

Local Structural Features Elucidate Crystallization of Complex Structures

Maya M. Martirosyan, Matthew Spellings, Hillary Pan, and Julia Dshemuchadse*



Cite This: *ACS Nano* 2024, 18, 14989–15002



Read Online

ACCESS |



Metrics & More



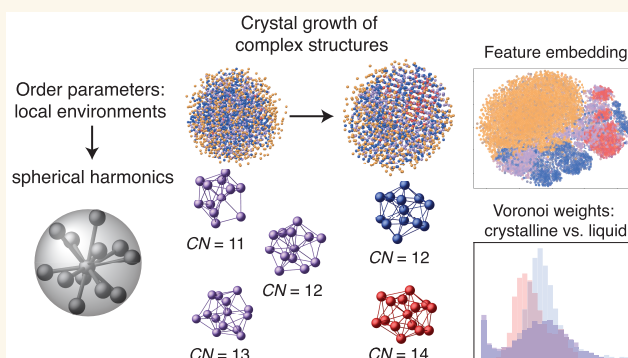
Article Recommendations



Supporting Information

ABSTRACT: Complex crystal structures are composed of multiple local environments, and how this type of order emerges spontaneously during crystal growth has yet to be fully understood. We study crystal growth across various structures and along different crystallization pathways, using self-assembly simulations of identical particles that interact via multiwell isotropic pair potentials. We apply an unsupervised machine learning method to features from bond-orientational order metrics to identify different local motifs present during a given structure's crystallization process. In this manner, we distinguish different crystallographic sites in highly complex structures. Tailoring this order parameter to structures of varying complexity and coordination number, we study the emergence of local order along a multistep crystal growth pathway—from a low-density fluid to a high-density, supercooled amorphous liquid droplet and to a bulk crystal. We find a consistent under-coordination of the liquid relative to the average coordination number in the bulk crystal. We use our order parameter to analyze the geometrically frustrated growth of a Frank–Kasper phase and discover how structural defects compete with the formation of crystallographic sites that are more high-coordinated than the liquid environments. The method presented here for classifying order on a particle-by-particle level has broad applicability to future studies of structural self-assembly and crystal growth, and they can aid in the design of building blocks and for targeting pathways of formation of soft-matter structures.

KEYWORDS: crystal growth, complex structures, machine learning, order parameters, molecular dynamics



INTRODUCTION

Crystal growth and nucleation constitute a rich field studied across a wide range of systems. Crystallization processes span length scales and materials families: from the biological—such as the crystallization of proteins in solution¹ or the assembly of gyroid structures in butterfly wings²—to the physical—on the mesoscale: the organization of colloids,³ nanoparticles,⁴ and block copolymers⁵ into various structures; and on the atomic scale: the formation of quasicrystals⁶ or more ubiquitous substances such as ice.⁷ The emergence of order is a ubiquitous phase transition, yet our understanding of spontaneous phenomena such as crystallization remains incomplete.⁸ Experimental and simulation studies of growth are largely isolated to specific systems,^{9,10} while theoretical models for studying, e.g., layer growth¹¹ or Wilson–Frenkel growth,^{12,13} or methods such as kinetic Monte Carlo¹⁴ make a number of simplifying assumptions that impede the extraction of generalized crystal growth principles for the study of more intricate crystal structures.

The applicability of mechanisms such as particle-by-particle attachment and classical nucleation theory^{15,16} has been

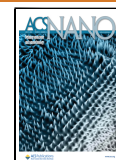
challenged by experimental evidence across various chemistries and length scales, such as the effects of order development in liquid precursor droplets prior to nucleation^{17,18} that may enhance nucleation and growth rates^{19,20} or other multistep crystallization pathways.^{21–24} Nonclassical nucleation and crystallization theory posit that prior to nucleation, an intermediate, metastable liquid phase (or prenucleation cluster) is formed^{25,26} lowering the barrier to nucleation,²⁷ yet a complete understanding of how liquid motifs affect crystallization remains elusive. In the case of complex crystals, growth models and experimentally observed mechanisms do not account for the presence of more than one local environment or a large periodic unit cell in the crystallizing structure. How do identical particles find their specific role to

Received: January 26, 2024

Revised: May 9, 2024

Accepted: May 17, 2024

Published: May 30, 2024



occupy in a structure with multiple local environments,^{28,29} and what role does the liquid play in crystal growth?^{30,31} As such one-component complex structures are being studied on the mesoscale^{32,33} and engineering such materials with DNA-functionalization is becoming an ever more powerful tool for design,^{34–36} achieving a better understanding of the structural transitions and pathways during crystal growth for complex structures is critical for directing self-assembly.

Probing these questions can prove experimentally challenging, whereas simulations targeting growth of complex structures^{37,38} can provide access to rich real-space positional data—the complete trajectories of all particles in the assembly. Systems of identical particles that interact via isotropic, multiwell pair potentials have been shown to spontaneously assemble into a diverse set of crystal structures.^{39–42} Such short-ranged potentials with tunable features are good models for simulating the self-assembly of soft-matter particles, resulting in structures as complex as clathrates³² and Frank–Kasper phases.^{43–45} While conclusions drawn from these pair potentials may not be directly transferable to atomic systems, this study attempts to address open questions regarding the growth of complex structures whose self-assembly cannot be probed with available atomistic interaction potentials.

In this paper, we simulate the crystal growth of model systems across a wide range of chemical interactions, ranging from low-coordinated to high-coordinated structures (with coordination numbers CN = 0–6 and 12–15, respectively) with varying degrees of complexity,^{41,42} as shown in Figure 1.

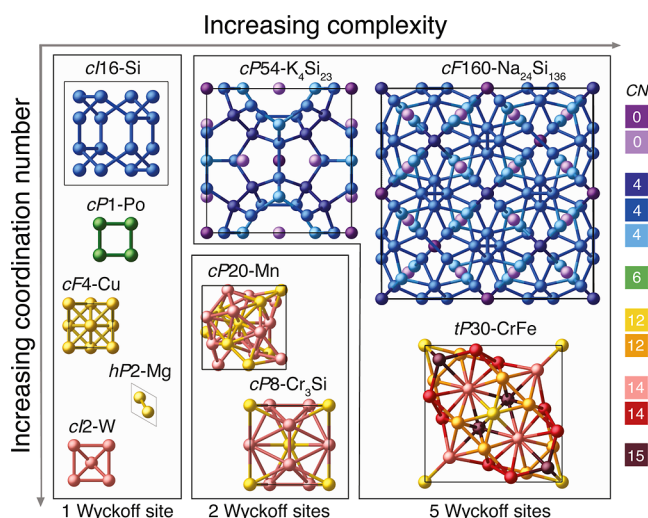


Figure 1. Unit cells of structures simulated in this study, plotted by their complexity (the number of Wyckoff sites) and CN. Color is used to represent the CN of individual particles, ranging from purple (CN = 0) and blue (CN = 4), through green (CN = 6), yellow/orange (CN = 12), to pink/red (CN = 14) and maroon (CN = 15). Average CNs for structures with more than one Wyckoff position are $\langle \text{CN} \rangle = 3.4$ for $cP54\text{-K}_4\text{Si}_{23}$ and $cF160\text{-Na}_{24}\text{Si}_{136}$, $\langle \text{CN} \rangle = 12.8$ for $cP20\text{-Mn}$, $\langle \text{CN} \rangle = 13.5$ for $cP8\text{-Cr}_3\text{Si}$ and $tP30\text{-CrFe}$.

We study crystallization across these structures through the evolution of local order throughout the growth process. To do so, we developed an order metric using unsupervised learning that can distinguish a large number of different local environments and be applied robustly along the self-assembly pathway across various structures. We discuss the interrelation

between the structure of precrystallization liquids and the lowest-coordinated sites in their respective crystalline solids, as well as how the liquid develops the various coordination environments of complex crystal structures in our study. Our exploration of crystal growth highlights the complexity of structural signatures present during self-assembly that depend on the growth pathway as well as the presence of defect motifs that arise when a crystal has multiple local environments.

LOCAL STRUCTURAL METRICS

For quantifying local order in three-dimensional crystal structures, the Steinhardt order parameter⁴⁶

$$Q_l(i) = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^l |q_{lm}(i)|^2 \right]^{1/2},$$

$$\text{where } q_{lm}(i) = \frac{1}{|N(i)|} \sum_{j \in N(i)} Y_{lm}(\theta_{ij}, \phi_{ij})$$

and $N(i) = \{\text{nearest neighbors of particle } i\}$

and its neighbor-averaged variant⁴⁷ have historically been used to fingerprint structural motifs in liquids,⁴⁶ distinguish liquid from crystalline order,³⁸ differentiate between simple sphere packings [body-centered cubic (bcc), cubic-close packed (ccp), and hexagonally close packed (hcp)],⁴⁷ distinguish structures formed in systems of hard shapes,⁴⁸ etc. However, in all of these applications, an informed choice of the value of l —the order of the spherical harmonic Y_{lm} —is required, which selects for the symmetries to which the parameter is sensitive. Moreover, the values of Q_l are highly dependent on the choice of neighborhood cutoff distance,⁴⁹ and its per-particle values are not sufficiently sensitive to distinguish between geometrically similar Wyckoff sites in a highly complex crystal structure.

Unsupervised Machine Learning for Local Environment Identification. More recently, Spellings and Glotzer⁵⁰ created a machine learning-based method for crystal structure identification (i.e., a global order parameter) using local features, which is able to accurately distinguish between similar, complex structures. The featurization is generated using the functional form

$$\bar{Y}_{lm}(i, k) = \frac{1}{k} \left| \sum_{j \in N_k(i)} Y_{lm}(\theta_{ij}, \phi_{ij}) \right|,$$

where $N_k(i) = \{k \text{ nearest neighbors of particle } i\}$

Values of $\bar{Y}_{lm}(i, k)$ are computed for a range of neighborhood sizes $4 \leq k \leq N_{\text{max}}$, and $0 \leq l \leq l_{\text{max}}$ values and $0 \leq m \leq l$, where N_{max} and l_{max} are chosen according to the investigated system. Like Steinhardt's parameter, spherical harmonics are used as a set of basis functions, but a key difference lies in how rotational invariance is established. Rather than summing over values of m , an orientation is generated from the principal axes of inertia, which allows the separation of symmetry elements for each spherical harmonic. Such a featurization is necessarily high-dimensional and requires machine-learning methods for interpretation. Unsupervised learning approaches^{51,52} have shown promise in uncovering phase transitions^{53,54} or learning atomic features from simulations.

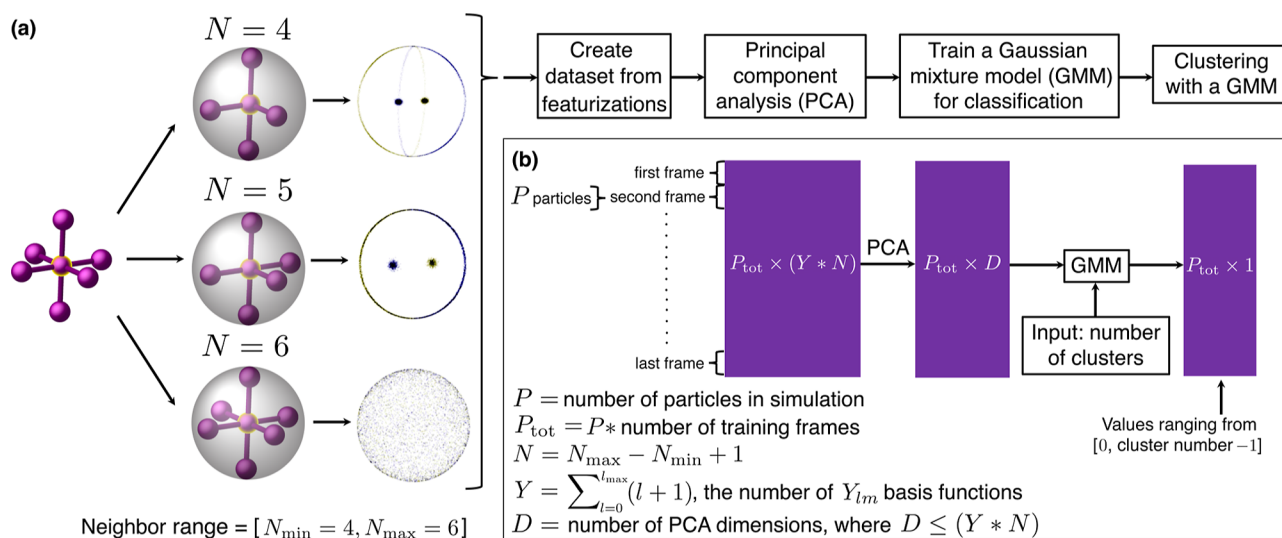


Figure 2. Structural analysis pipeline for Spellings' featurization. (a) For each particle in the training data set, a range of neighborhood sizes is chosen (in this example, the range [4, 6] is shown) and neighbors are selected based on their proximity to the central particle. The middle column illustrates the projection of the relative locations of a particle's neighbors onto a sphere. The right column shows the neighborhood's reorientation using the principal axes of its moment-of-inertia tensor, which is done for all particles in a system snapshot, such that their superposition encompasses all possible choices of N neighbors. The rotationally invariant representation shown in the right column thus also demonstrates permutation invariance—due to thermal noise, permutations of neighbor choices for the $N = 4$ and $N = 5$ neighborhood sizes are sampled, while the complete shell of $N = 6$ particles appears featureless. (b) The construction of the data set and its transformation throughout the analysis process: the featurization for each particle from each training frame is generated, and through dimensionality reduction less relevant features of the data set are discarded. After a Gaussian mixture model (GMM) is trained and the number of clusters optimized, particles can be assigned to clusters, representing local environments.

Put simply, Spellings' features are a higher-dimensional version of Steinhardt's Q_l features: they forego summation over the m values, odd-numbered l values are included, and neighborhood sizes of $N = 4$ to $N = N_{\max}$ are included, as opposed to a single neighborhood size for Steinhardt's features. We implemented both Steinhardt's Q_l and Spellings' featurization via the pythia⁵⁵ software to study the local environments present in simulations of crystallizing structures, using the features to build local order parameters. This process, as shown in Figure 2, is performed on individual simulations, creating an order parameter for each crystallizing structure of interest. We applied an unsupervised learning approach in order to avoid prescribing the local environments present during crystallization a priori as well as for ease of application to a variety of crystal structures. Similar approaches applying unsupervised methods to individual particles have been implemented with other spherical harmonics-based descriptors,^{56,57} but do not use varying neighborhood sizes in their data set construction and focus on the crystallization of simpler structures. The results presented here will describe the use of Spellings' featurization because it is higher-dimensional and contains more detailed information, but the treatment and training of the data will be nearly identical for Steinhardt's and Spellings' features (see "Benchmarking against existing order metrics" subsection in the Supporting Information).

For all structures, $l_{\max} = 12$ was used. Typically, $N_{\max}$ (for Steinhardt's featurization, $N = N_{\max}$) was set to the highest CN expected for the idealized structure; however, in some cases it was necessary to include more neighbors. Including part of (in the case of *t*P30-CrFe and *c*I16-Si) or the entire second-neighbor coordination shell (in the case of *c*P1-Po, *c*P54-K₄Si₂₃, and *c*F160-Na₂₄Si₁₃₆, which are all notably low-coordinated structures) showed better accuracy in distinguishing different local environments (further discussed in the

Results and the "Possible challenges for training unsupervised models" subsection of the Supporting Information). For each structure, the smallest sufficient $N_{\max}$ was the first free parameter fixed in the process of optimizing each structure's trained model. Additionally, we tested using subsets of neighborhood sizes rather than an entire range for Spellings' featurization, but this approach produced poor clustering results, suggesting that neighborhood sizes that did not necessarily correspond to a certain symmetry were important for fingerprinting local environments.

GMMs were implemented via the scikit-learn software⁵⁸ to cluster and classify particle environment data after projecting the data into 64-dimensional space using principal component analysis (PCA). For the Steinhardt featurization, since it is only 6-dimensional—one dimension per even value of l ranging from 2 to 12—we did not perform PCA. The choice of the number of clusters allows for flexibility in classifying order for each structure, as the "best" choice would depend on the use case for which a model is trained. For example, to distinguish between noncrystalline and crystalline order, it would typically be sufficient to simply train the model using two clusters. However, for a study of the growth of complex crystals, it is desirable to train the model to distinguish among different bulk environments, the surface of the crystallite, the liquid, and the gas. Therefore, it was often necessary to use a larger number of clusters in this study. Further details for training the models are provided in the "Unsupervised learning for various structures" section of the Supporting Information. We used the following criteria to select the number of clusters: (1) separation of noncrystalline environments from bulk crystalline sites, (2) separation of the liquid from the gas wherever possible, and (3) minimizing the number of clusters needed to differentiate as many local environments (Wyckoff positions) as possible, while maintaining good accuracy at low temperatures. We

Table 1. Structures Analyzed with Spellings' Features^a

structure	$N_{\max}$	clusters	local environments	expected CNs	calculated CNs
<i>cP54</i> -K ₄ Si ₂₃ (clathrate I)	24 (full second shell)	8	2× gas surface, liquid defects 24 <i>k</i> , 6 <i>c</i> 16 <i>i</i> 6 <i>d</i> 2 <i>a</i>	4, 4 4 0 0	0 0–2 3 4 4 0 0
<i>cF160</i> -Na ₂₄ Si ₁₃₆ (clathrate II)	28 (full second shell)	8	2× gas surface, liquid defects 96 <i>g</i> 32 <i>e</i> , 8 <i>a</i> 16 <i>c</i> 8 <i>b</i>	4 4, 4 0 0	0 0–2 3–4 4 4 0 0
<i>cI16</i> -Si (high-pressure silicon)	11 (partial second shell)	4	gas, surface liquid, subsurface 2 × 16 <i>c</i>	4	0–1 3–4 4
<i>cP1</i> -Po (simple cubic)	18 (full second shell)	3	gas liquid, surface 1 <i>a</i>	6	0–2 2–6 6
<i>cF4</i> -Cu (ccp/fcc)	18 (full second shell)	3	gas, partial surface liquid, partial surface 4 <i>a</i>	12	0–2 9–12 12
<i>hP2</i> -Mg (<i>hcp</i>)	18 (full second shell)	3	gas liquid, surface 2 <i>c</i>	12	0–1 7–12 12
<i>cP20</i> -Mn (β -manganese)	12 (partial first shell)	3	gas, surface, liquid 12 <i>d</i> 8 <i>c</i>	14 12	0–1, 6–9, 11–14 14 12
<i>cI2</i> -W (<i>bcc</i>)	14 (full first shell)	3	gas liquid, surface 2 <i>a</i>	14	0–2 7–14 14
<i>cP8</i> -Cr ₃ Si (FK A15 phase)	14 (full first shell)	4	gas liquid, surface 2 <i>a</i> , liquid 6 <i>c</i>	12, - 14	0–2 8–14 11–12 14
<i>tP30</i> -CrFe (FK σ -phase)	32 (partial second shell)	8	2× gas surface subsurface, liquid defects 2 <i>b</i> , 8 <i>i'</i> 8 <i>i</i> 8 <i>j</i> 4 <i>g</i>	12, 12 14 14 15	0–2 3–11 11–15 12 14 14 14–15

^aFor each structure, the $N_{\max}$ hyperparameter used to train the order parameter, the number of GMM clusters, the environments associated with each cluster, and the CNs of those environments (expected and calculated) are listed. The environments are described by the following: gas (low-density fluid), liquid (high-density fluid), surface(s), defects, and Wyckoff site in the bulk crystal. In almost all cases, results were highly similar using Steinhardt's features.

expect that the crystalline sites will have symmetries that may not be present in environments such as the surface, liquid, or gas and that complex crystals with multiple local environments in the bulk should also have different symmetries at those sites.

Optimizing Training Parameters. We study the growth process for both low- and high-coordinated structures, and in particular structures with more than one Wyckoff site, by applying our method to all 10 structures shown in Figure 1. The optimization of training parameters, or hyperparameters, was most sensitive for low-coordinated structures, which did not contain enough structural information in their first

neighbor shell. Highly complex structures proved challenging as well, as they were susceptible to inadvertently splitting low-density environments that were effectively identical into two clusters in order to allow for different local environments in the crystal to be distinguished. The separation of certain sites, especially with low multiplicity (i.e., number of equivalent positions in the unit cell) or highly similar neighborhood geometry or the same CN, as in the case of the highly complex structures, was not always possible.

The optimized hyperparameters— $N_{\max}$ and number of clusters—are reported in Table 1 for each structure analyzed.

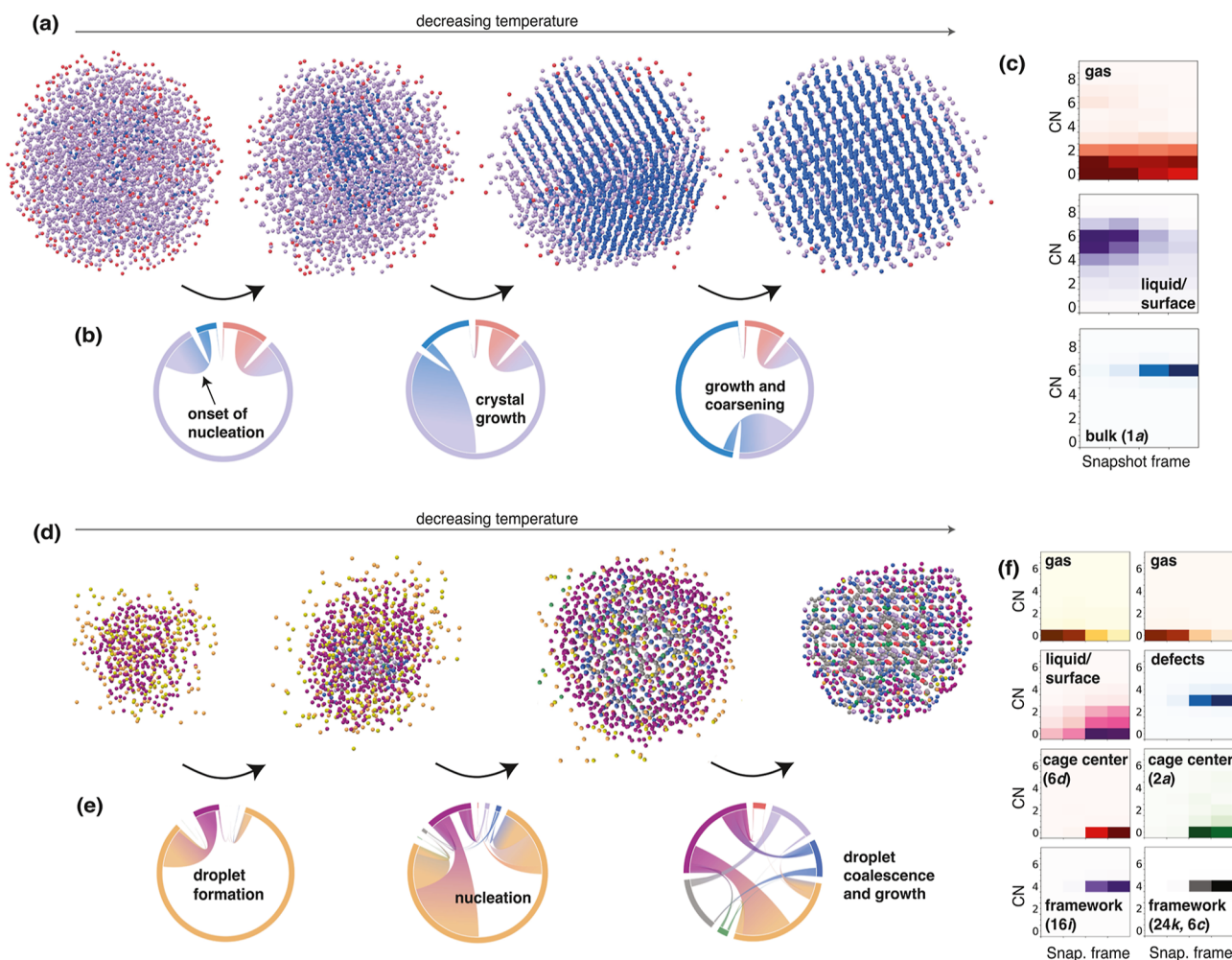


Figure 3. Snapshots from simulations of growing (a–c) simple cubic (*cP1-Po*) and (d–f) clathrate-I (*cP54-K₄Si₂₃*) crystal structures colored by particles' GMM clusters. Visualizations in (a,d) are created using the flowws-analysis⁶² software, and for ease of visualization, most fluid/surface particles have been removed iteratively from the convex hull. In (a–c) the colors correspond to gas/LDF (red), liquid/HDF and surface (lilac), and crystalline bulk (blue), while in (d–f) the colors correspond to the gas/LDF (yellow and orange), three-coordinated defects (blue), surface and liquid (magenta), the 0-coordinated cage centers (green: Wyckoff site 2a, red: Wyckoff site 6d), and the 4-coordinated cage framework sites (lilac: Wyckoff site 16i, gray: Wyckoff sites 24k and 6c). In (b,e), chord plots are used to visualize the change of cluster identity for individual particles from each simulation snapshot; these show the flow between states (i.e., GMM clusters) as well as the total fraction of each cluster out of the total number of particles (as represented by the circular sector assigned to each cluster). In (e), the two gas/LDF clusters are represented by a single grouping (orange) for simplicity. In (c,f), the CN histograms for each simulation snapshot shown in (a,d), respectively, are generated for each cluster and labeled accordingly.

These hyperparameters were chosen based on each structure's coordination and complexity. For high-coordinated, simple structures, this process was straightforward, as $N_{\max}$ was set to be equal to the singular motif's CN. Low-coordinated, high-complexity structures often needed to include neighborhood sizes corresponding to partial or even full second-neighbor shells. An initial guess for the cluster number can be made based on the expected number of local environments to be distinguished (i.e., Wyckoff sites, as well as fluid phases), but ultimately this number was determined by testing a range of cluster numbers. The identified local environments for Spellings' featurization are labeled by inspection of simulation trajectories and by comparing physical metrics of clusters, such as CN. For each cluster, the "calculated CN" indicates the CN that was computed from the simulation trajectory, while the "expected" CN is based on the environment(s) identified as belonging to each cluster, according to the ideal crystal structure that is ultimately being assembled.

Training time varied depending on the size of the data set (defined by the number of training frames provided and $N_{\max}$) as well as the number of GMM clusters tested, taking a few minutes on one CPU to train a mixture model on a data set of roughly 40,000 particles. The computational cost of cluster assignment for a system of 4096 particles was under a minute per simulation frame on a CPU, but it also depended highly on the number of Gaussian clusters in the model. In the [Supporting Information](#), we provide benchmarks for the performance of the order parameter—trained using Steinhardt's and Spellings' featurizations—against both the standard Steinhardt Q_l parameter and CN to show the improved level of structural detail that can be captured with the here-presented method. The ability to train and apply our unsupervised models without the use of high-performance or GPU resources is a significant advantage over more sophisticated machine learning methods with fewer tuned features.^{54,59–61} Furthermore, using features based on existing bond-orientational order

metrics allows for greater interpretability of our resulting machine-learned models.

The featurization and clustering approach outlined in this paper, with which we distinguish particles with geometrically distinct local environments, works well in application even to systems with some amount of thermal noise and disorder, such as those in our simulations. In almost all structures, the gas [i.e., low-density fluid (LDF)], liquid [i.e., high-density fluid (HDF)], and bulk crystal environments can be distinguished. The surface environment tended to be combined with the GMM cluster that included the liquid/HDF since both motifs exhibit high degrees of anisotropy and ambiguity. In some cases, the surface and subsurface environments could be separated, in which cases the subsurface environment was typically grouped with the liquid (dense fluid) environment, while the outermost surface environment was either entirely distinct or grouped with the gas (LDF). The two motifs, however, can often be separated according to the time at which they occur during the simulation, because the surface environment cannot form without the presence of the crystalline bulk and is distinct from the local motifs in the dense fluid. In all crystal structures with two Wyckoff sites, the sites were separated from one another; in all crystal structures with five Wyckoff sites, all but two sites (with the same number of neighbors) were separated from one another. For the clathrate structures, we identified a GMM cluster representing 3-coordinated defects which were abundant in the training set.

To our knowledge, no existing methods of quantifying bond-orientational order have been applied to identify many different local environments during self-assembly for a broad range of structures. The strength of this method for order classification is due to the sensitivity of spherical harmonics-based featurizations to symmetry because using different l values and choosing neighborhood sizes prior to dimensionality reduction generates the representations that are most important for a particular crystal structure. Differentiating various crystalline sites and fluid environments, which are not necessarily known a priori, would necessitate a manual trial-and-error investigation of each crystal structure. With appropriate training parameters, the featurization captures a sufficient amount of relevant information to allow for the identification of local environments without needing supervision during training. Further details on how to train a GMM for a given structure are included in the [Supporting Information](#). A trained model generalized well to different simulations that assembled the same crystal structure, even if those simulations were performed by modeling interactions with a different pair potential. It is worth noting that for Spellings' featurization, the variability in the data captured after PCA was typically 40–50%, suggesting a nonlinear distribution of the original high-dimensional data that can be harnessed further in the future, using other machine learning approaches.

RESULTS AND DISCUSSION

Our order parameter can be applied to study the growth of all ten structures shown in [Figure 1](#). For all structures, a simulation with condensation from a low-density gas phase to a dense fluid phase, followed by crystallization from a dense fluid droplet, was observed. The types of local environments present during this growth process are described in [Table 1](#) for all crystal structures. In [Figure 3](#), the application of the trained models to a simulation of nucleation and growth from a droplet is shown for both a simple structure (*cP1-Po*) and a

complex structure (*cP54-K₄Si₂₃*), with particles colored by their GMM cluster assignment. [Figure 3](#) shows how the cluster assignments and CN of particles belonging to each cluster change on a per-particle basis over the time elapsed, as illustrated by consecutive simulation snapshots. The two examples show how differently growth can proceed from the liquid phase: in the case of *cP54-K₄Si₂₃* (clathrate I), the dense fluid phase has barely formed when the crystalline environments become present and a crystal nucleates (apparent in the small size of the liquid droplet). We provide a more detailed analysis of the growth of the clathrate I structure in the [Supporting Information](#), and we show that the structure of the dense fluid/surface environment is structurally more similar to that of the LDF. We document later in this paper how this is not the case for other dense fluids in our study. In the case of *cP1-Po* (simple cubic), the fluid droplet constitutes the majority of particles in the simulation before nucleation and growth of the crystalline phase.

While the vast majority of our simulations crystallized via a dense fluid droplet, some exhibited growth directly from the gas phase. The latter crystallization pathway was significantly rarer across the hundreds of simulations in our data set, likely because a dense fluid lowers the nucleation free energy barrier.²⁷ Among the three types of pair potentials used in this study, the Lennard-Jones–Gauss potential stabilized liquids with the widest temperature range, the oscillating pair potential exhibited a narrower liquid phase stability region, and the Yukawa–Gauss potential was the only pair potential to exhibit nucleation and growth directly from the gas phase—or via an exceedingly narrow liquid phase that could only be captured by high frame-rate data. This is likely due to the increasing steepness of Lennard-Jones–Gauss potentials vs oscillating pair potentials vs Yukawa–Gauss potentials.

In several cases, we were able to compare particles interacting with different types of pair potentials that crystallized into the same structures. Here, we highlight the trends and observations of crystal growth across structures and growth pathways with a particular focus on the complex structures in our study such as the Frank–Kasper phases *cP8-Cr₃Si* and *tP30-CrFe*. In the [Supporting Information](#), we include a detailed analysis of the growth of *cI2-W* (*bcc*) and *cP54-K₄Si₂₃* (clathrate I). Application of our order parameter elucidates the structural complexity inherent in the process of crystallization and allows us to compare growth across different crystallization pathways, examine the similarity of the liquid environments to a subset of the crystalline coordination environments, and dissect the growth of a complex structure, including recrystallization.

Differing Crystallization Pathways. For the *cP8-Cr₃Si* structure in particular, growth along two different pathways was observed: particle-by-particle growth directly from the gas phase was observed with the Yukawa–Gauss potential, while crystallization proceeding via a dense fluid was observed with the oscillating pair potential. This set of simulations allows us to compare the performance of the order parameters for different crystallization pathways and consider the relevance of the dense fluid to crystal growth.

We trained two models, one on simulations of each pathway. The optimal hyperparameter $N_{\text{max}} = 14$ reported in [Table 1](#) for *cP8-Cr₃Si* was found by training on the simulation exhibiting growth from a HDF. Although we report an optimal cluster number hyperparameter in [Table 1](#), exploring different cluster numbers for each of the two models can elucidate the

distinctiveness of various environments present during growth. In both models, the 14-coordinated Wyckoff site ($6c$) is the first crystalline environment to be differentiated from the remaining particle environments, as early as with two clusters. Most of the HDF and all surface particles can be separated from the 12-coordinated Wyckoff site ($2a$) in a four-cluster model, having remained combined when using three clusters. A second model can be trained on simulation data exhibiting growth directly from the gas phase, also with $N_{\text{max}} = 14$. This model behaves similarly to the first model for the three-cluster case—and differently for the four-cluster case.

The application of both models is illustrated by the two-dimensional t -distributed stochastic neighbor embeddings (t -SNEs) of the respective training data colored by their GMM cluster in Figure 4. Proximity between data points in the two-dimensional embeddings suggests structural similarity as detected by the features used. Likewise, the separation of

groupings of data points in the two-dimensional embedding implies a separation in the original high-dimensional data. The structural dissimilarity of the 12- and 14-coordinated Wyckoff sites is apparent by the distinct separation of their GMM clusters along both growth pathways. In the case of growth via a liquid phase, the difference between the three-cluster and four-cluster models illustrates the structural similarity between the $cP8$ - Cr_3Si liquid and the icosahedral environment of the 12-coordinated Wyckoff site. While icosahedral order in liquids has been demonstrated,^{63,64} the icosahedra present in the liquid and crystalline environments are still sufficiently distinct as to be distinguished by our order parameter.

The four-cluster models can be directly compared using the CN histograms for each cluster as well as the space spanned by each cluster's embedded data. In the four-cluster model trained on the simulation exhibiting growth directly from the gas phase, the fourth cluster is formed by splitting the gas environment identified in the three-cluster model in two. The two partitions exhibit equivalent CN distributions, suggesting that the split is not structurally meaningful. This is unsurprising, given that no dense fluid was observed in this simulation.

For the four-cluster model trained on the simulation exhibiting growth via a dense fluid, the fourth cluster identified is structurally distinct from the other three clusters and represents most of the liquid and all surface local environments. Additional simulation snapshots and CN histograms for this model are provided in the Supporting Information. This pair of simulations and their respective order parameters show how our simulations can model structurally distinct growth pathways, which nonetheless can lead to the formation of the same final structure. Our analysis approach and low-dimensional embeddings highlight the complexity of local structural data across a crystallization trajectory. Moreover, the embeddings shown in Figure 4 underscore the different relationships between the Wyckoff sites and the dense fluid phase along both growth pathways, which will further be highlighted by a comparison of liquid and crystalline coordinations in the following section.

Liquid Under-Coordination. Short-range order in liquids of metallic elements has been studied both in simulation⁶⁵ as well as experiment,⁶⁴ identifying a relationship between the structure of the liquid and the respective solid that forms in the same system. With the multiwell potentials used in this study, we further probe how this relationship extends to the formation of complex crystal structures. We can distinguish the HDF phase (i.e., liquid) from the LDF (i.e., gas) and crystalline phases through classification with our order parameter(s), and we calculate CNs for each liquid cluster in the simulation frames immediately preceding crystallization. We find that particles in the high-density liquid-like droplet prior to crystallization have fewer nearest neighbors than the average in its solid, as seen in Figure 5, if calculating CN with a radial cutoff found from the solid's radial distribution function (RDF). It is worth noting the diversity in CNs exhibited by the precrystallization liquids—with respect to which our order parameter is robust—highlighting the difficulty of using straightforward CN analysis to distinguish the liquid droplet from the gas (LDF).

A subtler feature, illustrated in Figure 5, lies in the comparison of liquid under-coordination for simple structures (one Wyckoff site) to complex structures (two or more Wyckoff sites). All of the simple structures— $cI16$ -Si, $cP1$ -Po,

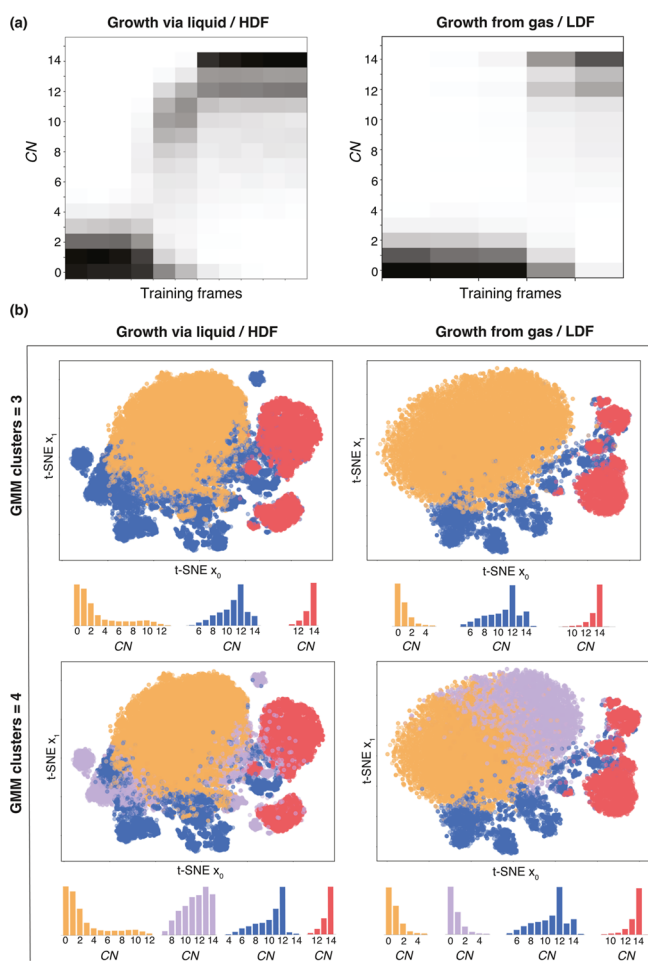


Figure 4. (a) CN histograms and (b) two-dimensional embeddings of particle training data from two different simulations in which $cP8$ - Cr_3Si forms via different crystallization pathways (left: via a HDF/liquid; right: via a LDF/gas). (a) Training frames are chosen before and after crystallization in each system, and the temperature in each frame decreases from left to right. (b) Embeddings are generated using t -SNE⁵⁸ on the PCA data and colored by each particle's GMM cluster assignment. The CNs for all particle training data in each GMM cluster are shown for each model, with the histograms left unscaled for ease of visualization due to particle count differences between clusters. Detailed CN histograms are provided in the Supporting Information.

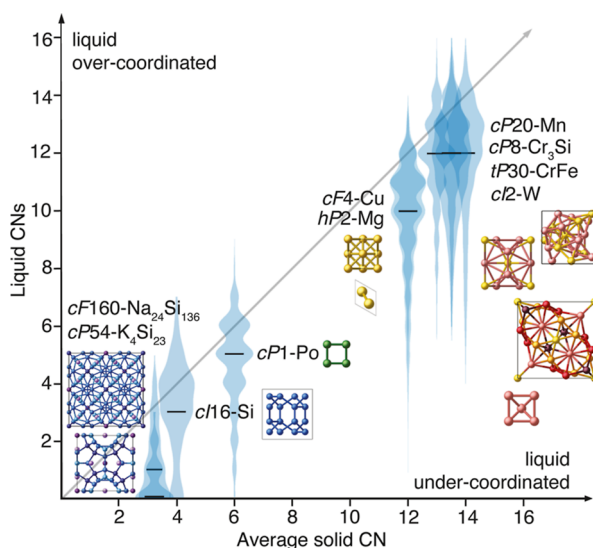


Figure 5. CNs of the liquid clusters for each crystal structure plotted against that structure's theoretical average CN in the solid, calculated using a spherical cutoff distance from the first minimum in the solid's RDF. Violin plots are used to show the distribution of (integer) CNs of the supercooled fluid immediately before crystallization, with the median marked as a horizontal line. All structures consistently exhibit under-coordination of their dense fluids relative to their bulk crystal phases.

cF4-Cu, *hP2-Mg*, and *cI2-W*—show under-coordination by one to two neighbors compared to their bulk crystals. On the other hand, the complex structures—*cF160-Na₂₄Si₁₃₆*, *cP54-K₄Si₂₃*, *cP20-Mn*, *cP8-Cr₃Si*, and *tP30-CrFe*—are not under-coordinated relative to the lowest coordination in their bulk crystal structures but only relative to the crystalline sites with higher CNs. The high-coordinated *cP20-Mn*, *cP8-Cr₃Si*, and *tP30-CrFe* all have a median of 12 neighbors in their dense liquid droplet, which is equivalent to the lowest-coordinated site(s) in their solids; the low-coordinated clathrates, *cP54-K₄Si₂₃* and *cF160-Na₂₄Si₁₃₆*, have a median of 0 and 1 neighbor(s) in their liquids, and both have zero-coordinated cage centers in their bulk crystals.

It may seem unsurprising to report liquid under-coordination relative to each structure's bulk crystal given that a change in density is often expected upon crystallization. Calculating a liquid's CN is highly sensitive to how a particle's neighbors are chosen, but because the relationship of the liquid's local structure to that of the solid is of primary interest in our study, we choose to use a radial cutoff derived from the solid's RDF. Prior literature on two-dimensional hard disks⁶⁶ and three-dimensional Lennard-Jones liquids (comprised of ccp, hcp, and icosahedral motifs)⁴⁶ did not report under-coordination of the liquids for simple structures, but this discrepancy is likely due to how CN cutoffs are calculated (see [Supporting Information](#) for calculation of under-coordination for a Lennard-Jones liquid). The under-coordination of liquids has, however, been theorized for Frank–Kasper phases. These structures have multiple high-coordinated environments in their solids (CN $\geq$ 12), and their liquids consist largely of icosahedral shells (CN = 12), under- and overcoordinated particles (with CN = 10 and CN = 14) forming disclination lines, and pairs of particles with CN = 11 and CN = 13 neighbor shells signifying the presence of vacancies and interstitials, respectively.⁶³ The formation of CN > 12 sites upon crystallization is attributed to

disclination networks forming in order to accommodate the geometric frustration from the CN = 12 icosahedral shells present in the crystal structure.

In more recent work, the Lennard-Jones–Gauss potential has been used to model geometrically frustrated amorphous calcium carbonate due to competing stabilizing interparticle distances from multiple potential wells.⁶⁷ Geometric frustration can explain our more general observation of liquid under-coordination prior to growth of complex structures including those that are not Frank–Kasper phases, such as *cP20-Mn* (β -manganese), *cP54-K₄Si₂₃* (clathrate I), and *cF160-Na₂₄Si₁₃₆* (clathrate II). This means that the liquids of highly complex structures may be structurally more similar to only certain bulk sites rather than all of them equally. This insight is not reflected in any existing models of crystal growth and suggests the need to incorporate the relative energetic favorabilities of possible local environments. Moreover, how the geometric frustration of local environments in the precristallization liquid is resolved by crystallizing into the environments of the solid is not yet broadly understood. From our analysis of precristallization liquids, it follows that the emergence of different local environments present in the complex structure must undergo different types of geometric transitions allowing them to “choose” their various roles in the bulk crystal.

Differentiation of Local Environments during Growth and Recrystallization. Changes in the local geometry for different Wyckoff sites during crystallization are investigated by applying physically interpretable metrics to our particles partitioned by their GMM cluster assignment. We investigate the growth via a dense fluid from a high frame-rate simulation of *tP30-CrFe* (Frank–Kasper σ -phase)—a structure with five Wyckoff sites in its unit cell ranging from CN = 12–15. As discussed in the previous section, the Frank–Kasper liquid phases are expected to have a high occurrence of CN = 12 particles, and in this section, we dissect how the higher coordinated sites (CN = 14 and 15) emerge from the dense fluid, and how the presence of CN = 13 defects can inhibit crystallization.

In particular, we analyze a highly frustrated crystal growth trajectory, signified by an increased presence of CN = 13 particles⁶³ (see [Supporting Information](#) for CN histograms). Applying GMM clustering to this simulation, coupled with Voronoi analysis, reveals that crystallization proceeds in three stages, as shown in [Figure 6](#). Five replica simulations were run at this state point at a high frame rate, with frustrated assembly—a high occurrence of CN = 13 during the formation of CN = 14 and 15 sites—seen in two of the five simulations. An example of assembly where a single crystal grows directly from the fluid without such frustration is provided in the [Supporting Information](#) for comparison.

The boundaries between the three stages illustrated in [Figure 6](#) are defined at times t_1 , ..., t_d , as delineated by jumps in the correlation function in [Figure 6a](#) at those times. In the first stage, between t_1 and t_2 , the droplet grows and densifies, with very few crystal-like environments appearing and disappearing via random thermal fluctuations. At time t_2 , both isoperimetric quotient and CN increase for the GMM clusters associated with the 4g (15-coordinated), 8i (14-coordinated), and 8j (14-coordinated) Wyckoff sites. In the second stage, between times t_2 and t_3 , densification and droplet size growth slow down and crystalline order begins to emerge, as noted by a differentiation in average local number density $\langle N/V \rangle_{\text{local}}$, average isoperimetric quotient $\langle Q \rangle$, and average CN_{Voronoi} shown in [Figure](#)

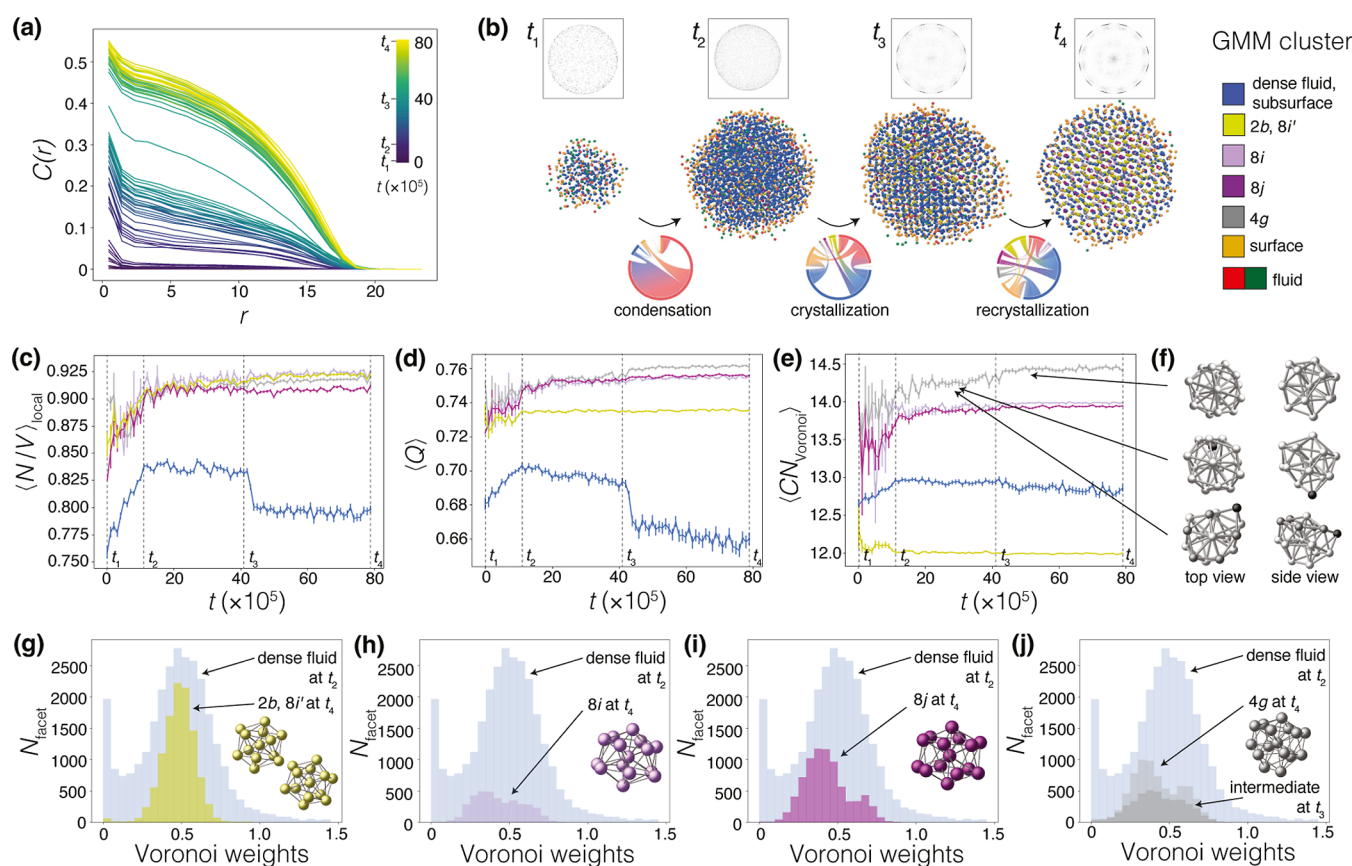


Figure 6. Simulation of the growth of a *t*P30-CrFe structure type. (a) Correlation function $C(r)$ calculated as a function of interparticle distance r and colored by simulation time. (b) Simulation snapshots colored by GMM cluster and bond-orientational order diagrams at times t_1 through t_4 , with gas particles omitted for clarity. Switches between clusters between each time marker are represented using chord plots. (c–e) Measures of order for each cluster (excluding the gas and surface clusters) over time. Left to right: (c) average local number density; (d) average isoperimetric quotient of the Voronoi polyhedra; (e) average CN calculated from the Voronoi polyhedra using a weight cutoff. Error bars represent the standard error of the mean. (f) Representative motifs extracted from the cluster associated with the 4g-site before and after recrystallization. The topmost row shows a typical CN = 15 site after recrystallization, while the middle and bottom rows show CN = 14 defects—which only appear prior to recrystallization—that are distorted CN = 15 sites. Both of the shown defects are characterized by the presence of a square pyramidal motif formed by the inclusion of a neighbor (colored in black) outside the Voronoi weight cutoff. In the bottom row, particles in the pentagonal motif are colored dark gray as a guide for the eye. (g–j) Histograms of Voronoi weights (i.e., polyhedra facet areas) for each of the crystalline clusters from the end of the simulation (time t_4) plotted against the histogram of the “dense fluid” cluster from immediately before the onset of crystalline order (time t_2). The Voronoi weight histogram at time t_3 is included for the cluster associated with the 4g-site in (j). Idealized Voronoi polyhedra for each Wyckoff site are included as insets.

6c–e. The bond-orientational order diagrams begin to show signs of global order between t_2 and t_3 (see Figure 6b), suggesting that the droplet is no longer structurally amorphous.

At t_3 , we see further differentiation in $\langle N/V \rangle_{\text{local}}$ for the crystalline GMM clusters (see Figure 6c), with particles in the “dense fluid” cluster dropping to lower $\langle N/V \rangle_{\text{local}}$ and $\langle Q \rangle$ (see Figure 6d). Also, at t_3 , the cluster associated with the 15-coordinated 4g site increases in $\langle Q \rangle$ (see Figure 6d) and in $\text{CN}_{\text{Voronoi}}$ (see Figure 6e). The final stage of assembly, from t_3 to t_4 , moves the frustrated system with multiple nucleating grains and grain boundaries to a single-crystalline assembly, representing recrystallization. The delayed differentiation of the 15-coordinated site in the calculated physical metrics is a key difference between the frustrated and nonfrustrated crystallization processes, and it is detected by the order parameter. Due to the direct crystallization of a single crystal, the nonfrustrated crystallization pathway lacks the intermediate stage between time t_2 and t_3 , suggesting that the proper formation of the CN = 15 site is inhibited by the geometric frustration from CN = 13 defects contained in the “dense fluid”

cluster. This is not to say that the 15-coordinated site is necessarily the only one inhibited in its formation by CN = 13 defects: the “dense fluid” cluster contains particles ranging from CN = 11–15 as noted in Table 1 (and in the Supporting Information).

The particles in the GMM cluster associated with the 12-coordinated Wyckoff site(s) undergo little change in their local environment after time t_2 , i.e., crystallization. The “dense fluid cluster” prior to crystallization is not under-coordinated relative to the lowest-coordinated sites in the *t*P30-CrFe crystal structure (see Figure 6e), and the distribution of Voronoi weights of the associated GMM clusters (the “dense fluid” cluster at t_2 and the CN = 12 cluster at t_4) are centered at the same peak position (shown in Figure 6g–j). The Voronoi weights for the cluster associated with the CN = 12 sites at time t_4 only show a narrowing of the unimodal distribution relative to that of the dense fluid at time t_2 . By contrast, the other Wyckoff sites with CN > 12 neighbors all exhibit geometric transitions to bimodal distributions of Voronoi weights—corresponding to the facets connected to either 5- or

6-coordinated vertices in the Frank–Kasper polyhedra.⁶⁸ In contrast, the dense fluid exhibits a unimodal Voronoi weight distribution, with the most abundant weights falling between the two modes. We provide additional unimodal Voronoi weight data for the “dense fluid” cluster for the growth trajectory in the [Supporting Information](#). The fact that the 12-coordinated sites are most similar in local structure to the dense fluid suggests that the mechanism for the emergence of 14- and 15-coordinated sites at crystallization occurs in conjunction with a reorganization of those coordination shells to accommodate more neighbors, which does not happen for the 12-coordinated crystalline sites.

At high temperature (near crystallization), our order classification method places particles into clusters associated with—but not necessarily identical to—the Wyckoff site at low temperatures, particularly because of motifs present during the frustrated assembly. This is evident in all crystalline clusters between times t_1 and t_2 , for the cluster associated with the 15-coordinated Wyckoff site 4g between times t_2 and t_3 , and vice versa for the dense fluid cluster between t_3 to t_4 (when the entire droplet has crystallized). In the case of the CN = 15 site, we see an “intermediate” local environment between times t_2 and t_3 , which is similar to that of the 4g Wyckoff site but with a lower $\langle \text{CN}_{\text{Voronoi}} \rangle$ value and a geometry (measured by $\langle Q \rangle$) that is more similar to that of the 14-coordinated sites. The Voronoi weight histogram of the 4g-site associated cluster at t_3 , as seen in [Figure 6j](#), lacks the characteristic bimodal distribution that appears at time t_4 . The 8i-site associated cluster at t_4 is also not quite bimodal. Interestingly, the modest increase in $\langle \text{CN}_{\text{Voronoi}} \rangle$ for the site at time t_3 also vanishes if faraway neighbors (with Voronoi weights < 0.05) are not discarded, and instead we observe a decrease in $\langle \text{CN}_{\text{Voronoi}} \rangle$ at time t_3 for the “dense fluid” cluster (see [Supporting Information](#)).

We used this insight to extract these distorted local environments between times t_2 to t_3 from the cluster associated with the 4g site, and we find that—in addition to the 4g local environment—the cluster also detects defects, such as distorted 4g sites with a pentagon and triangle motif replacing a planar hexagonal motif, as depicted in [Figure 6f](#). As a result, the $\langle \text{CN}_{\text{Voronoi}} \rangle$ value of the cluster associated with the 4g Wyckoff site is not quite equal to 15. In the [Supporting Information](#), we provide a further breakdown of Q as it changes with $\text{CN}_{\text{Voronoi}}$ over time for the particles in this cluster and the “dense fluid” cluster, and we also include the Voronoi histograms for the GMM clusters associated with crystalline sites at low temperature ($kT = 0.1$). The structural distortions and defects highlighted in this particular frustrated assembly and recrystallization process illuminate the need for a more structure-based approach for understanding crystal growth.

CONCLUSIONS

In this paper, we utilize a versatile unsupervised machine learning-based method for classifying particles by local environment in order to better understand the growth of crystal structures with different CNs and degrees of complexity. Tuning the model hyperparameters in accordance with a structure’s complexity and CN can allow for the distinction between different local geometries, which is not accomplished to the same degree by standard order metrics. This method can also distinguish between different phases—gas, liquid, and solid—even in cases where a liquid’s local environments are similar to those in the solid. The broad applicability to different

kinds of structures (demonstrated on 10 crystal structures with 1–160 particles per unit cell and local CN = 0–15), and the ease of training a model using standard unsupervised learning methods, make our method a good local order parameter for studying the growth of complex structures in simulation. An extension of this method to multicomponent systems would also be straightforward, as the descriptor is agnostic to particle-level features such as type, size, or shape. The unsupervised technique we present here can also further enable the exploration of other phenomena such as surface reorganization, coarsening, and dislocation motion—in simulations or experiments.

We used our method to analyze both fluid–fluid and fluid–solid structural transitions and extract general patterns in crystal growth. We observe under-coordination of the liquid across all structures relative to the crystal coordination, in contrast to hard anisotropic shapes, for which equal CN had been reported for fluids and their respective solids in simulation.⁶⁹ Our findings suggest that the geometric mechanisms for crystallization of complex structures from a dense fluid differ compared to those of simple structures, because certain crystallographic sites in a complex structure are not under-coordinated relative to their precrystallization liquid.

In systems of particles interacting with isotropic pair potentials, we capture the structural transitions during crystal growth via a pathway from a LDF to a HDF and ultimately to a crystal and investigate how these transitions can differ for different Wyckoff sites in a highly complex structure. We observe the differential effect of recrystallization and frustrated assembly on specific Wyckoff sites, finding that structural defects during growth, such as in *t*P30-CrFe are identified by our order parameter. Further investigation of other nucleation and crystallization pathways for complex structures is warranted, as we did not use methods such as umbrella sampling⁷⁰ to probe difficult-to-access states near the crystallization temperature of each structure.

The method we present could be useful for tracking nucleation in simulations with larger system sizes or as an order parameter to use in forward-flux sampling.⁷¹ Future simulation studies should also consider how the energy landscape and growth kinetics are altered by the presence of an intermediate supercooled liquid droplet. Significant work has already considered growth along different crystalline facets in simple atomic structures and even colloidal systems,⁷² or by considering liquid ordering near crystalline interfaces,⁷³ yet few models of crystal growth incorporate a dependence of thermodynamic quantities on local structure.⁹ The structural characteristics of different local environments captured by spherical harmonics-based features could help build more sophisticated models of crystal growth. Moreover, using structural descriptors—such as those developed in this study—that employ supervised machine learning can help establish formal connections between the shapes of particles’ interaction potentials and their assembled structures in future research. Extending growth models to explore complex structures can create a richer understanding of crystallization and aid in the design of growth units for experimental soft matter assembly in the quest to design materials on the mesoscale.

METHODS

Particle interactions adapted from previous studies were modeled with the following isotropic multiwell pair potentials: Lennard-Jones—

Gauss, oscillating pair potentials,⁴¹ and a Yukawa–Gauss-based functional form.⁴² Few structures exist in the intermediate-coordinated (CN = 7–11) range, which were, therefore, not included in this study. Further details of the specific interaction potential functions implemented and parameters used for assembling each structure can be found in the [Supporting Information](#).

Structures are labeled with their Pearson symbol, which specifies the Bravais lattice and the number of particles in the unit cell, and a prototypical compound. We denote the “complexity” of a structure by the number of its crystallographic Wyckoff sites, which represent symmetrically inequivalent positions in a structure’s unit cell, which can therefore adopt different local environments.

Simulations of Crystal Growth. Molecular dynamics simulations were performed using the open-source HOOMD-blue software.⁷⁴ Systems of $N = 4096$ particles were initialized with randomized velocities in an NVT (canonical) ensemble at effectively zero pressure and cooled using a linear temperature ramp from $kT = 1.0$ or $kT = 1.5$ to $kT = 0.1$, controlled by a Nosé–Hoover thermostat. For all simulations, either 10^7 or 10^8 molecular dynamics time steps were used with a step size of $dt = 0.005$ to ensure that the system was being cooled sufficiently slowly for single crystals to form. The signac and signac-flow software were used for workflow and data management.^{75,76}

Several replica simulations were performed at each state point, and simulation trajectories that exhibited a low defect occurrence in the formation of high-quality crystals were chosen for generating training data for structural analysis. For each simulation, 100 total frames were collected, and training frames were chosen near (before, during, and after) the crystallization event as well as from the final frames of the simulation. To capture more data in the crystallizing region of interest, simulations were restarted from an existing simulation frame in the temperature window where the crystallization occurred, and frames were then collected every 100 time steps (the temperature ranges tended to be sufficiently narrow for the data volume to remain on the order of a few gigabytes per simulation trajectory for the system sizes simulated in this study). We note that the process of restarting the simulation from an existing frame effectively creates a replica as our simulations are not deterministic.

Physical Order Metrics. We used physically interpretable metrics to analyze the structural properties of each GMM cluster. These include CN, Voronoi analysis and isoperimetric quotient, local number density, and correlation functions, which are all calculated using the freud software package.⁷⁷

Coordination Number. In any discussion of “bonding” and nearest neighbors, it is important to consider how those neighbors are defined. For the most part, we define the first neighbor shell using a spherical cutoff at the position of the first minimum in the RDF.⁷⁸ The CN values presented in this paper are calculated using a spherical cutoff at the first minimum in the RDF, unless otherwise noted. We note that the CN depends sensitively on the choice of RDF cutoff and that the presented data should be interpreted only in this context. The RDF is calculated using the freud software⁷⁷ and smoothed using a Gaussian filter with standard deviation $\sigma = 2$. While the RDF changes throughout the crystallization process, the locations of its maxima change only subtly so that the cutoff distance is calculated using only the last (coldest) frame of a simulation. CN histograms were generated for the training frames for each GMM cluster, which served as a check for the interpretation of different clusters as representing particular Wyckoff sites. We also apply CN analysis by the GMM cluster to the high frame-rate crystallization data and isolate specific frames prior to and during crystallization to extract phase-specific CN values.

Voronoi Analysis and Isoperimetric Quotient. Voronoi analysis⁷⁹—a geometric, parameter-free way of determining neighbors—is used for understanding local geometry and can serve as a proxy for bond angles, which are particularly susceptible to thermal noise for high-coordinated structures. Each neighbor in a particle’s Voronoi polyhedron is weighted by the area of its corresponding facet, and these Voronoi weights (facet areas) are determined by both a neighboring particle’s distance from the central particle and the

locations of other nearest neighbors adjacent to the neighboring particle, which define the respective bond angles.

For high-coordinated structures, the Voronoi polyhedra for ordered motifs correspond well to the coordination environments found using a spherical cutoff. For low-coordinated and disordered motifs, however, Voronoi polyhedra will overestimate the CN by including additional faraway neighbors. We resolve this discrepancy by imposing a Voronoi weight cutoff when calculating CN_{Voronoi} (here: 0.05), and for comparison, we provide figures that calculate CN_{Voronoi} without cutoffs imposed in the [Supporting Information](#). The isoperimetric quotient, $Q = 36\pi V^2/S^3$, can also be calculated from the volume of a Voronoi polyhedron V and its total surface area S as a way to quantify local geometry.⁸⁰

Local Number Density and Correlation Functions. We define a simple correlation function $C(r) = \langle s_1(0) \cdot s_2(r) \rangle$, calculated over all particle pairs as a function of radial distance r , with the values of s_1 and s_2 based on the GMM cluster assignment: $C = 1$ for particle pairs both assigned to “crystalline” (bulk solid) clusters and $C = 0$ otherwise. This correlation function can be used to observe how crystallinity propagates during the simulation. We also used average local number density, $\langle N/V \rangle_{\text{local}}$, to characterize the particles in each cluster throughout the growth process. The local number density is calculated based on the number of particle centers within the first coordination shell (using a fixed radial cutoff deduced from the RDF of the solid).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsnano.4c01290>.

Additional information including details of pair potentials used and the respective RDFs from simulations, discussion of how to implement the unsupervised learning method, benchmarking and comparison of the proposed method against existing metrics, details of neighbor counting, analysis of a Lennard-Jones liquid, analysis of bcc and clathrate-I growth, and further coordination-number and Voronoi analysis examining application of the here-presented unsupervised method to Frank–Kasper phases as well as to a nonfrustrated Frank–Kasper phase crystal growth trajectory ([PDF](#))

AUTHOR INFORMATION

Corresponding Author

Julia Dshemuchadse — Department of Materials Science and Engineering, Cornell University, Ithaca, New York 14853, United States; orcid.org/0000-0003-2310-6687; Email: jd732@cornell.edu

Authors

Maya M. Martirosyan — Department of Materials Science and Engineering, Cornell University, Ithaca, New York 14853, United States; orcid.org/0000-0002-6138-5296

Matthew Spellings — Vector Institute for Artificial Intelligence, Toronto, Ontario M5G 1M1, Canada; orcid.org/0000-0002-4061-4299

Hillary Pan — Department of Materials Science and Engineering, Cornell University, Ithaca, New York 14853, United States; orcid.org/0000-0002-4073-2180

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acsnano.4c01290>

Notes

The authors declare no competing financial interest.

A preprint of the submitted version of this manuscript is available on arXiv.⁸¹ Simulation results are available at the Materials Data Facility^{82,83} and the code is available on GitHub.⁸⁴

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grant no. DMR-2144094 and was supported by the Cornell Center for Materials Research with funding from the NSF MRSEC program (DMR-1719875), as well as the Camille and Henry Dreyfus Foundation through a Machine Learning in the Chemical Sciences and Engineering Award (ML-22-038). M.M.M. acknowledges support from the National Science Foundation Graduate Research Fellowship grant no. DGE-1650441 (2019–2021) and DGE-2139899 (2021–2024) and from the Dolores Zohrab Liebmann Fund Fellowship. M. S. acknowledges resources provided by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute. This work was performed at the Cornell University Center for Advanced Computing (CAC). The authors thank R. S. Skye, R. N. Scott, E. R. Dufresne, A. Amon, B. Z. Gregory, and S. K. Bright-Thonney for helpful discussions and feedback.

REFERENCES

- (1) Galkin, O.; Vekilov, P. G. Control of Protein Crystal Nucleation around the Metastable Liquid–Liquid Phase Boundary. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6277–6281.
- (2) Saranathan, V.; Osuji, C. O.; Mochrie, S. G. J.; Noh, H.; Narayanan, S.; Sandy, A.; Dufresne, E. R.; Prum, R. O. Structure, Function, and Self-Assembly of Single Network Gyroid (I4₁32) Photonic Crystals in Butterfly Wing Scales. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11676–11681.
- (3) Gasser, U.; Weeks, E. R.; Schofield, A.; Pusey, P. N.; Weitz, D. A. Real-Space Imaging of Nucleation and Growth in Colloidal Crystallization. *Science* **2001**, *292*, 258–262.
- (4) Sun, Y.; Ma, K.; Kao, T.; Spoth, K. A.; Sai, H.; Zhang, D.; Kourkoutis, L. F.; Elser, V.; Wiesner, U. Formation Pathways of Mesoporous Silica Nanoparticles with Dodecagonal Tiling. *Nat. Commun.* **2017**, *8*, 252.
- (5) Song, R.-Q.; Hoheisel, T. N.; Sai, H.; Li, Z.; Carloni, J. D.; Wang, S.; Youngman, R. E.; Baker, S. P.; Gruner, S. M.; Wiesner, U.; Estroff, L. A. Formation of Periodically-Ordered Calcium Phosphate Nanostructures by Block Copolymer-Directed Self-Assembly. *Chem. Mater.* **2016**, *28*, 838–847.
- (6) Senabulya, N.; Xiao, X.; Han, I.; Shahani, A. J. On the Kinetic and Equilibrium Shapes of Icosahedral Al₇₁Pd₁₉Mn₁₀ Quasicrystals. *Scr. Mater.* **2018**, *146*, 218–221.
- (7) Matsumoto, M.; Saito, S.; Ohmine, I. Molecular Dynamics Simulation of the Ice Nucleation and Growth Process Leading to Water Freezing. *Nature* **2002**, *416*, 409–413.
- (8) Gasser, U. Crystallization in Three- and Two-Dimensional Colloidal Suspensions. *J. Phys.: Condens. Matter* **2009**, *21*, 203101.
- (9) Freitas, R.; Reed, E. J. Uncovering the Effects of Interface-Induced Ordering of Liquid on Crystal Growth Using Machine Learning. *Nat. Commun.* **2020**, *11*, 3260.
- (10) Smeets, P. J. M.; Cho, K. R.; Kempen, R. G. E.; Sommerdijk, N. A. J. M.; De Yoreo, J. J. Calcium Carbonate Nucleation Driven by Ion Binding in a Biomimetic Matrix Revealed by in Situ Electron Microscopy. *Nat. Mater.* **2015**, *14*, 394–399.
- (11) Burton, W. K.; Cabrera, N.; Frank, F. C.; Mott, N. F. The Growth of Crystals and the Equilibrium Structure of Their Surfaces. *Phil. Trans. Roy. Soc. Lond. Math. Phys. Sci.* **1951**, *243*, 299–358.
- (12) Wilson, H. W. XX. On the velocity of solidification and viscosity of super-cooled liquids. *London, Edinburgh Dublin Phil. Mag. J. Sci.* **1900**, *50*, 238–250.
- (13) Frenkel, J. Note on a Relation between the Speed of Crystallization and Viscosity. *Physik. Zeit. Sowjetunion* **1932**, *1*, 498–510.
- (14) Gilmer, G. H.; Bennema, P. Simulation of Crystal Growth with Surface Diffusion. *J. Appl. Phys.* **1972**, *43*, 1347–1360.
- (15) Becker, R.; Döring, W. Kinetische Behandlung Der Keimbildung in Übersättigten Dämpfen. *Ann. Phys.* **1935**, *416*, 719–752.
- (16) Weeks, J. D.; Gilmer, G. H. *Advances in Chemical Physics*; John Wiley & Sons, Ltd, 1979; pp 157–228.
- (17) Vekilov, P. G. Dense Liquid Precursor for the Nucleation of Ordered Solid Phases from Solution. *Cryst. Growth Des.* **2004**, *4*, 671–685.
- (18) Kaissaratos, M.; Filobelo, L.; Vekilov, P. G. Two-Step Crystal Nucleation Is Selected Because of a Lower Surface Free Energy Barrier. *Cryst. Growth Des.* **2021**, *21*, 5394–5402.
- (19) Sleutel, M.; Van Driessche, A. E. S. Role of Clusters in Nonclassical Nucleation and Growth of Protein Crystals. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E546–E553.
- (20) Xie, Y.; Sohn, S.; Wang, M.; Xin, H.; Jung, Y.; Shattuck, M. D.; O'Hern, C. S.; Schroers, J.; Cha, J. J. Supercluster-Coupled Crystal Growth in Metallic Glass Forming Liquids. *Nat. Commun.* **2019**, *10*, 915.
- (21) De Yoreo, J. J.; Gilbert, P. U. P. A.; Sommerdijk, N. A. J. M.; Penn, R. L.; Whitelam, S.; Joester, D.; Zhang, H.; Rimer, J. D.; Navrotsky, A.; Banfield, J. F.; Wallace, A. F.; Michel, F. M.; Meldrum, F. C.; Cölfen, H.; Dove, P. M. Crystallization by Particle Attachment in Synthetic, Biogenic, and Geologic Environments. *Science* **2015**, *349*, aaa6760.
- (22) Li, D.; Nielsen, M. H.; Lee, J. R. I.; Frandsen, C.; Banfield, J. F.; De Yoreo, J. J. Direction-Specific Interactions Control Crystal Growth by Oriented Attachment. *Science* **2012**, *336*, 1014–1018.
- (23) Savage, J. R.; Dinsmore, A. D. Experimental Evidence for Two-Step Nucleation in Colloidal Crystallization. *Phys. Rev. Lett.* **2009**, *102*, 198302.
- (24) Fang, H.; Hagan, M. F.; Rogers, W. B. Two-Step Crystallization and Solid–Solid Transitions in Binary Colloidal Mixtures. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 27927–27933.
- (25) Russo, J.; Tanaka, H. Crystal Nucleation as the Ordering of Multiple Order Parameters. *J. Chem. Phys.* **2016**, *145*, 211801.
- (26) Walker, J. M.; Marzec, B.; Nudelman, F. Solid-State Transformation of Amorphous Calcium Carbonate to Aragonite Captured by CryoTEM. *Angew. Chem. Int. Ed.* **2017**, *56*, 11740–11743.
- (27) ten Wolde, P. R.; Frenkel, D. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Science* **1997**, *277*, 1975–1978.
- (28) Samanta, D.; Klajn, R. Clathrates Grow Up. *Science* **2017**, *355*, 912.
- (29) Doye, J. P. K.; Noya, E. G. Computational Explorations in the Space of One-Component Crystals. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2107024118.
- (30) Gebauer, D.; Kellermeier, M.; Gale, J. D.; Bergström, L.; Cölfen, H. Pre-Nucleation Clusters as Solute Precursors in Crystallisation. *Chem. Soc. Rev.* **2014**, *43*, 2348–2371.
- (31) de Graaf, J.; Manna, L. A Roadmap for the Assembly of Polyhedral Particles. *Science* **2012**, *337*, 417–418.
- (32) Lin, H.; Lee, S.; Sun, L.; Spellings, M.; Engel, M.; Glotzer, S. C.; Mirkin, C. A. Clathrate Colloidal Crystals. *Science* **2017**, *355*, 931–935.
- (33) Reddy, A.; Buckley, M. B.; Arora, A.; Bates, F. S.; Dorfman, K. D.; Grason, G. M. Stable Frank–Kasper Phases of Self-Assembled, Soft Matter Spheres. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 10233–10238.
- (34) Mirkin, C. A.; Letsinger, R. L.; Mucic, R. C.; Storhoff, J. J. A DNA-based Method for Rationally Assembling Nanoparticles into Macroscopic Materials. *Nature* **1996**, *382*, 607–609.

- (35) Macfarlane, R. J.; Lee, B.; Jones, M. R.; Harris, N.; Schatz, G. C.; Mirkin, C. A. Nanoparticle Superlattice Engineering with DNA. *Science* **2011**, *334*, 204–208.
- (36) Biffi, S.; Cerbino, R.; Bomboi, F.; Paraboschi, E. M.; Asselta, R.; Sciortino, F.; Bellini, T. Phase Behavior and Critical Activated Dynamics of Limited-Valence DNA Nanostars. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 15633–15637.
- (37) Iacovella, C. R.; Keys, A. S.; Glotzer, S. C. Self-Assembly of Soft-Matter Quasicrystals and Their Approximants. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 20935–20940.
- (38) Keys, A. S.; Glotzer, S. C. How Do Quasicrystals Grow? *Phys. Rev. Lett.* **2007**, *99*, 235503.
- (39) Engel, M.; Trebin, H.-R. Self-Assembly of Monatomic Complex Crystals and Quasicrystals with a Double-Well Interaction Potential. *Phys. Rev. Lett.* **2007**, *98*, 225505.
- (40) Engel, M.; Damasceno, P. F.; Phillips, C. L.; Glotzer, S. C. Computational Self-Assembly of a One-Component Icosahedral Quasicrystal. *Nat. Mater.* **2015**, *14*, 109–116.
- (41) Dshemuchadse, J.; Damasceno, P. F.; Phillips, C. L.; Engel, M.; Glotzer, S. C. Moving beyond the Constraints of Chemistry via Crystal Structure Discovery with Isotropic Multiwell Pair Potentials. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2024034118.
- (42) Pan, H.; Dshemuchadse, J. Targeted Discovery of Low-Coordinated Crystal Structures via Tunable Particle Interactions. *ACS Nano* **2023**, *17*, 7157–7169.
- (43) Lee, S.; Bluemle, M. J.; Bates, F. S. Discovery of a Frank-Kasper σ Phase in Sphere-Forming Block Copolymer Melts. *Science* **2010**, *330*, 349–353.
- (44) Cho, B.-K.; Jain, A.; Gruner, S. M.; Wiesner, U. Mesophase Structure-Mechanical and Ionic Transport Correlations in Extended Amphiphilic Dendrons. *Science* **2004**, *305*, 1598–1601.
- (45) Ungar, G.; Zeng, X. Frank–Kasper, Quasicrystalline and Related Phases in Liquid Crystals. *Soft Matter* **2005**, *1*, 95–106.
- (46) Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. Bond-Orientational Order in Liquids and Glasses. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1983**, *28*, 784–805.
- (47) Lechner, W.; Dellago, C. Accurate Determination of Crystal Structures Based on Averaged Local Bond Order Parameters. *J. Chem. Phys.* **2008**, *129*, 114707.
- (48) Du, C. X.; van Anders, G.; Newman, R. S.; Glotzer, S. C. Shape-Driven Solid–Solid Transitions in Colloids. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E3892–E3899.
- (49) Mickel, W.; Kapfer, S. C.; Schröder-Turk, G. E.; Mecke, K. Shortcomings of the Bond Orientational Order Parameters for the Analysis of Disordered Particulate Matter. *J. Chem. Phys.* **2013**, *138*, 044501.
- (50) Spellings, M.; Glotzer, S. C. Machine learning for crystal identification and discovery. *AIChE J.* **2018**, *64*, 2198–2206.
- (51) Wang, L. Discovering Phase Transitions with Unsupervised Learning. *Phys. Rev. B* **2016**, *94*, 195105.
- (52) Wetzel, S. J. Unsupervised Learning of Phase Transitions: From Principal Component Analysis to Variational Autoencoders. *Phys. Rev. E* **2017**, *96*, 022140.
- (53) Reinhart, W. F. Unsupervised Learning of Atomic Environments from Simple Features. *Comput. Mater. Sci.* **2021**, *196*, 110511.
- (54) Wang, Y.; Deng, W.; Huang, Z.; Li, S. Descriptor-Free Unsupervised Learning Method for Local Structure Identification in Particle Packings. *J. Chem. Phys.* **2022**, *156*, 154504.
- (55) Pythia. Glotzer Group, 2022. <https://github.com/glotzerlab/pythia> (accessed Nov 11, 2022).
- (56) Adorf, C. S.; Moore, T. C.; Melle, Y. J. U.; Glotzer, S. C. Analysis of Self-Assembly Pathways with Unsupervised Machine Learning Algorithms. *J. Phys. Chem. B* **2020**, *124*, 69–78.
- (57) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.
- (58) Pedregosa, F.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (59) Coli, G. M.; Dijkstra, M. An Artificial Neural Network Reveals the Nucleation Mechanism of a Binary Colloidal AB13 Crystal. *ACS Nano* **2021**, *15*, 4335–4346.
- (60) Aroboto, B.; Chen, S.; Hsu, T.; Wood, B. C.; Jiao, Y.; Chapman, J. Universal and Interpretable Classification of Atomistic Structural Transitions via Unsupervised Graph Learning. *Appl. Phys. Lett.* **2023**, *123*, 094103.
- (61) Banik, S.; Dhabal, D.; Chan, H.; Manna, S.; Cherukara, M.; Molinero, V.; Sankaranarayanan, S. K. R. S. CEGANN: Crystal Edge Graph Attention Neural Network for Multiscale Classification of Materials Environment. *npj Comput. Mater.* **2023**, *9*, 23.
- (62) Spellings, M. Flowws-Analysis, 2022. <https://github.com/klarh/flowws-analysis> (accessed 03 Feb 2023).
- (63) Nelson, D. R. Order, Frustration, and Defects in Liquids and Glasses. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1983**, *28*, 5515–5535.
- (64) Spaepen, F. Five-Fold Symmetry in Liquids. *Nature* **2000**, *408*, 781–782.
- (65) Hafner, J.; Kahl, G. The Structure of the Elements in the Liquid State. *J. Phys. F: Met. Phys.* **1984**, *14*, 2259–2278.
- (66) Nelson, D. R.; Halperin, B. I. Dislocation-Mediated Melting in Two Dimensions. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1979**, *19*, 2457–2484.
- (67) Nicholas, T. C.; Stones, A. E.; Patel, A.; Michel, F. M.; Reeder, R. J.; Aarts, D. G. A. L.; Deringer, V. L.; Goodwin, A. L. Geometrically Frustrated Interactions Drive Structural Complexity in Amorphous Calcium Carbonate. *Nat. Chem.* **2024**, *16*, 36–41.
- (68) Frank, F. C.; Kasper, J. S. Complex Alloy Structures Regarded as Sphere Packings. I. Definitions and Basic Principles. *Acta Crystallogr.* **1958**, *11*, 184–190.
- (69) Damasceno, P. F.; Engel, M.; Glotzer, S. C. Predictive Self-Assembly of Polyhedra into Complex Structures. *Science* **2012**, *337*, 453–457.
- (70) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (71) Allen, R. J.; Valeriani, C.; ten Wolde, P. R. Forward Flux Sampling for Rare Event Simulations. *J. Phys.: Condens. Matter* **2009**, *21*, 463102.
- (72) Li, L.; Goodrich, C.; Yang, H.; Phillips, K. R.; Jia, Z.; Chen, H.; Wang, L.; Zhong, J.; Liu, A.; Lu, J.; Shuai, J.; Brenner, M. P.; Spaepen, F.; Aizenberg, J. Microscopic Origins of the Crystallographically Preferred Growth in Evaporation-Induced Colloidal Crystals. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2107588118.
- (73) Spaepen, F. A Structural Model for the Solid-Liquid Interface in Monatomic Systems. *Acta Metall.* **1975**, *23*, 729–743.
- (74) Anderson, J. A.; Glaser, J.; Glotzer, S. C. HOOMD-blue: A Python Package for High-Performance Molecular Dynamics and Hard Particle Monte Carlo Simulations. *Comput. Mater. Sci.* **2020**, *173*, 109363.
- (75) Adorf, C. S.; Dodd, P. M.; Ramasubramani, V.; Glotzer, S. C. Simple Data and Workflow Management with the Signac Framework. *Comput. Mater. Sci.* **2018**, *146*, 220–229.
- (76) Ramasubramani, V.; Adorf, C. S.; Dodd, P. M.; Dice, B. D.; Glotzer, S. C. Signac: A Python Framework for Data and Workflow Management. In *Proceedings of the 17th Python in Science Conference*; SCIPY, 2018; pp 152–159.
- (77) Ramasubramani, V.; Dice, B. D.; Harper, E. S.; Spellings, M. P.; Anderson, J. A.; Glotzer, S. C. F. freud: A software suite for high throughput analysis of particle simulation data. *Comput. Phys. Commun.* **2020**, *254*, 107275.
- (78) Brunner, G. O.; Schwarzenbach, D. Zur Abgrenzung der Koordinationssphäre und Ermittlung der Koordinationszahl in Kristallstrukturen. *Z. Kristallogr.* **1971**, *133*, 127–133.
- (79) Lazar, E. A.; Lu, J.; Rycroft, C. H. Voronoi Cell Analysis: The Shapes of Particle Systems. *Am. J. Phys.* **2022**, *90*, 469–480.
- (80) Pólya, G. *Mathematics and Plausible Reasoning*; Princeton University Press, 2020; Vol. 1, pp 188–189.

(81) Martirosyan, M. M.; Spellings, M.; Pan, H.; Dshemuchadse, J. Local structural features elucidate crystallization of complex structures. 2024, <http://arxiv.org/abs/2401.13765>, arXiv:2401.13765 [cond-mat] (accessed May 02, 2024).

(82) Blaiszik, B.; Chard, K.; Pruyne, J.; Ananthakrishnan, R.; Tuecke, S.; Foster, I. The Materials Data Facility: Data Services to Advance Materials Science Research. *J. Miner. Met. Mater. Soc.* **2016**, 68, 2045–2052.

(83) Martirosyan, M. M.; Spellings, M.; Pan, H.; Dshemuchadse, J. Local structural features elucidate crystallization of complex structures. *DataCite Commons* **2024**, <https://commons.datacite.org/doi.org/10.18126/wy01-4e11>.

(84) Local structural features, 2024. <https://github.com/capecrystal/local-structural-features> (accessed May 05, 2024).