

On the challenge of sampling multiple nucleation pathways: A case study of heterogeneous ice nucleation on FCC (211) surface

Cite as: J. Chem. Phys. **158**, 124501 (2023); <https://doi.org/10.1063/5.0144712>

Submitted: 31 January 2023 • Accepted: 02 March 2023 • Accepted Manuscript Online: 06 March 2023
• Published Online: 22 March 2023

 Wanyu Zhao and  Tianshu Li



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Binary salt structure classification with convolutional neural networks: Application to crystal nucleation and melting point calculations

The Journal of Chemical Physics **157**, 204108 (2022); <https://doi.org/10.1063/5.0122274>

Molecular simulations of heterogeneous ice nucleation. I. Controlling ice nucleation through surface hydrophilicity

The Journal of Chemical Physics **142**, 184704 (2015); <https://doi.org/10.1063/1.4919714>

Kinetic coefficient for ice–water interface from simulated non-equilibrium relaxation at coexistence

The Journal of Chemical Physics **157**, 174502 (2022); <https://doi.org/10.1063/5.0124848>



Time to get excited.

Lock-in Amplifiers – from DC to 8.5 GHz



Find out more



Zurich
Instruments

On the challenge of sampling multiple nucleation pathways: A case study of heterogeneous ice nucleation on FCC (211) surface

Cite as: J. Chem. Phys. 158, 124501 (2023); doi: 10.1063/5.0144712

Submitted: 31 January 2023 • Accepted: 2 March 2023 •

Published Online: 22 March 2023



Wanyu Zhao and Tianshu Li^{a)}

AFFILIATIONS

Department of Civil and Environmental Engineering, George Washington University, Washington, District of Columbia 20052, USA

Note: This paper is part of the JCP Special Topic on Nucleation: Current Understanding Approaching 150 Years After Gibbs.

^{a)}Author to whom correspondence should be addressed: tsli@email.gwu.edu

ABSTRACT

Heterogeneous ice nucleation is ubiquitous but its microscopic mechanisms can be extraordinarily complex even on a simple surface. Such complexity poses a challenge in modeling nucleation using advanced sampling methods. Here, we investigate heterogeneous ice nucleation on an FCC (211) surface by a forward flux sampling (FFS) method to understand how the complexity in nucleation pathways affects the accuracy of FFS. We first show the commonly adopted, size-based order parameter fails to describe heterogeneous ice nucleation on the FCC (211) surface. Inclusion of geometric anisotropy of ice nucleus as an additional descriptor is found to significantly improve the quality of the size-based order parameter for the current system. Subsequent application of this new order parameter in FFS identifies two competing ice nucleation pathways in the system: a primary-prism-planed (PPP) path and a secondary-prism-planed (SPP) path, both leading to the formation of hexagonal ice but with different crystalline orientations. Although the PPP pathway dominates ice nucleation on the FCC (211) surface, the occurrence of the less efficient SPP pathway, particularly its strong presence at the early stage of FFS, is found to yield a significant statistical uncertainty in the calculated FFS rate constant. We develop a two-path model that enables gaining a general, quantitative understanding of the impact of initial finite sampling on the reliability of FFS calculations in the presence of multiple nucleation pathways. Our study also suggests a few general strategies for improving the accuracy of FFS when exploring unknown but complex systems.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0144712>

I. INTRODUCTION

Ice nucleation is a vital natural phenomenon, which plays a significant role in global climate change,¹ cell cryopreservation,² and drug production.³ Despite its ubiquity and importance, ice nucleation is also a complex process where the underlying microscopic mechanisms can be sophisticated, particularly in heterogeneous ice nucleation. A vivid manifestation of such complexity is the diversity in the physical and chemical nature of those identified efficient ice nucleation centers. From minerals to biological macromolecules, all these materials are known for being able to efficiently catalyze ice formation, despite a wide range of surface characters.⁴ Subtle

modifications on the surface can also drastically change the efficacy of ice nucleation,^{5–8} highlighting the delicacy of heterogeneous ice nucleation.^{9,10}

Despite the complexity and diversity in the microscopic factors controlling the formation of ice embryos, extensive modeling studies have suggested ice nucleation itself is rather a one-step process where nucleation occurs in a single activated step, including both homogeneous^{11–20} and heterogeneous nucleation.^{5,7,8,21–27} The size of ice nucleus, which is defined as the number of ice-like water molecules calculated based on local bond-order parameters,^{11,28} has been shown to serve as the best reaction coordinate for ice nucleation.^{17,23} The one-step nature and the applicability of

size-based order parameter strongly suggest ice nucleation is a process that can be well described by classical nucleation theory (CNT).^{11,21} Indeed, studies showed homogeneous ice nucleation leads to stacking disordered ice^{11,17,29–32} and when corrected in its driving force by including size-dependence, ice nucleation pathway can be satisfactorily depicted by CNT.¹⁷ For heterogeneous ice nucleation on the graphitic surface,^{21,23} organic crystals,²⁶ and ice-binding proteins,^{25,26} modeling studies based on the mW model³³ also show that ice nucleation follows a classical pathway on different substrates.

These findings provided a justification for using nucleus size as an effective order parameter to drive advanced sampling approaches such as forward flux sampling (FFS)³⁴ to acquire a quantitative description of ice nucleation (e.g., nucleation rate) under various thermodynamic conditions,^{8,11,12,15,18,20,21,24,27,35–39} particularly for those where nucleation event becomes inaccessible by standard molecular dynamics (MD) simulations. An advantage of FFS is that its accuracy does not require its order parameter to be exactly the actual reaction coordinate.⁴⁰ This flexibility enables the application of a less specific, but more inclusive order parameter in FFS that can include as many potential pathways as possible, which can be particularly beneficial for discovering unknown nucleation mechanisms or pathways. However, the exploratory benefit of FFS may come at an expense of compromised efficiency or even accuracy because sampling efficiency can be significantly decreased when an order parameter significantly deviates the actual reaction coordinates. In practice, with only a finite sampling that can be afforded, the most relevant pathway sometimes can be significantly under-sampled if phase points at the initial milestones are not sufficiently collected, particularly when there exist multiple competing pathways.¹⁶ In this regard, a commonly adopted order parameter in modeling crystal nucleation is the size of the nucleus, as it naturally measures the progress of nucleation.⁴¹ However, other factors, such as shape and structure, have been shown to play a key role in determining the critical state of crystal nucleation.^{42,43} Furthermore, as crystal size can be a loosely defined quantity and its determination critically depends on the identity of crystalline-like atoms/molecules, adopting crystal size as an order parameter in an unknown system can be of risk, even if the same order parameter has been proven effective and accurate in a closely related system. In this work, we demonstrate this ineffectiveness through a case study by investigating heterogeneous ice nucleation on a Lennard-Jones (LJ) FCC (211) surface using FFS.

Previous direct MD simulations employing the mW model showed ice nucleates spontaneously on an FCC (211) surface.^{7,10,22} The enhancement of ice nucleation was attributed to the topography of the FCC (211) surface that facilitates a structural match to the primary prism plane of hexagonal ice I_h ,⁷ thus promoting the formation of pure I_h . This constitutes an interesting system that contrasts the heterogeneous ice nucleation on graphene that facilitates the formation of the ice basal plane.⁴⁴ In a graphene–water system, the nucleation of ice, which yields a stacking disordered ice, was found to be accurately characterized by the size of the ice nucleus.^{21,23}

Surprisingly, we find that the size-based order parameter leads to extremely low nucleation rates or even completely fails to drive FFS to obtain the relevant ice nucleation trajectories on FCC (211) surface. The inapplicability of this size-based order parameter is

attributed to the fact that the FCC (211) surface promotes an interfacial ordering of water that is distinguished from ice I but cannot be properly differentiated from ice I by the local bond-order parameter commonly used to identify crystalline order.^{11,28} Employing a collective variable based on such local order parameter, thus, yields an ensemble of phase points that are virtually composed of “disk-like” seeds that cannot grow into ice. Inclusion of the geometric anisotropy of ice nucleus as an additional collective variable is found to significantly mitigate the issue. Using this new set of order parameters, we find that the nucleation of ice on FCC (211) with a lattice constant of 4.158 Å can proceed via two distinct pathways, both leading to hexagonal ice I_h . One pathway forms the primary prism plane (PPP) of I_h parallel to FCC (211) surface, consistent with the structure previously identified in direct MD simulations.^{7,22} The other pathway, which was absent from these direct MD simulations, yields the secondary prism plane (SPP) of I_h . Although the PPP path has a significantly higher nucleation rate than the SPP path, thus dominating ice nucleation, the less-efficient SPP path is found to overwhelm the ensemble of configurations collected at the early milestones of FFS modeling, which can lead to a significant variation in the calculated nucleation rate constants that exhibit an undesirable dependence on the choice of FFS parameters. To rationalize our finding, we subsequently develop a two-path model to understand the convergence issue of the FFS rate constant when two pathways with distinct rate constants co-exist. The model, which depicts the variation of the final FFS rate constant with respect to both the level of initial sampling of FFS and the difference in rate constant between the two co-existing paths, provides a general strategy for improving the reliability of FFS for modeling those systems.

II. METHOD

A. Molecular dynamics simulation

Molecular dynamics simulation is performed employing the coarse-grained mW model³³ through the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package.⁴⁵ The interaction between mW water and atoms in an FCC (211) substrate is described by a truncated LJ potential, with $\epsilon_{ws} = 0.48$ kcal/mol, $\sigma_{ws} = 2.582$ Å, and cutoff distance $r_c = 7.53$ Å. The lattice constant of the FCC substrate is set to be $a_{\text{FCC}} = 4.158$ Å with all FCC atoms frozen. With these parameters, previous direct MD simulations^{7,22} showed hexagonal ice I_h forms spontaneously with its primary prism plane parallel to the FCC (211) surface. A total of 4000 mW molecules and 960 surface atoms are included in a simulation cell with a dimension of $58.8 \times 57.6 \times 70$ Å³. With a periodic boundary condition applied, a water slab about 35 Å along the z axis is created between an FCC–water interface and a vacuum–water interface. The isothermal canonical ensemble (NVT) with a Nose–Hoover thermostat is employed throughout our simulations, with an equilibrium time of 0.5 ns at the target temperature of 230 K. A time step of 5 fs is used throughout this work.

B. Forward flux sampling method

Heterogeneous ice nucleation rates on the FCC (211) surface are calculated using our recent implementation of FFS in LAMMPS.

In FFS, the transition pathway is decomposed into consecutive segments through milestones based on order parameter λ .³⁴ Nucleation rate is then given by $R = \Phi_{\lambda_0} \prod_{i=1}^n P(\lambda_i|\lambda_{i-1})$,⁴⁶ where Φ_{λ_0} is the initial flux rate crossing the initial milestone λ_0 , and $P(\lambda_i|\lambda_{i-1})$ is the crossing probability for a trajectory starting from milestone λ_{i-1} and successfully reaching the next adjacent milestone λ_i . In practice, FFS is carried out through two steps. In the first step, the initial flux rate Φ_{λ_0} is obtained by N_0/t_0V , where N_0 is the number of successful crossings collected at λ_0 , t_0 is the total simulation time, and V is the volume of the water slab. Sufficient sampling of the initial flux has been demonstrated to be critical for ensuring the convergence of the final rate constant,¹⁶ and its role will be discussed in depth in the current system. To ensure sufficient sampling, a total of 1000 ~ 3000 configurations (~1 μ s of simulation time) are collected at the first milestone λ_0 . In the second step, the crossing probability $P(\lambda_i|\lambda_{i-1}) = N_i/M_{i-1}$ is computed by firing a large number (M_{i-1}) of MD trial shootings starting from the milestone λ_{i-1} and collecting N_i successful crossings (~120) at the milestone λ_i . The statistical uncertainty of $P(\lambda_i|\lambda_{i-1})$ is mainly attributed to both the variance of binomial distributions of the number of configurations at each milestone and the landscape variance of the configurations collected at the previous milestone.⁴⁷ However, for a system that exhibits multiple nucleation pathways, the statistical uncertainty due to finite sampling can be more significant in the final rate constant R . Therefore, multiple independent FFS runs (e.g., $m = 5$) are carried out to obtain the geometric mean $\langle R \rangle = (\prod_i^m R_i)^{1/m}$ and the standard error of $\ln \langle R \rangle$ is given by $\sqrt{\sum_i^m (\ln R_i - \ln \langle R \rangle)^2 / m}$. It should be noted that the heterogeneous nucleation rate on the surface should be measured by area, not by volume. However, since the volume of water is small and there is only one nucleation event occurring, a volume-based nucleation rate is used instead to enable a direct comparison between homogeneous and heterogeneous nucleation.

The milestones in FFS are defined by order parameter(s) λ . In the case of ice nucleation, λ has been conventionally chosen to be the number of the ice-like water molecules contained in the largest ice cluster.^{8,11,12,15,21,24,35,37} This definition has been shown to serve well to describe the reaction coordinates of both homogeneous and heterogeneous ice nucleation. In fact, prior investigations further showed that mW ice nucleation behaviors can be well described by classical nucleation theory that assumes nucleus size as the sole order parameter.^{8,11,17,21,23} However, in the current system with the FCC (211) surface, we find that this size-based order parameter fails to describe heterogeneous mW ice nucleation. As shown in Fig. 1, the calculated growth probability $P(\lambda|\lambda_0)$ in FFS employing the conventional size-based order parameter λ becomes extremely low, leading to an inability to grow ice. By analyzing those ice nuclei collected at the initial milestones of FFS runs, we find ice nuclei all exhibit a “disk-like” geometry that spreads out on an FCC (211) surface, as shown in Fig. 1. The 2-D nature of these ice seeds is attributed to both the complex surface topography of the FCC (211) surface and the convention used to define an “ice-like” water molecule. On the one hand, the arrangement of surface atoms on the FCC (211) surface creates a strongly ordered interfacial water layer that does not resemble the structure of ice I (either I_h or I_c).⁷ This is similar to the interfacial ordering previously identified on a phloroglucinol dihydrate surface that seamlessly connects ice and organic surface but

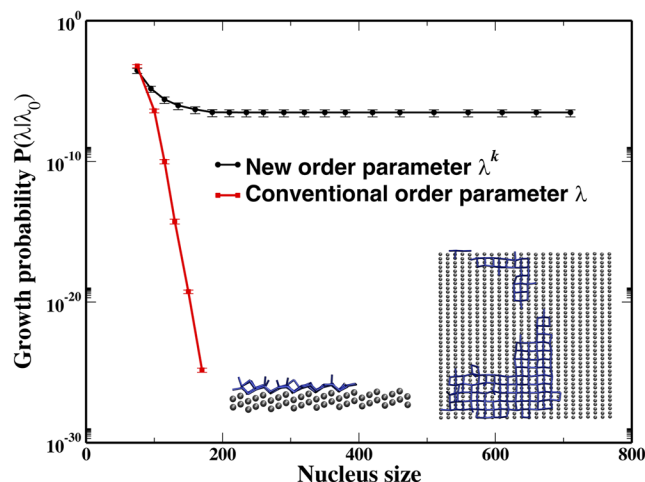


FIG. 1. Calculated growth probability $P(\lambda|\lambda_0)$ of ice nucleus forming on FCC (211) surface as a function of the conventional, size-based order parameter λ (red) and the new, size-based order parameter combined with a geometric constraint λ^k (black). Insets show the side view (left) and top view (right) of the “disk-like” ice nucleus, as the only structure identified through the conventional λ . FCC (211) atoms and ordered water molecules are colored as gray and blue, respectively.

does not resemble any ice face;²⁶ On the other hand, water molecules in this ordered interfacial layer are found to indeed satisfy the commonly used criterion for being “ice-like” water, i.e., a $q_6 > 0.5$ with four hydrogen-bonded neighboring water molecules.¹¹ Therefore, water molecules in this interfacial layer are incorrectly labeled as “ice-like” with ease, leading to a strong, artificial preference for the in-plane geometry of those ice seeds selected by the local bond-order parameter. In fact, a detailed analysis of the structures shows that only three out of 5000 seeds collected at λ_0 are truly “ice-like,” while the rest are all “disk-like.” The subsequent FFS shootings based on this ensemble can only lead to horizontal but not vertical growth because these “disk-like” seeds do not possess the correct structural template to grow into ice I.

Since the interfacial water molecules do exhibit an orientational order but are not “ice-like,” tuning the cutoff value of a local bond-order parameter like q_6 is unlikely to resolve their true non-“ice-like” identity. After all, the essence of the Steinhardt-type bond-order parameter is that it should be sensitive to a general but not specific crystalline order.⁴⁸ Therefore, the remedy to the issue should call for an additional descriptor that enables screening out these ice seeds that cannot truly grow. In this regard, an important structural attribute of these “disk-like” seeds is their shape, which can be reflected by the maximum ratio in the resolved components of the radius of gyration for an ice seed, i.e., $k = \max(R_g^\alpha/R_g^\beta)$, where $R_g^\alpha = 1/N \sqrt{\sum_{i>j}^N (r_{ij}^\alpha)^2}$. Here, N is the number of water molecules in an ice cluster, $\alpha, \beta = x, y, \text{ or } z$ and r_{ij}^α is the α component of the distance between molecule i and j . Therefore k reflects the geometric anisotropy of a cluster, as an isotropic geometry yields $k \sim 1$ and a large k corresponds to a strong geometric anisotropy. Using this definition, we indeed find the average k of ice nuclei collected at λ_0

to be around 8.5, indicating their strong geometric anisotropy (see Fig. S1 in the [supplementary material](#)).

Conceptually, a constraint on k thus can potentially prevent those “disk-like” seeds from “poisoning” the ensemble and allow identifying the true, “healthy” ice seeds that will grow. To implement this idea, we thus introduce a combined order parameter λ^k as follows: For a given configuration, the largest ice cluster is first identified, following the same procedure as described in Ref. 11. Then, shape screening is conducted so that if this cluster yields a k greater than a pre-defined cut-off k_{cut} , the next largest ice-like cluster is selected and then goes through the same shape screening. The procedure is repeated until the largest ice-like cluster with a $k \leq k_{\text{cut}}$ is identified, and its size is defined as λ^k .

III. RESULT

A. Effect of k_{cut} on ice nucleation rate

To understand whether the new order parameter λ^k mitigates the sampling issue, we compute heterogeneous ice nucleation rates as a function of k_{cut} by FFS at 230 K, as shown in Fig. 2(a). Clearly, a large k_{cut} (>5) restores the conventional, size-based order parameter, thus leading to an extremely low nucleation rate (not shown in Fig. 2). On the contrary, a small k_{cut} (<2) places a very strong geometric constraint that becomes incompatible with the natural geometry of a nucleus forming on a substrate via heterogeneous nucleation. For example, heterogeneous nucleation typically proceeds with the formation of nucleus with a spherical cap or a similar shape,^{21,23} which is inherently anisotropic. Therefore, a low k_{cut} can lead to a complete miss of the most relevant nucleation pathway. Indeed, the calculated nucleation rate $10^{11 \pm 1} \text{ m}^{-3} \text{ s}^{-1}$ for $k_{\text{cut}} = 1.5$ is found comparable to the homogeneous ice nucleation rate $10^{13} \text{ m}^{-3} \text{ s}^{-1}$ at the same temperature.¹¹ As shown in Fig. 2(b), the ice nucleus formed under $k_{\text{cut}} = 1.5$ contains cubic ice I_c and stacking disorder, which is typical in homogeneous ice nucleation.^{11,17,29} This is in contrast to the structure (I_h with its primary prism plane ($10\bar{1}0$) parallel to the substrate) identified in previous MD simulations on the FCC (211) surface.^{7,22} Although the bottom of the ice nucleus (in the vicinity of a surface) is indeed found to carry a structural signature resembling the primary prism plane of I_h (Fig. 2), the strong geometrical constraint imposed by k_{cut} restricts this layer from growing into I_h .

With a proper range of geometric constraints, the calculated nucleation rates are found to be nearly invariant with a k_{cut} between 3 and 5. The fact that there exists a range of k_{cut} where the calculated rate becomes independent of k_{cut} is desirable and indicates the potential effectiveness of the new order parameter. To further understand whether λ^k can serve as a good approximation to the actual reaction coordinate, we carry out p_B histogram analysis at the critical sizes identified at different k_{cut} . A committor p_B measures the probability of a configuration to reach state B (solid ice in this case); thus, by definition, a critical nucleus should have a p_B of 0.5.⁴⁹ For the identified critical milestone λ^{k*} , we select 120 configurations, each receiving 20 trial runs to compute p_B , to construct p_B histogram. As shown in Fig. 3(a), the calculated distribution for $k_{\text{cut}} = 3.5$ is peaked at the mean $\mu = 0.504$, with a standard deviation σ of 0.17. The overall agreement with the Gaussian distribution of the same μ and σ confirms λ^k with k_{cut} of 3.5 is indeed a good order parameter.⁵⁰

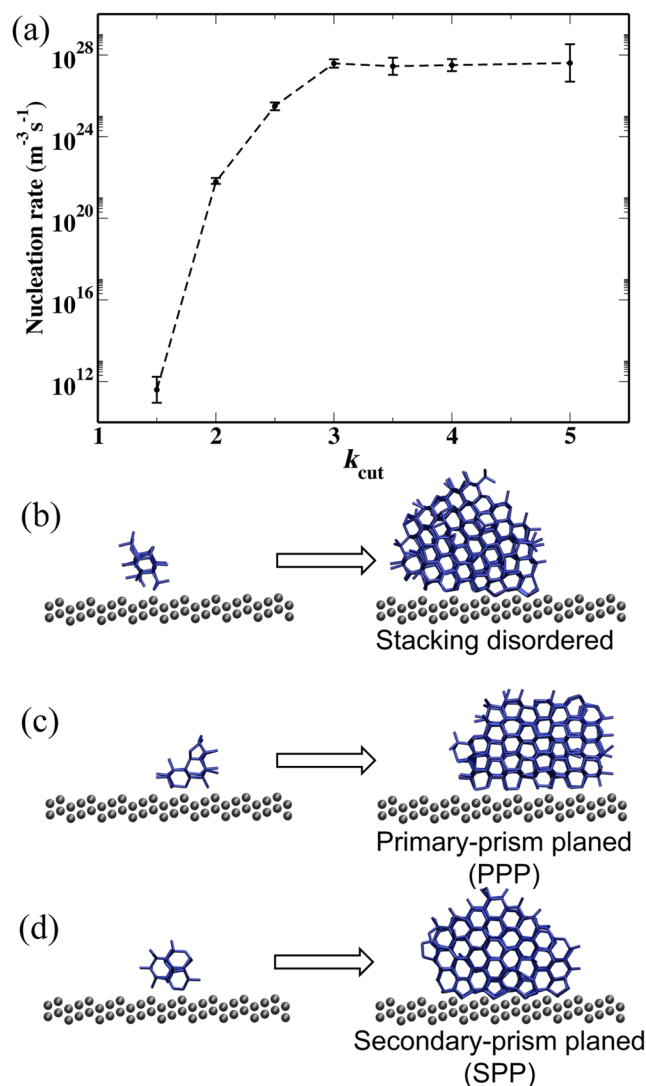


FIG. 2. (a) Calculated heterogeneous ice nucleation rate as a function of the degree of geometric constraint k_{cut} on an FCC (211) surface at 230 K. (b)–(d) Small, pre-critical ice nuclei containing ~ 40 water molecules (on the left) growing into large, post-critical ice clusters containing ~ 500 water molecules (on the right) with different structures. In (b), a stacking disordered ice nucleus forms under a strong geometric constraint $k_{\text{cut}} = 1.5$, accompanied by a low nucleation rate. With a softened constraint ($k_{\text{cut}} \geq 2$), ice nucleation is found to proceed via two pathways, forming I_h with its (c) primary-prism plane (PPP) and (d) secondary-prism plane (SPP) parallel to the FCC (211) surface.

In contrast, the calculated p_B histogram for $k_{\text{cut}} = 2$ is found to exhibit a clear deviation from a Gaussian distribution, suggesting λ^k with $k_{\text{cut}} = 2$ does not well describe the actual reaction coordinate. In particular, the abnormally high distribution of p_B between 0.9 and 1 suggests at $\lambda^{k*} = 550$, there already exist configurations that are strongly committed to basin B . To understand the nature of these configurations, we trace back all the configurations at λ^{k*} to their parents at the first milestone λ_0^k . Interestingly, this structural analysis

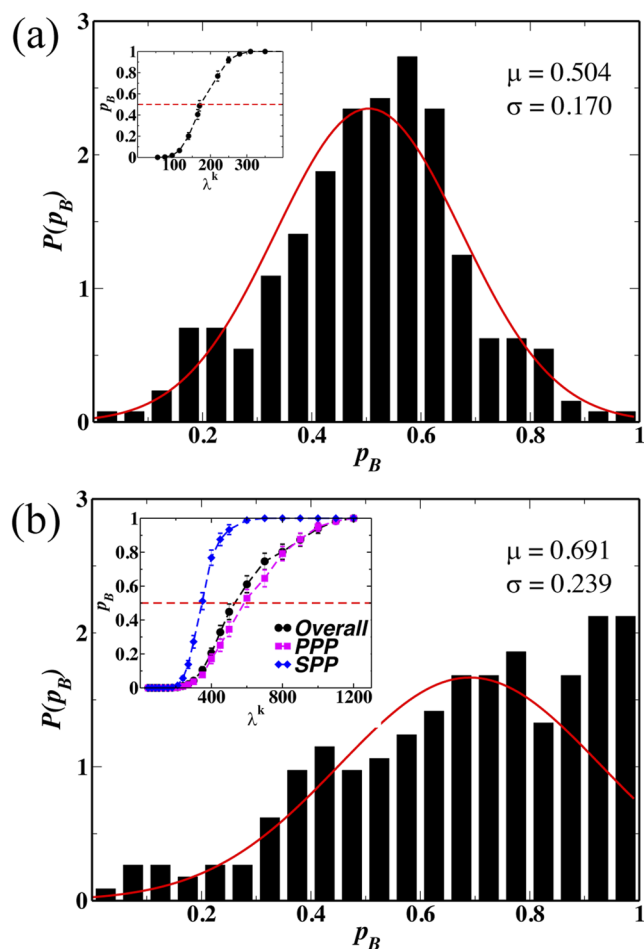


FIG. 3. Calculated p_B histogram for (a) $k_{\text{cut}} = 3.5$ and (b) $k_{\text{cut}} = 2$ at the corresponding critical milestones λ^{k*} . The Gaussian distribution with a mean μ and standard deviation σ determined from the corresponding p_B histogram is shown as the red line. Insets show the calculated committor p_B as a function of order parameter λ^k , which allows determining the critical size λ^{k*} through its intersection with the horizontal dash line corresponding to $p_B = 0.5$. For (a) $k_{\text{cut}} = 3.5$, the overall critical size is determined to be ~ 170 . For (b) $k_{\text{cut}} = 2$, the overall critical size is ~ 550 , but PPP and SPP yield very different critical sizes, being ~ 600 and ~ 350 , respectively.

of nucleation trajectories identifies the co-existence of two distinct nucleation pathways. One pathway leads to the formation of I_h with its primary prism plane (10 $\bar{1}0$) parallel to the surface, as shown in Fig. 2(c). This pathway was previously identified in direct MD simulations,^{7,22} where the enhancement of heterogeneous ice nucleation in this pathway was attributed to the strong geometric match of the surface topography of FCC (211) to the ice double layer. We accordingly denote this pathway as “primary-prism-planed” or PPP. Another pathway, which has not been found from direct MD simulations on FCC (211) surface with the lattice constant of 4.158 Å, is identified from the ensemble of FFS trajectories to form I_h but with its *secondary* prism plane (11 $\bar{2}0$) parallel to the FCC (211) surface [Fig. 2(d)]. This pathway is correspondingly denoted as

“secondary-prism-planed” or SPP. We note that SPP was indeed reported to form spontaneously in direct MD on FCC (211) surface, but with a much greater lattice constant of 4.66 Å.⁷ In fact, the primary and secondary prism planes of I_h share a structural similarity in that both planes exhibit double layers (see Fig. S3 in the [supplementary material](#)). However, the difference in the length of double layers means when growing on FCC (211) surface with a lattice constant of 4.158 Å, PPP structurally matches the periodicity of the surface, while SPP is under compression.

To understand the efficiencies of the two distinct nucleation pathways, we calculate the growth probability $P(\lambda^k|\lambda_0^k)$ for PPP and SPP with different k_{cut} . This is carried out through conducting the second step of FFS by using the PPP and SPP seeds as the only starting configurations at the first milestone λ_0^k , respectively. As shown in Fig. 4, under a strong geometric constraint $k_{\text{cut}} = 2$, the two pathways are found to yield comparable growth probabilities (PPP is only about 1 order of magnitude faster than SPP). However, the shape of the calculated growth probabilities suggests a significant difference between PPP and SPP in their corresponding critical sizes. Indeed, the calculated p_B as a function of λ^k [Fig. 3(b)] shows SPP yields a critical size of ~ 350 , much smaller than that (~ 600) of PPP. Such a large difference means that, for $k_{\text{cut}} = 2$, the critical size $\lambda^{k*} = 550$ determined through the “overall” growth probability becomes inadequate to define the true critical state: At this milestone, the SPP pathway has already well passed its own critical size (~ 350), thus, is strongly committed to the basin B, whereas the PPP pathway is just about to approach its critical state (~ 600). Therefore, in the case where the system contains two nucleation pathways with comparable probabilities to grow but very different critical sizes, the nucleus size itself becomes an ineffective order parameter.

Interestingly, under $k_{\text{cut}} = 2$, the faster route (PPP) is found to exhibit a critical size *larger* than that of the slower route (SPP). This

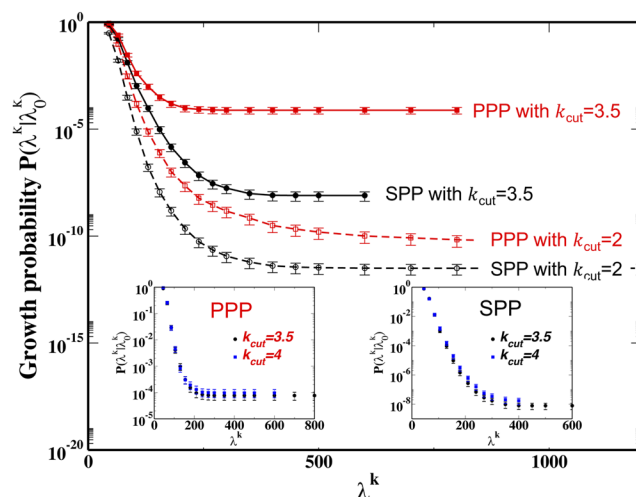


FIG. 4. Calculated growth probability $P(\lambda^k|\lambda_0^k)$ for PPP and SPP paths as a function of λ^k under different geometric constraints k_{cut} . Insets show the convergence of the calculated $P(\lambda^k|\lambda_0^k)$ with respect to applied geometric constraint k_{cut} for both pathways.

counterintuitive behavior is attributed to the applied strong geometric constraint artificially altering the landscape of the nucleation pathway. To recover the natural nucleation pathways, we repeat the partial FFS runs for SPP and PPP with a proper range of k_{cut} where the ice nucleation rate exhibits a plateau (see Fig. 2). As shown in Fig. 4, softening geometric constraint leads to a significant increase in growth probability for both pathways. For the PPP pathway, such increase amounts to 6 orders of magnitude, yielding a converged growth probability to 10^{-4} , consistent with the calculated growth probability from a full FFS simulation with a k_{cut} between 3 and 5. Correspondingly, the critical size for the true PPP pathway under this range of constraint is identified to be ~ 180 , which is also consistent with the critical size determined from the full FFS run [Fig. 3(a)]. For the SPP pathway, the growth probability is found to converge at 10^{-8} (see the insets of Fig. 4), yielding a critical size of ~ 300 . Therefore, as a proper k_{cut} restores the landscape of nucleation pathways, the faster route indeed leads to a smaller critical size. Importantly, the large difference in the growth probability between PPP and SPP (4 orders of magnitude) implies the nucleation of ice on the FCC (211) surface is dominated by PPP. Indeed, the fact that both the partial FFS run based on PPP seeds at the first milestone λ_k^0 and the full FFS run including all possible seeds yield identical growth probability simply means that the overall nucleation rate of mW ice on FCC (211) is determined by the fast PPP pathway. This also explains the absence of the SPP in the previous direct MD simulations,^{7,22} as the 4-orders-of-magnitude difference in rate constant implies only one out of 10 000 successful nucleation trajectories can be of SPP type at this temperature.

The preference of PPP over SPP is not surprising because the spacing between the steps of FCC (211) with $a_{\text{FCC}} = 4.158 \text{ \AA}$ matches well the double layer of the primary prism plane of I_h (see Fig. S3 in the supplementary material). Therefore, the epitaxial growth of the secondary prism plane on this surface will be negatively strained by about 5%, thus should be less unfavorable. However, despite the dominance of the PPP path in the final rate constant, the co-existence of a fast PPP and a slow SPP is found to yield an unexpected convergence issue when acquiring nucleation rate using FFS. In particular, the calculated rate constants from multiple, independent FFS runs exhibit an undesirable variability. Such an issue has not been encountered in our previous investigation of either homogeneous ice nucleation^{11,12} or heterogeneous nucleation on graphene.^{8,21,36} Therefore, it is important to understand the origin for the lack of convergence in the current system.

B. Effect of FFS parameters on ice nucleation rate

An advantage of FFS is its flexibility in defining milestones λ_A and λ_i in that the final rate constant can be virtually independent of the choices of these parameters if they are within a certain range. For example, basin A (liquid in this case) represents a well-defined region in phase space, but there is no unique way in defining its boundary λ_A in practice. One way for defining λ_A is through choosing a λ such that the cumulative probability below λ_A is 0.5, i.e., $\int_0^{\lambda_A} P(\lambda) d\lambda = 0.5$.⁵¹ Alternatively, λ_A can be chosen near the peak of the distribution obtained from a direct simulation near basin A. Despite the arbitrariness in defining λ_A , the calculated FFS rates stay nearly invariant, which renders both the robustness and convenience of the method. Indeed, as shown in Fig. S2

in the supplementary material, the calculated homogeneous nucleation rates for mW ice, each obtained based on ten independent FFS runs at 220 K, show little variation with respect to the position of λ_A .

However, in heterogeneous ice nucleation on the FCC (211) surface, the same test shows the position of λ_A^k does affect the FFS rate. As shown in Fig. 5(a), the calculated geometric mean ice nucleation rate based on five independent runs is found to decrease by one order of magnitude from $10^{26} \text{ m}^{-3} \text{ s}^{-1}$ for $\lambda_A^k = 10$ to $10^{25} \text{ m}^{-3} \text{ s}^{-1}$ for $\lambda_A^k = 13$. The variation of the mean rate may not seem to be significant given the intrinsic statistical uncertainty of the FFS rate constant being of a similar order of magnitude,⁴⁷ but the variability among the five independent FFS runs is found to substantially increase with λ_A^k . In particular, at $\lambda_A^k = 13$, the calculated rates are found to span over five orders of magnitude, ranging from

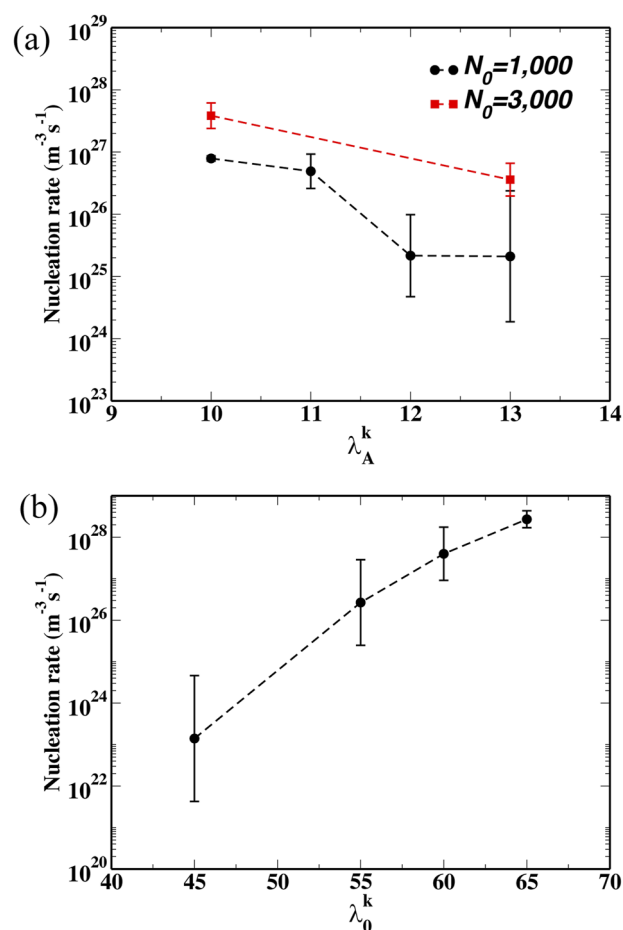


FIG. 5. Variation of the calculated ice nucleation rate on the FCC (211) surface through FFS with respect to the positions of (a) λ_A^k and (b) λ_0^k in FFS. Each data point represents the geometric mean nucleation rate based on five independent FFS runs, with error bars being the standard errors of the mean. k_{cut} is set for 3 and 5 for (a) and (b), respectively. Each FFS run is conducted by collecting 1000 (N_0) initial configurations at λ_0^k . To understand the role of N_0 , FFS tests with 3000 initial configurations (red) are also carried out in (a).

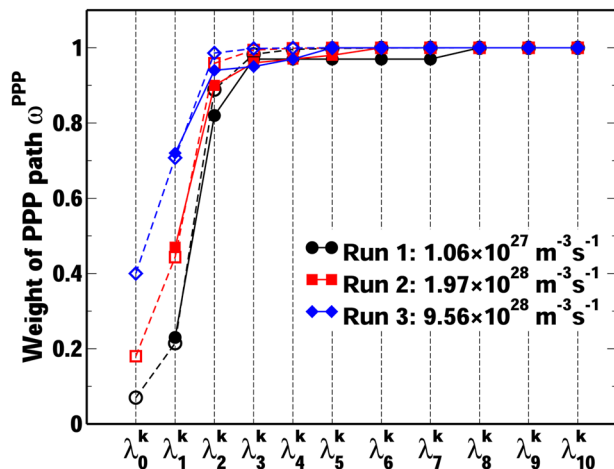


FIG. 6. Comparison between FFS simulations (solid symbols) and the two-path model (open symbols) on the weight of the PPP path (ω^{PPP}) at each milestone. Three independent FFS simulations are carried out using a $k_{\text{out}} = 5$. The corresponding nucleation rates obtained directly from FFS simulations are indicated in the legends.

10^{22} to $10^{27} \text{ m}^{-3} \text{ s}^{-1}$. Similarly, the choice of λ_0^k in preparing the ensemble of initial configurations is also found to significantly affect the final rate constant. Figure 5(b) shows a decreasing λ_0^k not only leads to a steady decrease of the *mean* nucleation rate but also a significant increase in the variability of the rates. The large uncertainty thus raises a concern about the convergence and reliability of the calculated FFS rate based on a single FFS run in the current system.

Since the convergence issue was virtually absent in the FFS studies of homogeneous ice nucleation^{11,12,35,39} and heterogeneous ice nucleation on a graphene surface,^{8,21,36} its origin is expected to be related to the intrinsic dynamics of ice nucleation on the FCC (211) surface. As demonstrated above, although the SPP pathway is significantly less efficient, its nucleation rate is not extremely far from that of the PPP pathway; thus, SPP still can make a non-negligible contribution to the overall ensemble of nucleation trajectories. In

particular, the less-effective SPP pathway is found to have a significant representation at the early stage of the FFS, e.g., λ_0^k and λ_1^k , despite that these SPP seeds have a lower probability than those PPP seeds to grow at a later stage.

To quantify the fraction of the PPP seeds, we carry out a structural analysis for the configurations collected at each milestone. The differentiation between PPP and SPP seeds is based on the difference in the distribution of the projected hydrogen-bond lengths (see the [supplementary material](#) for more details). Although such analysis can be less quantitative for small ice nucleus that barely contains one complete hexagonal core, e.g., those at λ_0^k , the quantification is robust for larger clusters. Indeed, the analysis reveals a clear trend, as shown in Fig. 6, that the overall nucleation rate is highly dependent on the fraction of PPP seeds at the early milestones of FFS. For the FFS run that yields the lowest rate, the entire ensemble is found completely composed of the SPP pathway; in contrast, the highest rate FFS run is found to proceed with a significant fraction of PPP seeds at the early stage, which then dominate the growth afterward. For other runs, a higher fraction of PPP at the early stages of FFS is found to lead to a higher nucleation rate.

Thus, consequently, how well an FFS calculation can capture the true nucleation rate in the current system strongly depends on how well the PPP pathway can be sufficiently sampled at the early stage of FFS. In an ideal situation where an unlimited sampling can be afforded, the weights of all pathways at the initial stage can be correctly sampled, but in practice where only a finite sampling can be obtained, the final FFS rate constant can depend on the “composition” of the fast route collected at the initial milestone λ_0 .

C. Two-path model

To understand how the collection of initial sampling at λ_0 can affect the final FFS rate constant quantitatively, we consider a two-path model where two competing nucleation pathways, path *a* (fast) and path *b* (slow), co-exist in one system and both pathways can be characterized by the same order parameter λ . If we know the full details of each pathway, say, $R_a = \Phi_0^a \prod_{i=0}^{n-1} P_a(\lambda_{i+1}|\lambda_i)$ for path *a*, then the total FFS rate constant R_{tot} can be obtained through the following (see the [supplementary material](#) for its derivations):

$$R_{\text{tot}} = (\Phi_0^a + \Phi_0^b) \prod_{i=1}^{n-1} \left[\frac{\omega_{i-1}^a P_a(\lambda_i|\lambda_{i-1}) P_a(\lambda_{i+1}|\lambda_i) + \omega_{i-1}^b P_b(\lambda_i|\lambda_{i-1}) P_b(\lambda_{i+1}|\lambda_i)}{\omega_{i-1}^a P_a(\lambda_i|\lambda_{i-1}) + \omega_{i-1}^b P_b(\lambda_i|\lambda_{i-1})} \right] \times [\omega_0^a P_a(\lambda_1|\lambda_0) + \omega_0^b P_b(\lambda_1|\lambda_0)], \quad (1)$$

where ω_i^a and ω_i^b are the weights of paths *a* and *b* at the milestone λ_i , respectively, and they are related to the weights ω_{i-1}^a and ω_{i-1}^b at the milestone λ_{i-1} through

$$\omega_i^a = \frac{\omega_{i-1}^a P_a(\lambda_i|\lambda_{i-1})}{\omega_{i-1}^a P_a(\lambda_i|\lambda_{i-1}) + \omega_{i-1}^b P_b(\lambda_i|\lambda_{i-1})}, \quad (2)$$

$$\omega_i^b = \frac{\omega_{i-1}^b P_b(\lambda_i|\lambda_{i-1})}{\omega_{i-1}^a P_a(\lambda_i|\lambda_{i-1}) + \omega_{i-1}^b P_b(\lambda_i|\lambda_{i-1})}, \quad (3)$$

with $\omega_0^a = \Phi_0^a / (\Phi_0^a + \Phi_0^b)$ and $\omega_0^b = \Phi_0^b / (\Phi_0^a + \Phi_0^b)$. Equation (1) can be simplified as

$$R_{\text{tot}} = \Phi_{\text{tot}} P_{\text{tot}} = \Phi_{\text{tot}} (\omega_0^a P_a + \omega_0^b P_b). \quad (4)$$

For ice nucleation on the FCC (211) surface, the fast route *a* and slow route *b* are the PPP and SPP pathways, respectively. Using Eqs. (1)–(3), the total FFS nucleation rate R_{tot} and the weights of both pathways at all milestones can be readily predicted through

the two-path model if one knows the initial weights ω_0 and the full growth probabilities $P(\lambda_i|\lambda_{i-1})$ for both path *a* and *b*. This potentially enables a comparison between the theoretical estimate and the actual FFS simulations to understand the impact of initial sampling on the FFS rate constant. However, a direct comparison may not be feasible because an accurate estimate of ω_0^{PPP} directly from simulation can be difficult. Instead, an indirect comparison can be achieved through fitting an initial weight α_0^{PPP} at λ_0 (as an approximation of the true weight ω_0^{PPP}) such that the predicted weight α_1^{PPP} at λ_1 through Eq. (2) matches ω_1^{PPP} obtained from simulation.

As shown in Fig. 6, the two-path model is found to compare well against the actual FFS runs. For example, an α_0^{PPP} of 0.18 ~ 0.19 from the two-path model is found to not only well reproduce the weights at all milestones (Fig. 6, red line) for FFS run-2, but also yield a predicted R_{tot} of $1.71 \times 10^{28} \text{ m}^{-3} \text{ s}^{-1}$, which is in a good agreement with the FFS rate constant ($1.97 \times 10^{28} \text{ m}^{-3} \text{ s}^{-1}$). Similar agreement can also be found for run-1 ($R_{\text{tot}} = 6.64 \times 10^{27} \text{ m}^{-3} \text{ s}^{-1}$, with $\alpha_0^{\text{PPP}} = 0.07$) and run-3 ($R_{\text{tot}} = 3.80 \times 10^{28} \text{ m}^{-3} \text{ s}^{-1}$, with $\alpha_0^{\text{PPP}} = 0.4$). These results thus confirm the variations in the calculated nucleation rates obtained from independent FFS runs are largely attributed to the difference in the weights of the PPP and SPP paths obtained from different initial samplings.

Importantly, the two-path model provides a comprehensive understanding of how the initial weight of the fast route ω_a^0 and the difference in the growth probability $\eta \equiv P_a/P_b$ affect the final rate constant in FFS. Re-arranging Eq. (4), one obtains $P_{\text{tot}} = P_a[(1 - \eta^{-1})\omega_a^0 + \eta^{-1}] = P_a f$, where the factor $f \equiv P_{\text{tot}}/P_a$ measures the fraction of the total growth probability relative to that of the fast route *a*. Figure 7 displays the variation of *f* with respect to both ω_a^0 and η through the contour of *f*. As shown in Fig. 7(a), when the fast route *a* is sufficiently represented at the initial sampling ($\omega_a^0 > 0.1$), the total growth probability P_{tot} is virtually indistinguishable from P_a (within one order of magnitude), irrespective of η . When the fast route is significantly underrepresented ($\omega_a^0 \ll 1$)

at the initial milestone λ_0 , the degree of underestimation for P_{tot} becomes more complex: for a large η ($\eta \gg 1$), $f \approx \omega_a^0$, as indicated by the vertical contour lines in Fig. 7(b). This means how well FFS reproduces the true rate depends only on how much the fast route is captured during initial sampling when the slow route has a negligibly small rate. This is sensible because, in such a scenario, the entire transition process proceeds essentially as if there exists only one pathway (path *a*) so that the fraction of initial sampling of path *a* ultimately determines *f*; For a small η ($\eta > 1$ but not by a lot), consequently $f \approx 1/\eta$, as indicated by the horizontal contour lines in Fig. 7(b). This is the case where the two pathways are comparable in rates (say path *a* is 100 times faster than path *b*), and if path *a* is severely missed from initial sampling ($\omega_a^0 \ll 1$), then FFS can only catches route *b*, which yields a final rate $R_{\text{tot}} = R_b = (1/\eta)R_a$; Finally, for an intermediate η , the scaling factor *f* depends on both ω_a^0 and η , whose domain is indicated by the shaded arrow in Fig. 7(b). In this case, an increasing ω_a^0 and a decreasing η will lead to an increase in *f*.

Applying the two-path model to the current system [FCC (211)] where $\eta = 10^4 \sim 10^5$, one finds *f* can exhibit a wide variation ($10^{-3} \sim 1$) when ω_a^0 varies between 10^{-3} to 1. It should be noted that the two-path model only considers the role of initial sampling while assuming the determination of weights and growth probabilities at other milestones is not subject to errors. In practice, errors in these quantities introduced due to finite sampling can exacerbate the issue, leading to an *f* even smaller than considered here. The model suggests any factor affecting ω_a^0 can contribute to the variation of the FFS rate. In the current system, these factors include the position of milestones λ_A^k and λ_0^k , and the number of configurations N_0 collected at λ_0^k . Although SPP is the slow route, its propensity near basin A is significantly higher than the fast route PPP, as reflected by the fact that most of the configurations collected at λ_0^k are from SPP. This indicates the free energy landscape for SPP near basin A is much less steep than PPP. Moving away from basin A, the propensity of SPP decreases quickly (Fig. 6) while PPP has a greater probability to

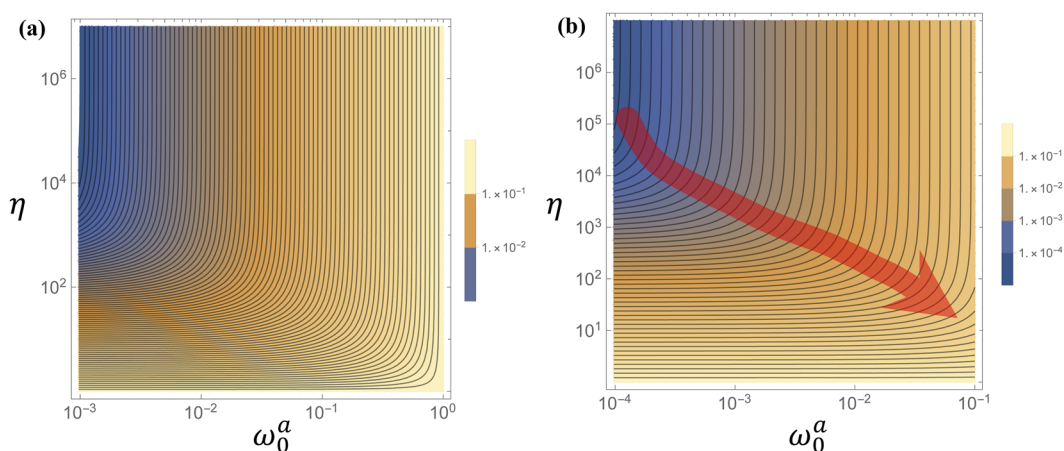


FIG. 7. Contour of $f = (1 - \eta^{-1})\omega_a^0 + \eta^{-1}$ in (a) full scale and (b) zoom-in for $\omega_a^0 \leq 0.1$. f (< 1) is the ratio of P_{tot} to P_a , η (> 1) is the ratio of P_a to P_b , and ω_a^0 ($\in [0, 1]$) is the weight of path *a* at λ_0 . The contour of *f* is featured by vertical lines for $\eta \gg 1$ and horizontal lines for $\eta \sim 1$. For intermediate η , *f* increases with a decreasing η and an increasing ω_a^0 , as indicated by the red arrow.

grow (Fig. 4). Therefore, positioning λ_0^k at a greater value increases the fraction (ω_a^0) of the PPP route sampled at λ_0^k , leading to a total FFS rate closer to R_{PPP} . On the other hand, the dependence of the FFS rate on λ_A^k seems less straightforward in its first look. However, since λ_A^k sets the “end point” of a trial run in FFS, having a smaller λ_A^k tends to keep a trial run “alive,” which allows it to continue exploring parameter space, thus enhancing the chance for sampling the more effective PPP pathway. In the same spirit, simply increasing the number of configurations N_0 can also help better explore the PPP pathway—even having a few more PPP seeds included in the ensemble will help reduce the error in the final rate. Indeed, as shown in Fig. 5, increasing N_0 from 1000 to 3000 is found to significantly reduce the variation in the final FFS rate.

IV. DISCUSSION

This study, which is carried out in a relatively simple system but contains rich physics, has enabled an investigation of a few key issues in modeling crystal nucleation using FFS. First and foremost, the quality of order parameters plays a paramount role in FFS. An advantage of FFS is its exploratory nature that allows probing nucleation pathways and dynamics that are usually largely unknown *a priori*. Therefore, an order parameter should be both general (inclusive of possible pathways) and accurate (close to actual reaction coordinates). In practice, however, striking a balance between the two can be challenging. For a given system, approaches have been developed to identify or optimize order parameters to enable a faithful description of the nucleation pathway.^{52,53} However, it is not guaranteed that an optimized order parameter in one system may serve well in another. This study presents a good example in this regard: even an order parameter that has been proven effective in both homogeneous and heterogeneous mW ice nucleation can still be problematic in the same system upon a subtle change in nucleation dynamics. In the same spirit, the developed order parameter λ_k with $k_{\text{cut}} \in [3, 5]$ in the current system may not always perform well in another system because the nucleus can carry a wide range of shapes depending on how heterogeneous nucleation is promoted by surface; thus, a different range of λ_k may be required. Nevertheless, the idea that enforcing some kind of constraint helps screening out non-relevant, “poisoning” seeds and facilitating the most relevant pathways can be more generically applicable to a broader range of nucleation modeling.

The issue may be more severe in a system where nucleation may proceed via multiple pathways. When an order parameter includes multiple pathways that can exhibit very different rates and free energy landscapes, how FFS can effectively capture the most relevant pathway can be an issue. In particular, when the slow route is much less steep than the fast route in the early portion of their free energy profiles, just as the SPP path in this study, the initial sampling can be nearly dominated by the slow path. The two-path model quantitatively shows how the level of sampling at the initial part of FFS affects the determination of the final rate constant. In practice, there can be several remedies for this issue. First, a specific order parameter can be tailored to focus on the main path by excluding the slow paths. For example, for the current system, crystal orientation⁵⁴ can be added as another metric in the order parameter to favor the PPP pathway over SPP pathway. However, this requires the knowledge

of nucleation dynamics in the system, which is, in general, unavailable unless already explored first by advanced approaches like FFS. Second, a less-specific order parameter can be used to include multiple pathways, and knowing that the final rate can be subject to convergence issues and errors due to finite sampling, FFS studies need to be carried out with carefully chosen FFS parameters and/or multiple FFS runs to gain sufficient statistics. In the case of choosing optimal FFS parameters, a general strategy is to ensure milestones are sufficiently separated, so as to both reduce correlations and “filter out” weak path from FFS path ensemble. Nevertheless, this leads to a significant increase in computational cost: for example, the cost increases from 2700 CPU hours for $\lambda_0^k = 45$ to 12 000 CPU hours for $\lambda_0^k = 65$ for a single FFS run in the current system. On the other hand, a similar level of statistics may be achieved by conducting multiple independent FFS runs to compensate the inaccuracy from individual FFS run. The variability itself from multiple runs can also reflect the hidden dynamics that may be missed otherwise, just as shown in this study. Ideally, multiple runs with well-designed FFS parameters should be conducted to achieve the optimal accuracy, but the cost can be prohibitive when the system itself is already expensive, for example, ice nucleation with all-atom models.^{15,24,37,38} It remains to be understood how to optimize the cost and efficiency for an unknown materials system.

V. CONCLUSION

In this study, we have investigated heterogeneous ice nucleation on an FCC (211) surface through the forward flux sampling method and mW water model. Our study shows the commonly adopted, size-based order parameter fails to describe heterogeneous ice nucleation on the FCC (211) surface, which promotes a unique 2D order incoherent to ice I. To overcome the issue, we develop a new order parameter λ_k by including a geometric constraint on ice seed by computing its maximum ratio in the resolved components of the radius of gyration k . The new order parameter is found to significantly improve the quality of FFS when a range of geometric constraints is applied, as demonstrated by p_B histogram analysis and the obtained ice nucleation rates.

Analysis of the ensemble of nucleation trajectories obtained from FFS identifies two competing nucleation pathways: a primary-prism-planed path and a secondary-prism-planed path. Although the nucleation of ice on the FCC (211) surface is dominated by the primary-prism-planed path, which has a nucleation rate 4 ~ 5 orders of magnitude greater than the secondary-prism-planed path at 230 K, the significant presence of the SPP path at the early stage of FFS can lead to large uncertainty in the calculated FFS rate constants that can exhibit an undesirable dependence on the chosen FFS parameters due to finite sampling. To understand the behavior in depth, we develop a two-path model that considers the co-existence of a fast and a slow route in one system and provides a general understanding of the dependence of the final rate constant in FFS on the initial weights and the difference in the growth probabilities of both routes. Through the application of the two-path model, we rationalize how the choice of FFS parameters, particularly pertaining to initial sampling, including, for example, initial milestones and the number of initial configurations collected, affects the final rate constant and its variability. Such dependence highlights the

challenge of modeling crystal nucleation in the presence of multiple pathways, particularly in pursuit of a quantitative description of the key rate constant that governs the true nucleation dynamics. Our study also suggests a few strategies for achieving this goal based on FFS.

SUPPLEMENTARY MATERIAL

The [supplementary material](#) includes (1) a description of the method for distinguishing the primary prism plane from the secondary prism plane in a hexagonal ice nucleus, (2) a derivation of the two-path model, (3) a figure showing the geometric anisotropy of crystalline seeds collected in the initial sampling of FFS using the conventional, size-based order parameter, and (4) a figure showing FFS homogeneous ice nucleation rate independent of λ_A .

ACKNOWLEDGMENTS

The authors thanked Davide Donadio for fruitful discussions. This work was supported by NSF (Award No. CBET 2053330).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Wanyu Zhao: Data curation (equal); Formal analysis (equal); Methodology (equal); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (equal); Writing – review & editing (equal). **Tianshu Li:** Conceptualization (lead); Formal analysis (equal); Funding acquisition (lead); Investigation (lead); Methodology (lead); Project administration (lead); Resources (lead); Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹T. Bartels-Rausch, *Nature* **494**, 27 (2013).
- ²G. John Morris and E. Acton, *Cryobiology* **66**, 85 (2013).
- ³J. Luoma, E. Ingham, C. Lema Martinez, and A. Allmendinger, *Processes* **8**, 1439 (2020).
- ⁴B. J. Murray, D. O'Sullivan, J. D. Atkinson, and M. E. Webb, *Chem. Soc. Rev.* **41**, 6519 (2012).
- ⁵S. J. Cox, S. M. Kathmann, B. Slater, and A. Michaelides, *J. Chem. Phys.* **142**, 184704 (2015).
- ⁶S. J. Cox, S. M. Kathmann, B. Slater, and A. Michaelides, *J. Chem. Phys.* **142**, 184705 (2015).
- ⁷M. Fitzner, G. C. Sosso, S. J. Cox, and A. Michaelides, *J. Am. Chem. Soc.* **137**, 13658 (2015).
- ⁸Y. Bi, R. Cabriolu, and T. Li, *J. Phys. Chem. C* **120**, 1507 (2016).
- ⁹M. Fitzner, P. Pedevilla, and A. Michaelides, *Nat. Commun.* **11**, 367 (2020).
- ¹⁰M. B. Davies, M. Fitzner, and A. Michaelides, *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2205347119 (2022).
- ¹¹T. Li, D. Donadio, G. Russo, and G. Galli, *Phys. Chem. Chem. Phys.* **13**, 19807 (2011).
- ¹²T. Li, D. Donadio, and G. Galli, *Nat. Commun.* **4**, 1887 (2013).
- ¹³E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. F. Abascal, and C. Valeriani, *J. Am. Chem. Soc.* **135**, 15008 (2013).
- ¹⁴J. R. Espinosa, E. Sanz, C. Valeriani, and C. Vega, *J. Chem. Phys.* **141**, 18C529 (2014).
- ¹⁵A. Haji-Akbari and P. G. Debenedetti, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10582 (2015).
- ¹⁶Y. Bi and T. Li, *J. Phys. Chem. B* **118**, 13324 (2014).
- ¹⁷L. Lupi, A. Hudait, B. Peters, M. Grünwald, R. Gotchy Mullen, A. H. Nguyen, and V. Molinero, *Nature* **551**, 218 (2017).
- ¹⁸E. Rosky, W. Cantrell, T. Li, and R. A. Shaw, *Chem. Phys. Lett.* **789**, 139289 (2021).
- ¹⁹P. M. Piaggi, J. Weis, A. Z. Panagiotopoulos, P. G. Debenedetti, and R. Car, *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2207294119 (2022).
- ²⁰S. Hussain and A. Haji-Akbari, *J. Chem. Phys.* **156**, 054503 (2022).
- ²¹R. Cabriolu and T. Li, *Phys. Rev. E* **91**, 052402 (2015).
- ²²M. Fitzner, G. C. Sosso, F. Pietrucci, S. Pipolo, and A. Michaelides, *Nat. Commun.* **8**, 2257 (2017).
- ²³L. Lupi, B. Peters, and V. Molinero, *J. Chem. Phys.* **145**, 211910 (2016).
- ²⁴G. C. Sosso, T. Li, D. Donadio, G. A. Tribello, and A. Michaelides, *J. Phys. Chem. Lett.* **7**, 2350 (2016).
- ²⁵Y. Qiu, A. Hudait, and V. Molinero, *J. Am. Chem. Soc.* **141**, 7439 (2019).
- ²⁶A. K. Metya and V. Molinero, *J. Am. Chem. Soc.* **143**, 4607 (2021).
- ²⁷S. Hussain and A. Haji-Akbari, *J. Chem. Phys.* **154**, 014108 (2021).
- ²⁸E. B. Moore, E. de la Llave, K. Welke, D. A. Scherlis, and V. Molinero, *Phys. Chem. Chem. Phys.* **12**, 4124 (2010).
- ²⁹E. B. Moore and V. Molinero, *Phys. Chem. Chem. Phys.* **13**, 20008 (2011).
- ³⁰T. L. Malkin, B. J. Murray, A. V. Brukhno, J. Anwar, and C. G. Salzmann, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1041 (2012).
- ³¹W. F. Kuhs, C. Sippel, A. Falenty, and T. C. Hansen, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21259 (2012).
- ³²T. L. Malkin, B. J. Murray, C. G. Salzmann, V. Molinero, S. J. Pickering, and T. F. Whale, *Phys. Chem. Chem. Phys.* **17**, 60 (2014).
- ³³V. Molinero and E. B. Moore, *J. Phys. Chem. B* **113**, 4008 (2009).
- ³⁴R. J. Allen, D. Frenkel, and P. R. ten Wolde, *J. Chem. Phys.* **124**, 024102 (2006).
- ³⁵A. Haji-Akbari, R. S. DeFever, S. Sarupria, and P. G. Debenedetti, *Phys. Chem. Chem. Phys.* **16**, 25916 (2014).
- ³⁶Y. Bi, B. Cao, and T. Li, *Nat. Commun.* **8**, 15372 (2017).
- ³⁷A. Haji-Akbari and P. G. Debenedetti, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10582 (2017).
- ³⁸G. C. Sosso, T. F. Whale, M. A. Holden, P. Pedevilla, B. J. Murray, and A. Michaelides, *Chem. Sci.* **9**, 8077 (2018).
- ³⁹A. Haji-Akbari, *J. Chem. Phys.* **149**, 072303 (2018).
- ⁴⁰S. Hussain and A. Haji-Akbari, *J. Chem. Phys.* **152**, 060901 (2020).
- ⁴¹G. C. Sosso, J. Chen, S. J. Cox, M. Fitzner, P. Pedevilla, A. Zen, and A. Michaelides, *Chem. Rev.* **116**, 7078 (2016).
- ⁴²D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, *Phys. Rev. Lett.* **94**, 235703 (2005).
- ⁴³W. Lechner, C. Dellago, and P. Bolhuis, *Phys. Rev. Lett.* **106**, 085701 (2011).
- ⁴⁴L. Lupi, A. Hudait, and V. Molinero, *J. Am. Chem. Soc.* **136**, 3156 (2014).
- ⁴⁵A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, *Comput. Phys. Commun.* **271**, 108171 (2022).
- ⁴⁶T. S. van Erp, D. Moroni, and P. G. Bolhuis, *J. Chem. Phys.* **118**, 7762 (2003).

- ⁴⁷R. J. Allen, D. Frenkel, and P. R. ten Wolde, *J. Chem. Phys.* **124**, 194111 (2006).
- ⁴⁸P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, *Phys. Rev. B* **28**, 784 (1983).
- ⁴⁹P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- ⁵⁰B. Peters, *J. Chem. Phys.* **125**, 241101 (2006).
- ⁵¹T. Li, D. Donadio, and G. Galli, *J. Chem. Phys.* **131**, 224519 (2009).
- ⁵²E. E. Borrero and F. A. Escobedo, *J. Chem. Phys.* **127**, 164101 (2007).
- ⁵³R. S. DeFever and S. Sarupria, *J. Chem. Phys.* **150**, 024103 (2019).
- ⁵⁴P. M. Piaggi, A. Z. Panagiotopoulos, P. G. Debenedetti, and R. Car, *J. Chem. Theory Comput.* **17**, 3065 (2021).