, such as evaporation/condensation,[31] have been tested,

we will focus here on surface diffusion dynamics in two-dimensions (2D):

$$\frac{\partial h}{\partial t} = M\sqrt{1 + h'^2}\frac{\partial^2 \mu}{\partial s^2} + f, \qquad (1)$$

where M is the surface mobility, $\partial/\partial s$ is the surface derivative, $h' = \partial h/\partial x$, and f is a vertical deposition flux. μ comprises a surface energy contribution, which is local, and an elastic energy contribution ρ_ε, which is non-local and stems from different sources such as film-substrate lattice mismatch, thermal mismatch, external stress, and dislocations.

The evaluation of ρ_ε requires in general the solution of the mechanical equilibrium problem by numerical approaches. Among these, the Finite Element Method (FEM) is particularly well suited for generic free-boundary problems, although efficient alternatives are available for more specific cases.[40] While single static calculations are not an issue for nowadays computational resources, accurate FEM solutions still pose severe limitations when time-integrating equations such as Eq. (1), as ρ_ε needs to be continuously updated, even for millions of iterations, thus multiplying the computational burden on simulation times. A few studies in the literature pursued this strategy by using advanced numerical approaches with optimized time-integration schemes[41–44] or resorted to computationally-cheap semi-analytical methods,[37,45] at the price of an approximated solution for the elastic problem.

In this work, we explore the possibility of by-passing the explicit solution of the elastic problem using a convolutional NN (CNN)[2,46–48] model to directly predict the ρ_ε contribution. CNNs involve numerical operations that are much cheaper than those required for a FEM solver and can be efficiently implemented on multi-threaded systems and GPUs, allowing for the approximation of complex, non-linear functions at a fraction of the computational cost of conventional approaches. Here, we train a CNN capable of providing a mapping of arbitrary surface profiles to their corresponding ρ_ε with the same accuracy as FEM. After validating the model and inspecting its extrapolation performances, we exploit it in the time integration of the thin-film evolution equations. The substantial speed-up offered by this deep learning approach enables the analysis of evolution regimes, which are usually expensive to tackle.

We remark that, in principle, an even stronger computational-time compression might be obtained by predicting the full driving force μ or the profile sequence directly.[16,17] The present choice of focusing only on the term that causes the major computational bottleneck, however, ensures full transferability, while retaining a better physical insight beyond the "blind" prediction of the required quantity. Our approach is also complementary to other ML methods devised to tackle differential equation problems. For example, the Physics Informed NN (PINN) method developed in Ref. 14 (and recently applied to linear elasticity[18]) could potentially substitute FEM solvers for the mechanical equilibrium problem in the whole domain. In this respect, the NN approach here proposed represents a data-driven alternative that does not require re-training for new profiles and completely bypasses the solution step by directly predicting ρ_ε at the film surface.

II. METHODS

A. Strain calculation and dataset definition

We describe the film morphology by a 1D profile function $y = h(x)$, evolving according to Eq. (1). With no loss of generality for the workflow, we refer to the prototypical case of a Ge film subject to a purely diagonal 3.99% eigenstrain ε^* [49,50] as the one produced by the (compressive) lattice mismatch in Ge/Si(001) epitaxy.[26,51]

As stated, the driving force for the surface dynamics is the local chemical potential $\mu(x)$ along the surface, defined as the variational derivative of the free energy F with respect to the profile itself:

$$\mu = \frac{\delta F}{\delta h} = \kappa\gamma + \mu_w + \rho_\varepsilon, \qquad (2)$$

where κ is the local profile curvature, which can be directly computed by finite differences as $-h''/\sqrt{1 + h'^2}$, γ is the (isotropic) surface energy density for the film, and μ_w is the film-substrate wetting interaction term. For thick films, μ_w can be neglected and γ reduces to a constant, that is, $\gamma_{\text{Ge}} = 6.0$ eV/nm² for Ge(001). In the thin-film case, representative of the early growth stages, instead, γ is a function of the film thickness.[52] Following Ref. 45, $\gamma(h) = \gamma_{\text{Ge}} + (\gamma_{\text{Si}} - \gamma_{\text{Ge}})\exp(-h/d)$, with the surface energy density of the Si(001) substrate $\gamma_{\text{Si}} = 8.7$ eV/nm² and the characteristic decay length $d = 0.27$ nm.[45] By this definition, we get that $\mu_w(h) = d\gamma/dh = (1/d)(\gamma_{\text{Ge}} - \gamma_{\text{Si}})\exp(-h/d)$, which provides a stabilization of the planar film configuration below a critical thickness $h_c \approx 1.4$ nm.

The elastic contribution $\rho_\varepsilon(x)$ is defined in the framework of linear-elasticity as

$$\rho_\varepsilon(x) = \frac{1}{2}\sigma : (\varepsilon - \varepsilon^*), \qquad (3)$$

where σ and ε are the (local) stress and strain tensors evaluated by the explicit solution of the mechanical equilibrium problem $\nabla\sigma = 0$.[53] For the sake of simplicity, the Ge film is approximated as an elastically isotropic medium with Young modulus $Y = 103$ GPa and Poisson ratio $\nu = 0.26$ and its elastic deformation state is numerically computed by the 2D FEM implementation described in Ref. 39. As sketched in Fig. 1, a triangular mesh is defined with $h(x)$ as the top boundary corresponding to the film free surface,
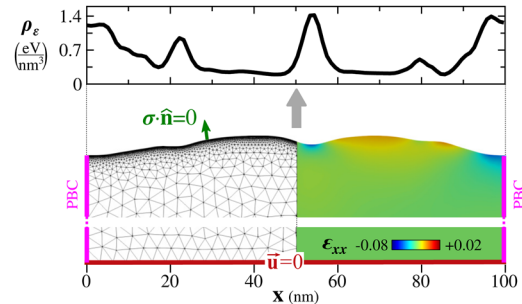
FIG. 1. Schematics of the FEM representation of a Perlin-noise surface profile from the validation set reporting the corresponding strain map for the ε_{xx} component and the elastic energy density ρ_ε at the surface.

where the natural, no-stress boundary condition $\sigma \cdot \hat{n} = 0$ holds. The domain is closed by a straight bottom boundary, far enough (≈ 100 nm) from the surface so as to mimic a semi-infinite film, with a zero-displacement ($\vec{u} = 0$) boundary condition. Periodic Boundary Conditions (PBCs) are imposed at the sides. An adaptive mesh, shown in Fig. 1, has been used, with typical element sizes ranging from ≈ 5 nm at the bottom boundary to 0.05 nm at the free-surface, where strain fields change in a shorter range. This guarantees the convergence of the FEM solution while keeping the number of collocation points as small as possible.

A domain size of $L = 100$ nm is considered since the characteristic length scale of morphology is expected to be of the order of a few tens of nanometers, as the fastest growing mode has wavelength $\lambda \approx 23$ nm,[32,33,35] given the aforementioned parameters.

The NN dataset is composed of $\sim 80\,000$ $[h(x), \rho_\varepsilon(x)]$ couples, which can be found in Ref. 54, and has been split at random into training (85%) and validation (15%) sets. Given the non-linear relationship between h and ρ_ε, two types of profiles are considered: (1) pure sine-waves with wavelengths $\lambda \in \{L/6, L/5, L/3, L/2, L\}$ and amplitudes in $[10^{-3}, 10]$ nm, representative of the Fourier components of a generic profile; (2) arbitrary undulated morphologies set by the kind of gradient noise known as Perlin noise[55,56] with amplitudes in the $[5 \times 10^{-3}, 15.4]$ nm range. An example of the latter is reported in Fig. 1.

B. Neural network architecture

A fully convolutional NN[1,2,57] implemented within the PyTorch framework[58] has been used to map the surface profile $h(x)$ to the corresponding elastic energy density $\rho_\varepsilon(x)$. The term "Fully Convolutional" refers to the fact that the NN comprises only convolutions and element-wise functions. As a result, the input of the network is not restricted in length, thus allowing for the generalization to computational domains of arbitrary size.

This architecture incorporates translational equivariance[2] by construction, which increases NN accuracy and reduces the number of examples required in training. Other symmetries can also be straightforwardly added by ad-hoc manipulations of the NN parameters and structure. Mirror symmetry is enforced by making convolution kernels reflection-invariant by construction and through data augmentation. While in principle only one of these strategies should be sufficient, models trained using both obtained a lower generalization error in our tests. The mean value of $h(x)$, $\bar{h}$, is also removed before passing a profile as input to the NN to enforce the vertical shift invariance proper of the considered system. The CNN adopted circular padding,[59,60] enforcing PBCs exactly and consistently with the FEM calculation scheme.

In terms of structure, the NN is composed of stacked, one-dimensional convolutional layers followed by *tanh* activation functions. Each layer contains 20 different kernels with a receptive field, that is, the spatial extent of the convolutional filter, of 21 "pixels," each corresponding to a collocation point used to discretize the $[h(x), \rho_\varepsilon(x)]$ pairs. A constant spacing of $dx = 1$ nm has been chosen to ensure accurate tracking of the surface dynamics while still allowing for large enough time-steps. A total of 5 convolution-*tanh* blocks are used, with a nominal total of $\approx 26\,000$ parameters. Thanks to reflection invariance, however, their effective number is reduced.

The NN is trained by minimizing the mean squared error loss function $\mathscr{L}$:

$$\mathscr{L}(\vartheta) = \frac{1}{N_{\mathrm{TS}}} \sum_{i=1}^{N_{\mathrm{TS}}} |\rho_\varepsilon^{\mathrm{NN}}(h_i|\vartheta) - \rho_\varepsilon^{\mathrm{FEM}}(h_i)|^2, \qquad (4)$$

where $\rho_\varepsilon^{\mathrm{NN}}$ is the NN prediction, $\rho_\varepsilon^{\mathrm{FEM}}$ is its ground-truth FEM counterpart, i indices a training instance within the N_{TS} of the training set, $|\cdot|^2$ represents the L^2 norm in $\mathbb{R}^{N_x}$ (N_x being the number of collocation points used for the discretization of profiles), and ϑ is the set of NN parameters. The standard PyTorch implementation of the Adam optimizer,[58,61] with a learning rate of 1×10^{-5} and a batch size of 512, was used.

III. RESULTS AND DISCUSSION

A. Neural network training and validation

Figure 2 reports the training and validation losses. These values steadily decrease during the 6000 training epochs, without any sign of overfitting. A comparison between the NN prediction and the FEM ρ_ε for the same validation Perlin-noise profile for Fig. 1 is reported in Fig. 3(a).

An additional proof of the reliability of the trained NN model is given in Figs. 3(b)–3(d), where we compare the NN prediction against the FEM solution for three profiles, qualitatively different from those present in the dataset (additional examples can be found in Ref. 54). The accurate reproduction provided by the NN asserts its generalization capabilities, as the NN was never exposed to such functional forms during training. Noticeably, the main discrepancies are present at the singular points corresponding to the sharp angles of the profile in panel (d), clearly indicating that faceted profiles exceed the extrapolation limits of the trained model.

For a more quantitative evaluation of the NN accuracy, we introduce the normalized Root Mean Squared Error (RMSE):

$$\sigma_\rho = \frac{1}{\rho_{\max}^{\mathrm{FEM}} - \rho_{\min}^{\mathrm{FEM}}} \sqrt{\sum_j^{N_x} \frac{\left(\rho^{\mathrm{NN}}(x_j) - \rho^{\mathrm{FEM}}(x_j)\right)^2}{N_x}}, \qquad (5)$$

where j runs on collocation points. We identify a value of $\sigma_\rho = 0.02$, that is, a 2% relative RMSE, as a reasonable threshold below which the NN predictions can be considered to accurately reproduce the
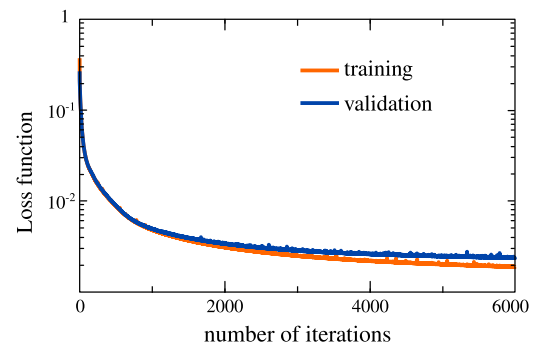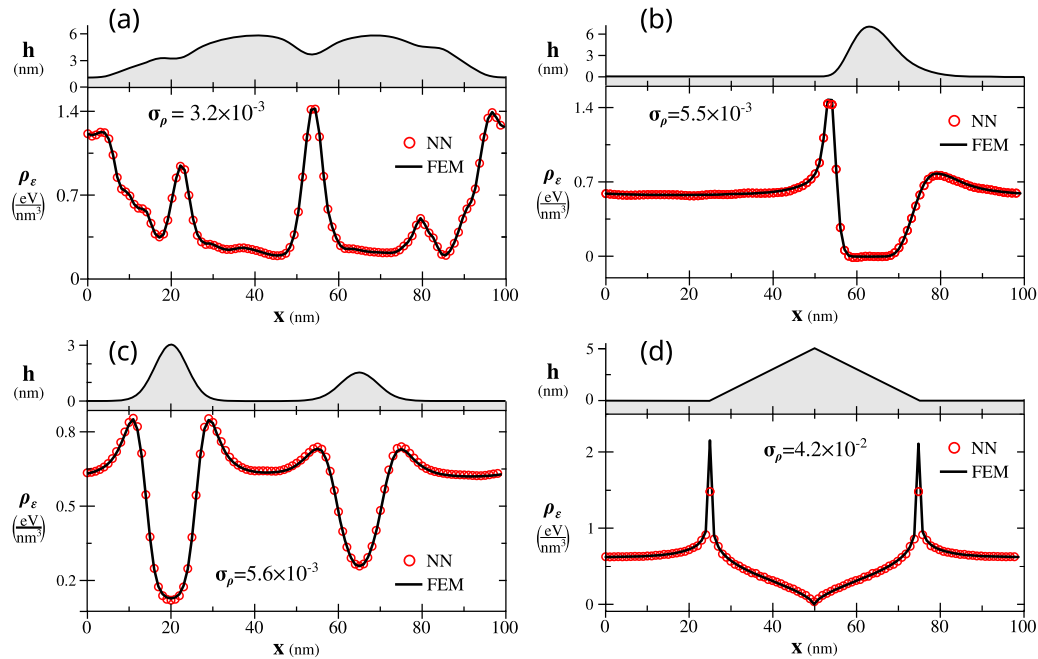
FIG. 2. Log-scale plot of the training and validation losses of Eq. (4) for the NN trained on FEM data.

FIG. 3. Comparison between the NN-predicted (red circles) elastic energy density ρ_ε and the corresponding FEM calculations (solid black line) for different profiles $h(x)$: (a) Perlin noise sample from the validation set; (b) asymmetric island; (c) two Gaussian islands superimposed on a flat profile; (d) a triangular island.

FEM calculations. Indeed, profiles reported in Figs. 3(a)–3(c) all return below-threshold σ_ρ errors, while the case (d) cannot be considered in one-to-one correspondence with FEM calculations due to the failures in predicting ρ_ε spikes. In the present work, as we focus on isotropic systems, such faceted morphologies do not pose critical issues. A suitable extension of the training set, however, should be considered for future extensions tackling more realistic cases with surface anisotropy.

B. Validation of time evolutions based on the NN approximation of ρ_ε

The NN-predicted elastic energy densities can be fruitfully exploited for the numerical integration of the surface evolution dynamics stated in Eq. (1) at a reduced computational cost. In particular, an ≈ 4 orders of magnitude speed-up with respect to FEM is found. An explicit-Euler scheme is implemented, with a fixed time-step δt. Given the present purpose of demonstrating the stability of the NN approach during time integration, we report the simulation time t by the number of time-steps. The absolute timescale can be recovered by multiplying the reported numbers by $M\delta t$, here set to 5×10^{-3} in all simulations. Since a typical evolution sequence occurs over hundreds of thousands of time-steps, one-to-one comparisons with the ground-truth provided by iterating the explicit FEM calculation of ρ_ε over time are possible only for small systems and limited time intervals. The NN approximation capabilities have been further confirmed in a quantitative study limited to small-slope profiles for which semi-analytical expressions of the strain field are available.[62]

As a first example, we consider the simplest case of annealing ($f = 0$) of a thick film with isotropic surface energy density γ, such that $\mu = \kappa\gamma + \rho_\varepsilon$ (κ being the profile curvature). In this case, any perturbation of the film exceeding a certain critical wavelength is expected to grow exponentially according to the well-known ATG-instability model.[32–35]

Figure 4(a) reports representative stages of the growth of a sinusoidal perturbation, setting an unstable wavelength of 25 nm that was on purpose excluded from the training set, so as to return a fair test on the NN predictive capabilities. A semi-log plot of the peak-to-valley profile height Δh as a function of time is also reported in panel (b). The linear behavior confirms the exponential growth regime predicted by the analytic ATG theory. In this specific case, a one-to-one comparison with the FEM evolution is provided. Notice, however, that the number of FEM evaluations required for tracking the dynamics is approximately equal to the size of the whole dataset used to train the NN, making it inconvenient to extensively repeat a similar analysis. A substantial overlap is obtained over the tens of thousands of integration steps of the simulation. Local discrepancies around the minima and maxima are recognizable only in the last evolution stages, when non-linear contributions induce the formation of cusp-points in between the peaks, leading to numerical divergence. To quantify the discrepancy between the NN predicted profiles and the true ones obtained by means of FEM, we resort to the normalized RMSE σ_h, defined as in Eq. (5) but now evaluated for the height profiles $h^{\mathrm{NN}}(x,t)$ and $h^{\mathrm{FEM}}(x,t)$. In panel (c), the σ_h error is reported as a function of time. As it can be observed, σ_h progressively increases, up to reaching the threshold value at the end of the reported simulation time, that is, close to the divergence
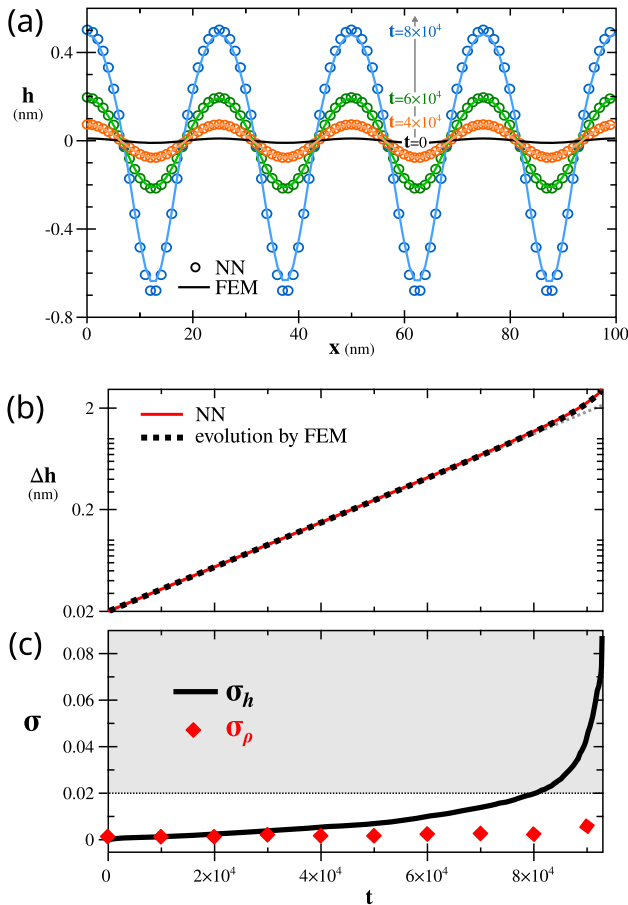
FIG. 4. Analysis of ATG-like dynamics by surface diffusion. (a) Evolution of a sinusoidal profile with $\lambda = 25$ nm and amplitude 0.01 nm as obtained by time integration using NN (circles) and FEM (solid line) evaluations of ρ_ε. (b) Evolution over time of the profile amplitude Δh in log-scale. (c) Prediction error for both the profile (solid line, σ_h) and the elastic energy density (red diamonds, σ_ρ).

[see also the curve at $t = 8 \times 10^4$ in Fig. 4(a)]. This increment is not due to a failure in the NN prediction of ρ_ε as made evident by the constant, below-threshold σ_ρ error, evaluated at representative evolution stages. It is instead the result of small-error accumulation in time integration, exacerbated by the exponential nature of the process considered.

A more appropriate description of the film growth dynamics, which also heals the divergent behavior of the ATG model allowing for the study of island formation and coarsening, can be achieved by introducing the substrate–film wetting interaction, as discussed in Sec. II A. Since the ML approach only applies to the elastic contribution, the same NN model can be used despite this modification. In Fig. 5, we report an example of a simulation of island growth, including the wetting energy term and a small deposition flux of $f/M = 2 \times 10^{-5}$, such to follow quasi-equilibrium conditions. Here and in the following figures, a 5× scaling is applied to the height axis for the sake of readability. As we want to provide a one-to-one comparison with the full FEM solution, at least for part of the dynamics,

we directly set the initial profile thickness above the critical value h_c, with an arbitrary modulation by random Fourier components with $\lambda \in [15, 100]$ nm to trigger unstable modes. As expected, in the early stages of the reported evolution, the corrugation amplifies similarly to the ATG model while, in the latest stages, that is, after 1×10^6 steps, stable islands grow and coarsen, resembling Stranski–Krastanov[30] behavior. The corresponding true evolution by FEM was carried on for the first 40 000 integration steps (still corresponding to half of the FEM evaluations required for generating the dataset). Again, close correspondence between the NN and FEM evolutions is achieved, as made evident by the overlap of profiles at both 20 000 and 40 000 steps in Fig. 5. Correspondingly, the σ_h error remains well below the threshold over the whole tested interval, certifying the accuracy of the NN approach. While the one-to-one comparison with FEM simulations cannot be extended to the last stage at 1×10^6 steps, it must be noted that, in contrast to the divergent behavior of the ATG case, the dynamics of island growth is quenched by the wetting interactions so that the error accumulation is expected to have a lesser impact on the long-time stages, preserving the reliability of the predicted profiles. As an indirect confirmation, in Fig. 5, we also compare the NN prediction of ρ_ε with its FEM counterpart for the three reported evolution stages. Evidently, the NN provides accurate evaluations of ρ_ε, with low σ_ρ error over the whole time interval, even after 1×10^6 steps.

## C. Growth simulations and generalization to large domains and long times

After having demonstrated the reliability of NN evolutions for islanding dynamics, we can use the trained model to produce more realistic simulations of island growth, starting from the bare substrate and depositing Ge for several nms. In particular, in Fig. 6, we report the evolution sequence for three different f/M ratios (for full animations, see supplementary material). In these cases, a random-Fourier-components-noise of amplitude 10^{-2} nm is reset at every step as long as the profile remains flat, so to trigger any unstable mode. As expected, for high f values [Fig. 6(a)], the film grows flat since the time scale of the instability is much slower than the one of deposition. By raising the deposition flux (b), the corrugation formation becomes concurrent with growth, yielding a final morphology with several island-like features. Finally, in the more thermodynamic, low-flux condition (c), the unstable modes develop faster than deposition and coarsen into a single island. Due to the constant supply of material by deposition, the morphology of the final island continuously adapts toward the minimum of energy, thus resulting in a sequence of quasi-equilibrium configurations. We remark that this last simulation required 30×10^6 integration steps, giving impressive proof of the stability of the proposed method. Also, it exceeds the training set size by two orders of magnitude, thus single-handedly justifying the computational effort of the dataset building.

Studying the equilibrium morphology of heteroepitaxial islands is relevant for assessing the long-time limits of simulations and provides further proof of the NN capability of returning meaningful outputs. This task can be conveniently treated as a free energy minimization problem constrained to the assigned amount of material. Since $\mu = \delta F / \delta h$, a steepest-descent scheme can be exploited, profiting from the cheap estimation of the elastic term ρ_ε provided by
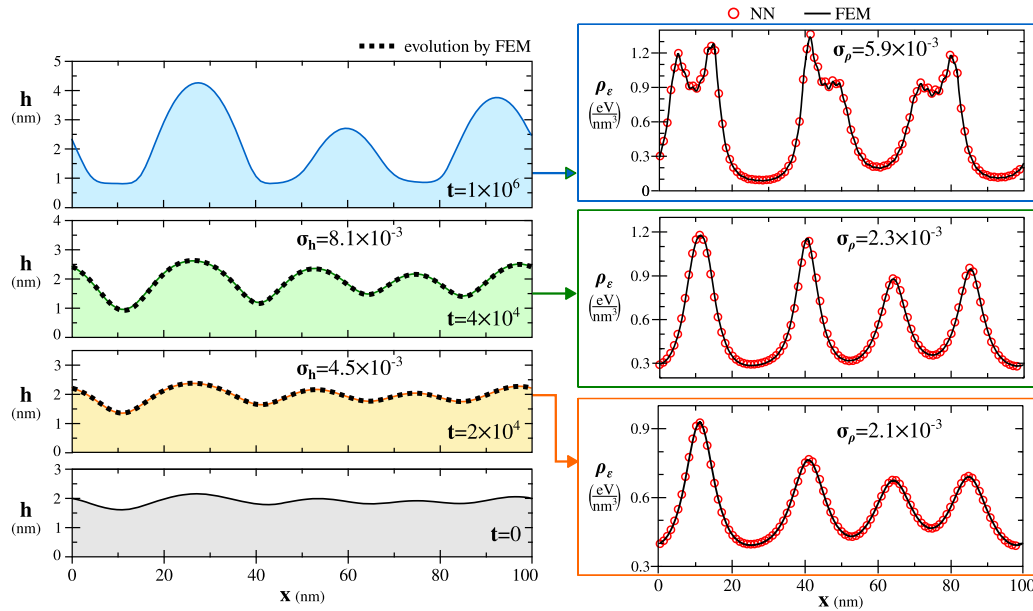
FIG. 5. Ge film growth simulation using the NN approximation of ρ_ε. Representative evolution stages are reported along with the corresponding ρ_ε profiles as computed by both NN (red circles) and FEM (black solid line). For the two stages after 20 000 and 40 000 integration steps, the NN predicted profiles are superimposed on the "true" ones (dotted lines) obtained by using the FEM evaluation of ρ_ε. Here, $f/M = 2 \times 10^{-5}$, yielding a deposition of 0.1 nm of material at the last stage.

the NN model. As initial conditions, we consider flat profiles with a Gaussian dimple to trigger the reshaping. Different film thicknesses h_0 are then considered to inspect the change in the equilibrium morphology as a function of the amount of material. The minimization
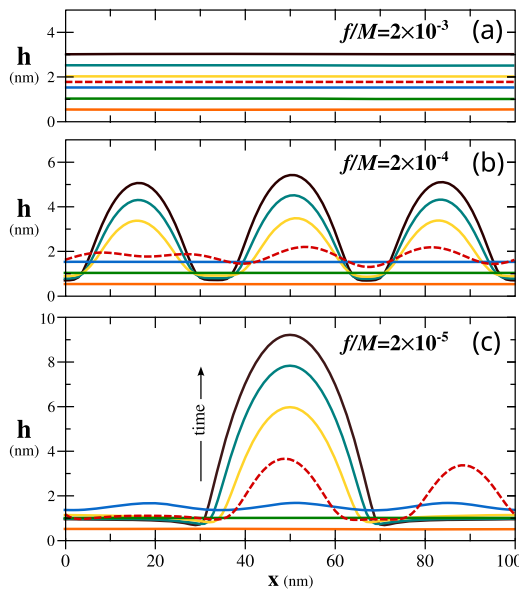
FIG. 6. Simulations of Ge film growth using the NN approximation of ρ_ε for three different deposition rates. Parameters are the same as in Fig. 5 but the simulation starts at zero film thickness and lasts until 3 nm of materials are deposited. Profiles are reported for every 0.5 nm of deposition but for the one dashed, which corresponds to a thickness of 1.75 nm.
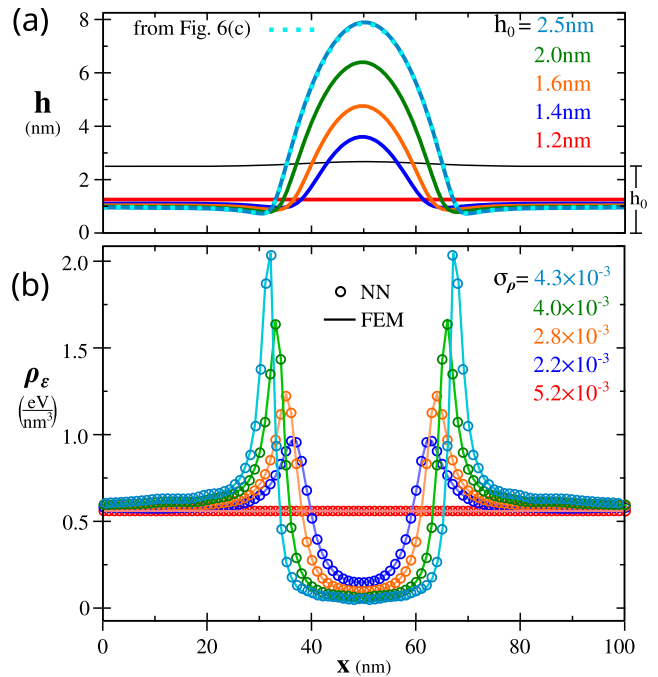
FIG. 7. Optimization of island profiles using the NN approximation of ρ_ε. (a) Equilibrium morphologies calculated for different amounts of Ge obtained by steepest-descent minimization starting from a flat profile of different thickness h_0, modulated by a small Gaussian (black line corresponds to $h_0 = 2.5$ nm). The largest profile for $h_0 = 2.5$ nm is superimposed with the growth stage of Fig. 6(c) with the same amount of material. (b) Comparison between the NN (circles) and FEM (solid line) elastic energy densities for the profiles in (a). σ_ρ values are also reported.

procedure is halted when the maximum deviation of $\mu(x)$ from its mean value is lower than 10^{-4} eV/nm^3. As shown in Fig. 7(a), for small enough h_0, due to the stabilization effect of the wetting contribution, the equilibrium configuration is the flat film. Vice versa, when considering thicker initial profiles, an island morphology yields the minimum energy. The largest island in the plot is also shown to overlap with the one obtained for the same amount of material from the surface diffusion evolution of Fig. 6(c), thus giving further proof of consistency of the NN predictions. To confirm that the computed profiles have a quantitative foundation, it is sufficient to verify the match of the NN evaluation of the elastic contribution ρ_ε with the corresponding FEM solution, as the other terms in μ are identical by construction. In Fig. 7(b), we show such a comparison for the profiles of panel (a). In all cases, one-to-one correspondence

can be observed, and the σ_ρ values are well below the 0.02 threshold. Once again, we remark that performing the same optimization via FEM would be a computationally demanding task, given the number of iterations required ($\gtrsim 10\,000$).

One of the key advantages offered by the fully convolutional architecture is that it is not restricted to the input size used in training. In Fig. 8, we take profit of this for inspecting the coarsening dynamics of islands formed on a 1-μm-long, that is, 10 times larger than the training examples, Ge film surface (see the video in the supplementary material). Even larger dimensions have been tested, up to 1 mm, but are not shown here for graphical reasons and can be found in Ref. 54. Given the large domain and the number of iterations, up to 7×10^6 steps, comparisons with explicit FEM calculations of ρ_ε are only possible at a few representative stages. The
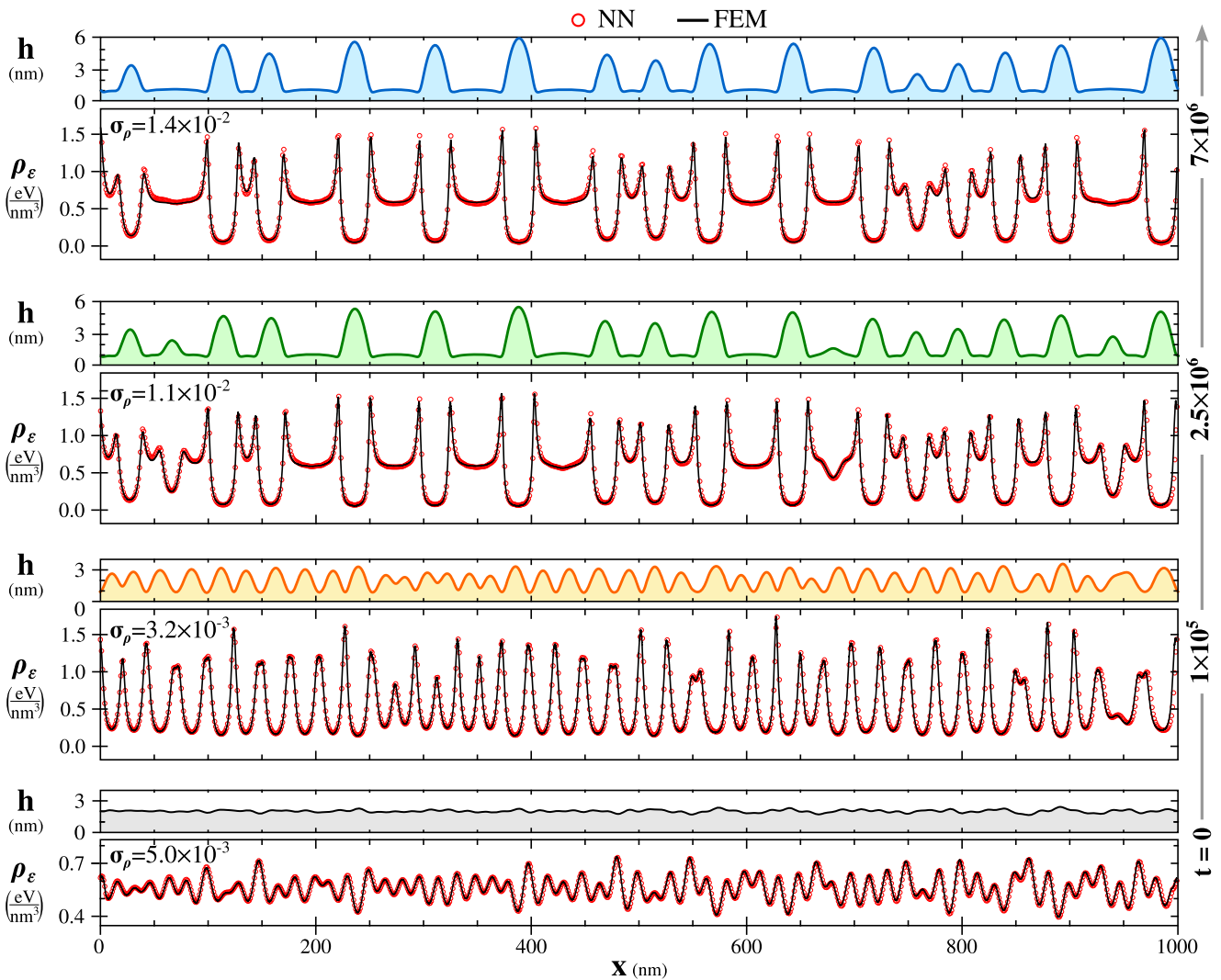
FIG. 8. Ge film annealing simulation using the NN approximation of ρ_ε. Representative evolution stages are reported along with the corresponding ρ_ε profiles as obtained by both NN (circles) and FEM (solid line). The same parameters in Fig. 5 are considered with a null deposition flux. Notice that the heights of the reported profiles are scaled-up by a factor of 2 with respect to previous figures.

NN approach returns remarkable accuracy, with below-threshold σ_ρ values at all stages. It must be however pointed out that these generalization capabilities are limited to configurations in which the profile is composed of features separated by distances comparable to the training set periodicity. Indeed, configurations exceeding such constraints are not expected to be reproduced using a dataset with a fixed size, as the trained model lacks any knowledge of the intrinsically long-range effects of the elastic field. A more detailed investigation of this effect is discussed in Ref. 62.

IV. CONCLUSIONS

The present study demonstrates how a Convolutional Neural Network can be fruitfully applied to learn one of the main driving forces of strained film growth, that is, the surface elastic energy density. A speed-up of ~4 orders of magnitude is found with respect to using the same FEM code exploited for the training, while retaining the same level of accuracy, with typical discrepancies of ~1% or less. Notably, such a gain is entirely ascribed to the calculation of ρ_ε irrespective of the adopted time-integration scheme. Here, the basic forward-Euler method has been implemented to ensure a fair comparison. However, the proposed approach could be profitably integrated into semi-implicit schemes, for example, leveraging automatic differentiation techniques, even when this would be non-trivial with conventional methods. This promises to push even further the spatial and temporal scales accessible to numeric investigation.

As the computational cost of a NN comes from its architecture and not from the number of examples in training, future efforts on building bigger FEM datasets (e.g., including faceted morphologies) promise to push the accuracy of the method even further while keeping evaluation computational costs low. Clearly, one major limitation of the current work is the simplifying assumption of 2D modeling. Besides the obviously larger computational effort, generalization to 3D description should be straightforward, and it is here left for future work.

While the present study focused on the prediction of the elastic energy density for simulating the growth of thin, strained Ge films, applications are not limited to it. The same approach could be applied to predict other quantities, for example, the displacement and stress/strain fields or other surface properties. Another interesting perspective concerns the possibility of tackling different heteroepitaxial systems. Due to the non-linear dependence of ρ_ε on elastic constants, in principle, a new dataset and model should be set for each new material. The only exception would be when considering a variation in the eigenstrain only, which is a reasonable approximation to describe SiGe/Si alloys. In this case, thanks to the quadratic dependence of ρ_ε on ε^* a simple rescaling is sufficient. Nonetheless, the current model could still be used as a starting point in a transfer-learning procedure. This is expected to reduce the number of examples required for the training when considering systems affine to Ge/Si. A more challenging possibility would be to condition the NN output on Y and ν explicitly and build an ad-hoc dataset exploring different elastic constant combinations, so as to build a comprehensive NN model for the ρ_ε of generic heteroepitaxial films.

Finally, the same paradigm presented here could also be used for tackling more complex physical effects (e.g., including surface and elastic anisotropy or plastic relaxation by dislocations) or could even be transferred to other dynamical problems, thus broadening the class of simulations that could be accelerated.

SUPPLEMENTARY MATERIAL

Videos of the complete profile evolution for the growth simulations in Fig. 6 are provided in the supplementary material as Video_Growth_High_Rate.avi for the fastest deposition rate of panel (a), Video_Growth_Intermediate_Rate.avi for the intermediate one of panel (b), and Video_Growth_Low_Rate.avi for the slowest one of panel (c). The video of the complete annealing simulation in Fig. 8 is also made available in the supplementary material as Video_Annealing_L1000.avi.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Daniele Lanzoni: Conceptualization (supporting); Investigation (equal); Software (lead); Writing – original draft (lead). **Fabrizio Rovaris**: Conceptualization (supporting); Investigation (equal); Software (equal); Writing – review & editing (equal). **Luis Martín-Encinar**: Investigation (equal); Software (supporting); Writing – original draft (equal). **Andrea Fantasia**: Investigation (supporting); Software (supporting). **Roberto Bergamaschini**: Conceptualization (supporting); Investigation (supporting); Supervision (equal); Writing – original draft (lead). **Francesco Montalenti**: Conceptualization (lead); Funding acquisition (lead); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are openly available in Materials Cloud at https://doi.org/10.24435/materialscloud:5r-9j, Ref. 54. The underlying code for this study is available in GitHub at https://github.com/mosegroup/nn4surf.

REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York NY, 2006).
[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge MA, 2016) http://www.deeplearningbook.org.
[3] P. Mehta, M. Bukov, C. H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to machine learning for physicists," Phys. Rep. **810**, 1–124 (2019).

[4] J. Sirignano and K. Spiliopoulos, "DGM: A deep learning algorithm for solving partial differential equations," J. Comput. Phys. **375**, 1339–1364 (2018).

[5] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine learning for fluid mechanics," Annu. Rev. Fluid Mech. **52**, 477–508 (2020).

[6] A. Scheinker and R. Pokharel, "Physics-constrained 3D convolutional neural networks for electrodynamics," APL Mach. Learn. **1**, 026109 (2023).

[7] W. G. Dettmer, E. J. Muttio, R. Alhayki, and D. Perić, "A framework for neural network based constitutive modelling of inelastic materials," Comput. Methods Appl. Mech. Eng. **420**, 116672 (2024).

[8] A. C. Ogren, B. T. Feng, K. L. Bouman, and C. Daraio, "Gaussian process regression as a surrogate model for the computation of dispersion relations," Comput. Methods Appl. Mech. Eng. **420**, 116661 (2024).

[9] Q. Chen, J. Cao, W. Lin, S. Zhu, and S. Wang, "Predicting dynamic responses of continuous deformable bodies: A graph-based learning approach," Comput. Methods Appl. Mech. Eng. **420**, 116669 (2024).

[10] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Phys. Rev. Lett. **98**, 146401–146404 (2007).

[11] J. Behler, "Representing potential energy surfaces by high-dimensional neural network potentials," J. Phys.: Condens. Matter **26**, 183001 (2014).

[12] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, "Machine learning a general-purpose interatomic potential for silicon," Phys. Rev. X **8**, 041048 (2018).

[13] L. Fulton, V. Modi, D. Duvenaud, D. I. Levin, and A. Jacobson, "Latent-space dynamics for reduced deformable simulation," Comput. Graphics Forum **38**, 379–391 (2019).

[14] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys. **378**, 686–707 (2019).

[15] X. Zhang and K. Garikipati, "Machine learning materials physics: Multi-resolution neural networks learn the free energy and nonlinear elastic response of evolving microstructures," Comput. Methods Appl. Mech. Eng. **372**, 113362–113424 (2020).

[16] K. Yang, Y. Cao, Y. Zhang, S. Fan, M. Tang, D. Aberg, B. Sadigh, and F. Zhou, "Self-supervised learning and prediction of microstructure evolution with convolutional recurrent neural networks," Patterns **2**, 100243 (2021).

[17] D. Lanzoni, M. Albani, R. Bergamaschini, and F. Montalenti, "Morphological evolution via surface diffusion learned by convolutional, recurrent neural networks: Extrapolation and prediction uncertainty," Phys. Rev. Mater. **6**, 103801 (2022).

[18] A. M. Roy, R. Bose, V. Sundararaghavan, and R. Arróyave, "Deep learning-accelerated computational framework based on physics informed neural network for the solution of linear elasticity," Neural Networks **162**, 472–489 (2023).

[19] A. Madanchi, M. Kilgour, F. Zysk, T. D. Kühne, and L. Simine, "Simulations of disordered matter in 3D with the morphological autoregressive protocol (MAP) and convolutional neural networks," J. Chem. Phys. **160**, 024101 (2024).

[20] A. Prakash and S. Sandfeld, "Automated analysis of continuum fields from atomistic simulations using statistical machine learning," Adv. Eng. Mater. **24**, 2200574 (2022).

[21] V. Shchukin, N. Ledentsov, and D. Bimberg, "Epitaxy of nanostructures," *NanoScience and Technology* (Springer, Berlin, Heidelberg, 2004).

[22] L. B. Freund and S. Suresh, *Thin Film Materials: Stress, Defect Formation and Surface Evolution* (Cambridge University Press, Cambridge, 2004).

[23] J. Evans, P. Thiel, and M. C. Bartelt, "Morphological evolution during epitaxial thin film growth: Formation of 2D islands and 3D mounds," Surf. Sci. Rep. **61**, 1–128 (2006).

[24] Y.-P. Zhao, G.-C. Wang, T.-M. Lu, G. Palasantzas, and J. T. M. De Hosson, "Surface-roughness effect on capacitance and leakage current of an insulating film," Phys. Rev. B **60**, 9157–9164 (1999).

[25] F. Rovaris, M. H. Zoellner, P. Zaumseil, A. Marzegalli, L. Di Gaspare, M. De Seta, T. Schroeder, P. Storck, G. Schwalb, G. Capellini, and F. Montalenti, "Dynamics of crosshatch patterns in heteroepitaxy," Phys. Rev. B **100**, 085307 (2019).

[26] G. Medeiros-Ribeiro, A. M. Bratkovski, T. I. Kamins, D. A. A. Ohlberg, and R. S. Williams, "Shape transition of germanium nanocrystals on a silicon (001) surface from pyramids to domes," Science **279**, 353–355 (1998).

[27] J. J. Zhang, F. Montalenti, A. Rastelli, N. Hrauda, D. Scopece, H. Groiss, J. Stangl, F. Pezzoli, F. Schäffler, O. G. Schmidt, L. Miglio, and G. Bauer, "Collective shape oscillations of SiGe islands on pit-patterned Si(001) substrates: A coherent-growth strategy enabled by self-regulated intermixing," Phys. Rev. Lett. **105**, 166102 (2010).

[28] F. Leroy, L. Borowik, F. Cheynis, Y. Almadori, S. Curiotto, M. Trautmann, J. C. Barbé, and P. Müller, "How to control solid state dewetting: A short review," Surf. Sci. Rep. **71**, 391–409 (2016).

[29] M. Albani, L. Ghisalberti, R. Bergamaschini, M. Friedl, M. Salvalaglio, A. Voigt, F. Montalenti, G. Tütüncüoglu, A. Fontcuberta i Morral, and L. Miglio, "Growth kinetics and morphological analysis of homoepitaxial GaAs fins by theory and experiment," Phys. Rev. Mater. **2**, 093404 (2018).

[30] R. Bergamaschini, M. Salvalaglio, R. Backofen, A. Voigt, and F. Montalenti, "Continuum modelling of semiconductor heteroepitaxy: An applied perspective," Adv. Phys.: X **1**, 331–367 (2016).

[31] W. W. Mullins, "Theory of thermal grooving," J. Appl. Phys. **28**, 333 (1957).

[32] R. J. Asaro and W. A. Tiller, "Interface morphology development during stress corrosion cracking: Part I. Via surface diffusion," Metall. Trans. **3**, 1789–1796 (1972).

[33] M. Grinfeld, "Instability of the separation boundary between a nonhydrostatically stressed elastic body and a melt," Sov. Phys.-Dokl. **31**, 831 (1986).

[34] M. Grinfeld, "The stress driven instability in elastic crystals: Mathematical models and physical manifestations," J. Nonlinear Sci. **3**, 35–83 (1993).

[35] D. Srolovitz, "On the stability of surfaces of stressed solids," Acta Metall. **37**, 621–625 (1989).

[36] B. J. Spencer, P. W. Voorhees, and J. Tersoff, "Morphological instability theory for strained alloy film growth: The effect of compositional stresses and species-dependent surface mobilities on ripple formation during epitaxial film deposition," Phys. Rev. B **64**, 235318 (2001).

[37] R. Bergamaschini, J. Tersoff, Y. Tu, J. J. Zhang, G. Bauer, and F. Montalenti, "Anomalous smoothing preceding island formation during growth on patterned substrates," Phys. Rev. Lett. **109**, 156101 (2012).

[38] F. Jonsdottir and L. Freund, "Equilibrium surface roughness of a strained epitaxial film due to surface diffusion induced by interface misfit dislocations," Mech. Mater. **20**, 337–349 (1995).

[39] F. Rovaris, R. Bergamaschini, and F. Montalenti, "Modeling the competition between elastic and plastic relaxation in semiconductor heteroepitaxy: From cyclic growth to flat films," Phys. Rev. B **94**, 205304 (2016).

[40] S. Lucarini, M. V. Upadhyay, and J. Segurado, "FFT based approaches in micromechanics: Fundamentals, methods and applications," Modell. Simul. Mater. Sci. Eng. **30**, 023002 (2021).

[41] Y. Zhang and A. Bower, "Numerical simulations of island formation in a coherent strained epitaxial thin film system," J. Mech. Phys. Solids **47**, 2273–2297 (1999).

[42] S. Wise, J. Lowengrub, J. Kim, and W. Johnson, "Efficient phase-field simulation of quantum dot formation in a strained heteroepitaxial film," Superlattices Microstruct. **36**, 293–304 (2004).

[43] A. Rätz, A. Ribalta, and A. Voigt, "Surface evolution of elastically stressed films under deposition by a diffuse interface model," J. Comput. Phys. **214**, 187–208 (2006).

[44] X. Zhang, Y. Wang, and W. Cai, "Anisotropy effect on strain-induced instability during growth of heteroepitaxial films," J. Mater. Sci. **53**, 5777–5785 (2018).

[45] J.-N. Aqua, T. Frisch, and A. Verga, "Nonlinear evolution of a morphological instability in a strained epitaxial film," Phys. Rev. B **76**, 165319 (2007).

[46] Y. Le Cun, L. Jackel, B. Boser, H. Graf, I. Guyon, D. Henderson, R. Howard, and W. Hubbard, "Handwritten digit recognition: Applications of neural network chips and automatic learning," IEEE Commun. Mag. **27**, 41–46 (1989).

[47] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Comput. **1**, 541–551 (1989).

[48]Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," IEEE Trans. Neural Networks Learn. Syst. **33**, 6999–7019 (2022).

[49]J. D. Eshelby, "The determination of the elastic field of an ellipsoidal inclusion, and related problems," Proc. R. Soc. London, Ser. A **241**, 376–396 (1957).

[50]T. Mura, *Micromechanics of Defects in Solids* (Springer, Dordrecht, 2012).

[51]L. Miglio and F. Montalenti, "Modeling the evolution of germanium islands on silicon(001) thin films," in *Silicon-Germanium (SiGe) Nanostructures: Production, Properties and Applications in Electronics, Woodhead Publishing in Materials*, edited by Y. Shiraki and N. Usami (Woodhead Publication LTD, Cambridge, England, 2011), pp. 211–246.

[52]G.-H. Lu and F. Liu, "Towards quantitative understanding of formation and stability of Ge hut islands on Si(001)," Phys. Rev. Lett. **94**, 176103 (2005).

[53]L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Butterworth–Heinemann Elsevier, 1986).

[54]D. Lanzoni, F. Rovaris, L. Martín-Encinar, A. Fantasia, R. Bergamaschini, and F. Montalenti, "A FEM dataset of Ge film profiles and elastic energies for machine learning approximation of strain state and morphological evolution," Mater. Cloud Arch. **59**, 2024 (2024).

[55]K. Perlin, "An image synthesizer," ACM SIGGRAPH Comput. Graphics **19**, 287–296 (1985).

[56]Python implementation for Perlin noise https://pypi.org/project/perlin-noise/ (2023).

[57]E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE Trans. Pattern Anal. Mach. Intell. **39**, 640–651 (2017).

[58]A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," arXiv:1912.01703 [cs.LG] (2019).

[59]S. Schubert, P. Neubert, J. Pöschmann, and P. Protzel, "Circular convolutional neural networks for panoramic images and laser data," in *2019 IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2019), pp. 653–660.

[60]O. Semih Kayhan and J. C. van Gemert, "On translation invariance in CNNs: Convolutional layers can exploit absolute spatial location," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2020), pp. 14262–14273.

[61]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 [cs.LG] (2017).

[62]L. Martín-Encinar, D. Lanzoni, A. Fantasia, F. Rovaris, R. Bergamaschini, and F. Montalenti, "Quantitative analysis of the prediction performance of a convolutional neural network evaluating the surface elastic energy of a strained film," arXiv:2405.03049 [physics.comp-ph] (2024).

29 August 2024 13:17:17