Uncovering atomistic mechanisms of crystallization using Machine Learning

Rodrigo Freitas* and Evan J. Reed
Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA
(Dated: September 16, 2019)

The process of solidification has been used through centuries to create materials with specific shapes and compositions, since much before microscopy was able to reveal the microstructure of materials. With the maturation of computational and microscopy techniques, a mechanism-based understanding of the crystallization process was developed. Yet, current methods fall short of delivering full atomistic understanding of solidification kinetics due to difficulties in capturing and classifying the complex atomic-scale processes at play. Here, we present an approach to uncover the fundamental atomistic events leading to silicon crystallization. Such events are classified according to the microstructure surrounding the crystallizing atoms using a Machine Learning (ML) algorithm to optimally categorize the local atomic structure. We discover that this approach reveals a local-structure dependent kinetic model for crystallization with predictive capabilities. From the model we also draw new insights on the role of interface-induced ordering of the liquid on the kinetics of crystallization.

Crystallization from the melt is a pervasive process in industry, from metal casting for structural applications to the Czochralski process for semiconductor wafer growth for electronics. It is important to control and understand the solidification process because it is at this stage that the material’s microstructure morphology is created, which in turn defines the material’s properties. Consequently, a great deal of effort has been put into understanding the complex interplay between thermodynamics and kinetics that governs the process of crystal growth.

The kinetics of crystallization at the solid-liquid interface is notoriously difficult to observe experimentally because the crystal interface to which atoms attach is buried under the melt. Moreover, the short timescale of atomic rearrangements leading to crystal growth makes these events difficult to capture. Computer simulations have been able to bypass many of these difficulties, effectively giving us direct access to the crystal growth kinetics – atom by atom and femtosecond by femtosecond. Nevertheless, with this abundance of details also comes the additional complication of sifting through the data in order to uncover relevant physical information. Historically, this problem has been circumvented by carefully designing the simulation geometry as to isolate certain microstructural elements, which are then probed separately. For example, the simulation geometry might be such as to only expose one specific crystal surface orientation. This approach makes the measurement of certain physical quantities straightforward (e.g. the growth rate of the exposed crystal surface), but it lacks the promise of full atomistic understanding of the solidification process because the interaction with other potentially important – or even dominant – microstructural elements is not included (e.g. spiral growth due to the presence of dislocations).

Here, we perform Molecular Dynamics (MD) simulations of silicon growth from its melt employing a simulation geometry akin to laboratory experiments of crystal growth: a crystalline seed is introduced in the liquid and its growth is monitored over the course of the simulation (see fig. 1a and Supplementary Video 1). This setup allows the different microstructural elements of the growing crystal to interact naturally (see Supplementary Videos 2 and 3), as they would in a crystal growth experiment. But, typically, it also greatly diminishes the amount of information that can be inferred from the simulation due to the lack of a crystal growth model that accounts for all microstructural elements present and their respective interactions. Moreover, it also becomes challenging to decipher the atomic events at play due to the sheer complexity of the environment they are embedded in. In this letter we show that these disadvantages can be overcome by employing ML algorithms to systematically encode and classify the microstructure surrounding the crystallizing atoms. Our approach builds on recently-proposed ML strategies for construction of a structural quantity (namely softness $S$) that captures the propensity for atomic rearrangements to occur in disordered atomic environments, such as in glasses and inside grain boundaries.

The structural characterization is realized by assigning to each atom $i$ a local-structure fingerprint $x_i$ constructed from a set of 21 radial structure functions $G(r)$, as illustrated in fig. 1b. Furthermore, atoms are labeled into three possible categories according to their first-neighbor’s arrangement: “liquid” and “crystal” atoms have arrangement patterns statistically identical to the bulk liquid and bulk crystal, respectively. Meanwhile, “crystallizing” atoms have arrangement patterns intermediary between the other two groups (see Methods and Supplementary Information for more details). It is possible to observe how these three groups of atoms are spread in the $\mathbb{R}^{21}$-space of local-structure fingerprints $x$, with

* freitas@stanford.edu
FIG. 1. a) Snapshots of a crystal growth simulation using Molecular Dynamics (see Supplementary Video 1 for the complete video and Supplementary Video 3 for the dislocation dynamics). The system initially contains a small crystalline seed (yellow atoms) surrounded by liquid (transparent blue atoms). Shown in the last frame is the dislocation network\(^{12,13}\) formed during the growth process (edge dislocations are colored blue while screw dislocations are shown in red). b) The local structure (atoms in yellow) surrounding a central atom (green) is encoded using a set of 21 radial structure functions \(G(r_n)\) evaluated at different radial distances \(r_n\) with \(n = 1, 2, 3, \ldots, 21\). Together, these functions comprise the atom’s local-structure fingerprint, denoted as \(x_i\) for the \(i\)th atom. c) Atomic trajectory during solidification as encoded by the time-evolution of \(x_i\). The \(\mathbb{R}^{21}\)-space was represented in two dimensions using the first two components of the Principal Component Analysis (PCA) method. d) Cross section of the initial stages of a Molecular Dynamics simulation of silicon growth. Each liquid atom is colored according to its softness (\(S\)) value and atoms in the crystalline phase are colored in gray.

Atoms assume varied local-structure fingerprints \(x_i\) depending on the surrounding microstructure at the time of crystallization. In order to quantify these variations in structure we proceed as follows. First, a ML algorithm known as Support Vector Machine\(^{15-17}\) (SVM) is employed to find the hyperplane that optimally separates the crystallizing atoms from the liquid atoms in the \(\mathbb{R}^{21}\)-space of \(x_i\). Then, the distance of each atom \(i\) from the hyperplane (\(S_i\), known as softness in other contexts\(^{9-11}\)) is measured: atoms with \(S_i > 0\) lie on the “crystallizing” side of the hyperplane, while \(S_i < 0\) atoms lie on the “liquid” side. This approach is found to correctly classify liquid and crystallizing atoms with an accuracy of 96%.

Shown in fig. 1d is a simulation snapshot with atoms colored according to their softness value (see also Supplementary Video 4 and 5). In this figure, \(S\) is seen to assume high values close to the crystal interface, decreasing quickly with distance. This suggests that the crystal surface affects the structure of the nearest liquid layers by imparting some amount of order\(^{18-20}\). Moreover, fig. 1d also indicates that \(S\) captures signs of dynamical heterogeneity in the supercooled liquid far from the crystal interface, with clear indications of strong spatial correlations. These fluctuating heterogeneities have recently been shown to be preferential sites for ice nucleation in water\(^{21,22}\), a liquid with a multitude of physical properties in common with silicon\(^{23}\).

The crystal growth rate measured in the MD simulations is shown as function of the interface temperature in fig. 2a. Notice that although each temperature in fig. 2a...
is characterized by a unique value of growth rate, it is possible to decompose this single value as a function of the local structure using $S$. For example, in fig. 2b it can be seen that the growth rate for $T \approx 1233$ K varies by almost four orders of magnitude as the local structure (as encoded by $S$) changes.

Qualitative understanding of the kinetics of crystal growth is given by the Wilson-Frenkel (WF) model:

$$r(T) = k(T) \left\{ 1 - \exp \left[ -\beta \Delta \mu(T) \right] \right\},$$

where $r(T)$ is the growth rate, $k(T) = k_0 \exp(-\beta \Delta E_a)$ is the kinetic factor, $\Delta E_a$ is the activation energy for solidification, and $\Delta \mu(T)$ is the difference in chemical potential between the liquid and solid phases. The free-energy landscape of this model is illustrated in fig. 2c. In the WF model the activation energy for solidification is taken as the energy barrier for diffusion in the liquid, $\Delta E_a = \Delta E_d$, because crystallizing atoms must undergo the same self-diffusion process that occurs in the associated liquid phase. For the geometry employed in our calculations (fig. 1a) we have $\Delta \mu(T) = \Delta G(T) - \kappa \gamma / \rho_s$, where $\Delta G(T)$ is the difference in free energy (per atom) between the two phases and the second term is due to the Gibbs-Thomson effect, with $\rho_s$ being the density of the solid, $\gamma$ the interfacial free energy, and $\kappa = 2/r_{\text{eff}}$ is a geometrical factor where $r_{\text{eff}}$ is the effective crystallite radius (see Methods). Shown in fig. 2d and 2e are the input parameters for WF the model: $\Delta E_d$ and $\Delta G(T)$ (see Methods for details on the calculation of these parameters). Comparison of the WF model predictions against the MD simulation results can be seen in fig. 2a. This notorious discrepancy between the WF model and simulation (or experimental) results, while largely unsolved, has been attributed to changes in mobility of the supercooled liquid in the vicinity of the crystal interface. For this reason, we propose to address the deficiencies of this model by taking into account the local structure surrounding the crystallizing atoms through an explicit dependence on $S$:

$$r(T, S) = k(T, S) \left\{ 1 - \exp \left[ -\beta \Delta \mu(T, S) \right] \right\},$$

with $\Delta \mu(T, S) = \Delta G(T) - \kappa \gamma (S) / \rho_s$. Indeed, including the information about the local structure present in plots such as fig. 2b results in a model with much better de-
FIG. 3. a) Arrhenius plot of the crystallization kinetic factor for different local structures, as encoded by softness ($S$). The Arrhenius dependence on temperature suggests that each value of $S$ can be interpreted as a thermally activated and independent channel for crystallization with a well-defined energy barrier. Notice that the glass transition temperature, $T_m/T_g \approx 1.68^{25}$, is beyond the temperature range of the figure. b) Dependence of the activation energy for solidification on the local structure. The energy barrier for liquid diffusion is found to be lower than $\Delta E_a(S)$ for all values of $S$. The experimental results ($\Delta E_a^{\text{exp}}$) from ref. 26 were obtained assuming a single activation energy. c) Arrhenius prefactor dependence on $S$. The decrease of about four orders of magnitude with $S$ implies that there are less rearrangement pathways leading liquid atoms to the activated crystallizing state as $S$ increases.

We now turn to investigate the ramifications of the local-structure dependent model $r(T, S)$. The kinetic factor $k(T, S)$, shown in fig. 3a, is a strong function of the local structure, varying by as much as three orders of magnitude with $S$. For each value of $S$ the kinetic factor has an Arrhenius-like temperature dependence $k(T, S) = k_0(S) \exp[-\beta \Delta E_a(S)]$. This striking outcome suggests that each value of $S$ corresponds to a thermally-activated and independent crystallization channel with well-characterized energy barrier. The dependence of the activation energy barrier of these channels with the local structure is displayed in fig. 3b, where it is seen that $\Delta E_a(S)$ varies over 1 eV with $S$. Additionally, the activation energy decreases monotonically with $S$ and seems to approach the energy barrier for diffusion $\Delta E_d$ (fig. 2d) asymptotically. Varying $S$ also has a pronounced effect on the Arrhenius prefactor $k_0(S)$, as indicated in fig. 3c, which decreases by over three orders of magnitude with $S$. Because $\log[k_0(S)]$ can be interpreted as the entropic contribution to the free energy barrier, the observed decrease in the prefactor indicates that there are less rearrangement pathways leading liquid atoms to the activated state (i.e. to crystallization) as $S$ increases. Hence, figs. 3b and 3c together indicate that from all observed local-structure arrangements surrounding crystallizing atoms, only very few lead to low energy barriers.

Next, we examine the solid-liquid interface free energy as a function of the local structure, shown in fig. 4a, as obtained from the Gibbs-Thomson term in eq. (1). The interface free energy decreases monotonically with $S$, starting at large values – corresponding to rough interfaces – and reaching interfacial free energy values characteristic
FIG. 4. a) Solid-liquid interface free energy dependence on the local structure as encoded by softness $S$. Large values of $S$ have interface free energies characteristic of faceted solid-liquid interfaces in silicon (illustrated in the inset). Similarly, negative values of $S$ are found to be associated with rough interfaces. Interfacial free energies $\gamma_{(100)}$, $\gamma_{(101)}$, $\gamma_{(111)}$ were obtained from ref. 27. b) Distribution of interfaces to which the crystallizing atoms attach, showing a strong preference for (111) vicinals. c) Step-step separation distances for steps on (111) surfaces. d) Illustration of how interface-induced ordering of the liquid alters the local structure around crystallizing atoms and affects the activation energy for solidification. The crystallizing atom (green) has its local structure (illustrated here only by its first neighbors) affected by the nearby crystal surface. This effect is anisotropic: atomically rough surfaces interact strongly with the liquid and cause significant ordering of the liquid, which becomes rigid, resulting in large activation energies $\Delta E_a$ when compared to the barrier for diffusion in the liquid $\Delta E_d$. e) Atomically smooth surfaces interact weakly with the liquid and cause very small ordering, resulting in low $\Delta E_a$. This finding implies that the decrease in Arrhenius prefactor $k_0(S)$ with softness (fig. 3c) leads to fewer rearrangement pathways because crystallization events with large positive values of $S$ happen at vicinal surfaces, which naturally offer less crystallization sites than atomically rough interfaces. Nevertheless, according to fig. 2b most crystallization events occur at these low-$\gamma(S)$ interfaces, with 70% of them taking place at interfaces with $S \geq 0.75$. Indeed, direct measurement of the distribution of crystal interfaces to which crystallizing atoms attach (fig. 4b) reveals strong preferential attachment to a wide variety of (111) vicinals. The high-intensity spot around (435) corresponds to step-step separation distances from 15 Å to 24 Å (fig. 4c), indicating that the majority of the crystallization events take place on vicinal surfaces with well-separated steps.

The crystal interface affects the adjacent liquid layers by imparting some amount of order on them, as seen in fig. 1d. This ordering, in turn, affects the kinetics of liquid atoms through changes in the activation barrier $\Delta E_a$. Comparison of figs. 3b and 4a reveals that the interface-induced ordering of the liquid is anisotropic, i.e., it depends on the crystal surface orien-
tation. The trend (illustrated in figs. 4d and 4e) is such that faceted surfaces (large positive $S$ in fig. 4a) cause weak ordering, resulting in activation energies $\Delta E_a(S)$ close to the energy barrier for atomic diffusion in the liquid bulk ($\Delta E_d$). Meanwhile, atomically rough surfaces (negative $S$ in fig. 4a) cause strong ordering of the liquid, which becomes rigid and results in activation energies much larger than $\Delta E_d$. But, even in the case of strong ordering the activation energy ($\approx 1.75$ eV) is still much smaller than the $\approx 4.6$ eV$^{29}$ barrier for vacancy-mediated self-diffusion in crystalline silicon. This indicates that interface-induced ordering is nowhere as substantial as crystalline order.

The observed anisotropy of the interface-induced ordering emerges from the fact that the intensity of the ordering is proportional to the amount of dangling bonds of crystal surface atoms, which is always lower for faceted surfaces when compared to rough surfaces (figs. 4d and 4e). Dynamical heterogeneities present in the liquid (fig. 1d) also affect the coordination of atoms$^{21,30,31}$ For this reason, it is reasonable to expect that they contribute to the $\approx 1$ eV dispersion in $\Delta E_a(S)$ observed in fig. 3b. Nonetheless, there is no evident reason to believe that their effect is anisotropic since dynamical heterogeneities have origin in random thermal fluctuations.

In conclusion, we have presented a predictive kinetic model for crystal growth with an explicit dependence on the local structure surrounding crystallizing atoms. The local-structure dependence was uncovered using ML algorithms to optimally inform the encoding of structural variations in a scalar quantity called softness ($S$). We have shown that different values of $S$ correspond to independent channels for the thermally-activated crystallization of atoms in regions with different microstructures (i.e. crystal surfaces orientation and surface defects). Moreover, with this novel approach we were able to obtain quantities of practical importance in materials science – such as activation energies, interface free energies, and growth rates – and derive new insights on the role of anisotropic interface-induced ordering of the liquid on the kinetics of crystal growth. In this letter we have employed ML tools as an aid to augment human intuition, rather than a substitute thereof. This innovative application of ML in materials science blends conventional scientific methods with data science tools to produce physically-consistent models and novel conceptual knowledge. Our approach was guided by the physical intuition that local structure is paramount for crystallization events, but without tools specialized in making sense of complexity the difficulties in creating a predictive model that takes into account the entire zoo of microstructural elements present would have been insurmountable.

ACKNOWLEDGMENTS

The authors would like to thanks Dr. Vasily Bula- tov for providing the script for projecting crystal direc-

n.
Methods

Crystal growth simulations. The MD simulations were performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software, with the interactions between silicon atoms described by the Stillinger-Weber\textsuperscript{33} interatomic potential. The timestep was selected as approximately $1/56$th of the period of the highest-frequency phonon mode of this system, or $\Delta t = 1$ fs. The crystal growth simulations contained 500,000 atoms and were initialized with a spherical crystalline seed of approximately 3,000 atoms in the diamond cubic structure. The lattice parameter for the atoms in the crystal seed was chosen taking into account dilation due to thermal expansion, then the remainder of the simulation cell was filled with randomly distributed atoms at the equilibrium liquid density for that temperature at zero pressure.

The system was equilibrated by first relaxing the liquid atoms using a Conjugate Gradient\textsuperscript{34} algorithm for 200 steps. Next, the liquid was equilibrated at finite temperature using the Bussi-Donadio-Parrinello\textsuperscript{35} (BDP) thermostat for 3 ps with a damping parameter of 0.1 ps. Finally, the liquid atoms were equilibrated for 2 ps at zero pressure and finite temperature using the same thermostat just described and a Nosé-Hoover barostat\textsuperscript{36-40} (with damping parameter of 1 ps) allowing only for isotropic dilation/contraction of the system. During the entirety of this equilibration process the crystalline seed atoms were kept frozen at their equilibrium crystal structure with fixed lattice parameters (i.e. they did not dilate/contract with the liquid atoms). After equilibration a BDP thermostat and Nosé-Hoover barostat were applied to the system, both with damping parameter of 1.0 ps, to maintain the system at finite temperature and zero pressure for a total of 3 ns during which snapshots were recorded every 1 ps. The crystal growth process can be seen in the Supplementary Video 1. Snapshots saved from the MD simulations were subsequently relaxed using 20 steps of the Steepest-Descent\textsuperscript{34} algorithm. The crystal growth simulation were performed at temperatures ranging from 1200 K to 1500 K in intervals of 25 K. The error bars in fig. 1 represent the 95% confidence interval as computed using the bootstrap method with 1000 samples of the same size as the original distribution.

The damping parameter for the thermostat was selected conservatively such that the liquid diffusivity was not affected by the presence of the thermostat, i.e. it had the same value within the statistical uncertainty as the diffusivity computed without a thermostat. Thus, the thermostating of the crystal growth simulation was performed gently as to not affect the kinetics of the system. See Supplementary Information sec. 6 for more details about the diffusivity calculations.

Phase identification. In order to identify to which phase (liquid or crystal) each particle belongs to, we used the order parameter introduced by Rein ten Wolde, Ruiz-Montero, and Frenkel\textsuperscript{41}. The complete description and analysis of the construction of this order parameter can be found in the Supplementary Information sec. 2. Ultimately, this method provides us with a parameter $\alpha_i(t)$ for each atom $i$ at the time $t$ of each MD snapshot. The physical interpretation of this parameter is that $\alpha_i$ is the fraction of bonds that atom $i$ makes that resemble bonds in a perfect crystal structure. As shown in the Supplementary Information sec. 2.1, the parameter $\alpha_i$ correctly identify atoms in the perfect crystal or bulk liquid with accuracy of 100% within the statistical uncertainty. It is important to notice that although $\alpha_i$ can discern between liquid and crystal atoms, it does not differentiate between crystalline structures. We confirm that the silicon atoms are indeed crystallizing in the diamond cubic structure by performing the Polyhedral Template Matching\textsuperscript{13,42} analysis.

Encoding atomic dynamics (ML labeling). The dynamics of each atom was encoded using the time evolution of the $\alpha_i(t)$ order parameter. A representative plot\textsuperscript{43} of $\alpha_i(t)$ is shown in figs. S4b and S4c. Notice that due to thermal fluctuations the instantaneous value of $\alpha_i(t)$ for atoms in the liquid and crystal phases might differ from their perfect values of 0.0 and 1.0 respectively, even after the short Steepest-Descent relaxation. Hence, we perform a moving-window average of $\alpha_i(t)$ with window length of 20 ps and use the window-averaged $\bar{\alpha}_i(t)$ to label the atomic dynamics as illustrated in fig. S4a. Atoms with $\bar{\alpha}_i(t) = 0$ for $t \in [t_0 - \tau, t_0 + \tau]$ receive label $y_i = -1$ at time $t_0$. Thus, these are atoms deep in the liquid phase that will not be transitioning to the crystal state in the near future, neither have tried to transition in the near past. From the analysis of curves such as in fig. S4c we choose $\tau = 15$ ps as a reasonable value. Next we identify atoms that have just started to move out of the bulk liquid (i.e. crystalizing atoms) as those within a 20 ps window from the point where $\bar{\alpha}_i = 0.25$, i.e. $y_i = 1$ for $t \in [t_0 - \tau, t_0 + \tau]$ where $\bar{\alpha}_i(t_0) = 0.25$ and $\tau = 10$ ps. See Supplementary Information sec. 2 for more details on the labeling process.

Local-structure fingerprint (ML features). The local structure surrounding each atom was characterized using a set of radial structure functions\textsuperscript{6,14}:

$$G_i(r) = \sum_{j=1}^{n(i)} \exp \left[-\frac{(r_{ij} - r)^2}{2\sigma^2}\right],$$

where $i$ is the atom whose local structure is being described, $n(i)$ is the number neighbors of $i$ within a cutoff radius $r_{\text{cut}}$, $r_{ij}$ is the distance between atom $i$ and one of its neighbors $j$, $r$ and $\sigma$ are two parameters that define the radial structure function. These smoothly varying functions of $r$ count the number of neighbors of $i$ at a distance $r$. In this interpretation, parameter $r$ represents the radial distance from $i$ at which we are counting the number of neighbors, while $\sigma$ adjusts how smoothly
the function varies as atoms move in and out of the distance vicinity. We have used a grid-search\textsuperscript{44} algorithm to perform the hyperparameters tuning (see Supplementary Information sec. 1 for more details), resulting in \( \sigma = 0.5\,\text{Å} \), \( r_{\text{cut}} = 10.8\,\text{Å} \), and \( r_n = (2.0 + 0.4n)\,\text{Å} \), with \( n = 0, 1, 2, \ldots, 20 \). With this set of 21 radial structure functions – one for each value of \( r \) – the local-structure fingerprint of each atom \( i \) was built as a vector:

\[
x_i = \left[ G_i(r_1), G_i(r_2), \ldots, G_i(r_{21}) \right].
\]

**Softness calculation.** The data was assembled by pairing the dynamic labels \( y_i \) with their corresponding structural fingerprint \( x_i \). Then, 10,000 \((y_i, x_i)\) pairs were randomly selected and equally divided between the \( y = -1 \) and \( y = 1 \) classes to train a SVM\textsuperscript{15–17} classifier. The SVM algorithm finds the hyperplane of the form \( \mathbf{w} \cdot \mathbf{x} - b = 0 \) that optimally separates the two classes, where \( \mathbf{w} \) and \( b \) are the parameters that define this hyperplane. Before training the SVM classifier the elements of the fingerprints were standardized\textsuperscript{44} to have zero mean and standard deviation of one. The optimal hyperplane found, denoted by the parameters \( \mathbf{w}^* \) and \( b^* \), correctly separates the two classes with an accuracy of 96%. See Supplementary Information sec. 1 for more details about how these parameters are found and an in-depth analysis of the quality of the classifier computed. The results shown in the letter were obtained using data from the crystal growth simulation at \( T = 1500\,\text{K} \) to train the SVM classifier. But, all results presented here are reproduced within the statistical uncertainty when training at any other temperature, as shown in the Supplementary Information sec. 3.

Once the SVM classifier has been trained it was applied to the entire data set, composed of roughly 27.5 million data points (excluding the data points used for training, hyperparameter tuning, and cross validation, see Supplementary Information sec. 1 for details). The value of softness\textsuperscript{10} for each data point (or atom) is the signed distance from the hyperplane, or \( \tilde{S}_i = \mathbf{w}^* \cdot \mathbf{x}_i - b^* \) for each atom \( i \).

**Principal Component Analysis.** Figure 1c was obtained by applying the PCA\textsuperscript{44} to a data set containing equal amounts of crystallizing, liquid, and crystal, for a total of 60,000 data points. Crystal atoms were defined as those for which \( \tilde{\alpha}_i(t) = 1.0 \) (bulk crystal atoms) or \( \tilde{\alpha}_i(t) = 0.75 \) (stacking fault atoms) for \( t \in [t_0 - \tau_x, t_0 + \tau_x] \) with \( \tau_x = 15\,\text{ps} \). From the PCA we obtained the components of each data point along the two eigenvectors with largest eigenvalues, which are used to plot fig. 1c. The atom trajectory was obtained by applying the same PCA transformation along a 3 ns atomic trajectory of a crystal growth simulation at 1500 K.

**Growth rate determination.** The number of atoms in the crystallite \( N(t) \) at any given time \( t \) was determined as the number of atoms with \( \tilde{\alpha}_i(t) > 0.25 \). From this information the effective crystallite radius \( r_{\text{eff}}(t) \) (shown in fig. S9a) was computed assuming a spherical shape (which results in \( \kappa = 2/r_{\text{eff}} \), where \( \kappa \) is the geometrical factor in the Gibbs-Thomson term). The growth rate for each temperature, shown in fig. 2a, was determined by a linear fit of \( N(t) \) over the time interval for which \( r_{\text{eff}} \in [80\,\text{Å}, 100\,\text{Å}] \). This interval is such that the crystallite is small enough to not be affected by finite-size effects, but large enough to give the system time to equilibrate into a steady-state growth condition. See Supplementary Information sec. 4.1 for a more detailed analysis.

**Interface temperature.** When studying crystal growth it is important to differentiate between the temperature of the supercooled liquid surrounding the crystal (but far from the interface) from the solid-liquid interface temperature. Under steady-growth conditions these two temperatures will differ because of the latent heat released at the interface and the finite rate of heat transport. Here, the interface temperature was computed by considering only the kinetic energy of atoms with \( \tilde{\alpha}_i \in (0.15, 0.75) \). This interval of \( \tilde{\alpha} \) was selected because it includes both, interfacial liquid and interfacial crystal atoms. Figure S9b shows the interface temperature under steady-state growth as a function of the surrounding liquid bath temperature. See Supplementary Information sec. 4.2 for more details.

**Crystal surface analysis.** Figure 4b was obtained by constructing a polyhedral surface mesh around the crystallite (i.e. atoms with \( \tilde{\alpha}_i(t) > 0.25 \)) using the algorithms in ref. 45 (as implemented in Ovito\textsuperscript{13}) with a probe-sphere radius of 3 Å and a smoothing level of 10. From this mesh the surface directions were inferred and averaged over the time interval for which the crystal growth occurs in a steady state. The data for constructing fig. 4b was obtained by finding the orientation of the closest surface to each crystallizing atom.

The only atomically smooth surfaces in silicon are \{111\} surfaces\textsuperscript{46}, consequently steps can only exist in these surfaces. Hence, the step-step separation distance shown in fig. 4c was computed assuming that \{111\} faceting occurs at all surface orientations\textsuperscript{46}.

**Solid and liquid free energies.** The accurate calculation of the solid and liquid free energies is crucial in crystal growth studies. As shown in the Supplementary Information sec. 5, employing approximations such as the quasi-harmonic approximation results in the under-estimation of the predicted growth rates by as much as 36%. For this reason, we performed the solid and liquid free energies calculations using state-of-the-art nonequilibrium thermodynamic integration methods that make no approximating assumptions on the physical characteristics of system. The crystal free energy was determined using the nonequilibrium Frenkel-Ladd\textsuperscript{47–49} (FL) and the Reversible Scaling\textsuperscript{49,50} (RS) methods, following.
closely the approach described in ref. 49. For both methods a system of 21,952 silicon atoms in the diamond cubic structure was employed. The thermodynamic switching was performed in 200 ps for each direction, before which the system was equilibrated for 20 ps. The FL switching was realized for temperatures ranging from 100 K to 2000 K in intervals of 100 K. For each temperature the switching was repeated in five independent simulations to estimate the statistical uncertainty. Similarly, the RS switching was also repeated five times. The S-shaped function was employed in the FL switching, while the RS switching was also repeated five times. The switching was repeated in five independent simulations to estimate the statistical uncertainty.

Similarly, for steps on the Cu (111) surface, Physical Review B 70, 85413 (2004).

[1] A. A. Chernov, Modern crystallography III: crystal growth, Vol. 36 (Springer Science & Business Media, 2012).

[2] A. Pimpinelli and J. Villain, Physics of crystal growth (Cambridge University Press, 1999).

[3] M. Asta, C. Beckermann, A. Karma, W. Kurz, R. Napolitano, M. Plapp, G. Purdy, M. Rappaz, and R. Trivedi, Solidification microstructures and solid-state parallels: Recent developments, future directions, Acta Materialia 57, 941 (2009).

[4] K. G. Libbrecht, Physical dynamics of ice crystal growth, Annual Review of Materials Research 47, 271 (2017).

[5] W.-K. Burton, N. Cabrera, and F. Frank, The growth of crystals and the equilibrium structure of their surfaces, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 243, 299 (1951).

[6] J. Hoyt and M. Asta, Atomistic computation of liquid diffusivity, solid-liquid interfacial free energy, and kinetic coefficient in Au and Ag, Physical Review B 65, 214106 (2002).

[7] G. Sun, J. Xu, and P. Harrowell, The mechanism of the ultrafast crystal growth of pure metals from their melts, Nature Materials 17, 881 (2018).

[8] R. Freitas, T. Frolov, and M. Asta, Step free energies at faceted solid surfaces: Theory and atomistic calculations for steps on the Cu (111) surface, Physical Review B 95, 155444 (2017).

[9] E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, and A. J. Liu, Identifying structural flow defects in disordered solids using machine-learning methods, Physical Review Letters 114, 108001 (2015).

[10] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, A structural approach to relaxation in glassy liquids, Nature Physics 12, 469 (2016).

[11] T. A. Sharp, S. L. Thomas, E. D. Cubuk, S. S. Schoenholz, D. J. Srolovitz, and A. J. Liu, Machine learning determination of atomic dynamics at grain boundaries, Proceedings of the National Academy of Sciences 115, 10943 (2018).

[12] A. Stukowski, V. V. Bulatov, and A. Arsenlis, Automated identification and indexing of dislocations in crystal interfaces, Modelling and Simulation in Materials Science and Engineering 20, 085007 (2012).

[13] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO—the Open Visualization Tool, Modelling and Simulation in Materials Science and Engineering 18, 015012 (2009), http://ovito.org.

[14] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Physical Review Letters 98, 146401 (2007).

[15] C. Cortes and V. Vapnik, Support-vector networks, Machine Learning 20, 273 (1995).

[16] 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, and E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12, 2825 (2011).

[17] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2, 27:1 (2011), http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[18] W. D. Kaplan and Y. Kaufmann, Structural order in liquids induced by interfaces with crystals, Annual Review of Materials Research 36, 1 (2006).

[19] S. H. Oh, Y. Kaufmann, C. Scheu, W. D. Kaplan, and M. Rühle, Ordered liquid aluminum at the interface with sapphire, Science 310, 661 (2005).

[20] S. H. Oh, M. F. Chisholm, Y. Kaufmann, W. D. Kaplan, W. Luo, M. Rühle, and C. Scheu, Oscillatory mass transport in vapor-liquid-solid growth of sapphire nanowires, Science 330, 489 (2010).

[21] M. Fitzner, G. C. Sosso, S. J. Cox, and A. Michaelides, Ice is born in low-mobility regions of supercooled liquid water, Proceedings of the National Academy of Sciences 116, 2009 (2019).
[22] E. B. Moore and V. Molinero, Structural transformation in supercooled water controls the crystallization rate of ice, Nature 479, 506 (2011).
[23] V. Molinero and E. B. Moore, Water modeled as an intermediate element between carbon and silicon, The Journal of Physical Chemistry B 113, 4008 (2008).
[24] G. Galvin, J. Mayer, and P. Peercy, Solidification kinetics of pulsed laser melted silicon based on thermodynamic considerations, Applied Physics Letters 46, 644 (1985).
[25] A. Hedler, S. L. Klaumünzer, and W. Wesch, Amorphous silicon exhibits a glass transition, Nature Materials 3, 804 (2004).
[26] P. Stolk, A. Polman, and W. Sinke, Experimental test of kinetic theories for heterogeneous freezing in silicon, Physical Review B 47, 5 (1993).
[27] W. Shou and H. Pan, Silicon-wall interfacial free energy via thermodynamics integration, The Journal of Chemical Physics 145, 184702 (2016).
[28] F. Panciera, J. Tersoff, A. D. Gamalski, M. C. Reuter, G. Bussi, D. Donadio, and M. Parrinello, Canonical sampling through velocity rescaling, The Journal of Chemical Physics 126, 044101 (2007).
[29] G. J. Martyna, D. J. Tobias, and M. L. Klein, Constant pressure molecular dynamics algorithms, The Journal of Chemical Physics 101, 4177 (1994).
[30] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, Journal of Computational Physics 117, 1 (1995), http://lammps.sandia.gov.
[31] F. Sciortino, A. Geiger, and H. E. Stanley, Effect of defects on molecular mobility in liquid water, Nature 354, 218 (1991).
[32] S. Sastry and C. A. Angell, Liquid–liquid phase transition in supercooled silicon, Nature Materials 2, 739 (2003).
[33] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, Journal of Computational Physics 117, 1 (1995), http://lammps.sandia.gov.
[34] V. Bulatov and W. Cai, Computer simulations of dislocations, Vol. 3 (Oxford University Press on Demand, 2006).
[35] G. Bussi, D. Donadio, and M. Parrinello, Canonical sampling through velocity rescaling, The Journal of Chemical Physics 126, 014101 (2007).
[36] G. J. Martyna, D. J. Tobias, and M. L. Klein, Constant pressure molecular dynamics algorithms, The Journal of Chemical Physics 101, 4177 (1994).
[37] M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, Journal of Applied Physics 52, 7182 (1981).
[38] M. E. Tuckerman, J. Alejandro, R. López-Rendón, A. L. Jochim, and G. J. Martyna, A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal–isobaric ensemble, Journal of Physics A: Mathematical and General 39, 5629 (2006).
[39] W. Shinoda, M. Shiga, and M. Mikami, Rapid estimation of elastic constants by molecular dynamics simulation under constant stress, Physical Review B 69, 134103 (2004).
[40] M. Tuckerman, Statistical mechanics: theory and molecular simulation (Oxford university press, 2010).
[41] P. Rein ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, Numerical calculation of the rates of crystal nucleation in a Lennard-Jones system at moderate undercooling, The Journal of Chemical Physics 104, 9932 (1996).
[42] P. M. Larsen, S. Schmidt, and J. Schiøtz, Robust structural identification via polyhedral template matching, Modelling and Simulation in Materials Science and Engineering 24, 055007 (2016).
[43] Unless otherwise noted, all figures were produced using Matplotlib56.
[44] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (MIT Press, 2016) http://www.deeplearningbook.org.
[45] A. Stukowski, Computational analysis methods in atomistic modeling of crystals, The Journal of The Minerals, Metals & Materials Society (JOM) 66, 399 (2014).
[46] K. Fujiwara, Crystal growth behaviors of silicon during melt growth processes, International Journal of Photoenergy 2012 (2012).
[47] D. Frenkel and A. J. Ladd, New Monte Carlo method to compute the free energy of arbitrary solids, Application to the fcc and hcp phases of hard spheres, The Journal of Chemical Physics 81, 3188 (1984).
[48] M. de Koning and A. Antonelli, Einstein crystal as a reference system in free energy estimation using adiabatic switching, Physical Review E 53, 465 (1996).
[49] R. Freitas, M. Asta, and M. de Koning, Nonequilibrium free-energy calculation of solids using LAMMPS, Computational Materials Science 112, 333 (2016).
[50] M. de Koning, A. Antonelli, and S. Yip, Optimized free-energy evaluation using a single reversible-scaling simulation, Physical Review Letters 83, 3973 (1999).
[51] T. Schneider and E. Stoll, Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions, Physical Review B 17, 1302 (1978).
[52] M. P. Allen and D. J. Tildesley, Computer simulation of liquids (Oxford university press, 2017).
[53] B. Fultz, Phase Transitions in Materials (Cambridge University Press, 2014).
[54] R. Paula Leite, R. Freitas, R. Azevedo, and M. de Koning, The uhlenbeck-ford model: Exact virial coefficients and application as a reference system in fluid-phase free-energy calculations, The Journal of Chemical Physics 145, 194101 (2016).
[55] R. P. Leite and M. de Koning, Nonequilibrium free-energy calculations of fluids using LAMMPS, Computational Materials Science 159, 316 (2019).
[56] J. D. Hunter, Matplotlib: A 2D graphics environment, Computing in Science & Engineering 9, 90 (2007).