The Journal of Chemical Physics

RESEARCH ARTICLE | JUNE 08 2023

Interplay of multiple clusters and initial interface positioning for forward flux sampling simulations of crystal nucleation

Special Collection: Nucleation: Current Understanding Approaching 150 Years After Gibbs

Katarina E. Blow ; Gareth A. Tribello ; Gabriele C. Sosso ; David Quigley

J. Chem. Phys. 158, 224102 (2023)
https://doi.org/10.1063/5.0152343

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Interplay of multiple clusters and initial interface positioning for forward flux sampling simulations of crystal nucleation

Cite as: J. Chem. Phys. 158, 224102 (2023); doi: 10.1063/5.0152343
Submitted: 29 March 2023 • Accepted: 19 May 2023 •
Published Online: 8 June 2023

Katarina E. Blow,[1,a] Gareth A. Tribello,[2] Gabriele C. Sosso,[3] and David Quigley[1,b]

AFFILIATIONS

[1] Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom
[2] Centre for Quantum Materials and Technologies, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, United Kingdom
[3] Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom

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:** k.blow@warwick.ac.uk
[b] d.quigley@warwick.ac.uk

ABSTRACT

Forward flux sampling (FFS) is a path sampling technique widely used in computer simulations of crystal nucleation from the melt. In such studies, the order parameter underpinning the progress of the FFS algorithm is often the size of the largest crystalline nucleus. In this work, we investigate the effects of two computational aspects of FFS simulations, using the prototypical Lennard-Jones liquid as our computational test bed. First, we quantify the impact of the positioning of the liquid basin and first interface in the space of the order parameter. In particular, we demonstrate that these choices are key to ensuring the consistency of the FFS results. Second, we focus on the frequently encountered scenario where the population of crystalline nuclei is such that there are multiple clusters of size comparable to the largest one. We demonstrate the contribution of clusters other than the largest cluster to the initial flux; however, we also show that they can be safely ignored for the purposes of converging a full FFS calculation. We also investigate the impact of different clusters merging, a process that appears to be facilitated by substantial spatial correlations—at least at the supercooling considered here. Importantly, all of our results have been obtained as a function of system size, thus contributing to the ongoing discussion on the impact of finite size effects on simulations of crystal nucleation. Overall, this work either provides or justifies several practical guidelines for performing FFS simulations that can also be applied to more complex and/or computationally expensive models.

I. INTRODUCTION

Crystal nucleation from the melt is a widespread and important phenomenon. Furthering the understanding of how this phase transition occurs has important implications for, e.g., climate modeling,[1,2] cryopreservation,[3] and the oil industry.[4] Of particular interest is the rate, per unit time and per unit volume, at which nucleation occurs, J. Simulations of nucleation provide an opportunity to study the freezing mechanism and obtain the nucleation rate under physical conditions and at a level of detail often unachievable through experiments.

Under typical conditions of interest, nucleation is a rare event obeying Poisson statistics. Thus, it is often impossible to perform unbiased molecular dynamics (MD) simulations and observe this phase transition directly. Therefore, enhanced sampling techniques are often needed to circumvent the timescale problem. In this work, we focus on forward flux sampling (FFS). FFS is a widely used tool in, e.g., studying transitions in biological systems,[5–7] as well as crystal nucleation.[4,8–14]

Despite the simplicity of FFS, there are a number of user-specified parameters that can lead to significantly different implementations. For example, the initial flux, Φ_0 (discussed in

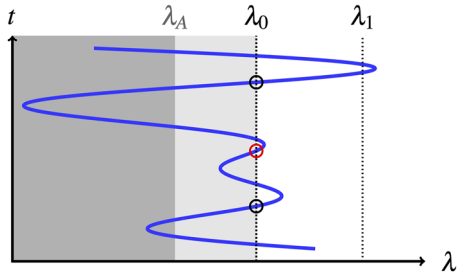

FIG. 1. Schematic of the calculation of the initial flux from a sample trajectory of an order parameter (λ, blue line), evolving in time (t, positive y direction). The gray areas represent the liquid basin. When the boundary of the liquid basin is placed at the first interface, λ_0 (lighter gray), every crossing back across λ_0 is a return to the liquid basin. Therefore, every positive crossing (circled) from the gray area across λ_0 is counted. When the boundary of the liquid basin, λ_A, is instead placed away from λ_0 (such that the liquid basin is represented by the darker gray area), when the order parameter falls below λ_0 it does not necessarily return to the liquid basin. Therefore, the red circled crossing of λ_0 is not counted when calculating the initial flux in this case.

Sec. II A), can be calculated by counting every positive crossing of λ_0 or only crossings when the trajectory is coming from within a liquid basin bounded by some value $\lambda_A \neq \lambda_0$ (illustrated in Fig. 1).[5,8,15] The impact of these different interpretations is often compounded by nomenclature, which makes the exact implementation unclear,[4,9,10,16] exacerbated by the fact that λ_A plays a significant role in calculating transition probabilities. It is therefore possible to define the edge of liquid basin but not consider its role in the calculation of initial flux. One of the aims of the present work is to investigate the robustness and reproducibility of initial flux calculations so as to determine whether or not the variability of the different FFS implementations in the literature is a concern. To this end, we investigate the effects of different placements of the first interface, λ_0, as well as the position of the interface marking the boundary of the liquid basin, λ_A.

Previous work on optimizing interface placement has centered on computational efficiency and error reduction and has neglected the impact of the initial flux (and therefore the locations of λ_A and λ_0) due to the minimal computational cost in computing the initial flux to a low variance compared to the full transition probabilities. In addition, these interface placement schemes rely either on adjusting interface positions as a result of the true transition probability or on the end location of several long trajectories. These can significantly increase the computational cost of FFS, performing several tests of transition probabilities at each interface in order to determine the "optimum" interface placement.[17,18] The placement of λ_0 has been investigated in the work of Velez-Vega *et al.* as a means to ensure that configurations stored upon crossings of λ_0 are uncorrelated. However, their technique is of limited applicability, especially in the context of crystal nucleation.[7] To the best of our knowledge, no investigation of the optimum placement of the edge of the liquid basin, λ_A, has been performed. Henceforth, we shall refer to λ_A, λ_0, and, to a lesser extent, λ_1 as the "initial interfaces" in FFS.

For almost all systems of practical interest, the system sizes relevant to real-life applications are several orders of magnitude larger than those that can be modeled using current computational resources. As such, most computational studies of nucleation rely on the use of periodic boundary conditions. These infinite repeats of the same simulation cell lead to spurious effects that would not be present in the corresponding macroscopic systems. These are often dependent on the size of the simulation cell and are known as finite size effects (FSEs). Several investigations of FSEs in the context of simulations of nucleation can be found in the literature. All of these studies have found evidence of finite size effects when a nucleus interacts with one or more periodic replicas of itself.[19–22] More recent studies have even found evidence of finite size effects in simulations including millions of atoms.[23] Of particular relevance to this work are the investigations of FSEs in the nucleation of mW water (on a surface) and Lennard-Jones (LJ) systems (in bulk) performed by Hussain and Haji-Akbari using jumpy FFS.[13,14] For the mW model, they found that, even in systems large enough for the surroundings of each cluster to behave like the bulk liquid phase, the nucleation rate is not constant with respect to volume [the inverse of a proxy for the side length of the simulation cell was found to be $\propto \log_{10}(J)$],[13] although this could not be corroborated by similar studies on the Lennard-Jones system due to a lack of data.[14]

In this work, we address some practical aspects of FFS implementations via systematic investigations utilizing the LJ model. We show that the placement of the initial interfaces is especially important. In particular, it is possible to use a simple and relatively cheap analysis of a MD simulation limited to the initial liquid basin as a means to pinpoint specific positions of the initial interfaces yielding consistent results. We propose a simple heuristic for the placement of initial interfaces in the context of crystal nucleation—placing λ_A at the most probable value of the largest cluster size in the metastable liquid, and the first interface at a location where interactions between nuclei are negligible (which reduces the potential to underestimate effective flux through subsequent interfaces due to merging of nuclei).

We also probe the effect of considering clusters other than the largest cluster (often utilized to define the order parameter)—with specific reference to the calculation of initial flux as well as of the subsequent crossing probabilities. It should also be noted that in the conventional implementation of FFS, it is impossible to consider multiple clusters when calculating crossing probabilities. Despite the effects of multiple clusters being accessible for the initial flux, counting only crossings of the single largest cluster appears to be almost universally done in the literature,[4,8–14,24] although the potential impacts of this choice have not, to the best of our knowledge, been investigated. It should be noted that according to work by Cheng and Ceriotti, the thermodynamically stable state of a nucleating liquid will switch from many smaller clusters to a single large cluster as the size of the largest cluster increases.[25] However, this switch may occur for cluster sizes much larger than the location of λ_0.

Finally, we provide evidence for the existence of substantial spatial correlations between different clusters. We find that these correlations lead to the merging of smaller clusters into larger ones—an occurrence that has a non-negligible impact at the supercooling condition considered in this work.

The paper is organized as follows: We begin by summarizing the relevant methodology and simulation details in Sec. II. Results are presented in Sec. III. These consist of relevant nuclei populations, an initial flux analysis of the systems with these populations, and a

brief consideration of the consistency of the flux through subsequent interfaces. Spatial correlations between clusters are also presented and discussed. The main findings of our work are then summarized in the context of the current literature in Sec. IV.

II. METHODOLOGY

A. Forward flux sampling

FFS is a path sampling technique that can be used for simulating crystallization under conditions where the timescale required for nucleation with brute force MD is intractable. The path between the liquid state, A, and the crystal, B, is described in terms of an order parameter (OP, λ). In FFS, we select a set of isosurfaces of the OP (labeled λ_i), henceforth referred to as interfaces, located along the space of the OP such that moving between successive interfaces leads to a complete path from A to B. Starting from the λ_{i-1} interface, when a run crosses λ_i the system configuration is saved. Many statistically independent trajectories are then initialized from these stored configurations. Whether these new trajectories reach the next interface at λ_{i+1} or return to a less crystalline state (usually λ_A, defined as the edge of the liquid basin) is then determined. This gives a probability that if the system is in state λ_i, it will progress to state λ_{i+1} before returning to state A, $P(\lambda_{i+1}|\lambda_i)$.

The interfaces are typically placed such that the probability of crossing interfaces can be computed with sufficient accuracy. If the probability of crossing the next interface is too low (i.e., the interfaces are too far away from each other), the lack of statistics leads to a high error as a result of insufficient coverage of phase space and sparsity of configurations at interfaces. On the other hand, if almost all trials cross the next interface, little information is gained and there is likely to be a high degree of correlation between configurations at successive interfaces (limiting the accuracy for these new starting configurations). For starting configurations that sufficiently sample the appropriate phase space at the interface, and are uncorrelated, the current consensus in the literature is that the choice of interface positions does not affect the value of the calculated flux although it may affect both the computational efficiency and the calculated uncertainty in the flux.[16–18,24,26–28]

The total probability, $P(\lambda_N|\lambda_0)$, of traveling from the λ_0 interface to the λ_N interface is given by

$$P(\lambda_N|\lambda_0) = \prod_{i=0}^{N-1} P(\lambda_{i+1}|\lambda_i). \tag{1}$$

This probability can then be multiplied by the flux across the first interface, Φ_0, to give the nucleation rate—the effective flux that leads to system solidifying,

$$J = \Phi_0 \times P(\lambda_N|\lambda_0). \tag{2}$$

Φ_0 is found by counting (per unit time and volume) the number of times the first interface is crossed when coming from the liquid basin in the direction of increasing OP, i.e., the flux of clusters through λ_0 moving from the liquid toward the crystalline configuration, see Fig. 1. In this work, Φ_0 will be used to refer to the initial flux generally, while $\Phi_{0|\lambda_A}$ is used for specific implementations considering the edge of the liquid basin at λ_A (Table I). Note that in other work using FFS, the definition of λ_A is often not explicit. In some studies, $\lambda_A = \lambda_0$ is used for the initial flux stage and the definition of λ_A for other stages may be ambiguous.[12,29,30] Alternatively, it may be unclear if (a sometimes unknown value of) λ_A is used for the flux stage or only when calculating transition probabilities—instead taking λ_0 as the edge of the liquid basin for the initial flux.[4,9,10] The initial flux is an integral component of calculating nucleation rates with FFS, and therefore care should be taken to ensure that this value is as accurate as possible. Assuming that the population distribution of nuclei is an extensive quantity, Φ_0 should be independent of simulation volume.

TABLE I. Summary of relevant nomenclature used in this work, for ease of reference.

Nomenclature	Interpretation
λ_A	Edge of the liquid basin
λ_0	First interface in FFS calculation
λ_1	Second interface in FFS calculation
"Initial interfaces"	λ_A, λ_0, and λ_1
λ'	Interface for direct and effective flux
$\lambda_{0'}$	First interface for effective flux
$\Phi_0/\Phi_{0\|\lambda_A}$	Initial flux/initial flux for given λ_A
$\Phi'_{\|\lambda_A}/\Phi_{\lambda'\|\lambda_A}$	Direct/effective flux for given λ' and λ_A
$g'(r)$	Minimum cluster separation histogram

B. Molecular dynamics simulations

In this work, we used Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) (23 June 2022)[31] to perform NpT simulations of LJ systems of several different system sizes, ranging from 4000 atoms to 108 000 atoms in cubic periodic boxes. Note that older versions of LAMMPS deal incorrectly with the computation of order parameters via dynamic groups and may therefore not give consistent results.[32] Atoms interacted through a shifted LJ forcefield with a cutoff of 3.5 σ and $m = \sigma = \varepsilon = 1$. Thermostatting was applied through a canonical velocity rescaling thermostat[33] set to enforce a temperature $T* = 0.86$ (here, $*$ shall represent LJ reduced units) with a damping parameter of $0.05t*$. This temperature corresponds to an approximate supercooling of 19%. At this supercooling condition, nucleation events can be considered as rare events from a Poisson statistics point of view[22] and the re-thermalization of the system should be rapid compared to the timescale of interest.

We also enforced a pressure of $p* = 5.68$ via a Nosé–Hoover barostat (with a Martyna–Tobias–Klein correction[34]), where barostat thermalization was applied through a chain of five Nosé–Hoover thermostats. Integration was performed using the velocity Verlet algorithm of time step $t* = 0.002$.

Crystalline atoms were identified using the sixth-order ten Wolde order parameter ($\mathbf{q_6}$).[35] This is a vectorial OP based on the combination of spherical harmonics proposed in the work of Steinhardt *et al.*[36] The maximum distance between neighbors was set to 1.432 σ; a connection was defined as solid if $\mathbf{q_6}(i) \cdot \mathbf{q_6}(j) > 0.5$; and an atom with eight or more solid connections was classified as solid. Grouping of crystalline atoms into clusters was performed internally in LAMMPS (see input file). These choices are obviously not unique and may have an impact on the absolute nucleation rates calculated. We do not expect this choice to have a significant effect on the conclusions drawn about initial fluxes, OPs which lead to a

difference in the diffusivity of cluster interfaces may, however, change the conclusions about the prevalence of cluster merging. However, as this work is concerned with the self-consistency of FFS calculations, investigations of the impacts of these parameters has not been performed.

It is important to consider the time resolution at which the OP is sampled. Here, we computed the OP at snapshots (or frames) taken every 100 MD time steps. This choice represents a balance between computational expense and time resolution. The consequences of this finite time resolution is discussed later.

To avoid methodological ambiguity originating from the occurrence of clusters of size exactly identical to the value of a given FFS interface, all interfaces were placed at non-integer values of cluster size.

The input files and analysis scripts we have used in this work are available on GitHub at https://github.com/keb721/FFS_Interfaces.

1. Initial flux analysis

For the initial flux analysis, 24 independent runs were performed for each system size. The systems were all generated by melting a solid system at $T^* = 2.4$ for 10 000 time steps and then cooling via a 1000 time step equilibration at $T^* = 1.2$ (close to but above the melting temperature), followed by a 2000 time step linear quench to the temperature of interest. Data at this temperature were generated by simulating for 1 000 000 time steps. Cluster sizes were output by LAMMPS, with additional post-processing for calculation of fluxes and spatial distribution functions performed by in-house code.

Spatial correlations were estimated by computing a weighted histogram of the minimum distances between clusters above a specified size, denoted $g'(r)$, within the regime where the minimum image convention is applicable (half of the side length of the cubic box). For each cluster of interest, the minimum distance between it and all of the images of all other clusters to consider was computed. The extent of the clusters was considered implicitly, as distances were computed between the location of atoms within the cluster, although atoms are considered point particles. Self-interactions were not included. Weighting was then performed with respect to the volume of the spherical shell of the histogram bin, such that clusters in close proximity were weighted more heavily, and the number of clusters in the frame for which the histogram was being computed—i.e., one less than the number of clusters of interest in the frame, to account for the excluded self-interactions. As the number of clusters of interest for the sizes considered here is not large, the difference between normalizing with respect to the total number of clusters in the frame and the number of considered clusters in the frame could be significant.

2. λ' test

Additionally, to test the robustness of the FFS approach, we calculated the flux through a given interface λ', in two different ways. (1) We calculated the "direct" flux through λ', i.e., by counting the number of crossings per unit time and volume to yield the flux Φ'. This can also be written as $\Phi'_{|\lambda_A}$ to make the dependence on liquid basin location explicit. (2) We defined a second interface at $\lambda_{0'} < \lambda'$; we computed the flux through $\lambda_{0'}$ by counting the number of crossings per unit time and volume to yield the flux $\Phi_{0'}$. Then, we calculated the "effective" flux through λ', by multiplying $\Phi_{0'}$ by $P(\lambda'|\lambda_{0'})$ to give $\Phi_{\lambda'}$. To account for the effect of λ_A in both $\Phi_{0'}$ (i.e., $\Phi_{0'|\lambda_A}$), and $P(\lambda'|\lambda_{0'})$, the effective flux can also be denoted $\Phi_{\lambda'|\lambda_A}$. Note that the choice of λ_A will always be applied consistently to both the initial flux and crossing probability components. This so-called "λ' test" allows us to establish whether the positioning of a given interface has an impact on the effective flux—which has an impact on the overall nucleation rate. Specifically, if the direct and effective fluxes computed via (1) and (2), respectively differ, we have a clear indication of the uncertainty associated with the positioning of λ_A and $\lambda_{0'}$ (and in fact, of any other FFS interface) on the nucleation rate calculated via Eq. (2).

To perform the λ' test, the direct fluxes were taken from the initial flux analysis, as described above. For calculation of the effective flux via method (2) above, configurations at a range of $\lambda_{0'}$ values between 20.5 and 40.5 were generated. Liquid systems at the temperature of interest were generated by melting a solid system and cooling it to the temperature of interest using the same procedure as above. Relevant configurations were stored every time the size of the largest cluster, λ, reached $(\frac{1}{2} + \lambda_{0'})$. Special care was taken in ensuring that these stored configurations were not correlated with each other. To this end, every time that a configuration at $\lambda_{0'}$ was stored, a simulation of 1000 MD time steps was performed, during which the value of λ was ignored. After these 1000 MD time steps, monitoring of λ was resumed, to check for the condition $\lambda < \lambda_P$. Note that λ_P is system size dependent, as illustrated in Table II. At this stage, we can be certain that the system was sufficiently de-correlated with respect to the previous crossing of $\lambda_{0'}$, and a new configuration was stored when λ reached the appropriate value.

We accumulated 500 configurations at each $\lambda_{0'}$, generated according to the procedure above. Then, we ran 10 000 MD "trials" starting from a random choice of these de-correlated configurations, redrawing particle velocities from a Maxwell–Boltzmann distribution relative to the temperature of interest.

TABLE II. System volumes, averaged over all frames of all simulations, $\langle V* \rangle$, and λ_P values, defined as the peak of the incidences of primary nuclei size—$\frac{1}{2}$, for different numbers of atoms in the system.

Number of atoms	$\langle \mathbf{V}* \rangle$	$\boldsymbol{\lambda_P}$
4 000	4 080	11.5
5 324	5 430	12.5
6 912	7 050	14.5
10 976	11 194	16.5
16 384	16 710	19.5
32 000	32 636	24.5
108 000	110 159	34.5

III. RESULTS

A. Population distributions of crystalline clusters

For the initial fluxes obtained at different system sizes to be equivalent, they must exhibit the same population of nuclei in the metastable liquid state. This population is a function of supercooling. Figure 2 represents the populations of nuclei in the simulation

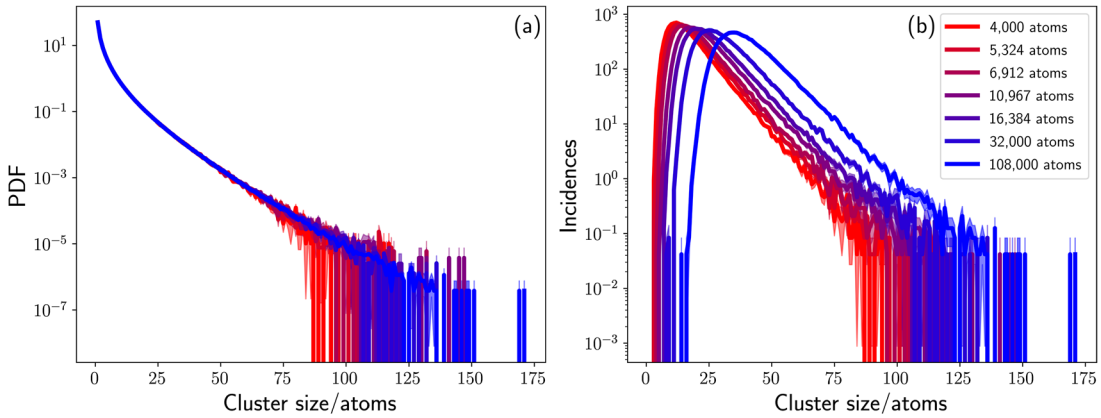

FIG. 2. Descriptors of the size of crystalline clusters within the metastable liquid at $T^* = 0.86$ and $p^* = 5.68$ before the onset of nucleation. We have considered 10 000 snapshots, separated by 100 MD time steps. (a) Probability distribution function of the size of all crystalline clusters (normalized to unit volume), and (b) number of incidences of the largest cluster per snapshot being of a given size. The solid lines represent the mean values and the shaded regions represent the standard error on the mean—at larger cluster sizes, these can present as sharp vertical lines, and where not visible, they are less than the linewidth. The legend applies to both panels. The uncertainty observed for large cluster sizes is due to poor statistics (as these large clusters correspond to the result of rare fluctuations).

(before the onset of the nucleation, monitored by means of the order parameter described in Sec. II B) for all system sizes studied. Figure 2(a) shows, on a log-linear scale, the probability density function (PDF) of all nuclei in a simulation, regardless of the size of any other nuclei in the system. This PDF has been normalized with respect to unit simulation volume (see Table II for the relative volumes of the different simulation cells). The colored lines represent the mean value of the PDF and the shaded regions represent the standard error on the mean. At low cluster sizes, the errors are negligible and the distribution of nuclei is the same for all simulation volumes. As cluster size increases, the errors become more significant as a result of insufficient statistics as large clusters correspond to large fluctuations, which are less accessible to brute force simulations. The longer tails for the larger simulations simply reflect the fact that as there are more nuclei in the system. It is also more likely that one of these nuclei undergoes statistical fluctuations to grow to a larger size. It should be noted that the largest sized cluster in all simulations was under 175 atoms, indicating that no nucleation events have taken place in any of the simulations.

In many nucleation studies, the OP used is the size of the largest crystalline cluster.[4,8–14,24] In this work, the largest cluster in any frame of the simulation will also be referred to as the primary cluster (or primary nucleus) of that frame. Figure 2(b) displays the number of incidences of primary clusters of given sizes, again on a log-linear scale, with the solid line representing the mean and the shaded region showing the standard error. No normalization has been performed, as by definition the number of primary nuclei is independent of the simulation volume. Incidences under one are possible due to averaging of multiple runs. Similar to the total PDF [Fig. 2(a)], the error increases with cluster size due to poor statistics at higher cluster. Importantly, the distribution of primary nuclei is not comparable between volumes for any cluster size, despite the equivalence of the total nuclei PDF. This can be explained via order statistics, as the same distribution is being sampled more times in larger simulation cells. The probability of the largest observed cluster having size n is given by

$$P_k(n) = k \times \mathrm{PDF}(n) \times (\mathrm{CDF}(n))^{k-1}, \tag{3}$$

where k is the number of clusters in the system and PDF and CDF represent the probability distribution function and the cumulative distribution function of nuclei sizes, respectively.[37] As the number of clusters in the system increases (i.e., the simulation volume becomes larger), it is more likely that larger clusters will be observed. This also explains the increase in number of atoms at the peak in the primary cluster distribution with increasing simulation volume.

These results demonstrate that although the distribution of total nuclei populations in the liquid basin is volume-independent, the modal size of the primary cluster is not as it is related to the average number of clusters in the simulation volume. This has implications for the definition of the liquid basin in studies that use primary cluster size as an OP. Carefully defining the location of the liquid basin is important in path sampling techniques when determining interface crossing probabilities and, in some cases, the initial flux. Using the same value of λ_A for different system sizes may lead to trajectories committing to the A basin not being treated as such. This can cause both large differences in calculated nucleation rate due to an artificially increased crossing probability and much longer simulation times (see Sec. III C).

B. Initial flux

As the population of nuclei per unit volume was comparable across different system sizes (see Fig. 2), we also expect the flux through the first interface, Φ_0, to be independent of system size, as it is normalized with respect to simulation volume. In order to compare Φ_0, it is important to ensure collection of sufficient statistics to calculate a stationary flux, the number of crossings through λ_0 should be approximately linear with time. The crossings of $\lambda_0 = 36.5$ are presented in Fig. S1, see the supplementary material.

Figure 3 offers an analysis of the flux through the first interface λ_0. Here, the solid line represents the mean, with the error bars (less than the linewidth) representing the standard error on the mean. Figures 3(a) and 3(b) correspond to calculating the flux by counting every positive crossing of λ_0, $\Phi_{0|\lambda_0}$ (i.e., any negative crossing of λ_0 is a return to the liquid basin). In contrast, Figs. 3(c) and 3(d) represent the flux, $\Phi_{0|\lambda_P}$, when the edge of the liquid basin is placed away from λ_0. Again, crossings are counted only if they originate from the liquid basin, but in this implementation this requires the cluster size to fall below $\lambda_P \neq \lambda_0$ between crossings of λ_0 (see Fig. 1). This is the procedure used in some, but not all, FFS implementations.[5,8,15] The λ_Ps are defined as the peaks in the primary incidences, reported in Table II. Recall that in order to prevent ambiguity regarding the treatment of clusters characterized by size exactly equal to an interface value, both λ_A and λ_0 were chosen to contain a non-integer number of atoms.

When considering the flux of the largest cluster only, the time series of the size of the primary cluster was considered. For the total positive flux [i.e., $\Phi_{0|\lambda_0}$, Fig. 3(b)], the number of positive crossings of λ_0 in this time series were counted, before being divided by the length of the simulation and the average volume of the simulation box. A positive crossing is one in which λ_0 is crossed in the direction of increasing OP. As λ_0 defines not only the first interface but also the edge of the liquid basin, this is equivalent to counting all crossings originating therefrom. For Fig. 3(d), the edge of the liquid basin was instead placed at $\lambda_P \neq \lambda_0$. Under this definition, not all positive crossings of λ_0 originated from the liquid basin. Therefore, more care must be taken to ensure that only the first crossing of λ_0 after leaving the liquid basin is counted. The influence of noise, which could lead to a large increase in the number of positive crossings due to a rapid fluctuation of cluster size, was mitigated by subsampling the time series (as described in Sec. II).

It can be seen in Fig. 3(a) that there is a strong dependence on the initial flux of all clusters with system size (when the edge of the liquid basin is placed at λ_0), despite the comparable nuclei populations. This dependence is most pronounced at small λ_0 and is partly due to the method used for calculating the fluxes of all clusters. At each snapshot, the number of clusters of size $>\lambda_0$ is counted. If, at

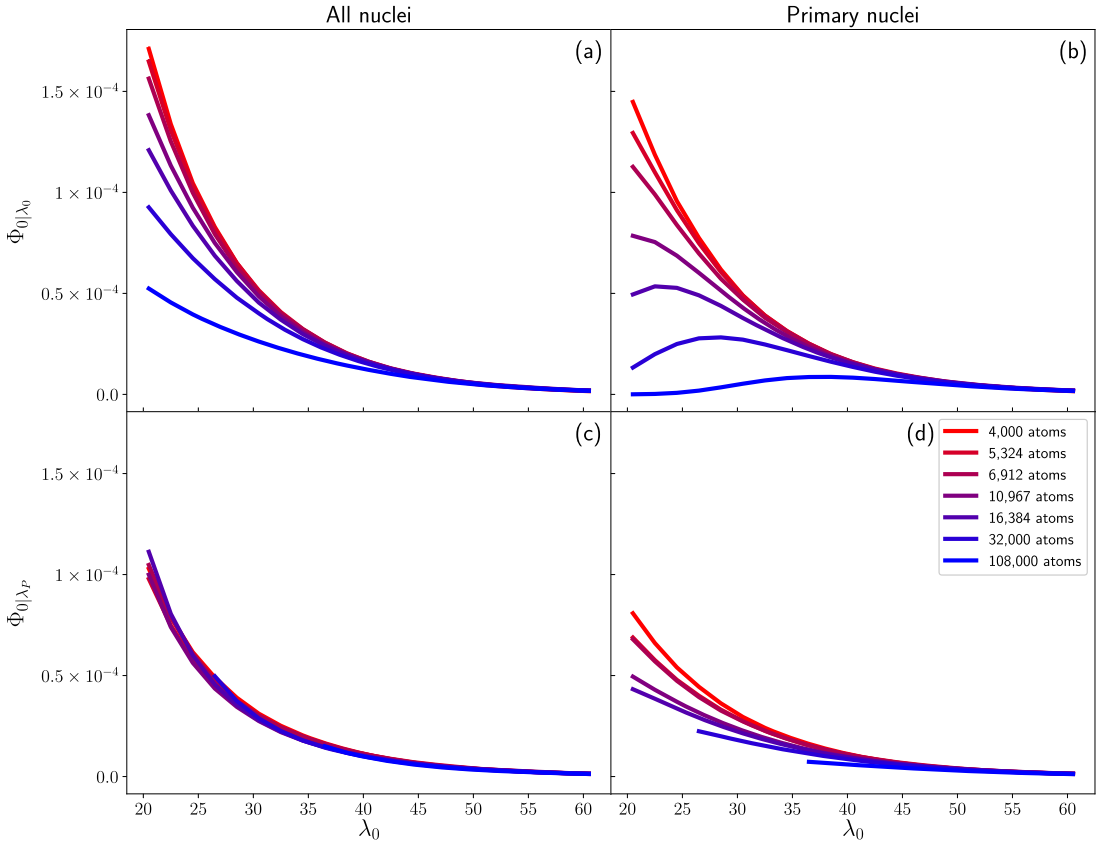

FIG. 3. Initial fluxes through the first interface λ_0 for different system sizes. (a) All positive crossings of all clusters, (b) all positive crossings of primary clusters only, (c) positive crossings of all clusters only after they have returned to λ_P, and (d) positive crossings of only primary clusters, only after they have returned to λ_P. Thus, $\Phi_{0|\lambda_0}$ (top panels) and $\Phi_{0|\lambda_A}$ (bottom panels) refer to the flux obtained for every crossing following the return of the system to below λ_0 and the flux obtained for the crossings that followed the return of the system to below λ_P, respectively (see Sec. II A). Note that, where not taken as λ_0, the edge of the liquid basin is defined in a volume-dependent manner (see Table II). The extent of the statistical uncertainty with respect to the flux is smaller than the linewidth. The legend applies to all panels.

the next snapshot there are more clusters of size $>\lambda_0$, then the flux is increased by this difference. This is equivalent to counting the total number of net positive crossings between frames. Positive crossings of λ_0 will therefore be missed if negative crossings of λ_0 have also occurred between snapshots.

Consider an ordered list of all of the solid clusters for every frame. Let the Mth and $M-1$th largest clusters in frame i be identified as X and Y, respectively.

$$\text{Clusters}(i) = \begin{cases} M\text{ th}, & X > \lambda_0, \\ M-1\text{ th}, & Y < \lambda_0. \end{cases}$$

If cluster Y grows concurrently with cluster X shrinking, then at the next point at which the cluster sizes are examined, frame $i+1$, cluster X may have shrunk back across the boundary to be of a size less than λ_0, while cluster Y has grown to exceed λ_0. Assume that they are still at the Mth and $M-1$th position in the new ordered list of cluster sizes given by

$$\text{Clusters}(i+1) = \begin{cases} M\text{ th}, & Y > \lambda_0, \\ M-1\text{ th}, & X < \lambda_0. \end{cases}$$

As we know that the Mth and $M-1$th largest clusters have swapped identities between frames i and $i+1$, we know that there has been a crossing of λ_0. However, as the Mth largest cluster is still above λ_0 and the $M-1$th largest cluster is still below λ_0, the net number of clusters above λ_0 is unchanged and therefore this crossing makes no contribution to Φ_0. The difference between no crossings occurring and the crossing of a growing nucleus that has been masked by concurrent shrinking of another nucleus cannot be resolved without infinitely fine time resolution. However, in the case of crystal nucleation, a degree of coarseness in the temporal resolution is potentially desirable due to inherent noise on the OP. Alternatively, distinguishing between no crossings and no net crossings can be attempted by tracking the identities of the clusters crossing the interfaces, although there are many difficulties and ambiguities involved in this. Due to the exponential decrease in PDF with nuclear size, the effect of concurrent growing and shrinking of nuclei becomes less important as λ_0 increases. For sufficiently large λ_0, the probability of sampling more than one cluster with size close to λ_0 is negligible.

The inability to identify concurrent growing and shrinking of clusters is also likely to lead to an underestimate of Φ_0 for the flux of primary nuclei, although there are now other important considerations. Comparing Figs. 3(a) and 3(b), it can be seen that (for the same liquid basin, bounded by λ_0) the flux of primary nuclei is lower than the flux of all nuclei for all system sizes at low λ_0. The primary flux is equal to 0 for the 108 000 atom system below $\lambda_0 \approx 25$ despite there being appreciable flux when all clusters are considered. The spread of Φ_0 as a function of system size is also larger at lower λ_0 in Fig. 3(b) than Fig. 3(a), and the stratification persists up to relatively large λ_0. Additionally, it is of interest that the peaks in the flux in Fig. 3(b) do not correspond to the values of the peaks in the primary incidences [Fig. 2(b)], instead occurring at slightly higher cluster sizes. This shows that (for $\lambda_A = \lambda_0$) most flux at small λ_0 occurs as a result of non-primary nuclei for large systems. This is not unexpected, as the probability of having multiple large clusters increases rapidly with the simulation volume. If λ_0 is chosen to be small, and/or the simulation volume is large, the contribution of non-primary clusters is significant.

Figure 3(a) shows only the positive crossings of λ_0, with no restrictions on how far a cluster must shrink back into the liquid basin before a subsequent boundary crossing—i.e., the edge of the liquid basin is taken to be λ_0 itself. In contrast, Fig. 3(c) shows $\Phi_{0|\lambda_P}$, the initial flux when the edge of the liquid basin is placed away from λ_0, here λ_P (see Table II). Instead of counting the net crossings of λ_0, as in Fig. 3(a), an ordered list of the sizes of all solid clusters is considered. If the size of the Mth largest cluster has fallen below λ_A, then the next crossing of λ_0 by the Mth largest cluster will be counted. Any subsequent crossing of λ_0 by the Mth largest cluster will be ignored until it has again fallen below λ_A. Importantly, the Mth largest cluster is *not* a fixed identity linked to a certain combination of atoms. Instead, it simply gives the size of the Mth largest cluster in each particular frame. Cluster identities of the largest clusters may therefore vary completely between successive snapshots. In practice, it may be useful to consider only the first N elements of this ordered list, where N is chosen such as to ensure that the last element of the subset never exceeds λ_0 over the course of the simulation, which will be system specific. Note that for the two largest system sizes, the lines do not begin at $\lambda_0 = 20.5$ due to $\lambda_P > 20.5$.

Separating λ_0 and the liquid basin decreases Φ_0 for all clusters and removes the dependence on simulation volume when all clusters in the simulation volume are considered [Fig. 3(c)]. When the edge of the liquid basin is placed at λ_0 such that all positive crossings are counted [Fig. 3(a)], then fluctuations of a cluster's size around λ_0 will artificially increase Φ_0. The impact of these fluctuations can be reduced by subsampling, which also reduces crossings as a result of noise in the OP, although this is unlikely to be sufficient for removing variation in the OP as a result of realistic fluctuations in the cluster size. Instead, ensuring that a cluster must vary in size more significantly to return to the liquid basin means that it is much more likely that the next crossing of λ_0 will be as a result of the growth of a different, independent cluster, increasing the robustness of the measure of the flux. This separation of λ_A and λ_0 has the most significant impact on the initial flux for the smallest system sizes. This is likely to be due to a combination of two related effects. First, for a smaller simulation volume there are fewer nuclei and therefore it is more likely that a rapid recrossing of λ_0 occurs as a result of the same cluster changing size. Additionally, for a smaller system the liquid basin is further away (due to the lower number of clusters) and therefore ensuring a cluster has returned there is more likely to remove a greater number of crossings.

Figure 3(d) displays $\Phi_{0|\lambda_P}$ for only primary clusters. Unlike $\Phi_{0|\lambda_P}$ for all clusters [Fig. 3(c)], a size-dependence in the initial flux is still observed, despite removing the contributions of small fluctuations in cluster size by ensuring a return to $\lambda_P \neq \lambda_0$ between subsequent crossings. This demonstrates the importance of non-primary nuclei in accurately calculating the initial flux. Especially in larger simulation volumes and at lower values of λ_0, the difference between primary flux and (size-independent) total flux is pronounced. As the population of primary clusters is size dependent [see Fig. 2(b)], this is not unexpected.

Despite having the same liquid basin (as described by nuclei populations), it can be seen that different methods of calculating the

initial flux can lead to different, volume-dependent results as they neglect the, possibly appreciable, fraction of significant non-primary nuclei and because they incorrectly define the edge of the liquid basin. It can be seen that incorrect placement of λ_A leads to unphysical effects in the flux—either artificially increasing it by inclusion of insignificant fluctuations if placed too high, as in Fig. 3(a), or underestimating Φ_0 as a result of real crossings not being counted as the edge of the liquid basin has been placed at a value that makes observing a return unlikely. If the λ_A chosen is not tailored to both the specific volume and composition of the system being simulated, either of these is a possibility and it can be impossible to spot from the value of the initial flux even with rigorous testing.

In the context of initial flux, the error introduced by incorrect placement of interfaces is dependent on both the λ_0 and λ_A values chosen. However, uncertainty in the nucleation rate is usually considered to be dominated by uncertainties in the transition probability, and the uncertainties considered here are small on the scale of variability of the results of FFS calculations. How the of placement of λ_A and λ_0 propagates to changes in transition probabilities through subsequent interfaces is discussed in more detail below.

It should be noted that there are preexisting heuristics for placing the initial interfaces in crystal nucleation, although to the best of our knowledge these are not the conclusions of in-depth studies. A relevant heuristic for crystal nucleation is that used in Ref. 8, which suggests placing λ_0 between the top 1% and top 0.1% of the OP distribution [for our system, this distribution is shown in Fig. 2(b)] and λ_A in the range $[\mu, \mu + \sigma]$ where μ and σ are the mean and standard deviation of the OP distribution, respectively. Our results indicate that this placement of λ_0 will ensure that the vast majority of the observed flux will be contributed by primary nuclei only, although the effect of non-primary nuclei on the transition probability through subsequent interfaces will be investigated in Sec. III C.

We also find that placing λ_A at the mode of our observed OP distribution leads to a consistent initial flux (on the chosen scale) when considering all nuclei. However, this is also the case at high λ_0 for $\lambda_A = \lambda_0$ and represents consistency only in a small part of the initial flux, not necessarily the entire FFS calculation. In addition to showing its consistency, we have posited that the rationale for using λ_A as the most probable value of the largest cluster size in the metastable liquid is to best ensure sufficient counting of independent crossings of different clusters. We acknowledge that this is not a perfect method to ensure that all independent crossings are counted, but such a method does not exist. While we have not explicitly studied the case of $\mu + \sigma$, the possibly significant overestimate of the liquid basin may lead to an assumption of independence between clusters where it does not exist and a premature termination of phase space exploration when computing transition probabilities—similar to what is expected for $\lambda_A = \lambda_0$, although to a lesser extent. In addition, the shape of the OP distribution may be sensitive to changes in cluster definition, and therefore the mean and the mode may not be coincident and the standard deviation may become significant. While changes in cluster definition may also exacerbate other differences between observed clusters, such as arrangement[38] and polymorph,[39] which may also influence the nucleation rate, this is not always the case.

C. λ' test

In a real FFS run, the initial flux Φ_0 is simply a component that is then combined with the probabilities of crossing subsequent interfaces to give the nucleation rate. The location of λ_0 is of no consequence to the crossing probabilities starting at subsequent interfaces (i.e., λ_1 and above). In contrast, the choice of λ_A is relevant to all interfaces, as it defines the point at which a trajectory is considered to have returned to the liquid basin.

The current consensus is that the placement of interfaces in FFS is important only in terms of the efficiency of and uncertainty in the rate calculation—it does not affect the value of the calculated rate, assuming that interfaces are sufficiently well-spaced that stored configurations are uncorrelated and that they are sufficient in number to sample the relevant phase space.[16–18,24,26–28] If this is true, then it should be possible to retrieve the direct flux through λ_1 by multiplying the flux through $\lambda_0 < \lambda_1$ by the probability that (uncorrelated) configurations with $\lambda = \lambda_0$ will progress to λ_1 before returning to λ_A for any consistent choice of λ_A, λ_0, and λ_1. The methodology used in this section is outlined in detail in Sec. II B 2. For consistency and clarity, we shall use the nomenclature outlined there, although it is worth noting that in practice the interface λ' corresponds to both the λ_0 (when computing the direct flux) and λ_1 interfaces (when computing the effective flux).

The direct and effective fluxes of primary nuclei through two different values of λ', for two choices of λ_A, are given in Fig. 4. Similar to the initial fluxes, these results should be volume-independent. In the initial flux, consideration of non-primary nuclei was required to eliminate significant stratification with system size. Conversely, for the effective flux, including the contribution of non-primary nuclei in $\Phi_{\lambda'}$ reduces agreement between system sizes (see Fig. S2 in the supplementary material, noting the larger shaded region of the effective flux, indicating that there is more spread between different volumes when considering all nuclei). Unlike in the initial flux stage, non-primary nuclei cannot be explicitly considered when calculating transition probabilities. Although they cannot be directly accounted for, significant non-primary nuclei are likely to be present in stored configurations with $\lambda = \lambda_{0'}$ and may therefore influence the flux. Any growth of an initially non-primary cluster, X, above the size of the initially primary cluster, Y, will be attributed to cluster Y. This will spuriously increase the transition probability as a result of the growth of clusters smaller than those present at the $\lambda_{0'}$ interface, although these are ostensibly the only relevant clusters in the system. As previously discussed, the prevalence of significant non-primary nuclei is more likely in larger simulation volumes. The greatest increase in transition probabilities—implicit in effective flux calculations—will be under the conditions where the prevalence of non-primary nuclei leads to the greatest decrease in initial flux, as shown in Fig. 3 (i.e., large simulation volumes and low λ_0). It is encouraging that the impact of non-primary nuclei in both of these computations appears to offset each other, leading to similar effective fluxes for all simulation volumes. It should also be noted that at sufficiently large primary clusters, the size of any non-primary cluster becomes negligible,[25] and therefore these need not be considered.

The impact of atomic velocities on the effective fluxes was also investigated (Fig. S3 in the supplementary material). There was little

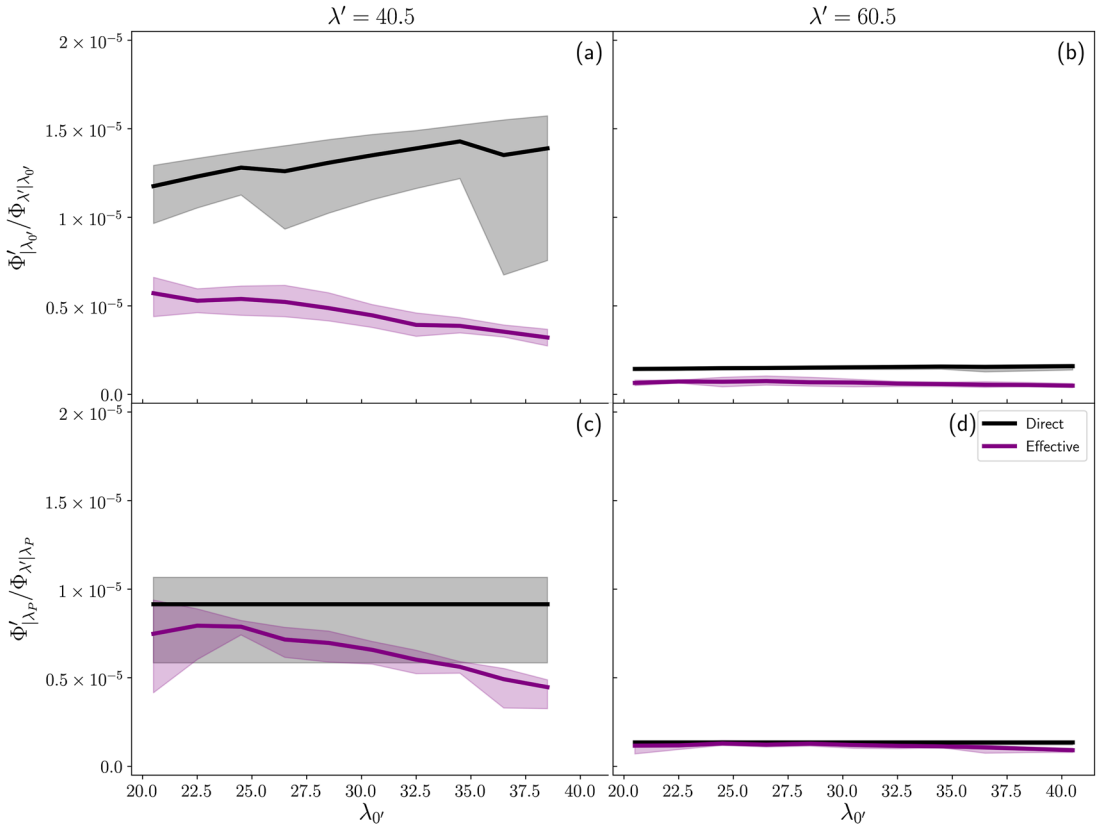

FIG. 4. Fluxes using different values of λ_A and λ'; both the direct flux, $\Phi'_{|\lambda_A}$, and the effective flux, $\Phi_{\lambda'|\lambda_A}$, are present on the same scale. (a) $\lambda_A = \lambda_{0'}$, $\lambda' = 40.5$, (b) $\lambda_A = \lambda_{0'}$, $\lambda' = 60.5$, (c) $\lambda_A = \lambda_P$, $\lambda' = 40.5$, (d) $\lambda_A = \lambda_P$, $\lambda' = 60.5$. The black line represents the mean of the direct flux of primary nuclei through λ' (requiring a return to λ_A between subsequent crossings). The purple line represents the mean of the effective flux of primary nuclei through λ' [$P(\lambda'|\lambda_{0'}) \times \Phi_{0'|\lambda_A}$, where $\Phi_{0'|\lambda_A}$ only includes contributions from primary nuclei]. The shaded region represents the uncertainty on the fluxes as a result of the different volumes of simulations considered here. The 32 000 and 108 000 atom systems only contribute to relevant fluxes for $\lambda_{0'} > \lambda_P$, even for the panels where the edge of the liquid basin is defined as $\lambda_A = \lambda_{0'}$. For effective fluxes, this is a result of the procedure used to generate initial configurations. For direct fluxes with $\lambda_A = \lambda_{0'}$, this is for a more consistent comparison. Note that the direct flux for $\lambda_A = \lambda_P$ is the only flux independent of $\lambda_{0'}$ and as such the 32 000 and 108 000 atom systems contribute for the entire range. The legend applies to all panels.

difference observed when the atomic velocities used when initializing trajectories at $\lambda_{0'}$ were those stored during positive crossings of $\lambda_{0'}$ during the flux stage or reinitialized randomly, indicating that the OP dynamics are overdamped with respect to particle momenta on the timescale of interface crossing.

Figure 4(a) corresponds to the flux through $\lambda' = 40.5$ using $\lambda_A = \lambda_{0'}$, whereas Fig. 4(b) corresponds to the flux through $\lambda' = 60.5$, again with $\lambda_A = \lambda_{0'}$. Note that in Fig. 4(b), the variation of effective flux with λ'_0 is too small to be seen on this scale and is lower than the variation in Fig. 4(a) due to fewer crossings of 60.5 compared to 40.5. For both values of λ', the direct flux is significantly above the effective flux for all values of $\lambda_{0'}$. This choice of λ_A leads to an artificially reduced transition probability—small deviations of the OP below $\lambda_{0'}$ lead to immediate termination of trajectories, underestimating the true likelihood of reaching λ'. This underestimation is likely to be most pronounced at interfaces in the vicinity of A, where the amount of natural nuclear variation possible before a cluster has been assigned as committed to the liquid basin is greatly reduced. In addition, the low probability of crossing λ' may lead to a lack of stored configurations at the next interface, which has severe implications for the accuracy of the results of further stages.

Conversely, if λ_A were chosen to be significantly below the modal value of the primary cluster, the transition probability would be overestimated (shown in the supplementary material in Fig. S4). In order for a crossing to λ' to be determined to have failed, all other clusters in the simulation must also be below λ_A when the primary cluster shrinks back into the melt. For a choice where this is unlikely (due to concurrent growing and shrinking of clusters), there is a large probability that another cluster will grow to exceed λ' before all clusters fall below the chosen λ_A value. As this is impossible to account for, there will be a significant and unrepresentative increase in the calculated transition probability as a result of this. In addition, there is likely to be an increase in simulation time for calculating transition probabilities—either due to waiting for the statistically improbable situation of all clusters being below λ_A or due to waiting for a new cluster to grow to cross λ_{i+1}, possibly

from below λ_i. The overcounting of crossings as a result of requiring over-commitment to the liquid basin is likely to be an insidious problem that will significantly affect all interfaces below the critical nuclear size.

Figure 4(d) shows the direct and effective fluxes through $\lambda' = 60.5$ using $\lambda_A = \lambda_P$. In comparison to Fig. 4(b), there is now a good agreement between direct and effective flux observed through $\lambda' = 60.5$. However, in Fig. 4(c), where $\lambda_A = \lambda_P$ but $\lambda' = 40.5$ does not display this consistency. There is a volume-independent (as indicated by a lack of significant broadening in the shaded region representing the spread of results at different volumes) and systematic decrease of effective flux with increasing $\lambda_{0'}$. It should be noted that this is also visible toward the right-hand side of Fig. 4(d). There are four possible reasons for this.

First, FFS assumes that the time taken to move between successive interfaces is negligible compared to the time taken for the initial crossings, although this will not always be true. For the 4000 atom system with $\lambda_A = \lambda_P, \lambda_{0'} = 38.5$, and $\lambda' = 40.5$ (where the discrepancy between direct and effective flux is largest), the average time taken between subsequent crossings of $\lambda_{0'}$ is $18.2 \pm 0.3t^*$, with the average transition time between $\lambda_{0'}$ and λ' being $0.774 \pm 0.015t^*$. This shows that the time taken to move between interfaces is of the same order of magnitude as the uncertainty in the mean time between crossings of $\lambda_{0'}$. It can hence be neglected. Therefore, effects from slow dynamics in the crossing stage are unlikely to be the cause of the systematic decrease in effective flux with $\lambda_{0'}$.

Second, this discrepancy may indicate that the delay between snapshots is too long and important crossings are therefore being missed. As the separation between $\lambda_{0'}$ and λ' decreases, fewer additions to the nucleus are required for the cluster to cross λ'. Assuming that the average rate of monomer addition is proportional to cluster surface area, then not only do clusters with a small $\lambda' - \lambda_{0'}$ need fewer monomer additions to exceed λ', but (assuming constant λ') these additions will happen on a faster timescale. If there is a long interval between snapshots, it is probable that these short-time fluctuations will be missed, while longer timescale fluctuations are still observed. Increasing the separation between $\lambda_{0'}$ and λ' does not eliminate the problem of missing crossings due to sampling intervals, although it may become less meaningful as a result of several factors: the larger chance to explore the phase space; the larger expected crossing time; and the decreased probability of being able to cross λ' at all. However, decreasing the sampling interval not only results in capturing legitimate interface crossings as a result of growing clusters; fluctuations in the cluster size due to an imperfect OP are also likely to be captured and counted as they are indistinguishable from real transitions. Again, these fluctuations are less important for a trajectory truly returning to the A basin than for one crossing λ' (or $\lambda_{0'}$)—small variations are more meaningful in the rarer region of OP space away from the metastable basin and therefore OP fluctuations are more liable to unavoidable misinterpretation. In addition to any impact on crossing probabilities, changing the sampling interval will necessarily also influence the observed flux. Although this change will be most pronounced for $\lambda_A = \lambda_{0'}$, due to the inclusion of more observed rapid and statistically meaningless fluctuations of nucleus size (whether occurring as a result of noise on the OP or not), it will have an effect in all other cases as well, e.g., by capturing additional forays back into the liquid basin. Decreasing the sampling interval was not able to resolve the systematic decrease in effective flux with increasing $\lambda_{0'}$ (see supplementary material, Fig. S6), although it does exhibit the expected increase in crossing probability.

Third, in a realistic FFS run with a finite sampling frequency, the configurations stored at interfaces are likely to be those that have crossed an interface,[7,8,16,17,24,28] which do not necessarily lie exactly on that interface. This means that the ensemble of stored configurations at an interface is likely to include clusters significantly larger than the interface size. Numerical work by Haji-Akbari has indicated that neglecting these configurations is likely to lead to an underestimation of the flux.[40] For our work, including configurations that have crossed $\lambda_{0'}$ but have a cluster size above $\lambda_{0'} + \frac{1}{2}$ at $\lambda_{0'} = 30.5$ (shown in Fig. S3 in the supplementary material) did not resolve the discrepancy between direct and effective flux, despite the large proportion of stored configurations with sizes much larger than 31, deemed to be configurations "at the boundary."

Finally, for crystal nucleation in LJ there is evidence that cluster growth occurs not only as a result of addition of single particles to a cluster but also through merging of multiple clusters (two or more) into a larger cluster.[41–43] This merging of clusters would violate the assumption of FFS that the only pathway between A and B is one that passes through every intermediate interface sequentially.[44] If this sequential crossing were not the case, then the effective flux would underestimate the true flux, as it would not be able to take into account the possible pathway of crossing λ' by merging of multiple clusters of a size less than $\lambda_{0'}$. Such an event would be included in the calculation of direct flux through λ' but omitted from the effective flux as at no time is there a cluster with $\lambda = \lambda_{0'}$ along this pathway. This would explain the steady decrease in effective flux in Fig. 4(c) with increasing $\lambda_{0'}$—the smaller the difference between $\lambda_{0'}$ and λ', the more likely it is that a merging event would allow for a crossing of λ' without requiring a crossing of $\lambda_{0'}$. This would also explain the small deviation at high $\lambda_{0'}$ in Fig. 4(d). The possibility of merging clusters in our system is explored in more detail in Sec. III D.

The potential influence of merging clusters on the effective flux can be minimized in a number of ways. First, $\lambda_{0'}$ can be placed at a sufficiently rare value of the OP that there is only ever one significant cluster at that size. This also means that the contribution of non-primary nuclei in the initial flux can be neglected, although crossings of $\lambda_{0'}$ may then be so rare that there is a large uncertainty in the flux, and that it is difficult to store configurations that sample a sufficiently broad region of phase space. Alternatively, $\lambda_{0'}$ and λ' can be spaced such that the likelihood of not crossing $\lambda_{0'}$ on the way to λ' is negligible. Although this is unlikely to cause sampling issues at $\lambda_{0'}$, the low probability of reaching λ' may lead to insufficient sampling at subsequent interfaces. Additionally, instead of using conventional FFS to determine the nucleation rate, the jumpy forward flux sampling (jFFS) algorithm proposed by Haji-Akbari can be used in order to take into account an OP that may vary significantly enough that not all interfaces may be crossed. However, to minimize computational cost, the suggested implementation of jFFS involves adaptively placing interfaces, such that they are sufficiently far apart for the jumpiness of the OP to be neglected.[40] Again, care must be taken to ensure adequate sampling at higher interfaces.

We have shown that consistency in the choices of λ_A, λ_0, and λ_1 does not automatically lead to agreement between direct and effective flux. Although we do not explicitly consider the case of

considering a different liquid basin when computing crossing probabilities and the initial fluxes, the differences in initial fluxes for $\lambda_0 = 20.5$ for $\Phi_{0|\lambda_0}$ and $\Phi_{0|\lambda_P}$ are of the order of 2 (see Fig. 3), but this difference is significantly less pronounced for $\lambda_0 = 60.5$. It is therefore obvious that, as well as the methodological inconsistency in the definition of the liquid basin, the results obtained when using $\Phi_{0|\lambda_0}$ and $\lambda_A = \lambda_P$ for the crossing stage are also inconsistent. We also expect this conclusion to hold for other inconsistent choices.

When attempting to minimize the errors in the crossing probabilities in FFS, the major issue considered is simply one of uncertainty due to a lack of crossing statistics. This can be minimized by increasing the number of trial runs performed. However, here we demonstrate the existence of significant systematic errors due to poor consideration of interface placement, which cannot be eliminated by generating more statistics. In the worst case, these systematic errors will be present for every subsequent interface, meaning that a relatively small error on a transition probability can be compounded into several orders of magnitude for the overall nucleation rate. Assuming a relatively modest difference in observed transition probability of 10%, similar to that which we have found for different choices of λ_A for the first four interfaces (see the supplementary material, Fig. S5), could lead to an uncertainty in the nucleation rate of 45%, which cannot be rectified by the inclusion of more trials at interfaces. This can be very significant in some cases, e.g., when comparing the dominance of two process with closely competing rates, as in Ref. 39.

Again, we have shown that a correct definition of the liquid basin is critical to ensuring consistency. We suggest that merging of clusters is a potentially important pathway to be considered when performing FFS of nucleation, the effect of which can be minimized in several ways. We also give evidence that the contributions of non-primary nuclei to transition probability are implicitly considered.

D. Spatial correlations

In Sec. III C, we discuss the possibility that the systematic difference between effective and direct flux for small separations between $\lambda_{0'}$ and λ' may arise as a result of clusters merging. In order for clusters to merge within a timescale that would impact the flux, they must be located near each other to some degree. As such, we investigated potential spatial correlations between the crystalline clusters in the system. To do so, we have computed a weighted histogram of all of the minimum distances between clusters, here denoted as $g'(r)$, for all clusters above a given size. Note that the actual extent of the clusters has been taken into account when determining these distances. These were binned into spherical shells of thickness 0.1σ, with the furthest extent of the final bin being given in Table III to ensure that configurations that violate the minimum image convention are never included. Using this procedure, the incidence of single solid particles in the melt gives a $g'(r)$ similar to the radial distribution function, $g(r)$, expected for a liquid (see the supplementary material, Fig. S7). This shows that single solid particles are reasonably common and will occur with approximately the same probability at any location in the simulation box—as expected.

These weighted histograms are given in Fig. 5 for several different cluster sizes. As the size of clusters increases, the distribution

TABLE III. Edge sizes in lattice units (for generating configurations in LAMMPS) and minimum edge sizes considered for $g'(r)$ analysis.

Number of atoms	Edge size (lattice)	Minimum edge size (σ)
4 000	10	15.8
5 324	11	17.4
6 912	12	19
10 976	14	22
16 384	16	25.4
32 000	20	31.8
108 000	30	47.4

retains some of the character of the $g'(r)$ of single solid particles, still showing well-defined peaks, which become sharper and more pronounced due to the decrease in number of clusters. Unlike single solid particles, nuclei containing 20 or more atoms are expected to be sufficiently rare that there should be no obvious correlation between their observed locations. As can be seen in Fig. 5(a), we observe distinct peaks within the short/medium-range order region of the $g'(r)$ across all system sizes. It should be noted that these $g'(r)$ are intrinsically biased toward smaller distances, as only the minimum distances between clusters are considered. However, if the peaks in the $g'(r)$ were occurring solely as a result of this, we would observe a flattening of the peaks as the simulation volume increases due to rise in the maximum possible distance between nuclei. This is not the case, as the position of the peaks does not change as a function of the system size. In fact, this result indicates that there is some degree of clustering of nuclei around other nuclei, which can also be clearly seen when visualizing the trajectories [see Fig. 6, obtained via the visual molecular dynamics (VMD)[45] software]. Although there would inevitably be incidences of clustering for truly randomly distributed clusters, the combination of peaks in the $g'(r)$ and the apparent frequency of the observed clustering in visualized trajectories indicates that these clusters are not completely independent. In addition, the observed shape of the peaks maps reasonably closely to a hard sphere model with two preferred, related cluster separations (see the supplementary material, Fig. S8).

When the size of the clusters increases, the peaks become less pronounced. Considering Fig. 5(b), it is clear that although the $g'(r)$ of the smaller systems is much noisier, the peaks are still well defined for the larger simulations. As the cluster sizes increase still further to 60 and 80 atoms [Figs. 5(c) and 5(d) respectively], the noise continues to increase and the $g'(r)$ goes to 0 for the smaller volumes, as there are no frames containing more than one cluster of the relevant size. The cluster sizes considered in Figs. 5(c) and 5(d) are very large in the context of a unbiased simulation under these conditions. Although having multiple large clusters in the same frame becomes more likely as simulation volume increases, the probability of this happening is still very small. This makes it unclear to what extent the apparent loss of clustering is simply due to a lack of statistics collected and how much is due to other, unknown effects.

Although there is evidence of clustering (see Figs. 5 and 6), the exact origin of this is unknown. A possible explanation is that clustering arises due to the breakup of clusters. If an originally

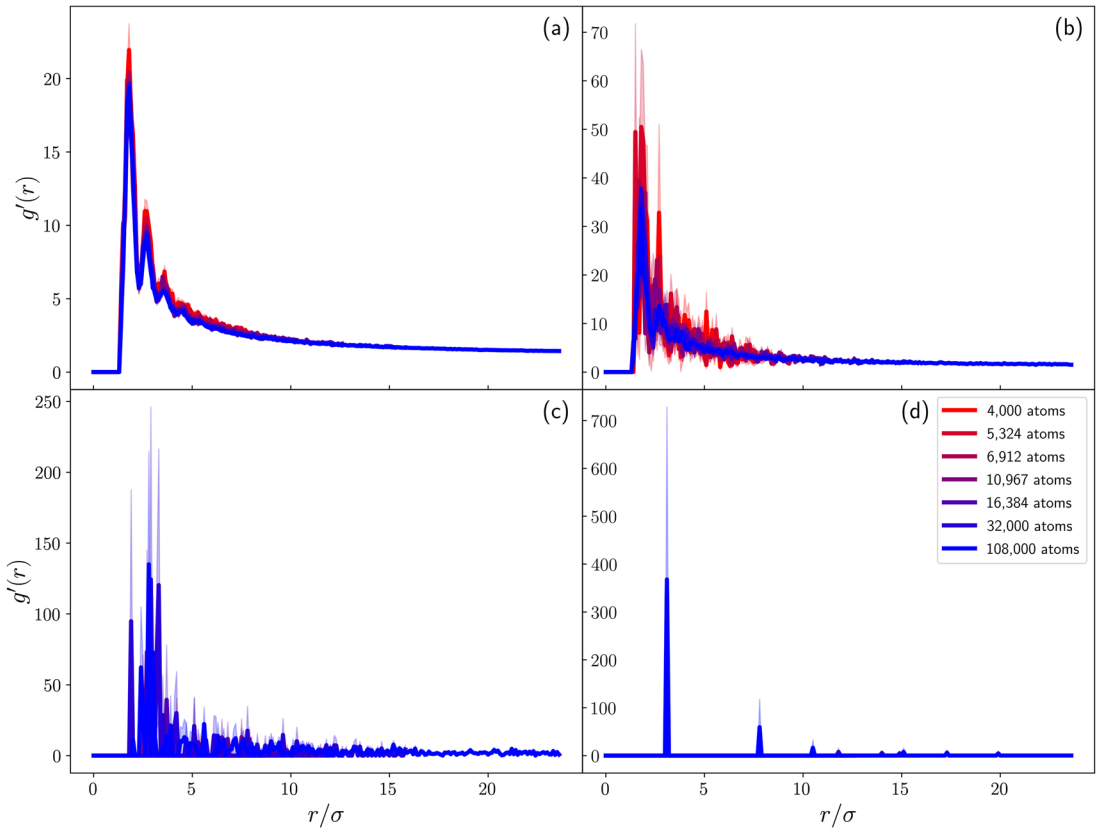

FIG. 5. $g'(r)$ of the minimum distances between clusters of size (a) 20 and above, (b) 35 and above, (c) 60 and above, (d) 80 and above. The legend applies to all panels—for panel (c) the $g'(r)$ of the systems containing fewer than 6912 atoms are omitted as there are no observed frames containing multiple clusters above the threshold. In panel (d), only the 108 000 atom system is presented, for the same reason. Shaded regions represent the statistical uncertainty and are less than the linewidth where not visible.

dumbbell-like solid cluster were to lose solid atoms from the bridging section of the nucleus, the result would be two nearby clusters of appreciable size. This is difficult to directly observe without detailed tracking of cluster identities.

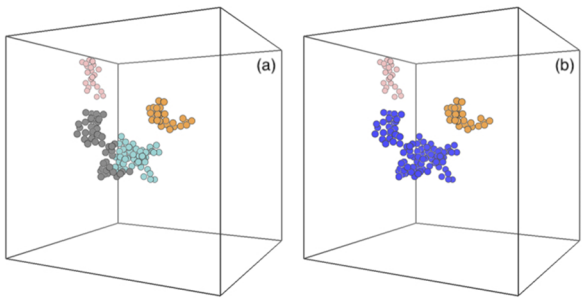

FIG. 6. VMD snapshots of two successive frames of a simulation of 32 000 atoms ($T^* = 0.86$, $p^* = 5.68$), showing clustering of nuclei of size 20 atoms or larger. (a) Four nuclei of size greater than 20 are present in the simulation box (black) of the first snapshot. (b) $0.002t^*$ later, the cyan and gray nuclei have moved slightly and therefore merged to form the blue nucleus, while the other nuclei are unchanged.

Alternatively, clustering may suggest structuring of the LJ melt around an established nucleus, potentially due to the diffuse interface of crystalline clusters, in a way that increases the propensity for other clusters to form in the vicinity. Hussain and Haji-Akbari found evidence for nuclei structuring the liquid around critical nuclei, showing distinct peaks and troughs in the liquid density before reaching a plateau.[14] Although the results are not comparable due to different simulation conditions and the fact that the nuclei considered here are far from critical, this indicates that the presence of nuclei does structure the surrounding liquid, perhaps making it easier for other nuclei to form.

In addition, the increased presence of significant solid clusters in the liquid around other solid clusters may indicate a propensity for nuclei to grow by amalgamation, since growth or movement of the cluster in any direction may lead to an encounter with another nucleus, with which it would then merge. This is shown in Fig. 6. It should be noted that this occurs at too fast a timescale to be observed when sampling at every 100 time steps (here additional snapshots have been generated with single time step separation), although the effects of merging occurring at faster timescales are still important to results obtained at higher sampling separations.

IV. DISCUSSION AND CONCLUSIONS

We have demonstrated the importance of careful interface placement when studying crystal nucleation using FFS. Our data clearly show that the correct placement of λ_A is critical to ensuring a consistent and volume-independent flux. We show that the placement of λ_A and λ_0 in nucleation studies is and should be simulation size dependent. We have given evidence for cluster merging with significant influences on effective flux, although this can be minimized by judicious placement of λ_1 and λ_0. We also illustrated the importance of non-primary nuclei but indicated that their contribution is likely to be implicitly considered when performing a complete FFS run. The importance of these findings also extends to other enhanced sampling techniques using interface placement, e.g., transition interface sampling.[46]

The initial implementation of FFS did not include a λ_A interface.[44] This was instead included in a later paper offering refinements, although it was stated that $\lambda_0 = \lambda_A$ was always a valid choice, and often a convenient one.[28] Some implementations of FFS have indicated that there is an advantage to placing λ_A away from λ_0 but usually only as a way of removing correlations between stored configurations at λ_0.[27,47] In contrast to this earlier work, we have shown that the choice of λ_A is crucial to ensure a consistent effective flux across λ_1, which is likely to extend to every subsequent interface. Unfortunately, the nomenclature often used in the literature to describe how the initial flux was calculated is ambiguous[4,9]—sometimes with reference to counting crossings leaving the A basin[10,16]—which makes it unclear if the crossings of λ_0 being counted are simply all positive crossings or positive crossings that have occurred after a return to the A basin (i.e., if what is being computed is $\Phi_{0|\lambda_0}$ or $\Phi_{0|\lambda_A}$ for $\lambda_A \neq \lambda_0$). This is often compounded by the fact that the value of λ_A used, if any, is omitted. This then extends the methodological confusion to the treatment of crossing of subsequent interfaces, as the condition for determining an unsuccessful transition attempt is undefined.

Previous investigations of interface placement have centered on computational efficiency and reducing the variance in the final nucleation rate and have neglected the location of λ_A and λ_0 due to the minimal computational cost in computing Φ_0 to within a small variance.[17,18] In the work of Velez-Vega *et al.*, the optimum placement of λ_0 has been explored by finding the minimum of the average simulation time needed to generate uncorrelated crossings. However, the need for an additional OP that can be cheaply computed *and* is distinct from λ significantly limits the applicability of their technique, especially in the context of crystal nucleation where OP choices are generally variants of the Steinhardt parameters.[8–14] In addition, they do not attempt to refine the placement of λ_A, even using $\lambda_0 < \lambda_A$, which is likely to lead to increased correlations between configurations stored at λ_0.[7] In this work, we have proposed a simple alternative heuristic for the placement of initial interfaces in the context of crystal nucleation—placing the edge of the liquid basin at the peak of the incidences of primary nuclei, and the first interface at a location where the effects of non-primary nuclei are negligible (which reduces the potential underestimate of effective flux through subsequent interfaces due to merging of nuclei). We note that the value of the edge of the liquid basin can continue to have a significant effect on transition probabilities at interfaces beyond λ_0, as it marks the failure criterion for a cluster to reach the next interface (see the supplementary material, Fig. S5). We find that the flux through λ_3 can differ by approximately an order of magnitude depending on where λ_A is placed, which is significant on the scale of uncertainty in a FFS calculation.

In systems where nuclei do not merge, it is only important that λ_0 is not placed in a region where primary flux is negligible, and in these cases placing λ_0 between the top 1% and the top 0.1% of primary incidences (as in Ref. 8 and suggested in Ref. 27) will be more than sufficient. However, when nuclei can merge, it is not only nuclei of a comparable size to the primary nucleus that are important—any significantly sized nucleus can be involved in a merging event that can bypass an interface and thus influence the rate. Therefore, for only primary nuclei to be relevant, λ_0 should be placed after it becomes more thermodynamically favorable to have a single nucleus rather than several smaller nuclei.[25] Depending on the properties of the system of interest, this system-specific condition is likely to be difficult to confirm and occur at a larger value of λ_0 than is practicable. In these scenarios, the effects of merging can be mitigated in a number of other ways.

While this work was performed under conditions where multiple nuclei are common, we believe that the results are applicable to systems where this is not the case—e.g., at low supercooling, low interfacial free energy, or for heterogeneous nucleation. A λ_A that is placed too high will count rapid, unstatistical recrossings of λ_0 regardless of why these occur. Similarly, it ensures that the counted flux is indeed the flux of an attempted nucleation event. It is interesting to note that the conclusions presented here regarding interface placement are similar to those presented in recent work of Zhao and Li in heterogeneous mW ice nucleation occurring through two pathways of substantially different rate. In their case, positioning λ_0 at higher values was to increase sampling of the reaction pathway that dominated at high λ, which is somewhat analogous to avoiding growth through the pathway of merging which is impossible at sufficiently large λ.[38]

We also investigated the effects of non-primary nuclei on direct and effective flux. Although a contribution from non-primary nuclei is not unexpected, to the best of our knowledge it has not been explored or discussed in the literature. This is somewhat concerning as the size of primary crystalline nucleus is a commonly used OP in studies of nucleation. In this work, we have demonstrated that there is an appreciable FSE in the initial flux that is caused by non-primary clusters and that non-primary clusters can affect crossing probabilities as they are present in stored configurations. However, we have shown that these effects combine destructively and explicit consideration of non-primary clusters is therefore not necessary.

Finally, we have presented strong evidence for clustering of nuclei around one another. We have found indications that these clusters merge and that this can have an appreciable impact on the observed flux through an interface.

It is important to note that this detailed investigation of FFS implementations is possible due to the low computational cost associated with the LJ potential, especially compared to more realistic models whose results are of greater interest. We have been able to systematically investigate the effects of initial interface placement and, from this, have given a procedure for determining the placement we have found to be necessary *without* requiring costly additional computation, which is of practical use when investigating more complex models.

We also acknowledge that the choice of OP made may also influence the nucleation rate, although the potential effects of this are significantly less important than a lack of self-consistency within the FFS calculation utilizing the same OP.

SUPPLEMENTARY MATERIAL

The supplementary material for this paper includes graphs of the number of boundary crossings as a function of time and the radial distribution function of single solid particles; several graphs showing different implementations of the λ' test—the influence of λ_A placement, the effect of using all nuclei for the fluxes, and the impact of keeping velocities and of storing configurations not "on the interface"; and change in direct and effective fluxes as a result of sampling interval. In addition, it also shows the influence of different values of λ_A on interfaces up to λ_3.

ACKNOWLEDGMENTS

The authors would like to acknowledge the use of the computational facilities provided by the University of Warwick Scientific Computing Research Technology Platform. We would also like to acknowledge the EPSRC Centre for Doctoral Training in Modeling of Heterogeneous Systems (EPSRC Grant No. EP/S022848/1). D.Q. is funded by EPSRC Program Grant No. EP/R018820/1. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Katarina E. Blow: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Gareth A. Tribello**: Methodology (supporting); Writing – review & editing (equal). **Gabriele C. Sosso**: Conceptualization (equal); Formal analysis (supporting); Methodology (equal); Supervision (equal); Writing – review & editing (equal). **David Quigley**: Conceptualization (equal); Formal analysis (supporting); Methodology (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data generated by the simulations reported in this manuscript are openly available at http://wrap.warwick.ac.uk/175784.

REFERENCES

[1]D. J. Cziczo, K. D. Froyd, C. Hoose, E. J. Jensen, M. Diao, M. A. Zondlo, J. B. Smith, C. H. Twohy, and D. M. Murphy, "Clarifying the dominant sources and mechanisms of cirrus cloud formation," Science **340**, 1320–1324 (2013).
[2]U. Lohmann and J. Feichter, "Global indirect aerosol effects: A review," Atmos. Chem. Phys. **5**, 715–737 (2005).
[3]A. Baskaran, M. Kaari, G. Venugopal, R. Manikkam, J. Joseph, and P. V. Bhaskar, "Anti freeze proteins (Afp): Properties, sources and applications—A review," Int. J. Biol. Macromol. **189**, 292–305 (2021).
[4]Y. Bi and T. Li, "Probing methane hydrate nucleation through the forward flux sampling method," J. Phys. Chem. B **118**, 13324–13332 (2014).
[5]E. E. Borrero and F. A. Escobedo, "Folding kinetics of a lattice protein via a forward flux sampling approach," J. Chem. Phys. **125**, 164904 (2006).
[6]A. Vijaykumar, P. R. ten Wolde, and P. G. Bolhuis, "Rate constants for proteins binding to substrates with multiple binding sites using a generalized forward flux sampling expression," J. Chem. Phys. **148**, 124109 (2018).
[7]C. Velez-Vega, E. E. Borrero, and F. A. Escobedo, "Kinetics and reaction coordinate for the isomerization of alanine dipeptide by a forward flux sampling protocol," J. Chem. Phys. **130**, 225101 (2009).
[8]A. Haji-Akbari, R. S. DeFever, S. Sarupria, and P. G. Debenedetti, "Suppression of sub-surface freezing in free-standing thin films of a coarse-grained model of water," Phys. Chem. Chem. Phys. **16**, 25916–25927 (2014).
[9]T. Li, D. Donadio, G. Russo, and G. Galli, "Homogeneous ice nucleation from supercooled water," Phys. Chem. Chem. Phys. **13**, 19807–19813 (2011).
[10]R. Cabriolu and T. Li, "Ice nucleation on carbon surface supports the classical theory for heterogeneous nucleation," Phys. Rev. E **91**, 052402 (2015).
[11]J. A. van Meel, A. J. Page, R. P. Sear, and D. Frenkel, "Two-step vapor-crystal nucleation close below triple point," J. Chem. Phys. **129**, 204505 (2008).
[12]L. Filion, M. Hermes, R. Ni, and M. Dijkstra, "Crystal nucleation of hard spheres using molecular dynamics, umbrella sampling, and forward flux sampling: A comparison of simulation techniques," J. Chem. Phys. **133**, 244115 (2010).
[13]S. Hussain and A. Haji-Akbari, "How to quantify and avoid finite size effects in computational studies of crystal nucleation: The case of heterogeneous ice nucleation," J. Chem. Phys. **154**, 014108 (2021).
[14]S. Hussain and A. Haji-Akbari, "How to quantify and avoid finite size effects in computational studies of crystal nucleation: The case of homogeneous crystal nucleation," J. Chem. Phys. **156**, 054503 (2022).
[15]E. E. Borrero and F. A. Escobedo, "Reaction coordinates and transition pathways of rare events via forward flux sampling," J. Chem. Phys. **127**, 164101 (2007).
[16]R. J. Allen, C. Valeriani, and P.-R. ten Wolde, "Forward flux sampling for rare event simulations," J. Phys.: Condens. Matter **21**, 463102 (2009).
[17]K. Kratzer, A. Arnold, and R. J. Allen, "Automatic, optimized interface placement in forward flux sampling simulations," J. Chem. Phys. **138**, 164112 (2013).
[18]E. E. Borrero and F. A. Escobedo, "Optimizing the sampling and staging for simulations of rare events via forward flux sampling schemes," J. Chem. Phys. **129**, 024115 (2008).
[19]H. E. A. Huitema, J. P. van der Eerden, J. J. M. Janssen, and H. Human, "Thermodynamics and kinetics of homogeneous crystal nucleation studied by computer simulation," Phys. Rev. B **62**, 14690–14702 (2000).
[20]J. D. Honeycutt and H. C. Andersen, "Small system size artifacts in the molecular dynamics simulation of homogeneous crystal nucleation in supercooled atomic liquids," J. Phys. Chem. **90**, 1585–1589 (1986).
[21]D. Quigley and P. M. Rodger, "A metadynamics-based approach to sampling crystallisation events," Mol. Simul. **35**, 613–623 (2009).
[22]J.-M. Leyssale, J. Delhommelle, and C. Millot, "A molecular dynamics study of homogeneous crystal nucleation in liquid nitrogen," Chem. Phys. Lett. **375**, 612–618 (2003).
[23]A. Mahata and M. A. Zaeem, "Size effect in molecular dynamics simulation of nucleation process during solidification of pure metals: Investigating modified embedded atom method interatomic potentials," Modell. Simul. Mater. Sci. Eng. **27**, 085015 (2019).
[24]T. Li, D. Donadio, and G. Galli, "Nucleation of tetrahedral solids: A molecular dynamics study of supercooled liquid silicon," J. Chem. Phys. **131**, 224519 (2009).
[25]B. Cheng and M. Ceriotti, "Bridging the gap between atomistic and macroscopic models of homogeneous nucleation," J. Chem. Phys. **146**, 034106 (2017).
[26]S. W. Hall, G. Díaz Leines, S. Sarupria, and J. Rogal, "Practical guide to replica exchange transition interface sampling and forward flux sampling," J. Chem. Phys. **156**, 200901 (2022).

[27]S. Hussain and A. Haji-Akbari, "Studying rare events using forward-flux sampling: Recent breakthroughs and future outlook," J. Chem. Phys. **152**, 060901 (2020).
[28]R. J. Allen, D. Frenkel, and P. R. ten Wolde, "Simulating rare events in equilibrium or nonequilibrium stochastic systems," J. Chem. Phys. **124**, 024102 (2006).
[29]V. G. Baidakov and K. R. Protsenko, "Spontaneous crystallization of a supercooled Lennard-Jones liquid: Molecular dynamics simulation," J. Phys. Chem. B **123**, 8103–8112 (2019).
[30]S. L. Meadley and F. A. Escobedo, "Thermodynamics and kinetics of bubble nucleation: Simulation methodology," J. Chem. Phys. **137**, 074109 (2012).
[31]A. 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, "LAMMPS—A flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," Comput. Phys. Commun. **271**, 108171 (2022).
[32]See https://matsci.org/t/lammps-users-dynamic-group-membership-atom-style-varia ble-consistency/35454 for a discussion of the lack of consistency when dynamic groups are used to determine number of solid atoms in LAMMPS.
[33]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys. **126**, 014101 (2007).
[34]G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," J. Chem. Phys. **101**, 4177–4189 (1994).
[35]P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, "Simulation of homogeneous crystal nucleation close to coexistence," Faraday Discuss. **104**, 93–110 (1996).
[36]P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," Phys. Rev. B **28**, 784–805 (1983).
[37]G. Casella and R. L. Berger, *Statistical Inference* (Cengage Learning, 2021).
[38]W. Zhao and T. Li, "On the challenge of sampling multiple nucleation pathways: A case study of heterogeneous ice nucleation on FCC (211) surface," J. Chem. Phys. **158**, 124501 (2023).
[39]D. Mandal and D. Quigley, "Kinetic control of competing nuclei in a dimer lattice-gas model," J. Chem. Phys. **157**, 214501 (2022).
[40]A. Haji-Akbari, "Forward-flux sampling with jumpy order parameters," J. Chem. Phys. **149**, 072303 (2018).
[41]W. C. Swope and H. C. Andersen, "10^6-particle molecular-dynamics study of homogeneous nucleation of crystals in a supercooled atomic liquid," Phys. Rev. B **41**, 7042–7054 (1990).
[42]X.-M. Bai and M. Li, "Test of classical nucleation theory via molecular-dynamics simulation," J. Chem. Phys. **122**, 224510 (2005).
[43]D. T. Yarullin, B. N. Galimzyanov, and A. V. Mokshin, "Direct evaluation of attachment and detachment rate factors of atoms in crystallizing supercooled liquids," J. Chem. Phys. **152**, 224501 (2020).
[44]R. J. Allen, P. B. Warren, and P. R. ten Wolde, "Sampling rare switching events in biochemical networks," Phys. Rev. Lett. **94**, 018104 (2005).
[45]W. Humphrey, A. Dalke, and K. Schulten, "VMD—Visual molecular dynamics," J. Mol. Graphics **14**, 33–38 (1996).
[46]T. S. van Erp, D. Moroni, and P. G. Bolhuis, "A novel path sampling method for the calculation of rate constants," J. Chem. Phys. **118**, 7762–7774 (2003).
[47]V. Thapar and F. A. Escobedo, "Simultaneous estimation of free energies and rates using forward flux sampling and mean first passage times," J. Chem. Phys. **143**, 244113 (2015).