How to Quantify and Avoid Finite Size Effects in Computational Studies of Crystal Nucleation

Sarwar Hussain\textsuperscript{1} and Amir Haji-Akbari\textsuperscript{1,*}

\textsuperscript{1}Department of Chemical and Environmental Engineering, Yale University, New Haven, CT 06520

(Dated: August 25, 2020)

Computational studies of crystal nucleation can be impacted by finite size effects, primarily due to unphysical interactions between crystalline nuclei and their periodic images. It is, however, not always feasible to systematically investigate the sensitivity of nucleation kinetics and mechanism to system size due to large computational costs of nucleation studies. Here, we use jumping forward flux sampling to accurately compute the rates of heterogeneous ice nucleation in the vicinity of square-shaped model structureless ice nucleating particles (INPs) of different sizes, and identify three distinct regimes for the dependence of rate on the INP dimension, $L$. For small INPs, the rate is a strong and non-monotonic function of $L$ due to artificial spanning of critical nuclei across the periodic boundary. Intermediate-sized INPs, however, give rise to the emergence of non-spanning 'proximal' nuclei that are close enough to their periodic images to fully structure the intermediary liquid. While such proximity can facilitate nucleation, its effect is offset by the compression of the intermediary liquid by the growing non-proximal nuclei, leading to artificially small nucleation rates overall. The critical nuclei formed at large INPs are neither spanning nor proximal. Yet, the rate is a weak function of $L$, with its logarithm scaling linearly with $1/L$. The key heuristic emerging from these observations is that finite size effects will be minimal if critical nuclei are neither spanning nor proximal, and if the density of the unstructured part of the intermediary liquid is statistically indistinguishable from the supercooled liquid density under the same conditions.

I. INTRODUCTION

The main premise of molecular simulations is to use the information obtained from simulating finite-sized systems to predict their behavior in the thermodynamic limit. The accuracy of such predictions, however, can depend strongly on the size of the simulated system, as estimates of thermodynamic,\textsuperscript{1–8} structural,\textsuperscript{9} and transport\textsuperscript{10} properties and nucleation rates\textsuperscript{11} in small systems can deviate from those in the thermodynamic limit in a statistically significant manner. Such a dependence on system size is typically referred to as finite size effects, which can be fairly strong for very small systems, while being mostly unnoticed for larger systems. Therefore, finite size effects can, in principle, be mitigated by simulating very large systems, a task that is only computationally feasible for simple model systems.\textsuperscript{12,13} Fortunately, this is not always necessary as similar conclusions can usually be obtained from simulating sufficiently large\textsuperscript{14} computationally tractable systems.\textsuperscript{12,14}

What constitutes sufficiently large, however, is subject to the property that is being estimated or the process that is being studied. For instance, a system comprised of several hundred molecules is usually large enough for accurately estimating thermodynamic, structural and transport properties of liquids,\textsuperscript{1–3,9,10,15} but might be too small for studying collective phenomena such as cavitation,\textsuperscript{16} condensation\textsuperscript{17} and crystal nucleation.\textsuperscript{1–8,18,19} It is therefore critical to develop heuristics for determining what qualifies as sufficiently large\textsuperscript{18} for studying such collective phenomena, in order to ensure the accuracy and reliability of the conducted simulations.

One such collective phenomenon that has been extensively studied using molecular simulations is crystal nucleation. As such, understanding the role of finite size effects on the thermodynamics and kinetics of crystal nucleation has been a topic of interest for decades.\textsuperscript{11} Nucleation is a process in which a sufficiently large nucleus of the new phase forms within the old metastable phase, and is usually the rate-limiting step of a first-order phase transition when the underlying thermodynamic driving force is small.\textsuperscript{20} Finite size effects in nucleation primarily arise due to periodic boundary conditions, which can result in an unphysical confinement of the metastable phase between the nucleus and its periodic images,\textsuperscript{15,18,19,21} or the formation of nuclei that span across the periodic boundary.\textsuperscript{16,22} In the case of crystal nucleation, the effect of periodic boundaries might be stronger due the extension of the diffuse crystal-liquid interface beyond the nucleus.\textsuperscript{12} However, finite size effects can also arise due to other factors such as solute depletion in multi-component systems,\textsuperscript{23,24} or peculiarities of the employed ensemble.\textsuperscript{14,17} These effects can collectively lead to unphysical nucleation rates in both homogeneous and heterogeneous nucleation, and has also been found to impact crystal growth.\textsuperscript{25–27} Indeed, the findings of several high-profile computational studies of nucleation are believed to have been strongly impacted by finite size effects. For instance, Matsumoto et al.'s observation\textsuperscript{28} of homogeneous ice nucleation in a system of 512 water molecules represented using the fully atomistic TIP4P\textsuperscript{29} model has never been reproduced in larger systems, and was later shown to be an artifact of strong finite size effects.\textsuperscript{30,31}

*Electronic address: amir.hajiakbaribalou@yale.edu
lier computational studies of surface freezing— or surface-induced homogeneous ice nucleation \cite{31,32}— by Vrbka and Jungwirth \cite{33,34} and Pluharova et al. \cite{35} are also believed to be impacted by finite size effects. \cite{36}

Early efforts to characterize finite size effects in crystal nucleation focused on homogeneous nucleation in the simple Lennard-Jones (LJ) liquid. \cite{37} For instance, Honeycutt and Andersen \cite{18,19} simulated systems of up to 1,500 LJ particles at a reduced density of 0.95 and a reduced temperature of 0.45 and concluded that the occurrence of the "catastrophic crystal growth" observed in earlier simulations of the deeply supercooled LJ liquid is not due to the emergence of a critical nucleus and is instead an artifact of periodic boundaries. Indeed, they later demonstrated that critical nuclei form way earlier than the catastrophic growth, but their average sizes and the time needed for their formation both tend to increase with system size, pointing to strong finite size effects even prior to catastrophic growth. \cite{15} Later, Swope and Andersen \cite{12} conducted large-scale MD simulations of 15,000 and $10^{6}$ LJ particles under similar conditions, and observed that the properties of the 15,000-particle system were similar to the average properties of the 64 subsystems within the million-particle simulation box. They therefore concluded that the 15,000-particle system is large enough to be devoid of finite size effects. A similar conclusion was reached by Huitema et al. \cite{38}, who examined homogeneous nucleation in systems with as many as 10,000 particles. According to these studies, avoiding finite size effects requires simulating systems that are at least three orders of magnitude larger than the characteristic critical nucleus size, a very stringent requirement that can only be satisfied for simple model systems. Consequently, such large-scale simulations of finite size effects in other systems are rare. \cite{13} This heuristic, however, based on observations in the high-rate regime, i.e., where nucleation can occur during a computationally tractable MD trajectory, and therefore the liquid structure is pre-disposed to freezing. It is therefore plausible to expect finite size effects to be weaker in the low-rate regime in which nucleation events are spatially isolated. Moreover, this heuristic can only apply to homogeneous nucleation at best, as the dependence of rate on system size might be completely different for heterogeneous nucleation.

Unfortunately, many of these questions are yet to be investigated in a systematic manner. Consequently, there are no rigorous guidelines or heuristics for avoiding finite size effects in crystal nucleation studies, and different authors have resorted to different ad hoc approaches to avoid finite size artifacts. While some have tested the robustness of their findings by repeating their simulations in computationally tractable larger systems, \cite{19-21} others have argued that finite size effects will be absent if the average distance between the critical nucleus and its periodic images is larger than half the box dimensions. \cite{22-24} This latter heuristic usually translates to critical nuclei that are at least an order of magnitude smaller than the system size in homogeneous nucleation, a heuristic that is sometimes referred to as the "10% rule". There are, however, reasons to doubt the adequacy of these approaches. The former approach can only be conclusive if rate calculations are conducted for a wide range of system sizes, an undertaking that is usually not feasible. The size and the distance thresholds invoked in the latter approach, on the other hand, are not based on any rigorous analysis, and can be fairly sensitive to the particulars of the algorithm utilized for detecting crystalline nuclei.

Here, we attempt to address some of these questions by systematically investigating how the rate and mechanism of heterogeneous ice nucleation within supercooled supported water nanofilms in the vicinity of a model structureless ice nucleating particle (INP) is affected by system size. Due to the surface-dominated nature of heterogeneous nucleation, the relevant "system size" is the dimensions of the INP. We use our recently developed jumpy forward flux sampling (jFFS) \cite{44} algorithm to compute nucleation rates for 16 different system sizes, comprising between 1,600 and 50,176 water molecules. We identify three distinct regimes for the dependence of rate and mechanism on system size, and identify the critical nuclei characteristics that signify each regime. Based on our observations, we devise a rigorous set of heuristics for assessing whether a particular nucleation rate calculation is impacted by finite size effects. Moreover, we provide a scaling approach for estimating the rate in the thermodynamic limit for system sizes where finite size effects are minimal but the rate is still a weak function of system size.


two.

II. METHODS

A. System Description and Molecular Dynamics Simulations

We consider heterogeneous nucleation in supported nanofilms of supercooled water in the vicinity of a model structureless INP at a temperature of 235 K. Unlike some earlier studies \cite{14,45} of heterogeneous ice nucleation in which the liquid film is sandwiched between the INP and its periodic image, the films considered in this work only touch the INP on one side, and are exposed to vacuum at the other interface. Despite being more expensive computationally, we believe that our setup constitutes a more faithful representation of heterogeneous nucleation in nature, where isolated INPs are in contact with a sea of the supercooled liquid. The water nanofilms considered in this work are approximately 4.8 nm thick, which makes it extremely unlikely for the free interface to impact the kinetics and mechanism of nucleation in a meaningful manner.

We model water molecules using the monoatomic water (mW) potential \cite{46}, a popular coarse-grained model of water obtained via re-parameterizing the Stillinger-Weber (SW) potential originally developed for modeling Group IV elements such as carbon and silicon. \cite{47} The
model INP is square-shaped and is located in the \(xy\) plane. It interacts with water molecules via the Lennard-Jones (LJ) 9-3 potential\(^{18}\) with \(\epsilon = 1.2\) kcal \cdot mol\(^{-1}\) and \(\sigma = 0.32\) nm and with a cutoff of 0.8 nm. These parameters are chosen so that the underlying INP does not induce ice nucleation very strongly. The utilized \(\epsilon\) value, in particular, is the smallest \(\epsilon\) for which heterogeneous nucleation is observed in unbiased 50-ns long MD simulations at 215 K. All MD simulations are performed in the canonical (NVT) ensemble using LAMMPS.\(^{49}\) Equations of motion are integrated using the velocity-Verlet algorithm with a time step of 5 fs, while temperature is controlled using the Nose-Hoover\(^{50,51}\) thermostat with a time constant of 0.5 ps.

As mentioned above, we utilize \(L\), the dimension of the structureless INP, as a proxy for system size. We consider a total of 16 different \(L\)’s, ranging from 3.19 nm to 17.85 nm, and use the following procedure for initializing the water nanofilms. We first generate a properly sized slab of cubic ice comprised of \(n \times n \times 8\) unit cells. The oxygen-oxygen distance in each unit cell, \(r_{OO}\) is adjusted so that the target \(L\) is an integer multiple of \(L_c = 4r_{OO}/\sqrt{3}\), the unit cell dimension of cubic ice.\(^{52}\) For most \(L\)’s, we use the value of \(r_{OO} = 0.276\) nm, while for a few smaller system sizes, \(r_{OO}\) is slightly adjusted in order to fit an integer number of unit cells within the box. These include boxes with lateral dimensions of 3.38, 3.61, 4.06 and 4.28 nm, for which the \(r_{OO}\) values of 0.24, 0.26, 0.25 and 0.26 nm are utilized, respectively. Each ice film is then melted at a temperature of 300 K. A minimum of 150 configurations are saved along the melted trajectory every 50 ps, which are then quenched to the target temperature of 235 K at a cooling rate of 7.69 ps \cdot K\(^{-1}\). A list of all \(L\)’s as well as the number of water molecules within each film is given in Table 1.

### TABLE I: Heterogeneous ice nucleation rates computed at 235 K using jFFS. \(N_p\) refers to the number of water molecules within each film. All error bars correspond to 95% confidence intervals.

| \(N_p\) | \(L\) [nm] | \(N^s\) | \(\log_{10} \lambda\) [m\(^{-2}\) \cdot s\(^{-1}\)] |
|-------|-------|-------|-----------------|
| 1,600 | 3.1869 | 145.3 ± 0.6 | 12.0243 ± 0.0941 |
| 2,304 | 3.3803 | 152.7 ± 1.0 | 11.5403 ± 0.0667 |
| 2,304 | 3.6056 | 177.8 ± 0.5 | 9.9900 ± 0.0734 |
| 2,304 | 3.8243 | 195.7 ± 0.6 | 9.1136 ± 0.0723 |
| 3,136 | 4.0563 | 230.2 ± 0.7 | 8.9292 ± 0.0651 |
| 3,136 | 4.2817 | 256.0 ± 0.8 | 7.7432 ± 0.0854 |
| 3,136 | 4.4617 | 260.1 ± 0.7 | 10.2308 ± 0.0928 |
| 4,096 | 5.0991 | 271.5 ± 0.6 | 8.3486 ± 0.0901 |
| 5,184 | 5.7365 | 275.6 ± 0.8 | 8.1461 ± 0.0737 |
| 9,216 | 7.6487 | 277.4 ± 0.6 | 8.5010 ± 0.0695 |
| 12,544 | 8.9234 | 268.0 ± 1.9 | 8.5019 ± 0.0668 |
| 16,384 | 10.1983 | 270.8 ± 0.6 | 8.8893 ± 0.0689 |
| 23,104 | 12.1104 | 272.5 ± 0.6 | 9.0289 ± 0.0623 |
| 30,976 | 14.0226 | 274.5 ± 0.8 | 9.2391 ± 0.0605 |
| 40,000 | 15.9348 | 271.6 ± 0.6 | 9.7301 ± 0.0589 |
| 50,176 | 17.8470 | 273.7 ± 0.6 | 9.9583 ± 0.0580 |

### B. Rate Calculations

#### 1. Order Parameter

We compute nucleation rates using our recently developed jFFS algorithm,\(^{44}\) which is a generalized variant of the forward flux sampling algorithm (FFS)\(^{53}\) that has been extensively utilized to study rare events.\(^{54}\) Similar to most other advanced sampling techniques, conducting an FFS calculation requires an order parameter, a mathematical function \(\lambda: Q \rightarrow \mathbb{R}\) that quantifies the progress of the corresponding rare event, in this case, heterogeneous ice nucleation. Here, \(Q\) is the configuration space that includes all the microscopic degrees of freedom of the corresponding system, i.e., the positions of all water molecules. In this work, we use the number of molecules in the largest crystalline nucleus as the order parameter. Each molecule \(i\) is classified as solid-like or liquid-like based on \(q_6(i)\), the Steinhart bond order parameter,\(^{55}\) given by,

\[
q_6(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} \frac{q_6(i) \cdot q_6^*(j)}{|q_6(i)||q_6(j)|}.
\]

Here, \(N_b(i)\) is the number of water molecules within a distance of \(r_c = 0.32\) nm from \(i\). \(q_6(i)\), however, is a \((2l + 1)\)-component complex-valued vector, and its components are given by,

\[
q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(\theta_{ij}, \phi_{ij}), \quad -l \leq m \leq +l
\]

with \(\theta_{ij}\) and \(\phi_{ij}\) the polar and azimuthal angles corresponding to the separation vector \(r_{ij} = r_j - r_i\), and \(Y_{lm}(\cdot, \cdot)\’s\), the spherical harmonic functions. Consistent with our earlier studies,\(^{36,44,56-58}\) we classify molecules with \(q_6 \geq 0.5\) as solid-like, cluster the solid-like molecules that are within a distance of \(r_c\) into crystalline nuclei, and apply the chain exclusion algorithm of Reinhardt et al.\(^{59}\)

We use \(q_6(i)\) to distinguish between solid-like molecules that have the local structure of cubic and hexagonal ice using the cutoffs given in Ref. 60.

#### 2. jFFS Calculations

With \(\lambda(\cdot)\) at hand, the transition region between the supercooled liquid basins \(A := \{x \in Q : \lambda(x) < \lambda_A\}\) and the crystalline basins \(B := \{x \in Q : \lambda(x) \geq \lambda_B\}\) is partitioned into \(N\) non-overlapping regions using \(N\) milestones \(\lambda_A < \lambda_0 < \lambda_1 \cdots \lambda_N = \lambda_B\), which are level sets of \(\lambda(\cdot)\). As demonstrated earlier,\(^{44}\) our utilized \(\lambda(\cdot)\) is jumpy as it undergoes high-frequency high-amplitude fluctuations along an MD trajectory. We therefore need to use jFFS in order to accurately capture the potentially subtle changes in rate upon changing \(L\). In order to minimize the number of FFS iterations, we follow the approach...
The statistical uncertainty of the computed $J$ is accurately determined from the committor probability $\Phi_0$, the flux of trajectories that leave $A$ and cross $\lambda_0$, is given by,

$$\Phi_0 = \frac{N_0}{T L^2},$$

with $T$ the combined duration of the MD trajectories and $L^2$ the surface area of the underlying INP. The next step is to compute the probability that a trajectory initiated from $C_0$ reaches $B$ by recursively computing the transition probabilities between successive milestones. In the simplified scheme of jFFS utilized here, the intermediate milestones are chosen so that it is beyond $\lambda_{k-1, \text{max}} = \max_{x \in C_{k-1}} \lambda(x)$, wherein $C_{k-1}$ contains all the configurations obtained upon a first crossing of $\lambda_{k-1}$. After choosing the next target milestone $\lambda_k$, a large number of trial trajectories are initiated from the configurations in $C_{k-1}$ with their momenta sampled from the Maxwell-Boltzmann distribution, and each trajectory is terminated upon crossing $\lambda_k$ or returning to $A$. The transition probability $P(\lambda_k | \lambda_{k-1})$ is then computed as the fraction of trial trajectories that cross $\lambda_k$. Note that $\lambda_B$ is not known a priori. Instead, the calculation is terminated when $P(\lambda_k | \lambda_{k-1})$ is statistically indistinguishable from unity. The nucleation rate $J$ is then computed from,

$$J = \Phi_0 \prod_{k=1}^{N} P(\lambda_k | \lambda_{k-1}),$$

The statistical uncertainty of the computed $J$ is estimated using the approach described in Ref. 53. All reported error bars correspond to 95% confidence intervals, i.e., twice the standard errors obtained from this approach.

It has been previously demonstrated in numerous studies\(^{43,61}\) that $\lambda(\cdot)$ is a good reaction coordinate for crystal nucleation. This implies that the critical nucleus size can be accurately determined from the committor probability given by,

$$p_c(\lambda_k) = \prod_{q=k+1}^{N} P(\lambda_q | \lambda_{q-1}),$$

More precisely, $N^*$, the critical nucleus size, is estimated by fitting $p_c(\lambda)$ to the following expression,

$$p_c(\lambda) = \frac{1}{2} \left\{ 1 + \text{erf} \left[ a(\lambda - N^*) \right] \right\},$$

with the reported error bar corresponding to the 95% confidence interval of the $N^*$ estimate within such a fit. In probing the properties of critical configurations, we identify $\lambda_{k^*}$, the closest milestone to $N^*$, and denote all $x \in C_{k^*}$ as critical. For most systems, this results in configurations with $32\% \leq p_c(\lambda(x)) \leq 68\%$. In two systems, i.e., $L = 12.11$ and $14.02$ nm, $N^*$ is not sufficiently close to any of the milestones, so we choose the configurations obtained from first crossings of the two closest milestones, which results in committor probabilities no smaller than $16\%$ and no larger than $78\%$.

3. Analysis of Nucleation Mechanism and Identification of Bottlenecks

Since a successful nucleation pathway in jFFS is generated sequentially by concatenating the trial trajectories between successive milestones, we can trace back the ancestry of any configuration in $C_N$, and identify the surviving configurations in earlier milestones, i.e., those with some progeny at $\lambda_B$. We denote the surviving subset of $C_k$ as $C^*_k$. By comparing the properties of $C^*_k$ and $C_k$, we can identify the important features that play a key role in successful nucleation. More specifically, for a given mechanical observable $\mu : Q \to \mathbb{R}$, one can compute $\mu_s(\lambda_k) = \langle \mu(x) \rangle_{x \in C^*_k}$ and $\mu_a(\lambda_k) = \langle \mu(x) \rangle_{x \in C_k}$. We denote the milestone at which $\mu_s(\lambda)$ and $\mu_a(\lambda)$ cross each other a bottleneck for $\mu(\cdot)$, and the corresponding average $\mu_B$. As demonstrated in our prior applications of this approach,\(^{36,44,57,62}\) the existence of a noticeable bottleneck indicates that $\mu(\cdot)$ is an important feature (orthogonal to the order parameter) that determines the likelihood of an early-milestone configuration to succeed in having progeny at $\lambda_B$.

III. RESULTS AND DISCUSSIONS

A. Summary of Nucleation Rates

We first explore the dependence of nucleation rate on $L$, the dimension of the square-shaped INP. The computed rates are shown in Fig. 1A and listed in Table I. We can, in particular, identify three distinct regimes for the dependence of rate on $L$. For small INPs, i.e., for $L \leq 4.46$ nm, the rate is a strong and non-monotonic function of $L$, and changes by as much as five orders of magnitude, indicative of strong finite size effects and potentially spurious behavior. For large INPs, i.e., for $L \geq 7.65$ nm, the rate is a weak and monotonic function of $L$ and increases by less than two orders of magnitude upon increasing $L$. These two regimes are highlighted in Fig. 1 with shaded green and red, respectively. There is, however, a third intermediate regime that is highlighted with a gradient shade in Fig. 1, and that also exhibits a weak and monotonic dependence of rate on $L$. Unlike the second regime, however, the rate decreases slightly upon increasing $L$ in this intermediate regime.

In order to understand the origin of these contrasting behaviors, we focus on the geometric features of the
critical nuclei. (See Section II B 2 for our definition of critical configurations.) We first determine the fraction of the critical nuclei that span across the periodic box, i.e., that have at least one constituent solid-like molecule with a nearest neighbor within their periodic images. Such spanning nuclei exhibit artificial directionality and spurious long range crystalline order along the $x$ and/or $y$ dimensions of the simulation box. Fig. 1B depicts the fraction of critical configuration with spanning crystalline nuclei. While such nuclei are prevalent for very small $L$’s, their fraction decreases upon increasing $L$ and eventually drops to zero at $L = 5.74$ nm. The presence of an appreciable number of spanning critical nuclei is the main feature that distinguishes the small-INP regime from the other two. We will therefore refer to the shaded green region of Fig. 1 as the spanning regime. The other two regimes, however, lack an appreciable fraction of spanning critical nuclei, and thus exhibit a weaker dependence of rate on $L$.

Finite size effects can, however, exist in the absence of spanning critical nuclei, as the proximity of a critical nucleus to its periodic image can still render noticeable finite size effects even in the absence of spanning. In order to devise a more rigorous measure of such proximity, we identify $\mathbf{u}$, the shortest vector connecting a crystalline nucleus to its closest periodic image, and compute the liquid density profile along $\mathbf{u}$. Note that for a spanning nucleus, $\mathbf{u} = \mathbf{0}$ by definition. For a non-spanning crystalline nucleus $\mathcal{N}$ comprised of $m$ water molecules with

![FIG. 1: (A) The dependence of heterogeneous nucleation rate on $L$, the dimension of the square-shaped INP. (B) The fraction of critical nuclei that are spanning (circles) and proximal (squares). Error bars correspond to 95% confidence intervals and are smaller than the symbols.](image)

![FIG. 2: $\rho_L(r)$, the inter-image liquid density as a function of the distance from the critical cluster for a few representative values of $L$.](image)

positions $\mathcal{N} \equiv \{ \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m \}$, $\mathbf{u}$ is given by:

$$[\hat{i}, \hat{j}, \hat{p}] = \arg\min_{i,j \leq m, p} |\mathbf{r}_i - \mathbf{r}_j^p|,$$

$$\mathbf{u} = \mathbf{r}_j^p - \mathbf{r}_i.$$  

Here, $\mathbf{p} \in \{(\pm L, 0, 0), (0, \pm L, 0)\}$ is a displacement vector that translates $\mathcal{N}$ into its four closest periodic images, and $\mathbf{r}_j^p = \mathbf{r}_j + \mathbf{p}$. The density profile is computed from the molecules that lie within the intersection of a cylinder of radius 0.32 nm along $\mathbf{u}$ and the INP. We call the part of the liquid that resides therein the "inter-image" liquid, and the associated density profile $\rho_L(r)$ the inter-image density profile. Here, $r$ refers to the closest $\mathbf{u}$-projected distance from either $\mathbf{r}_i$ or $\mathbf{r}_j^p$. Fig. 2 depicts $\rho_L(r)$ as a function of $r$ for several system sizes where the spanning fraction is negligible. The inter-image liquid is fairly structured in the immediate vicinity of the nucleus, as evident in the three peaks of $\rho_L(r)$ for $r < 0.77$ nm. Note that the heights and loci of these peaks are independent of the system size, so their emergence is unlikely to be impacted by finite size effects. Beyond this threshold, which we denote by $r_{c,p}$, the liquid density reaches a plateau. We therefore call the nuclei with $|\mathbf{u}| \leq 2r_{c,p}$ "proximal" since the inter-image liquid confined between them and their closest periodic images clearly lacks bulk-like behavior. Since $|\mathbf{u}| = 0$ for a spanning cluster, all spanning clusters are also proximal. Fig. 1B depicts the fraction of proximal critical nuclei as a function of $L$. Clearly, the presence of an appreciable fraction of proximal critical nuclei in the intermediate regime, and their absence in the large-INP regime is the main feature that distinguishes those two regimes. We refer to the large-INP regime as the non-spanning regime. The intermediate regime, however, constitutes a transition from lack of spanning and proximity in the non-spanning regime, to a preponderance of spanning and proximal critical nuclei in the spanning regime. We will analyze these three regimes separately and will explain in detail how the differences
We first compute order to understand the origin of such non-monotonicity, clei (i.e., those with $u_L$ strictly decreasing function of $L$). This indicates that the ability to form spanning nuclei artificially promotes nucleation. The fraction of spanning critical nuclei, however, is a magnitude and the nature of finite size effects.

### B. The Spanning Regime

Due to the close proximity of critical nuclei and their periodic images in the spanning regime, the kinetics and mechanism of nucleation exhibit a strong dependence on system size. Overall, we observe that an increase in the fraction of spanning configurations results in an increase in rate (Fig. 1) and a decrease in the size of the critical nucleus (Table 1). This indicates that the ability to form spanning nuclei artificially promotes nucleation. The fraction of spanning critical nuclei, however, is a strictly decreasing function of $L$ and thus fails to fully explain the non-monotonic dependence of rate on $L$. In order to understand the origin of such non-monotonicity, we first compute $d_{\text{min}}^{\text{proj},s}$ for non-spanning crystalline nuclei (i.e., those with $u \neq 0$) where $d_{\text{min}}^{\text{proj}} = |(u_x, u_y, 0)|_2$ is the minimum $xy$-projected distances between a crystalline nucleus and its closest periodic image. Focusing on this lateral distance enables us to detect and quantify peculiarities that arise prior to spanning. Fig. 3 shows $d_{\text{min}}^{\text{proj},s}$ as a function of crystalline nucleus size for all seven films in the spanning regime. In addition to all non-spanning configurations at a given milestone, we compute $d_{\text{min}}^{\text{proj},s}$ for the surviving non-spanning configurations as well. As expected, both $d_{\text{min}}^{\text{proj},s}(\lambda)$ and $d_{\text{min}}^{\text{proj},a}(\lambda)$ are strictly decreasing functions of $\lambda$, but their decline is faster in smaller films. At earlier FFS milestones and prior to the bottlenecks (the green circles in Fig. 3), however, $d_{\text{min}}^{\text{proj},s}(\lambda)$ is consistently larger than $d_{\text{min}}^{\text{proj},a}(\lambda)$ irrespective of the system size. This suggests that at initial stages of nucleation, a less spread-out nucleus is better suited to survive and to contribute to the nucleation pathway. Beyond the bottlenecks, which are always smaller than $N^*$, $d_{\text{min}}^{\text{proj},s}(\lambda)$ remains consistently smaller than $d_{\text{min}}^{\text{proj},a}(\lambda)$. This suggests that the more compact surviving nuclei of earlier milestones reach a geometry at the bottleneck that facilitates their further growth towards their periodic images to form spanning or proximal critical nuclei. The bottlenecks of $d_{\text{min}}^{\text{proj}}(\lambda)$ therefore constitute important milestones in the nucleation process, and the properties of the crystalline nuclei therein are likely to play an important role in the kinetics and mechanism of nucleation.

We first explore our hypothesis that a smaller $d_{\text{min}}^{\text{proj},s}$ will likely make the non-spanning crystalline nuclei at the bottleneck more prone to span at later milestones, and will thus result in faster nucleation. Indeed, a weak linear correlation (with a $R^2$ of 0.71) exists between $d_{\text{min}}^{\text{proj},s}$ and $\log_{10} J$ (Fig. 4A). This indicates that while $d_{\text{min}}^{\text{proj},s}$ is a decent predictor of $J$ in the spanning regime, it does not fully explain the observed variations in rate. In particular, the two pairs of systems with almost identical $d_{\text{min}}^{\text{proj},s}$ values—namely $L = 3.61$ and 3.82 nm, and 4.06 and 4.28 nm—have almost identical $d_{\text{min}}^{\text{proj},s}$ values but rates that differ by approximately one order of magnitude. One possible explanation for the subpar predictive power of $d_{\text{min}}^{\text{proj},s}$ is that it does not contain contributions from the spanning configurations, which, as can be seen in Fig. 5B, comprise between 3% to 18% of all configurations at the bottleneck. Indeed, the two anomalous pairs mentioned above only differ in their fraction of spanning nuclei at the bottleneck, with the system with the higher...
Another factor that can impact the ease of crystal nucleation in small systems is the ability of new crystalline motifs to fit within the inter-image region without any mechanical strain. For instance, if the spanning of a crystalline nucleus requires the incorporation of a certain number of crystalline unit cells in the inter-image region, a projected minimum distance that deviates significantly from an integer multiple of the relevant dimension of the unit cell will make spanning more difficult, as the new unit cells will have to be mechanically strained to fit within the inter-image region. In order to assess the importance of such effects, we first use the local $q_1$ order parameter to determine the dominant polymorph of ice favored by the structureless INP. We focus on the largest system, and compute the fraction of solid-like molecules within each layer of the critical crystalline nuclei that have local environments commensurate with cubic ice. As can be seen in Fig. 6A, the first two layers of such nuclei are considerably less cubic than the remaining layers. We reach similar conclusions upon applying the cage detection algorithm of Ref. 57 to identify double-diamond and hexagonal cages, which are the topological building blocks of cubic and hexagonal ice, respectively. As can be seen in Figs. 6B-C for a representative critical nucleus, the first two layers are almost exclusively comprised of hexagonal cages, which touch the INP through their basal planes (Fig. 6C). These observations suggest that the structureless INP considered in this work strongly favors the formation of the basal plane of hexagonal ice ($I_h$).

The prevalence of the hexagonal polymorph at the INP implies that the ease by which a non-spanning configuration can span across the simulation box will depend on whether the inter-image region can be filled with strain-free motifs of hexagonal ice. This latter assertion follows from the fact that spanning usually occurs within the first few layers of the nucleus where $I_h$ is preferred, and will only be possible if $d_{\text{proj}}$ is an integer multiple of a discrete set of "magic distances" corresponding to the unit cell of hexagonal ice. Unfortunately, the precise values of such "magic distances" will depend on the nature of the crystal-liquid interface at $i$ and $j$ given by Eq. (7), i.e., the endpoints of the inter-image vector $\mathbf{u}$. It is therefore not possible to identify universal magic distances and analyze how deviations from those affect span-
Cubicity (%)

Cooled liquid within the plateau region of Fig. 2. Regimes is likely governed by the properties of the supercooled liquid at its vicinity, and make its density a function of \( z \), the distance from the INP. In particular, the density profile of Fig. 7B indicates the existence of three distinct liquid layers. As such, the average liquid density within the inter-image plateau region should be compared to the average liquid density along \( u \), which can, in general, be different from the bulk density. It is, however, difficult to accurately compute these quantities in situations where \( i \) and \( j \) are within different layers of the liquid. We therefore simplify our analysis by computing inter-image plateau densities for configurations with \( i \)'s and \( j \)'s simultaneously lying within the first three liquid layers. For a configuration \( x \) with \( r_i \) and \( r_j \) in the same layer \( L \), the inter-image plateau density is computed by enumerating \( N_p(x) \) the number of water molecules that reside within a cuboid of length \( \rho_{\text{proj}} \), width \( w = 0.64 \) nm and height \( h \) along \( u_{xy} = (u_x, u_y, 0) \) and centered at the midpoint between the cluster and its closest periodic image. We choose \( w \) so that all in-layer nearest neighbors of molecules along \( u_{xy} \) are included within the cuboid, while \( h \) is chosen as the thickness of \( L \) according to Fig. 7B. A typical cuboid for a configuration with \( u \) within the first liquid layer is shown from three different angles in Fig. 7A. The average inter-image plateau density for each layer is then estimated as the weighted average of individual densities

\[
\rho_p(x) = \frac{\sum_{x \in L} N_p(x) \rho_p(x)}{\sum_{x \in L} N_p(x)},
\]

The average liquid density within \( L \) is estimated as,

\[
\rho_L = \frac{\int_L \rho_s(z)dz}{h_L},
\]
Liquid Density (g cm$^{-3}$)
Dimension of the Square-shaped INP (nm)

where $\rho_l(z)$ is the local liquid density at a distance $z$ from the INP.

Fig. 7C depicts the $\rho_p^C$'s of the three layers highlighted in Fig. 7B for the critical configurations in the intermediate and non-spanning regimes. For the two films in the intermediate regime, the inter-image plateau densities are significantly larger than the per-layer supercooled liquid densities. This effect likely arises from the negatively sloped melting curve of water, i.e., the fact that the liquid is denser than the crystal under ambient conditions. More precisely, the incorporation of new water molecules into the growing crystalline nucleus results in an increase in its volume, and the compression of the inter-image liquid during the out-of-equilibrium nucleation process. Since the thermodynamic driving force for crystallization is a decreasing function of density in water, larger inter-image densities will result in larger nucleation barriers and smaller nucleation rates. While this effect is partly offset by the existence of an appreciable fraction of proximal critical nuclei, it results in rates that are generally smaller than those in the other two regimes. (There is only one system size in the spanning regime with a smaller nucleation rate than those in the intermediate regime.)

This difference almost disappears in the non-spanning regime, presumably due to the larger size of the inter-image region, which better "absorbs" the growing nucleus front. Indeed, the inter-image plateau densities are statistically indistinguishable from the corresponding liquid densities for $L \geq 8.92$ nm. For the smallest non-spanning system, i.e., $L = 7.65$ nm, there is only a small—but statistically significant—difference between the inter-image plateau density for configurations with $i$ and $j$ in the first layer, and the corresponding liquid density. These findings generally indicate that the inter-image region is large enough to not be impacted by strong finite size effects in the non-spanning regime. While these findings are based on the configurations with us entirely within one of the first three layers, we expect a qualitatively similar behavior if configurations with inter-layer minimum image connections are also included.

Despite lack of any quantifiable structural signatures, the nucleation rate is still a weak function of $L$ in the non-
spanning regime. Indeed, the rate increases by around 1.5 orders of magnitude upon increasing \( L \) from 7.65 nm and 17.85 nm. The dependence of rate on system size can be satisfactorily described using a linear correlation between \( \log_{10} J \) and \( 1/L \), as depicted in Fig. 8. This correlation is satisfied only for the six larger systems, i.e., those for which the plateau inter-image density is statistically indistinguishable from the liquid density. It has indeed been reported that the averages of properties such as critical temperature,\(^5\) free energy difference,\(^6\) structure factor,\(^9\) diffusivity,\(^10\) surface tension\(^4\) and percolation threshold\(^63\) have a power law dependence on the system size. The linear relationship between \( \log_{10} J \) and \( 1/L \) can therefore be rationalized by invoking the formalism of classical nucleation theory (CNT)\(^64\) according to which the nucleation rate is given by,

\[
J = A \exp \left( -\frac{\Delta G_{\text{nuc}}}{k_B T} \right),
\]

Here, \( A, k_B \) and \( T \) are the kinetic prefactor, the Boltzmann constant, and temperature, respectively. \( \Delta G_{\text{nuc}} \) is the nucleation barrier and is given by:

\[
\Delta G_{\text{nuc}} = \frac{4\pi \gamma_l^3 (1 - \cos \theta)^2 (2 + \cos \theta)}{3 \rho_s^2 |\Delta \mu|^2},
\]

with \( \gamma_l, \Delta \mu, \rho_s \) and \( \theta \), the solid-liquid surface tension, the chemical potential difference between the two phases, the solid number density, and the solid-liquid-INP contact angle, respectively. All these quantities are known to be impacted by finite size effects,\(^4\)\(^-\)\(^8\) and some of them exhibit a power-law dependence on system size.\(^4\)\(^6\) Consequently, the linear correlation between \( \log_{10} J \) and \( 1/L \) is a likely consequence of the power-law dependence of some of these underlying thermodynamic properties. This scaling enables us to use the rates computed in sufficiently large finite systems to estimate the heterogeneous nucleation rate in the thermodynamic limit. According to the correlation depicted in Fig. 8, the nucleation rate at the thermodynamic limit (\( L \to \infty \)) is given by \( \log_{10} J_{\infty} = 11.22 \pm 0.24 \).

The smallest system in the non-spanning regime (\( L = 7.65 \) nm) does not satisfy this linear scaling. This is not surprising considering the statistically significant difference between the plateau inter-image density and the supercooled liquid density within the first liquid layer. In other words, the 7.65-nm system shares certain features of the intermediate regime despite having zero proximal critical nuclei. This suggests that the transition from the intermediate to the non-spanning regime is neither sharp nor well-defined, and its precise boundary will depend on how ”proximity“ is defined.

**IV. CONCLUSION**

In this work, we use MD simulations and jumpy forward flux sampling to explore the sensitivity of the heterogeneous ice nucleation kinetics and mechanism to \( L \), the size of a model structureless INP, and find that finite size effects that arise due to periodic boundary conditions can significantly affect heterogeneous nucleation rates over a wide range of INP sizes. We observe that nucleation rates can change by as much as five orders of magnitude within the range of \( L \)'s considered in this work, and are also non-monotonic functions of \( L \). We identify three distinct regimes for the dependence of rate on \( L \) based on whether critical crystalline nuclei span across the periodic boundary, and if not, whether they are proximal, i.e., that the liquid that separates them from their closest periodic images is fully structured. In the spanning regime, an appreciable fraction of critical nuclei span across the periodic boundary, while the overwhelming majority of those that do not span are proximal. This results in a strong and non-monotonic dependence of rate on \( L \), which we explain by analyzing the average minimum projected distance of subcritical crystalline nuclei at the bottleneck FFS milestones. We demonstrate that the fraction of spanning bottleneck configurations, the average minimum projected distance for the non-spanning ones, and the overall cubicity of such nuclei collectively explain the variations in rate in the spanning regime. The second regime, which is observed for intermediate-sized INPs and is thus called the intermediate regime, is characterized by the emergence of proximal— but non-spanning—critical nuclei. Within this regime, the rate is a weak function of \( L \), and its variations is governed by the fraction of critical nuclei that are proximal, and the inter-image plateau liquid density value for those that are not. In the third regime, which is observed for large INPs and which we denote as the non-spanning regime, critical nuclei are neither spanning nor proximal, and the rate is a weakly increasing function of \( L \). We are able to demonstrate that there is a linear scaling between \( \log_{10} J \) and \( 1/L \), which we use to estimate the nucleation rate in the thermodynamic limit.

The key heuristic that emerge from this work is that finite size effects are minimal or absent in systems where the critical nuclei are neither spanning nor proximal, and the inter-image liquid density in the plateau region does not deviate significantly from that of the liquid under the same conditions. These are criteria that can be unambiguously tested for both homogeneous and heterogeneous crystal nucleation in all systems and irrespective of the particular definition of the order parameter. While whether a nucleus spans along the simulation box will depend on the classification algorithm utilized for detecting solid-like molecules, its proximity to its closest periodic image— or lack thereof— and its inter-image plateau density— if it is not proximal— are independent of such details. It must be noted that the particular system size beyond which these conditions are satisfied will depend on the type of the system and the thermodynamic state. It might therefore be risky to use *ad hoc* heuristics, such as the ”10% rule“ discussed in Section I to determine whether a rate calculation is impacted by finite size effects. Determining how these heuristics translate into
The system size is too small. Indeed, the rates computed in this regime have important implications for identifying finite size effects in systems where long-range electrostatic interactions are present. The strong sensitivity of rate to $L$ in the spanning regime has important implications for identifying finite size artifacts in computational studies of nucleation, as it points to the unpredictable nature of such effects when the system size is too small. Indeed, the rates computed for certain $L$’s in this regime are fortuitously identical to those computed for much larger $L$’s in the other two regimes. This suggests that the common practice of validating nucleation rates by comparing them to rates computed for a larger system might yield misleading results since both systems might be affected by finite size effects— but for different reasons.

**Acknowledgments**

A.H.-A. gratefully acknowledges the support of the National Science Foundation CAREER Award (Grant No. CBET-1751971). These calculations were performed on the Yale Center for Research Computing. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant no. ACI-1548562.

[1] K. Binder, Ferroelectrics **73**, 43 (1987), URL https://doi.org/10.1080/00150198708227908.
[2] K. Mon and K. Binder, J. Chem. Phys. **96**, 6989 (1992), URL https://doi.org/10.1063/1.462557.
[3] J. Horbach, W. Kob, K. Binder, and C. A. Angell, Phys. Rev. E **54**, R5897 (1996), URL https://doi.org/10.1103/PhysRevE.54.R5897.
[4] A. Aguado, W. Scott, and P. A. Madden, J. Chem. Phys. **115**, 8612 (2001), URL https://doi.org/10.1063/1.1410394.
[5] P. Orea, J. López-Lemus, and J. Alejandre, J. Chem. Phys. **123**, 114702 (2005), URL https://doi.org/10.1063/1.2018640.
[6] E. A. Mastny and J. J. de Pablo, J. Chem. Phys. **127**, 104504 (2007), URL https://doi.org/10.1063/1.2753149.
[7] F. Biscay, A. Ghoufi, F. Goujon, V. Lachet, and P. Malfeyt, J. Chem. Phys. **130**, 184710 (2009), URL https://doi.org/10.1063/1.3132708.
[8] R. Burt, G. Birkett, M. Salanne, and X. Zhao, J. Phys. Chem. C **152**, 15244 (2016), URL https://doi.org/10.1021/acs.jpcc.6b04696.
[9] J. Salacuse, A. Denton, and P. Engelstaff, Phys. Rev. E **53**, 2382 (1996), URL https://doi.org/10.1103/PhysRevE.53.2382.
[10] S. H. Jamali, L. Wolff, T. M. Becker, A. Bardow, T. J. Vlugt, and O. A. Moultos, J. Chem. Theory Comput. **14**, 2667 (2018), URL https://doi.org/10.1021/acs.jctc.8b00170.
[11] G. C. Sosso, J. Chen, S. J. Cox, M. Fitzner, P. Pedevilla, A. Zen, and A. Michaelides, Chem. Rev. **116**, 7078 (2016), URL http://dx.doi.org/10.1021/acs.chemrev.5b00744.
[12] W. C. Swope and H. C. Andersen, Phys. Rev. B **41**, 7042 (1990), URL https://doi.org/10.1103/PhysRevB.41.7042.
[13] N. J. English and J. S. Tse, Phys. Rev. E **92**, 032132 (2015), URL https://doi.org/10.1103/PhysRevE.92.032132.
[14] J. Wedekind, D. Reguera, and R. Strey, J. Chem. Phys. **125**, 214505 (2006), URL https://doi.org/10.1063/1.2402167.
[15] M. J. Mandell, J. Stat. Phys. **15**, 299 (1976), URL https://doi.org/10.1007/BF01023055.
[16] S. L. Meadley and F. A. Escobedo, J. Chem. Phys. **137**, 074109 (2012), URL http://dx.doi.org/10.1063/1.4745082.
[17] A. Pérez and A. Rubio, J. Chem. Phys. **135**, 244505 (2011), URL https://doi.org/10.1063/1.3672063.
[18] J. Honeycutt and H. C. Andersen, Chem. Phys. Lett. **108**, 535 (1984), URL https://doi.org/10.1016/0009-2614(84)85049-6.
[19] J. D. Honeycutt and H. C. Andersen, J. Phys. Chem. **90**, 1585 (1986), URL https://doi.org/10.1021/j10399a026.
[20] H. Falahati and A. Haji-Akbari, Soft Matter **15**, 1135 (2019), URL https://doi.org/10.1039/C8SM02285B.
[21] L. J. Peng, J. R. Morris, and R. S. Aga, J. Chem. Phys. **133**, 084505 (2010), URL https://doi.org/10.1063/1.3472301.
[22] A. Statt, P. Virnau, and K. Binder, Phys. Rev. Lett. **114**, 026101 (2015), URL https://doi.org/10.1103/PhysRevLett.114.026101.
[23] M. Salvaglio, C. Perego, F. Giberti, M. Mazzotti, and M. Parrinello, Proc. Natl. Acad. Sci. USA **112**, E6 (2015), URL https://doi.org/10.1073/pnas.1421192111.
[24] C. Liu, G. P. Wood, and E. E. Santiso, Mol. Phys. **116**, 2908 (2018), URL http://dx.doi.org/10.1080/00268976.2018.1482016.
[25] E. Burke, J. Q. Broughton, and G. H. Gilmer, J. Chem. Phys. **89**, 1030 (1988), URL https://doi.org/10.1063/1.455254.
[26] B. O’Malley and I. Snook, Phys. Rev. Lett. **90**, 085702 (2003), URL https://doi.org/10.1103/PhysRevLett.90.085702.
[27] F. H. Streitz, J. N. Glosli, and M. V. Patel, Phys. Rev. Lett. **96**, 225701 (2006), URL https://doi.org/10.1103/PhysRevLett.96.225701.
[28] M. Matsumoto, S. Saito, and I. Ohmine, Nature **416**, 409 (2002), URL http://dx.doi.org/10.1038/416409a.
[29] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983), URL http://dx.doi.org/10.1063/1.445869.
[30] E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, and A. Zen, A. Michaelides, Chem. Rev. **116**, 2667 (2018), URL https://doi.org/10.1021/acs.chemrev.7b00266.
J. L. F. Abascal, and C. Valeriani, J. Am. Chem. Soc. 135, 15008 (2013), URL http://dx.doi.org/10.1021/ja4028814.

[31] A. Tabazadeh, Y. S. Djikaev, and H. Reiss, Proc. Natl. Acad. Sci. USA 99, 15873 (2002), URL http://dx.doi.org/10.1073/pnas.252640699.

[32] A. Haji-Akbari and P. G. Debenedetti, J. Chem. Phys. 147, 060001 (2017), URL https://doi.org/10.1063/1.4985879.

[33] L. Vrbka and P. Jungwirth, J. Phys. Chem. B 110, 18126 (2006), URL http://dx.doi.org/10.1021/jp060402c.

[34] L. Vrbka and P. Jungwirth, J. Mol. Liq. 134, 64 (2007), URL http://dx.doi.org/10.1016/j.molliq.2006.12.011.

[35] E. Pluhaťová, L. Vrbka, and P. Jungwirth, J. Phys. Chem. C 114, 7831 (2010), URL http://dx.doi.org/10.1021/jp9090238.

[36] A. Haji-Akbari and P. G. Debenedetti, Proc. Natl. Acad. Sci. USA 114, 3316 (2017), URL http://dx.doi.org/10.1073/pnas.1620999114.

[37] J. E. Lennard-Jones, Proc. R. Soc. Lond. A 106, 463 (1924), URL https://doi.org/10.1098/rspa.1924.0082.

[38] H. Huitema, J. van der Eerden, J. Janssen, and H. Hueman, Phys. Rev. B 62, 14690 (2000), URL https://doi.org/10.1103/PhysRevB.62.14690.

[39] S. J. Cox, Z. Raza, S. M. Kathmann, B. Slater, and A. Michaelides, Faraday Discuss. 167, 389 (2013), URL http://dx.doi.org/10.1039/C3FD00059A.

[40] S. J. Cox, S. M. Kathmann, B. Slater, and A. Michaelides, J. Chem. Phys. 142, 184704 (2015), URL http://dx.doi.org/10.1063/1.4919714.

[41] S. J. Cox, S. M. Kathmann, B. Slater, and A. Michaelides, J. Chem. Phys. 142, 184705 (2015), URL https://doi.org/10.1063/1.4919715.

[42] G. C. Sosso, T. Li, D. Donadio, G. A. Triolo, and A. Michaelides, J. Phys. Chem. Lett. 7, 3250 (2016), URL http://dx.doi.org/10.1021/acs.jpcl.6b01013.

[43] L. Lupi, B. Peters, and V. Molinero, J. Chem. Phys. 145, 211910 (2016), URL https://doi.org/10.1063/1.4961652.

[44] A. Haji-Akbari, J. Chem. Phys. 149, 072303 (2018), URL https://doi.org/10.1063/1.5018303JCP.2018.ESMS2018.issue-1.

[45] Y. Bi, R. Cabriolu, and T. Li, J. Phys. Chem. C 120, 1507 (2016), URL http://dx.doi.org/10.1021/acs.jpcc.5b09740.

[46] V. Molinero and E. B. Moore, J. Phys. Chem. B 113, 4008 (2009), URL http://dx.doi.org/10.1021/jp805227c.

[47] F. H. Stillinger and T. A. Weber, Phys. Rev. B 31, 5262 (1985), URL http://dx.doi.org/10.1013/PhysRevB.31.5262.

[48] J. Magda, M. Tirrell, and H. Davis, J. Chem. Phys. 83, 1888 (1985), URL https://doi.org/10.1063/1.449375.

[49] S. J. Plimpton, J Comp. Phys. 117, 1 (1995), URL http://dx.doi.org/10.1063/1.479520.

[50] S. Nosé, Mol. Phys. 52, 255 (1984), URL http://dx.doi.org/10.1080/00268978400101201.

[51] W. G. Hoover, Phys. Rev. A 31, 1695 (1985), URL http://dx.doi.org/10.1103/PhysRevA.31.1695.

[52] A. Haji-Akbari, in Antifreeze Proteins Volume 1 (Springer, 2020), pp. 13–51, URL https://doi.org/10.1007/978-3-030-41929-5_3.

[53] R. J. Allen, D. Frenkel, and P. R. ten Wolde, J. Chem. Phys. 124, 024102 (2006), URL http://dx.doi.org/10.1063/1.2140273.

[54] S. Hussain and A. Haji-Akbari, J. Chem. Phys. 152, 060001 (2020), URL https://doi.org/10.1063/1.5127780.

[55] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Phys. Rev. B 28, 784 (1983), URL http://dx.doi.org/10.1103/PhysRevB.28.784.

[56] A. Haji-Akbari, R. S. DeFever, S. Sarupria, and P. G. Debenedetti, Phys. Chem. Chem. Phys. 16, 25916 (2014), URL http://dx.doi.org/10.1039/C4CP03948C.

[57] A. Haji-Akbari and P. G. Debenedetti, Proc. Natl. Acad. Sci. USA 112, 10582 (2015), URL http://dx.doi.org/10.1073/pnas.1509267112.

[58] M. M. Gianetti, A. Haji-Akbari, M. P. Longinotti, and P. G. Debenedetti, Phys. Chem. Chem. Phys. 18, 4102 (2016), URL http://dx.doi.org/10.1039/C5CP06535F.

[59] A. Reinhardt, J. P. K. Doye, E. G. Noya, and C. Vega, J. Chem. Phys. 137, 194504 (2012), URL http://dx.doi.org/10.1063/1.4766362.

[60] T. Li, D. Donadio, G. Russo, and G. Galli, Phys. Chem. Chem. Phys. 13, 19807 (2011), URL http://dx.doi.org/10.1039/C1CP22167A.

[61] L. Lupi, A. Hudaït, B. Peters, M. Grünwald, R. G. Mullen, A. H. Nguyen, and V. Molinero, Nature 551, 218 (2017), URL http://dx.doi.org/10.1038/nature24279.

[62] Y. E. Altabet, A. Haji-Akbari, and P. G. Debenedetti, Proc. Natl. Acad. Sci. USA 114, E2548 (2017), URL http://dx.doi.org/10.1073/pnas.1620335114.

[63] A. Haji-Akbari and R. M. Ziff, Phys. Rev. E 79, 021118 (2009), URL http://dx.doi.org/10.1103/PhysRevE.79.021118.

[64] D. Turnbull, J. Chem. Phys. 18, 198 (1950), URL http://dx.doi.org/10.1063/1.1747588.

[65] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, et al., Comput. Sci. Eng. 16, 62 (2014), URL http://dx.doi.org/10.1109/MCSE.2014.80.