Crystal nucleation of hard spheres using molecular dynamics, umbrella sampling, and forward flux sampling: A comparison of simulation techniques

L. Filion,[a)] M. Hermes, R. Ni, and M. Dijkstra

Soft Condensed Matter, Debye Institute for NanoMaterials Science, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

Over the last number of years several simulation methods have been introduced to study rare events such as nucleation. In this paper we examine the crystal nucleation rate of hard spheres using three such numerical techniques: molecular dynamics, forward flux sampling, and a Bennett–Chandler-type theory where the nucleation barrier is determined using umbrella sampling simulations. The resulting nucleation rates are compared with the experimental rates of Harland and van Megen [Phys. Rev. E **55**, 3054 (1997)], Sinn *et al.* [Prog. Colloid Polym. Sci. **118**, 266 (2001)], Schätzel and Ackerson [Phys. Rev. E **48**, 3766 (1993)], and the predicted rates for monodisperse and 5% polydisperse hard spheres of Auer and Frenkel [Nature **409**, 1020 (2001)]. When the rates are examined in units of the long-time diffusion coefficient, we find agreement between all the theoretically predicted nucleation rates, however, the experimental results display a markedly different behavior for low supersaturation. Additionally, we examined the precritical nuclei arising in the molecular dynamics, forward flux sampling, and umbrella sampling simulations. The structure of the nuclei appears independent of the simulation method, and in all cases, the nuclei contains on average significantly more face-centered-cubic ordered particles than hexagonal-close-packed ordered particles. © *2010 American Institute of Physics*. [doi:10.1063/1.3506838]

I. INTRODUCTION

Nucleation processes are ubiquitous in both natural and artificially-synthesized systems. However, the occurrence of a nucleation event is often rare and difficult to examine both experimentally and theoretically.

Colloidal systems are almost ideal model systems for studying nucleation phenomena. Nucleation and the proceeding crystallization in such systems often take place on experimentally accessible time scales, and due to the size of the particles, they are accessible to a wide variety of scattering and imaging techniques, such as (confocal) microscopy,[1] holography,[2] and light and x-ray scattering. Additionally, progress in particle synthesis,[3] solvent manipulation, and the application of external fields[4] allows for significant control over the interparticle interactions, allowing for the study of a large variety of nucleation processes.

One such colloidal system is the experimental realization of "hard" spheres comprised of sterically stabilized polymethylmethacrylate (PMMA) particles suspended in a liquid mixture of decaline and carbon disulfide.[5] Experimentally, the phase behavior of such a system has been examined by Pusey and van Megen[6] and maps well onto the phase behavior predicted for hard spheres. Specifically, when the effective volume fraction of their system is scaled to reproduce the freezing volume fraction of hard spheres ($\eta = 0.495$) the resulting melting volume fraction is $\eta = 0.545 \pm 0.003$ (Ref. 6) which is in good agreement with that predicted for hard

spheres.[7] The nucleation rates have been measured using light scattering by Harland and van Megen,[5] Sinn *et al.*,[8] Schätzel and Ackerson[9] and predicted theoretically by Auer and Frenkel.[10]

On the theoretical side, hard-sphere systems are one of the simplest systems which can be applied to the study of colloidal and nanoparticle systems, and generally, toward the nucleation process itself. As such, it is an ideal system to examine various computational methods for studying nucleation, and comparing the results with experimental data. Such methods include, but are not limited to, molecular dynamics (MD) simulations, umbrella sampling (US), forward flux sampling (FFS), and transition path sampling (TPS). It is worth noting here that Auer and Frenkel[10] used umbrella sampling simulations to study crystal nucleation of hard spheres and found a significant difference between their predicted rates and the experimental rates of Refs. 5, 8, and 9. However, it was unclear where this difference originated. In this paper, we compare the nucleation rates for the hard-sphere system from MD, US, and FFS simulations with the experimental results of Refs. 5, 8, and 9. We demonstrate that the three simulation techniques are consistent in their prediction of the nucleation rates, despite the fact that they treat the dynamics differently. Thus, we conclude that the difference between the experimental and theoretical nucleation rates identified by Auer and Frenkel is not due to the simulation method.

A nucleation event occurs when a statistical fluctuation in a supersaturated liquid results in the formation of a crystal nucleus large enough to grow out and continue crystallizing the surrounding fluid. In general, small crystal nuclei are

---

a)Author to whom correspondence should be addressed. Electronic mail: L.C.Filion@uu.nl.

continuously being formed and melt back in a liquid. However, while most of these small nuclei will quickly melt, in a supersaturated liquid a fraction of these nuclei will grow out. Classical nucleation theory (CNT) is the simplest theory available for describing this process. In CNT it is assumed that the free energy for making a small nucleus is given by a surface free-energy cost which is proportional to the surface area of the nucleus and a bulk free-energy gain proportional to its volume. More specifically, according to CNT the Gibbs free-energy difference between a homogeneous bulk fluid and a system containing a spherical nucleus of radius R is given by

$$\Delta G(R) = 4\pi\gamma R^2 - \tfrac{4}{3}\pi |\Delta\mu| \rho_s R^3, \qquad (1)$$

where $|\Delta\mu|$ is the difference in chemical potential between the fluid and solid phases, ρ_s is the density of the solid, and γ is the interfacial free energy of the fluid–solid interface. This free-energy difference is usually referred to as the nucleation barrier. From this expression, the radius of the critical cluster is found to be $R^* = 2\gamma / |\Delta\mu| \rho_s$ and the barrier height is $\Delta G^* = 16\pi\gamma^3/3\rho_s^2 |\Delta\mu|^2$.

Umbrella sampling[11,12] is a method to examine the nucleation process from which the nucleation barrier is easily obtained. The predicted barrier can then be used in combination with kinetic Monte Carlo (KMC) or MD simulations to determine the nucleation rate.[10] In US an order parameter for the system is chosen and configuration averages for sequential values of the order parameter are taken. In order to facilitate such averaging, the system is biased toward particular regions in configuration space. The success of the method is expected to depend largely on the choice of order parameter and biasing potential. Note that the free-energy barrier is only defined in equilibrium, and thus is only applicable to systems which are in (quasi-) equilibrium.

Forward flux sampling[13–15] is a method of studying rare events, such as nucleation, in both equilibrium and non-equilibrium systems. Using FFS, the transition rate constants (e.g., the nucleation rate) for rare events can be determined when brute force simulations are difficult or even not possible. In FFS, a reaction coordinate Q (similar to the order parameter in US) is introduced which follows the rare event. The transition rate between phase A and B is then expressed as a product of the flux ($\Phi_{A\lambda_0}$) of trajectories crossing the A state boundary, typically denoted λ_0, and the probability ($P(\lambda_B|\lambda_0)$) that a trajectory which has crossed this boundary will reach state B before returning to state A. Thus the transition rate constant is written as

$$k_{AB} = \Phi_{A\lambda_0} P(\lambda_B|\lambda_0). \qquad (2)$$

Forward flux sampling facilitates the calculation of probability $P(\lambda_B|\lambda_0)$ by breaking it up into a set of probabilities between sequential values of the reaction coordinate. Little information regarding the details of the nucleation process is required in advance, and the choice of reaction coordinate is expected to be less important than the order parameter in US. Additionally, unlike US, FFS utilizes dynamical simulations

TABLE I. Packing fraction ($\eta = \pi\sigma^3 N/6V$), reduced pressure ($\beta p \sigma^3$), reduced chemical potential difference between the fluid and solid phases ($\beta |\Delta\mu|$) and reduced number density of the solid phase ρ_s of the state points studied in this paper. The chemical potential difference was determined using thermodynamic integration Ref. 17, and the equations of state for the fluid and solid are from Refs. 18 and 19, respectively.

η	$\beta p \sigma^3$	$\beta |\Delta\mu|$	$\rho_s \sigma^3$
0.5214	15.0	0.34	1.107
0.5284	16.0	0.44	1.122
0.5316	16.4	0.48	1.128
0.5348	16.9	0.53	1.135
0.5352	17.0	0.54	1.136
0.5381	17.5	0.58	1.142
0.5414	18.0	0.63	1.148
0.5478	19.1	0.74	1.161
0.5572	20.8	0.90	1.178

and hence this technique does not assume that the system is in (quasi-)equilibrium.

Molecular dynamics and Brownian dynamics (BD) simulations are ideal for studying the time evolution of systems, and, when possible, they are the natural techniques to study dynamical processes such as nucleation. Unfortunately, available computational time often limits the types of systems, which can be effectively studied by these dynamical techniques. Brownian dynamics simulations, which would be the natural choice to use for colloidal systems, are very slow due to the small time steps required to handle the steep potential used to approximate the hard-sphere potential. Event driven MD simulations are much more efficient to simulate hard spheres and enable us to study spontaneous nucleation of hard-sphere systems over a range of volume fractions. The main difference between the two simulation methods regards how they treat the short-time motion of the particles. Fortunately, the nucleation rate is only dependent on the long-time dynamics which are not sensitive to the details of the short-time dynamics of the system.[16]

In this paper we study in detail the application of US and FFS techniques to crystal nucleation of hard spheres, and predict the associated nucleation rates. Combining these nucleation rates with results from MD simulations, we make predictions for the nucleation rates over a wide range of packing fractions $\eta = 0.5214 - 0.5572$, with corresponding pressures and supersaturations shown in Table I. We compare these theoretical nucleation rates with the rates measured experimentally by Refs. 5, 8, and 9.

This paper is organized as follows: in Sec. II we discuss the model, in Sec. III we describe and examine the order parameter used to distinguish between solid-like and fluidlike particles throughout this paper, in Sec. IV we calculate essentially the "exact" nucleation rates using MD simulations, in Secs. V and VI we calculate the nucleation rates of hard spheres using US and FFS respectively, and discuss difficulties in the application of these techniques, in Sec. VII we summarize the theoretical results and compare the predicted nucleation rates with the measured experimental rates of Harland and van Megen,[5] Sinn *et al.*,[8] and Schätzel and Ackerson[9] and Sec. VIII contains our conclusions.

II. MODEL

In this paper, we examine the nucleation rate between spheres with diameter σ which interact via a hard-sphere pair potential given by

$$\beta U^{HS}(r_{ij}) = \begin{cases} 0 & r_{ij} \geq \sigma \\ \infty & r_{ij} < \sigma, \end{cases} \tag{3}$$

where r_{ij} is the center-to-center distance between particles i and j and $\beta = 1/k_B T$ with k_B Boltzmann's constant and T the temperature. This is in contrast to several studies on "hard" spheres where the hard-sphere potential is approximated by a slightly soft potential (e.g., Refs. [20] and [21]) so that Brownian dynamics simulations or traditional molecular dynamics simulations (i.e., molecular dynamics which is not event driven), which require a continuous potential, can be used. We would like to emphasize this distinction here as the hardness of the interaction has previously been shown to play a significant role in nucleation rates.[22,23]

III. ORDER PARAMETER

In this paper, an order parameter is used to differentiate between liquid-like and solid-like particles and a cluster algorithm is used to identify the solid clusters. We have chosen to use the local bond-order parameter introduced by ten Wolde *et al.* [24,25] in the study of crystal nucleation in a Lennard-Jones system. This order parameter has been used in many crystal nucleation studies, including a previous study of hard-sphere nucleation by Auer and Frenkel.[10]

In the calculation of the local bond order parameter a list of "neighbors" is determined for each particle. The neighbors of particle i include all particles within a radial distance r_c of particle i, and the total number of neighbors is denoted $N_b(i)$. A bond orientational order parameter $q_{l,m}(i)$ for each particle is then defined as

$$q_{l,m}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{l,m}(\theta_{i,j}, \phi_{i,j}), \tag{4}$$

where $Y_{l,m}(\theta, \phi)$ are the spherical harmonics, $m \in [-l, l]$ and $\theta_{i,j}$ and $\phi_{i,j}$ are the polar and azimuthal angles of the center-of-mass distance vector $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ with $\mathbf{r}_i$ the position vector of particle i. Solid-like particles are identified as particles for which the number of connections per particle $\xi(i)$ is at least ξ_c and where

$$\xi(i) = \sum_{j=1}^{N_b(i)} H(d_l(i,j) - d_c), \tag{5}$$

H is the Heaviside step function, d_c is the dot-product cutoff, and

$$d_l(i,j) = \frac{\sum_{m=-l}^{l} q_{l,m}(i) q_{l,m}^*(j)}{\left(\sum_{m=-l}^{l} |q_{l,m}(i)|^2\right)^{1/2} \left(\sum_{m=-l}^{l} |q_{l,m}(j)|^2\right)^{1/2}}. \tag{6}$$

A cluster contains all solid-like particles which have a solid-like neighbor in the same cluster. Thus, each particle can be a member of only a single cluster.
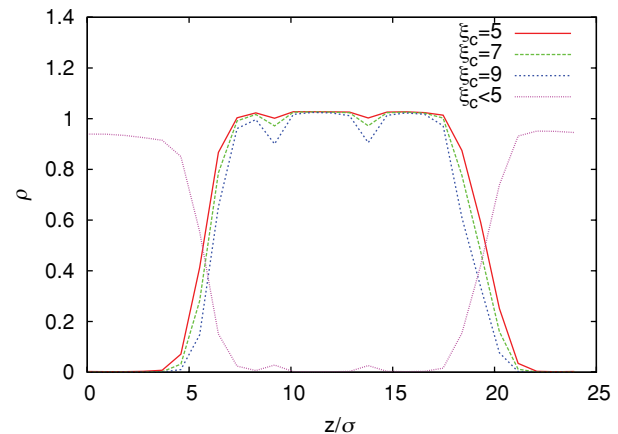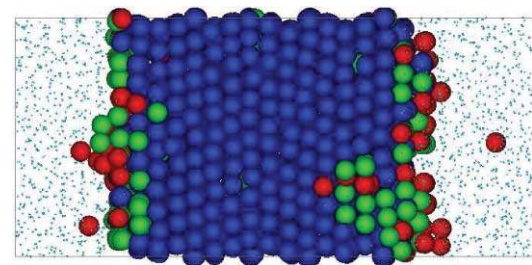

FIG. 1. Top: A typical configuration of an equilibrated random-hexagonal-close-packed (RHCP) crystal in coexistence with an equilibrated fluid. The crystalline particles are labeled according to three different crystallinity criteria: the red particles have $\xi = 5$ or 6 crystalline bonds, the green particles have $\xi = 7$ or 8 crystalline bonds and the blue particles have $\xi \geq 9$ crystalline bonds. The fluidlike particles ($\xi < 5$) are denoted by dots. Bottom: The density profile of particles with a minimum number of neighbors ξ as labeled. Note that the dips in the density profile correspond to HCP stacked layers. This implies that near the interface, the order parameter is slightly more sensitive to FCC ordered particles than to HCP ordered particles.

The parameters contained in this algorithm include the neighbor cutoff r_c, the dot-product cutoff d_c, the critical value for the number of solid-like neighbors ξ_c, and the symmetry index for the bond orientational order parameter l. The solid nucleus of a hard-sphere crystal is expected to have random hexagonal order, thus the symmetry index is chosen to be 6 in all cases in this study. Note that this order parameter does not distinguish between FCC and HCP ordered particles.

To investigate the effect of the choice of ξ_c, we examined the number of correlated bonds per particle at the liquid–solid interface. To this end, we constructed a configuration in the coexistence region in an elongated box by attaching a box containing an equilibrated random-hexagonal-close-packed (RHCP) crystal to a box containing an equilibrated fluid. Note that the RHCP crystal was placed in the box such that the hexagonal layers were parallel to the interface. The new box was then equilibrated in an NPT MC simulation. We then examined the density profile of solid-like particles as determined by our order parameter using $r_c = 1.4\sigma$, $d_c = 0.7$, and $\xi_c = 5$, 7, and 9. Figure 1 presents the density profiles along with a typical configuration of the RHCP crystal in coexistence with the fluid phase. As shown in Fig. 1, for all values of ξ_c that we examined the

order parameter appears to consistently identify the particles belonging to the bulk fluid and solid regions. The solid-like particles as defined by the order parameter are labeled according to the number of solid-like neighbors while the fluidlike particles are denoted by dots. The main difference between these order parameters relates to distinguishing between fluidlike and solid-like particles at the fluid–solid interface. Unsurprisingly, the location of the interface seems to shift in the direction of the bulk solid as ξ_c is increased. We note that the dips in the density profile correspond to HCP stacked layers, which are more pronounced for higher values of ξ_c.

IV. MOLECULAR DYNAMICS

A. Nucleation rates

In MD simulations the equations of motion are integrated to follow the time evolution of the system. Since the hard-sphere potential is discontinuous the interactions only take place when particles collide. Thus, the particles move in straight lines (ballistic) until they encounter another particle with which they perform an elastic collision.[26] These collision events are identified and handled in order of occurrence using an event driven simulation.

In theory, using an MD simulation to determine nucleation rates is quite simple. Starting with an equilibrated fluid configuration, an MD simulation is used to evolve the system until the largest cluster in the system exceeds the critical nucleus size. The MD time associated with such an event is then measured and averaged over many initial configurations. The nucleation rate is given by

$$k = \frac{1}{\langle t \rangle V},\tag{7}$$

where V is the volume of the system and $\langle t \rangle$ is the average time to form a critical nucleus. Measuring this time is relatively easy for low supersaturations where the nucleation times are relatively long compared to the nucleation event itself, which corresponds with a steep increase in the crystalline fraction of the system. However, for high supersaturations pinpointing the time of a nucleation event is more difficult. Often many nuclei form immediately and the critical nucleus sizes must be estimated from CNT or US simulations. Additionally, the precise details of the initial configuration can play a role at high supersaturations since the equilibration time of the fluid is of the same order of magnitude as the nucleation time. Hence, for each individual MD simulation we used a new initial configuration, which was created by quenching the system very quickly.

For the results in this paper, we performed MD simulations with up to 100 000 particles in a cubic box with periodic boundary conditions in an NVE ensemble. Time was measured in MD units $\sigma\sqrt{m/k_BT}$. The order parameter was measured every 10 time units and when the largest cluster exceeded the critical size by 100% we estimated the time τ_{nucl} at which the critical nucleus was formed using stored previous

TABLE II. The average nucleation time, obtained from MD simulations, to form a critical cluster that grew out and filled the box. The last column contains the rate (k) in units of $(6D_L)/\sigma^5$.

Volume fraction (η)	Average nucleation time ($t\sqrt{k_BT/(m\sigma^2)}$)	Rate $[k\sigma^5/(6D_L)]$
0.5316	1×10^6	5×10^{-9}
0.5348	1.7×10^4	3.6×10^{-7}
0.5381	1.4×10^3	5.3×10^{-6}
0.5414	2.0×10^2	4.3×10^{-5}
0.5478	42	3.0×10^{-4}
0.5572	10	2.4×10^{-3}

configurations. We performed up to 20 runs for every density and averaged the nucleation times.

The results are shown in Table II. The nucleation times shown here are for a system of 2.0×10^4 particles and in MD time units. Note that for $\eta = 0.5381$ we also checked the effect of system size on the nucleation rate by performing MD simulations with 100 000 particles and did not find a significant difference. To compare with other data we converted the MD time units to units of $\sigma^2/(6D_L)$ where D_L is the long-time diffusion coefficient measured in the same MD simulations. We were not able to measure the long-time diffusion coefficients for high densities because our measurements were influenced by crystallization. We used the fit obtained by Zaccarelli *et al.*[27] who used polydisperse particles to prevent crystallization. For $\eta < 0.54$, we find good agreement between our data for D_L and this fit.

V. UMBRELLA SAMPLING

A. Gibbs free-energy barriers

Umbrella sampling is a technique developed by Torrie and Valleau to study systems where Boltzmann-weighted sampling is inefficient.[11] This method has been applied frequently to study rare events, such as nucleation,[12] and specifically has been applied in the past to study the nucleation of hard spheres.[10] In general, umbrella sampling is used to examine parts of configurational space, which are inaccessible by traditional schemes, e.g., Metropolis Monte Carlo simulations. Typically, a biasing potential is added to the true interaction potential causing the system to oversample a region of configuration space. The biasing potential, however, is added in a manner such that it is easy to "un"-bias the measurables.

In the case of nucleation, while it is simple to sample the fluid, crystalline clusters of larger sizes will be rare, and as such, impossible to sample on reasonable time scales. The typical biasing potential for studying nucleation is given by[24,28]

$$U_{\text{bias}}(n(\mathbf{r}^N)) = \frac{\lambda}{2}(n(\mathbf{r}^N) - n_C)^2,\tag{8}$$

where λ is a coupling parameter, $n(\mathbf{r}^N)$ is the size of the largest cluster associated with configuration $\mathbf{r}^N$, and n_C is the targeted cluster size. By choosing λ carefully, the simulation will fluctuate around the part of configurational space with
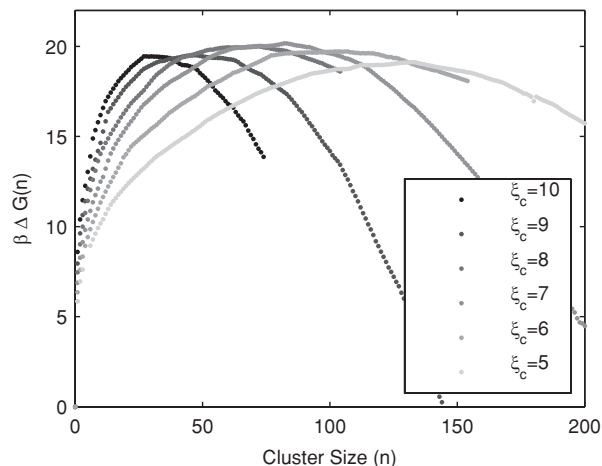
FIG. 2. Gibbs free-energy barriers $\beta \Delta G(n)$ as a function of cluster size n as obtained from umbrella sampling simulations using varying critical number of solid-like neighbors ξ_c as labeled at a reduced pressure of $\beta p \sigma^3 = 17$. For $\xi_c = 5, 7$, and 9, the neighbor cutoff is $r_c = 1.4$ and for $\xi_c = 6, 8$, and 10, $r_c = 1.3$. In all cases the dot product cutoff is $d_c = 0.7$.

$n(\mathbf{r}^N)$ in the vicinity of n_C. The expectation value of an observable A is then given by

$$\langle A \rangle = \frac{\langle A / W(n(\mathbf{r}^N)) \rangle_{\text{bias}}}{\langle 1 / W(n(\mathbf{r}^N)) \rangle_{\text{bias}}}, \quad (9)$$

where

$$W(x) = e^{-\beta U_{\text{bias}}(x)}. \quad (10)$$

Using this scheme to measure the probability distribution $P(n)$ for clusters of size n, the Gibbs free-energy barrier can be determined by[29]

$$\beta \Delta G(n) = \text{constant} - \ln(P(n)). \quad (11)$$

Many more details on this method are given elsewhere.[17,29]

For a pressure of $\beta p \sigma^3 = 17$, corresponding to a supersaturation of $\beta |\Delta \mu| = 0.54$, we examine the effect of one of the order parameter variables, namely ξ_c, on the prediction of the nucleation barriers. The barriers predicted by US using $\xi_c = 5, 6, 7, 8, 9$, and 10 are shown in Fig. 2. Note that the height of the barriers does not depend on ξ_c. In general, for larger values of ξ_c more particles are identified as fluid as compared with smaller values of ξ_c. This is consistent with the differences between these order parameters as demonstrated in Fig. 1. Thus, the radius measured in our simulation will depend on the definition of the order parameter. However, from classical nucleation theory [Eq. (1)], there exists a unique definition of the liquid–solid interface and thus a unique radius associated with CNT which we define as R_{CNT}. To a first approximation, for each definition of the order parameter, this radius (R_{CNT}) differs from that measured by our simulation
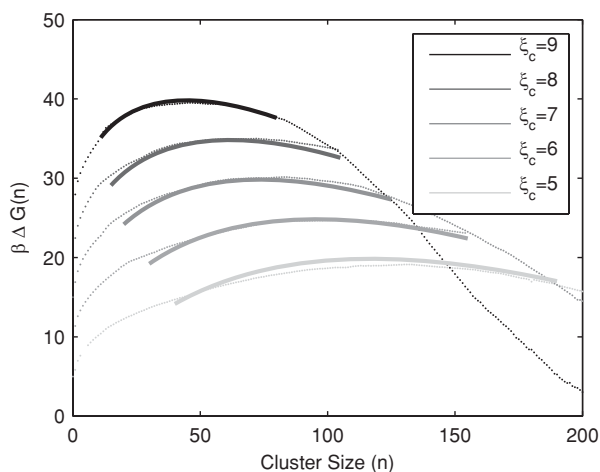

FIG. 3. Classical nucleation theory fits (thick lines) to the Gibbs free-energy barriers obtained from umbrella sampling simulations using varying ξ_c as labeled at a reduced pressure of $\beta p \sigma^3 = 17$. Note that the CNT radius (R_{CNT}) is related to the radius ($R(\xi_c)$) measured by umbrella sampling via $R(\xi_c) = R_{\text{CNT}} + \alpha(\xi_c)$, where $\alpha(\xi_c)$ is a constant that corrects for the different ways the various order parameters identify the particles at the fluid–solid interface. The fit parameters are given in Table III. We have shifted the barriers for $\xi_c = 6 - 9$ by $5, 10, 15$, and 20 $k_B T$, respectively, for clarity.

($R(\xi_c)$) by a constant which we denote as $\alpha(\xi_c)$, which is also dependent on ξ_c. Thus, we fit the barriers corresponding to $\xi_c = 5, 6, 7, 8$, and 9 using CNT where we have

$$R(\xi_c) = R_{\text{CNT}} + \alpha(\xi_c). \quad (12)$$

Note that we have assumed that the cluster size n can be related to the cluster radius $R(\xi_c)$ by

$$n(\xi_c) = \frac{4\pi R(\xi_c)^3 \rho_s}{3}. \quad (13)$$

Only the top part of the free-energy barriers are expected to fit to classical nucleation theory, so we take the top of the barrier corresponding to the region where the difference between $\beta \Delta G(n)$ and $\beta \Delta G(n^*)$ is approximately 5. Fitting all barriers simultaneously for the interfacial free energy γ, the classical nucleation theory radius R_{CNT}, and the various $\alpha(\xi_c)$, we obtain the fits displayed in Fig. 3. From the various values of α, the associated critical CNT radius (R_{CNT}^*) can be determined. We find $R_{\text{CNT}}^* = 2.49\sigma$. Additionally, we find an interfacial free energy of $\beta \gamma \sigma^2 = 0.76$ which roughly agrees with the results of Auer and Frenkel who obtained $\beta \gamma \sigma^2 = 0.699, 0.738$, and 0.748 for pressures $\beta p \sigma^3 = 15$, 16, and 17, respectively.[10] However, recent calculations by Davidchack et al.,[30] of the interfacial free energy at the fluid-solid coexistence find $\beta \gamma \sigma^2 = 0.574, 0.557$, and 0.546 for the crystal planes (100), (110), and (111), respectively. For a spherical nucleus, the interfacial free energy is expected to

TABLE III. Numerical values for the parameters associated with the fits in Figs. 3 and 4 for classical nucleation theory and the adjusted classical nucleation theory presented in this paper.

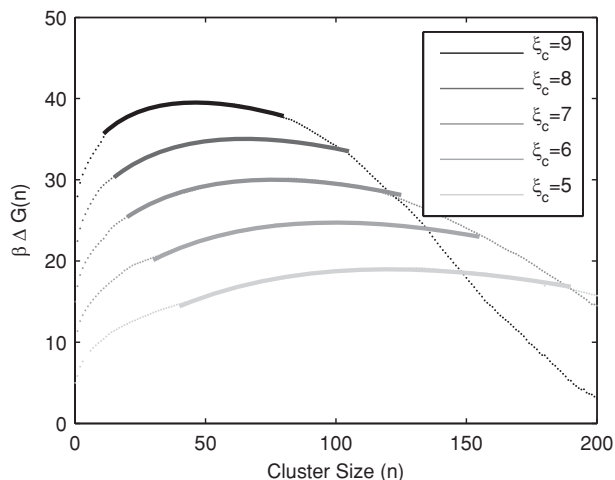|	$\beta \lvert \Delta \mu \rvert$	$\beta \gamma \sigma^2$	R_{CNT}^*	$\alpha(5)$	$\alpha(6)$	$\alpha(7)$	$\alpha(8)$	$\alpha(9)$	$c(5)$	$c(6)$	$c(7)$	$c(8)$	$c(9)$
CNT	0.54	0.76	2.49	-0.425	-0.231	-0.000	0.139	0.380					
ACNT	0.54	0.63	2.06	-0.879	-0.698	-0.464	-0.335	-0.076	7.80	8.56	8.84	8.87	8.34

FIG. 4. Fits of an adjusted classical nucleation theory (ACNT) presented in Sec. V A to the Gibbs free-energy barriers predicted using umbrella sampling simulations and using varying ξ_c as labeled at a reduced pressure of $\beta p \sigma^3 = 17$. Note that the CNT radius (R_{CNT}) is related to the radius measured by umbrella sampling by $R(\xi_c) = R_{CNT} + \alpha(\xi_c)$, where $\alpha(\xi_c)$ is a constant. The fit parameters are given in Table III. We have shifted the barriers for $\xi_c = 6 - 9$ by 5, 10, 15, and 20 $k_B T$, respectively, for clarity.

be an average over the crystal planes and was found to be $\beta \gamma \sigma^2 = 0.559$.[30] Thus, our result for the interfacial energy and that of Ref. 10 appear to be an overestimate.

There have been a number of papers discussing possible corrections to CNT (e.g., Refs. 31 and 32). Recent work on the 2D Ising model, a system where both the interfacial free energy and supersaturation are known analytically, demonstrated that in order to match a nucleation barrier obtained from US to CNT, two correction terms were required, specifically a term proportional to $\log(N)$ as well as a constant shift in ΔG which we define as c.[31] The US barrier is only expected to match CNT near the top of the barrier where the $\log(N)$ term is almost a constant. Thus, we propose fitting the barrier to an adjusted expression for CNT (ACNT), by adding a constant c to Eq. (1). Fitting the US barriers with this proposed form for the Gibbs free-energy barrier, where we assume c is a function of ξ_c, we obtain the fits displayed in Fig. 4. In this case we find an interfacial free energy $\beta \gamma \sigma^2 = 0.63$, and the values for $\alpha(\xi_c)$ and $c(\xi_c)$ are given in Table III. We note that this fit is much better than the fits in Fig. 3. The difference in the various $c(\xi_c)$ is around $1 k_B T$ and corresponds well to the difference in heights of the barriers. More strikingly, the interfacial free energy predicted from this proposed free-energy barrier is in much better agreement with recent calculations of Davidchack *et al.*,[30] than the interfacial free energy we calculate using classical nucleation theory directly. For a more thorough examination on the interfacial free energies of the hard-sphere model, see Ref. 35. We would like to point out here that due to the simple form of the nucleation barrier, it is difficult to be certain of any fit with more than one fitting parameter, as there are many combinations of parameters which fit almost equally well. In order to quantify the accuracy of these fits, we have calculated the root mean square of the residual for the two fits which we denote as σ_{RMSR}. In the case of the CNT fit we find $\sigma_{RMSR} = 0.50$ while for the

TABLE IV. Nucleation rates k in units of $6 D_L / \sigma^5$ with D_L the long time diffusion coefficient as a function of reduced pressure ($\beta p \sigma^3$) as predicted by umbrella sampling. $G''(n^*)$ is the second order derivative of the Gibbs free energy at the critical nucleus size n^*.

$\beta p \sigma^3$	ξ_c	n^*	$\beta \Delta G(n^*)$	$\beta \Delta G''(n^*)$	$f_{n^*}/6 D_L$	$k \sigma^5 / 6 D_L$
15	8	212	42.1 ± 0.2	-9.6×10^{-4}	2150	1.4×10^{-17}
16	8	112	27.5 ± 0.6	-1.6×10^{-3}	1950	3.5×10^{-11}
17	6	102	19.6 ± 0.3	-1.2×10^{-3}	3980	1.7×10^{-7}
17	8	72	20.0 ± 0.4	-2.0×10^{-3}	2620	9.9×10^{-8}
17	10	30	19.4 ± 0.7	-9.4×10^{-3}	1760	2.5×10^{-7}

ACNT fit we find $\sigma_{RMSR} = 0.11$ indicating that the ACNT fit is much better than the CNT fit. Additionally, we examined the ACNT fits for various interfacial free energies γ. Fixing the interfacial free energy in the ACNT fit to the value found by CNT ($\beta \gamma \sigma^2 = 0.76$), we find $\sigma_{RMSR} = 0.27$ and when we use interfacial free energy at coexistence[30] ($\beta \gamma \sigma^2 = 0.559$) we find $\sigma_{RMSR} = 0.18$.

Using either expressions for the Gibbs free-energy barrier, namely CNT and ACNT, we were unable to fit the barrier corresponding to $\beta p \sigma^3 = 17$ and $\xi_c = 10$ simultaneously with the other predicted barriers for the same pressure. We speculate that our difficulty in fitting the barrier at $\xi_c = 10$ stems from an "over-biasing" of the system. Specifically, by using $\xi_c = 10$ the biasing potential could cause the system to sample more frequently more ordered clusters, and hence change slightly the region of phase space available to the US simulations. In general, the least biased systems would be expected to explore the largest region of phase space resulting in the best results. It should be noted that, in fact, this problem is simply an equilibration and measuring problem, but it does emphasize the difficulty caused by using an overly strong biasing potential.

In conclusion, with the exception of $\xi_c = 10$, the value of ξ_c used in the order parameter did not appear to have an effect on the nucleation barriers once the difference in their measurements of the solid–liquid interface was taken into consideration. Finally, for use in our nucleation rate calculations (Sec. V B) we also determined the Gibbs free energy $\Delta G(n)$ for reduced pressures $\beta p \sigma^3 = 15$ and 16 using umbrella sampling simulations. We present the barrier heights in Table IV.

B. Umbrella sampling nucleation rates

The nucleation barriers as obtained from US simulations can be used to determine the nucleation rates. The crystal nucleation rate k is related to the free-energy barrier ($\Delta G(n)$) by[10]

$$k = A e^{-\beta \Delta G(n^*)}, \tag{14}$$

where

$$A \approx \rho f_{n^*} \sqrt{\frac{|\beta \Delta G''(n^*)|}{2\pi}}, \tag{15}$$

n^* is the number of particles in the critical nucleus, ρ is the number density of the supersaturated fluid, f_{n^*} is the rate particles are attached to the critical cluster, and $\Delta G''$ is the second derivative of the Gibbs free-energy barrier. Auer

TABLE V. Probabilities $P(\lambda_{i+1}|\lambda_i)$ for the first 8 interfaces at a pressure of $\beta p \sigma^3 = 15$, where the KMC simulations step size (Δ_{KMC}) and the number of MC steps between measuring the order parameter Δt_{ord} are varied. The following interfaces were used: $\lambda_2 = 20$, $\lambda_3 = 26$, $\lambda_4 = 32$, $\lambda_5 = 38$, $\lambda_6 = 44$, $\lambda_7 = 54$, $\lambda_8 = 65$, and $\lambda_9 = 78$. In all cases, 100 configurations were started in the fluid and reached the first interface, and at each interface, $C_i = 10$ copies of each successful configuration were used.

Δ_{KMC}	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Δt_{ord}	2	2	2	2	2	2	1	1	1	10	10	10
$P(\lambda_2|\lambda_1)$	0.112	0.103	0.139	0.101	0.105	0.132	0.112	0.146	0.138	0.122	0.127	0.146
$P(\lambda_3|\lambda_2)$	0.096	0.117	0.090	0.104	0.093	0.112	0.115	0.097	0.079	0.103	0.081	0.080
$P(\lambda_4|\lambda_3)$	0.128	0.117	0.074	0.116	0.111	0.161	0.151	0.110	0.110	0.121	0.091	0.116
$P(\lambda_5|\lambda_4)$	0.180	0.159	0.082	0.156	0.115	0.241	0.209	0.189	0.173	0.121	0.073	0.150
$P(\lambda_6|\lambda_5)$	0.167	0.154	0.149	0.225	0.148	0.256	0.274	0.151	0.189	0.189	0.121	0.187
$P(\lambda_7|\lambda_6)$	0.071	0.074	0.060	0.128	0.093	0.118	0.121	0.052	0.092	0.169	0.077	0.064
$P(\lambda_8|\lambda_7)$	0.104	0.078	0.051	0.109	0.091	0.109	0.119	0.077	0.126	0.132	0.087	0.064
$P(\lambda_9|\lambda_8)$	0.100	0.100	0.105	0.083	0.075	0.089	0.101	0.081	0.129	0.101	0.109	0.068
$P(\lambda_9|\lambda_1)$	3×10^{-8}	2×10^{-8}	4×10^{-9}	5×10^{-8}	1×10^{-8}	2×10^{-7}	2×10^{-7}	1×10^{-8}	6×10^{-8}	8×10^{-8}	6×10^{-9}	1×10^{-8}

and Frenkel[10] showed that the attachment rate f_{n^*} could be related to the mean square deviation of the cluster size at the top of the barrier by

$$f_{n^*} = \frac{1}{2} \frac{\langle \Delta n^2(t) \rangle}{t}. \tag{16}$$

The mean square deviation (MSD) of the cluster size $\Delta n^2(t) = \langle (n(t) - n^*)^2 \rangle$ can then be calculated by either employing a kinetic Monte Carlo (KMC) simulation or a MD simulation at the top of the barrier. For simplicity, in the remainder of this paper the nucleation rate determined using this method will be referred to as the umbrella sampling (US) nucleation rate, although to calculate the nucleation rates both US simulations and dynamical simulations (KMC or MD) are necessary. Note that information on KMC simulations can be found in, e.g., Refs. 33 and 34.

The mean square deviation, or variance, in the cluster size appearing in Eq. (16) has both a short-time and long-time behavior. At short times, fluctuations are due to particles performing Brownian motion around their average positions while the long-time behavior is caused by rearrangements of particles required for the barrier crossings. The slope of the variance is large at short times where only the fast rattling is sampled. However, the longer the time the further the system has diffused away from the critical cluster size at the top of the nucleation barrier. Auer[36] states that runs need to be selected that remain at the top of the barrier. However, when this is done the attachment rate is lower than when the average over all runs is taken since it excludes the runs that move off the barrier fast and have the largest attachment rate. This problem is analogous to determining the diffusion constant of a particle performing a random walk. By only including walks which remain in the vicinity of the origin, the measurement is biased and excludes trajectories which quickly move away from the origin. This results in an underestimation of the diffusion constant, and similarly, in this case, an underestimation of the attachment rate. Hence, in this paper we do not attempt to prevent the trajectories from falling off the barrier and we include all trajectories. In Fig. 5 we demonstrate how, starting from a critical cluster, the size of the nucleus fluctuates as a function of time and, in fact, can completely disappear or

double in size within $0.3\tau_L$ where τ_L is the time that it takes a particle to diffuse on average a distance equal to its diameter, i.e., $\tau_L = \sigma^2/(6D_L)$.

The kinetic prefactor was determined using KMC simulations with 3000 particles in an NVT ensemble in a cubic box with periodic boundary conditions. The initial configurations were taken from US simulations in one of the windows at the top of the barrier. We examined the results from both Gaussian and uniformly distributed Monte Carlo steps and found agreement within the statistical errors. For all the simulations, the MC step size was between 0.01 and 0.1σ. The variance of the cluster size for a typical system is shown in Fig. 6. We observed a large variance in the attachment rates calculated for different nuclei. Specifically, some nuclei have attachment rates more than an order of magnitude higher than other nuclei of similar size. The nuclei with low attachment rates appeared to have a smoother surface than the nuclei with a high attachment rate. In calculating the attachment rates we used ten independent configurations on the top of the barrier and followed ten trajectories from each.

Our results for the kinetic prefactors and nucleation rates for pressures $\beta p \sigma^3 = 15, 16$, and 17 are reported in Table IV.
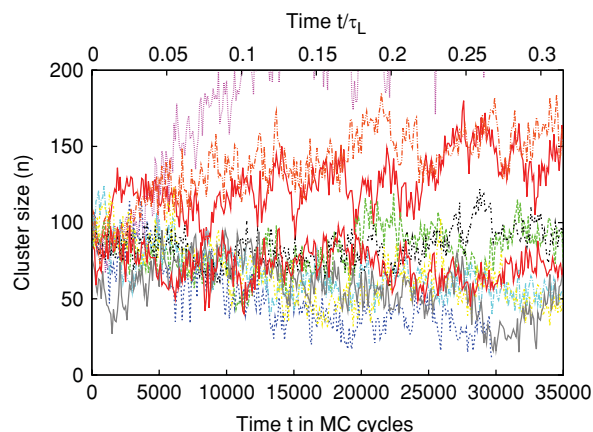

FIG. 5. The cluster size ($n(t)$) as a function of time in MC cycles for a random selection of clusters that start at the top of the nucleation barrier.
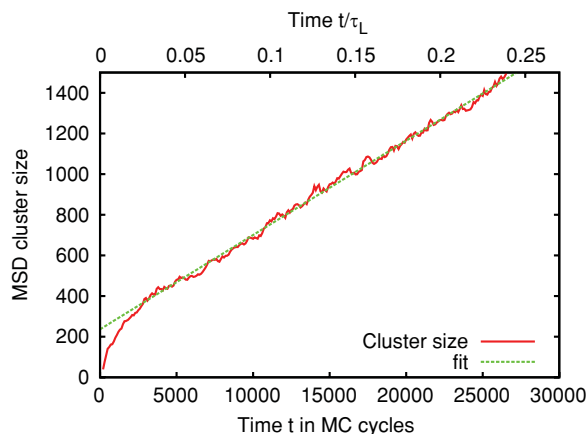
FIG. 6. The mean square deviation (MSD) of the cluster size $\langle \Delta n^2(t) \rangle$ as a function of time t in MC cycles. The cluster size has been measured every cycle and averaged over 100 cycles to reduce the short-time fluctuations. The slope of this graph is twice the attachment rate [Eq. (16)].
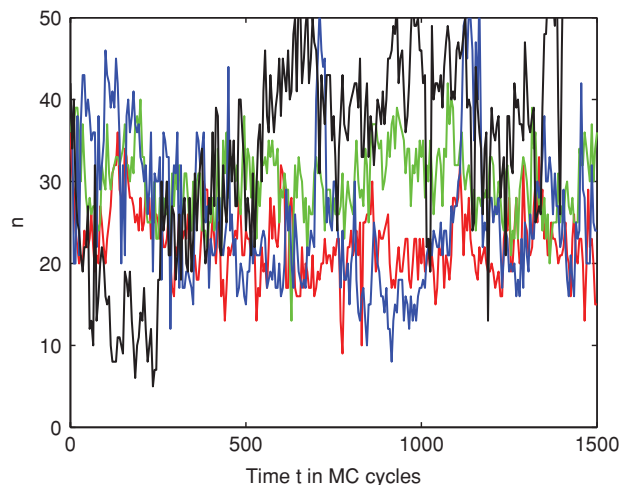

FIG. 7. The cluster size as a function of time t in MC cycles for 4 random trajectories at pressure $\beta p \sigma^3 = 17$ starting with a cluster size of $n = 43$ using kinetic MC simulations with step size $\Delta_{KMC} = 0.1\sigma$ and measuring the order parameter every $\Delta t_{ord} = 5$ MC steps.

VI. FORWARD FLUX SAMPLING

A. Method

The forward flux sampling method was introduced by Allen *et al.*[13] in 2005 to study rare events and has since been applied to a wide variety of systems. Two review articles (Refs. 37 and 38) on the subject have appeared recently and provide a thorough overview of the method. In the present paper, we discuss FFS as it pertains to the liquid to solid nucleation process in hard spheres. In general, FFS follows the progress of a reaction coordinate during a rare event. For hard-sphere nucleation, a reasonable reaction coordinate (Q) is the number of particles in the largest crystalline cluster in the system (n). For the remainder of this paper, for all FFS calculations, we take the reaction coordinate to be the order parameter discussed in Sec. III with $\xi_c = 8$, $r_c = 1.3\sigma$, and $d_c = 0.7$. In general, the reaction coordinate is used to divide phase space by a sequence of interfaces ($\lambda_0, \lambda_1, \ldots \lambda_N$) associated with increasing values $n(\mathbf{r}^N)$ such that the nucleation process between any two interfaces can be examined. In our case the liquid is composed of all states with $n < \lambda_0$ and the solid contains all states with $n > \lambda_N$. While the complete nucleation event is rare, the interfaces are chosen such that the part of the nucleation process between consecutive interfaces is not rare, and can thus be thoroughly studied.

In the FFS methodology, the nucleation rate from the fluid phase A to the solid phase B is given by

$$k_{AB} = \Phi_{A\lambda_0} P(\lambda_N | \lambda_0) \tag{17}$$

$$= \Phi_{A\lambda_0} \prod_{i=0}^{N-1} P(\lambda_{i+1} | \lambda_i), \tag{18}$$

where $\Phi_{A\lambda_0}$ is the steady-state flux of trajectories leaving the A state and crossing the interface λ_0 in a volume V, and $P(\lambda_{i+1} | \lambda_i)$ is the probability that a configuration starting at interface λ_i will reach interface λ_{i+1} before it returns to the fluid (A).

If we apply this method directly to a hard-sphere system a number of difficulties arise. As shown in Fig. 5, on short

times the size of a cluster measured by the order parameter fluctuates wildly. The variance in the cluster size displays two different types of behavior, short-time fluctuations related to surface fluctuations of the cluster, and a longer time cluster growth (Fig. 6). Thus, if we try to measure the flux $\Phi_{A\lambda_0}$ directly, we encounter difficulties due to these short-time surface fluctuations. In theory, FFS should be able to handle these types of fluctuations, however, they increase the amount of statistics necessary to properly measure the flux and the first probability window. In the second part of FFS calculations, probabilities of the form $P(\lambda_{i+1} | \lambda_i)$ need to be determined. In calculating these probabilities it is important to be able to determine if a cluster has returned to the fluid (A). For precritical clusters we find large fluctuations of the order parameter, as shown in Fig. 7, which can lead to a cluster being misidentified as the fluid (A). Specifically, in this figure the darkest trajectory (black) shows a cluster containing 43 particles that shrinks to 5 particles before it returns to 40, and finally reaches a cluster size of 60 particles. Hence, if we had set $\lambda_0 = 5$, this trajectory would have been identified as melting back to the fluid phase (A). However, since the growth of a cluster from size 5 to 60 is a rare event in our system, we presume that this was simply a short-time fluctuation of the cluster and not a "real" melting of the instantaneously measured cluster. For precritical clusters, these fluctuations result in cluster sizes that are smaller than the cluster "really" is. We suggest that these fluctuations are largely related to the difficulty that this order parameter has in distinguishing between solid-like and fluidlike particles at the fluid–solid interface. For larger clusters, where the surface to volume ratio is small, this problem is minimal. However, for elongated or rough precritical clusters, where the surface to volume ratio is large, these surface fluctuations and rearrangements are important, and can cause problems in measuring the order parameter.

Thus, to try and address these problems, in this paper, we apply forward flux sampling in a novel way. We regroup the

TABLE VI. Nucleation rates predicted using forward flux sampling in units of the long-time diffusion coefficient (D_l). The probabilities $P(\lambda_B|\lambda_1)$, number of steps between the order parameter measurements Δt_{ord}, and kinetic MC step size Δ_{KMC} are as in Tables VII, VIII, and IX. At each interface, C_i copies of each successful configuration were used as displayed in Tables VII, VIII, and IX.

$\beta p\sigma^3$	λ_1	$\tilde{\Phi}_{A\lambda_1}/6D_L$	$P(\lambda_B|\lambda_1)$	$k\sigma^5/6D_L$
17	27	2.66×10^{-5}	7.6×10^{-3}	2.0×10^{-7}
17	27	2.68×10^{-5}	1.4×10^{-2}	3.7×10^{-7}
16	20	8.57×10^{-6}	3.1×10^{-7}	2.6×10^{-12}
16	20	8.57×10^{-6}	2.1×10^{-7}	1.8×10^{-12}
15	15	8.72×10^{-6}	1.9×10^{-15}	1.6×10^{-20}

elements of the rate calculation such that

$$k_{AB} = \tilde{\Phi}_{A\lambda_1} \prod_{i=1}^{N-1} P(\lambda_{i+1}|\lambda_i), \qquad (19)$$

where

$$\tilde{\Phi}_{A\lambda_1} = \Phi_{A\lambda_0} P(\lambda_1|\lambda_0). \qquad (20)$$

We note that if λ_1 is chosen such that it is a relatively rare event for trajectories starting in A to reach λ_1, then

$$\tilde{\Phi}_{A\lambda_1} \approx \frac{1}{\langle t_{A\lambda_1} \rangle V} \qquad (21)$$

where $\langle t_{A\lambda_1} \rangle$ is the average time it takes a trajectory in A to reach λ_1. The approximation made here, in contrast to normal FFS simulations, is that the time the system spends with an order parameter greater than λ_1 is negligible. Since even reaching this interface is a rare event, this approximation should have a minimal effect on the resulting rate. Additionally, in this way we are relatively free to place the first interface (λ_0) anywhere under λ_1.[45] We choose to use $\lambda_0 = 1$ to minimize the effect of fluctuations, as seen in Fig. 7, on the probability of reaching the following interface. Here we assume that any crystalline order in a system with $\lambda_0 = 1$ does not likely arise from a fluctuation of a much larger cluster, but rather is very close to the fluid, and is expected to fully melt and not grow out to the next interface. In this manner we are able to start several parallel trajectories from the fluid in order to measure $\langle t_{A\lambda_1} \rangle$, stopping whenever the trajectory first hits interface λ_1.

In our implementation of FFS, we employ kinetic Monte Carlo (KMC) simulations at fixed pressure to follow the trajectories from the liquid to the solid. The KMC simulations are characterized by two parameters, the maximum step size

TABLE VII. Probabilities $P(\lambda_{i+1}|\lambda_i)$ for the interfaces used in calculating the nucleation rate for pressure $\beta p\sigma^3 = 17$ with step size $\Delta_{KMC} = 0.1\sigma$ and measuring the order parameter every $\Delta t_{ord} = 5$ MC cycles.

i	λ_i	Trial 1		Trial 2	
		C_{i-1}	$P(\lambda_i|\lambda_{i-1})$	C_{i-1}	$P(\lambda_i|\lambda_{i-1})$
2	43	10	0.137	10	0.157
3	60	10	0.272	10	0.312
4	90	10	0.350	10	0.414
5	150	2	0.594	2	0.691
6	250	2	0.988	2	0.988

TABLE VIII. Same as Table VII but for $\beta p\sigma^3 = 16$.

i	λ_i	Trial 1		Trial 2	
		C_{i-1}	$P(\lambda_i|\lambda_{i-1})$	C_{i-1}	$P(\lambda_i|\lambda_{i-1})$
2	28	10	0.105	10	0.110
3	38	10	0.075	10	0.077
4	50	10	0.070	10	0.089
5	70	10	0.114	10	0.089
6	90	10	0.095	10	0.101
7	110	10	0.339	10	0.278
8	250	10	0.152	10	0.112
9	350	1	1.000	1	1.000

(Δ_{KMC}) per attempt to move each particle, and the frequency with which the order parameter (reaction coordinate) is measured Δt_{ord}. However, during an FFS simulation, it is expected that the order parameter is known at all times such that it is possible to identify exactly when and if a given simulation reaches an interface. Thus, it is possible that Δt_{ord} introduces an additional error into our measurement of the rate.

To examine the effects of (i) the approximation associated with our method for calculating $\tilde{\Phi}_{A\lambda_1}$, (ii) the short-time fluctuations of the order parameter (which could be considered as an error in the measurement of the cluster size), and (iii) the frequency of measuring the order parameter, we examined the nucleation rate for a simple one-dimensional model system in the presence of such features. Details of these simulations are given in the Appendix. In this simple model system, we find that none of these features have a large effect on the rate. In fact, for most cases, the difference is too small to see within our error bars.

B. Simulation details and results

All simulations were performed with 3000 particle in a cubic box with periodic boundary conditions. Initial configurations were produced using NPT MC simulations of a liquid phase with a packing fraction of $\eta \approx 0.4$ and then simulated at a reduced pressure of $\beta p\sigma^3 = 1000$. The simulations were

TABLE IX. Same as Table VII but for $\beta p\sigma^3 = 15$ and with $\Delta t_{ord} = 2$.

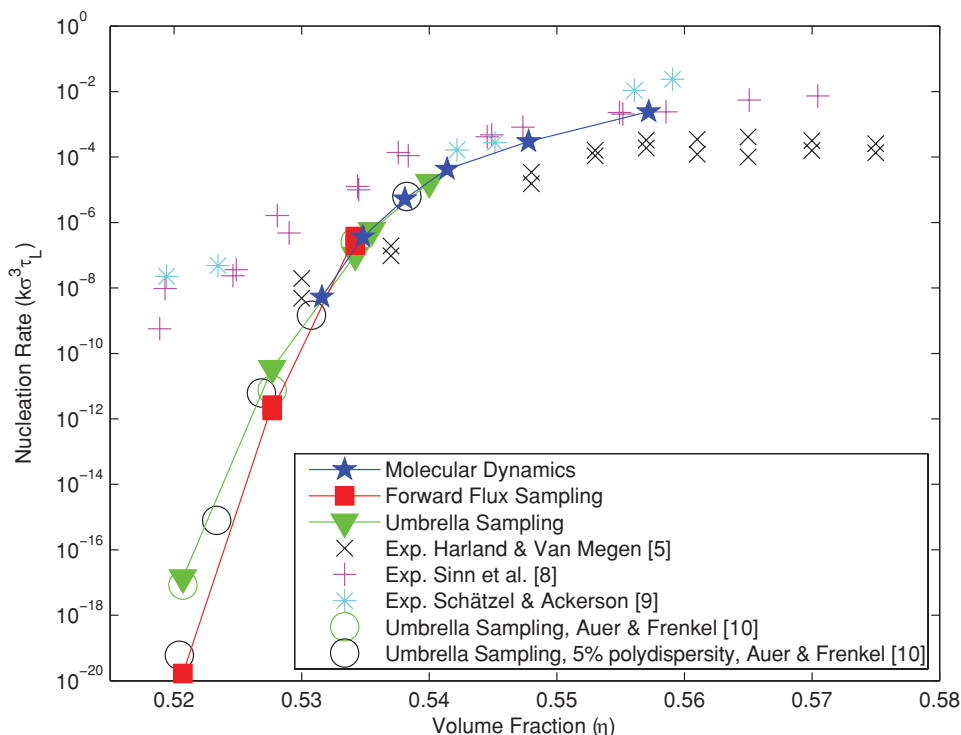| i	λ_i	C_{i-1}	$P(\lambda_i|\lambda_{i-1})$
2	20	10	0.101
3	26	10	0.104
4	32	10	0.116
5	38	10	0.156
6	44	10	0.225
7	54	10	0.128
8	65	10	0.109
9	78	10	0.083
10	92	10	0.101
11	110	10	0.085
12	135	10	0.062
13	160	10	0.131
14	190	10	0.131
15	230	10	0.134
16	400	10	0.058

FIG. 8. A comparison of the crystal nucleation rates of hard spheres as determined by the three methods described in this paper FFS, US, and MD with the experimental results from Refs. 5, 8, and 9 and previous theoretical results from Ref. 10. The nucleation rates are in units of τ_L where where $\tau_L = \sigma^2/(6D_L)$. Note that error bars have not been included in this plot but are discussed in the main text. Within these estimated error bars, all the simulated nucleation rates are in agreement, while the experimentally obtained rates show a markedly different behavior, particularly for low supersaturations where the difference between the simulations and experiments can be as large as 12 orders of magnitude.
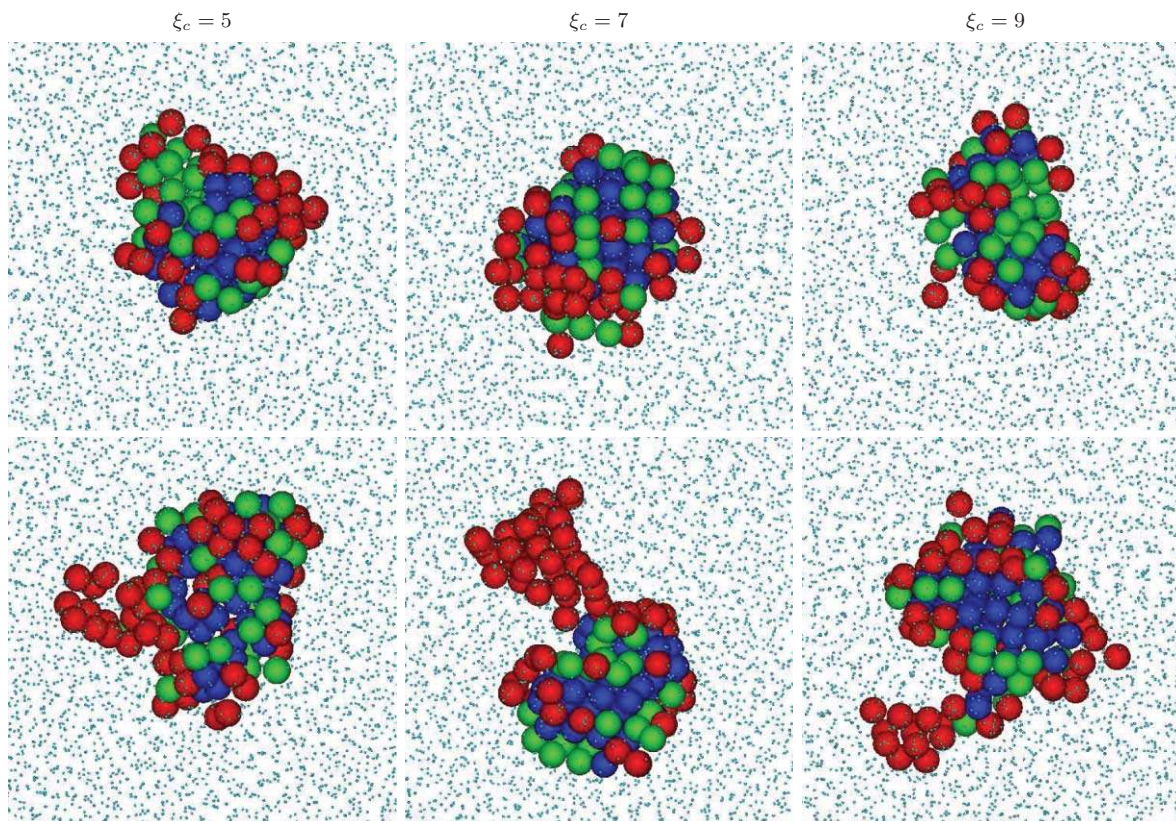

FIG. 9. Two typical snapshots (top and bottom) of the critical nuclei as obtained with US at a volume fraction $\eta = 0.5355$ using different values of the critical number of crystalline bonds $\xi_c = 5$ (left), 7 (middle), and 9 (right) in the biasing potential. The clusters are analysed with three different crystalline order parameters. The blue particles are found by all three cluster criteria, the green particles have $\xi = 7$ or 8 crystalline bonds and the red particles have only $\xi = 5$ or 6 crystalline bonds.
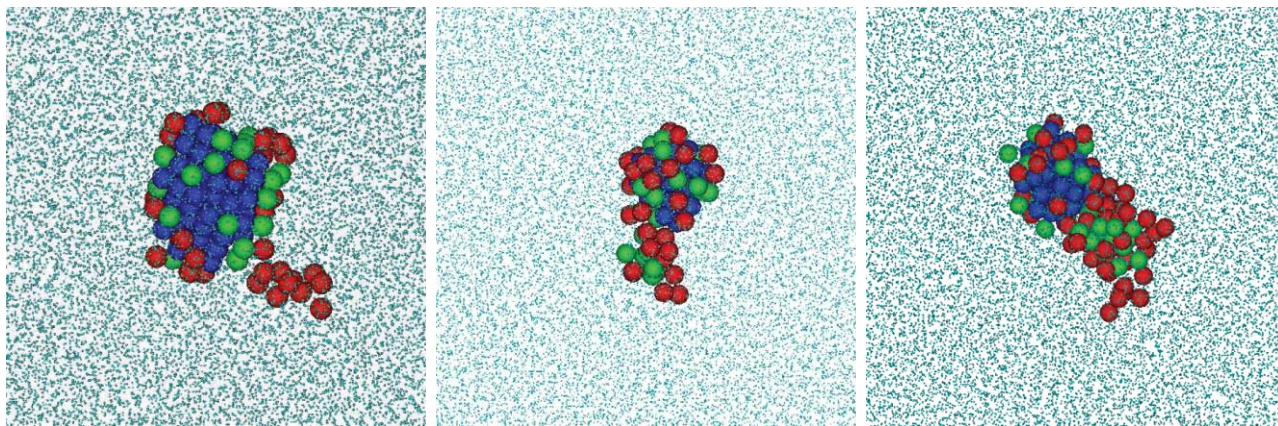
FIG. 10. Snapshots of nuclei formed spontaneously during an MD simulation at a volume fraction of $\eta = 0.537$. The snapshots were taken just before the nuclei grew. The color coding of the particles is the same as in Fig. 9.

stopped when the packing fraction associated with the pressure of interest was reached. In this way the system volume decreased rapidly to the target density. This initial configuration was then relaxed using an NPT simulation at the pressure of interest ($\beta p \sigma^3 = 15, 16, 17$). The relaxation consisted of at least 10 000 MC cycles, after which the simulation continued until a measurement of the order parameter found no crystalline particles in the system.

In order to determine the flux and the probabilities, 100 trajectories were started in the liquid and terminated when $n(\mathbf{r}^N) = \lambda_1$. These trajectories were produced using KMC simulations. The probability $P(\lambda_2|\lambda_1)$ was then found by making C_1 copies of the configurations that reached λ_1, and following these configurations until they either reached λ_2 or returned to the fluid. By taking different random number seeds, the various copies of the same configurations follow different trajectories. The fraction of successful trajectories corresponds to the required probability. The successful trajectories were then copied C_2 times to determine $P(\lambda_3|\lambda_2)$. The remaining $P(\lambda_{i+1}|\lambda_i)$'s are calculated similarly.

To study the effect of the two KMC parameters, namely Δ_{KMC} and Δt_{ord}, on the nucleation rates, we examined the first

8 FFS windows for $\beta p \sigma^3 = 15$ for various values of the number of MC steps between the order parameter measurements Δt_{ord} and the maximum displacement Δ_{KMC} for the KMC simulations. The results are shown in Table V. As shown in this table we do not find a significant effect on the rate from either parameter. Thus, for numerical efficiency, unless otherwise indicated, the rates in this section come from $\Delta t_{\mathrm{ord}} = 5$ MC cycles and $\Delta_{\mathrm{KMC}} = 0.2\sigma$.

For pressures $\beta p \sigma^3 = 16$ and 17 we have performed two separate FFS calculations to determine the nucleation rates, and for pressure $\beta p \sigma^3 = 15$ we have the result from a single FFS simulation. A summary of the results are given in Table VI. A complete summary of the results for $P(\lambda_{i+1}|\lambda_i)$ for each simulation is given in Tables VII, VIII, and IX.

VII. SUMMARY AND DISCUSSION

A. Nucleation rates

In this section we examine hard-sphere nucleation rates predicted using US simulations, MD simulations and FFS
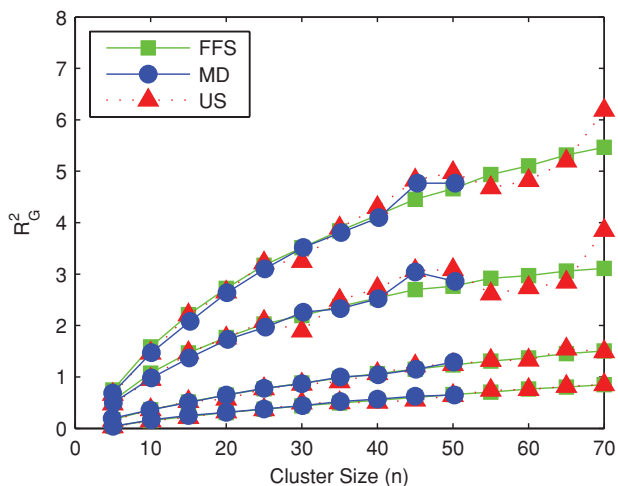

FIG. 11. A comparison of the three components of the radius of gyration tensor as a function of cluster size n, as well as the sum of the three components, for clusters produced using FFS, MD, and US simulations.
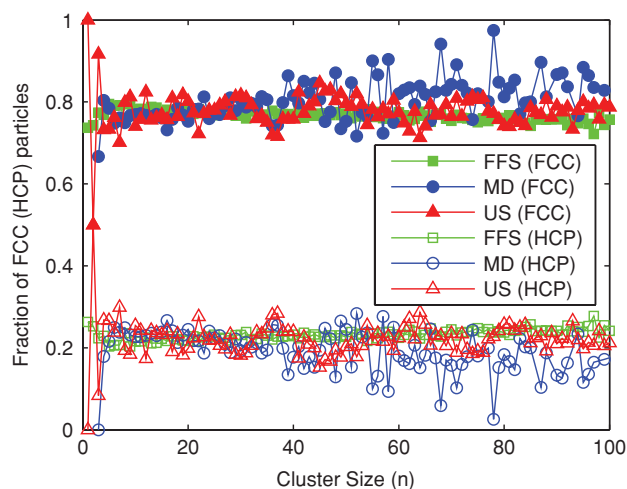

FIG. 12. Fraction of particles identified as either FCC or HCP respectively, in the clusters produced via molecular dynamics (MD), forward flux sampling (FFS), and umbrella sampling (US) simulations as a function of cluster size n. All three methods agree and find the precritial clusters predominately FCC.

TABLE X. Nucleation rates for the one-dimensional potential given by Eq. (A1) and shown in Fig. 13 for Δt_{ord} as indicated. For each Δt_{ord}, we performed 10 independent FFS simulations. The average rate and associated standard deviation is also as indicated. In all cases, 100 configurations were started in the fluid, and at each interface $C_i = 10$ copies of the successful configurations were used to calculate the proceeding probabilities. The interfaces were placed at $\lambda_0 = 0$, $\lambda_1 = 1.5$, $\lambda_2 = 1.7$, $\lambda_3 = 1.9$, $\lambda_4 = 2.2$, $\lambda_5 = 2.6$, $\lambda_6 = 3.3$, and $\lambda_7 = 4.0$ and the flux was calculated using Eq. (21).

Δt_{ord}	1	2	5	10	50
	1.2723×10^{-12}	1.0589×10^{-12}	1.8075×10^{-12}	1.5455×10^{-12}	1.3835×10^{-12}
	1.3780×10^{-12}	1.7217×10^{-12}	1.3314×10^{-12}	1.4461×10^{-12}	1.0666×10^{-12}
	1.2364×10^{-12}	1.2924×10^{-12}	1.4847×10^{-12}	1.1482×10^{-12}	1.6134×10^{-12}
	1.6942×10^{-12}	1.6422×10^{-12}	1.9482×10^{-12}	1.4383×10^{-12}	1.7550×10^{-12}
	1.2662×10^{-12}	1.2340×10^{-12}	1.5692×10^{-12}	1.6060×10^{-12}	1.2908×10^{-12}
	1.6918×10^{-12}	1.3530×10^{-12}	1.6238×10^{-12}	1.6244×10^{-12}	1.4012×10^{-12}
	1.4646×10^{-12}	1.1788×10^{-12}	1.6928×10^{-12}	1.0191×10^{-12}	1.3403×10^{-12}
	1.6809×10^{-12}	1.5860×10^{-12}	1.1903×10^{-12}	1.6227×10^{-12}	1.0582×10^{-12}
	1.4602×10^{-12}	1.7018×10^{-12}	1.3191×10^{-12}	1.3850×10^{-12}	2.3732×10^{-12}
	1.7459×10^{-12}	1.9154×10^{-12}	1.5638×10^{-12}	1.2378×10^{-12}	1.2692×10^{-12}
Avg. Rate	1.5×10^{-12}	1.5×10^{-12}	1.6×10^{-12}	1.4×10^{-12}	1.5×10^{-12}
Std. Error	6.0×10^{-14}	8.4×10^{-14}	7.0×10^{-14}	6.3×10^{-14}	1.2×10^{-13}

simulations together with the experimental results of Harland and van Megen,[5] Sinn *et al.*[8] and Schätzel and Ackerson[9] and the US simulations of monodisperse and 5% polydisperse hard-spheres systems examined by Auer and Frenkel.[10] The experimental volume fractions have been scaled to yield the coexistence densities of monodisperse hard spheres.[16] Similarly, we scale the polydisperse results of Auer and Frenkel with the coexistence densities determined in Ref. 39. Inspired by the recent work of Pusey *et al.*,[16] we plot the nucleation rates in units of the long-time diffusion coefficient. In experiments with colloidal particles, the influence of the solvent on the dynamics cannot be ignored. Specifically, the system slows down due to hydrodynamic interactions when the density is increased. However, by presenting the nucleation rates in terms of the long-time diffusion coefficient, we expect our simulated nucleation rates from the hard-sphere model without an explicit solvent to be in agreement with the experimental rates with a solvent. The time in experiments is typically measured in units of D_0, the free diffusion at low density. We convert the short-time diffusion coefficient D_0 to the long-time diffusion coefficient D_L using

$$\frac{D_L(\eta)}{D_0} = \left(1 - \frac{\eta}{0.58}\right)^{\delta}. \tag{22}$$

Harland and van Megen[5] claim that $\delta = 2.6$ gives a good fit to their system and Sinn *et al.*[8] use $\delta = 2.58$. Since the system that Schätzel and Ackerson[9] examine is very similar to the other two, we use $\delta = 2.6$ to convert their nucleation rates in terms of D_L. We note that both $\delta = 2.58$ and $\delta = 2.6$ give very similar results. The results for both the theoretical and experimental rates in terms of $\tau_L = \sigma^2/6D_L$ are shown in Fig. 8. Note that for clarity reasons the error bars have not been included in this plot. In general, the error bars of the simulated nucleation rates are largest for lower supersaturations (i.e., lower volume fractions), as the barrier height is higher. For the FFS and US simulations, the error for $\beta p \sigma^3 = 15$ ($\eta = 0.5214$) is between 2 and 3 orders of magnitude, and for $\beta p \sigma^3 = 17$ ($\eta = 0.5352$) is approximately one to two orders of magnitude. The MD results are quite accurate around

TABLE XI. Nucleation rates for the one-dimensional potential given by Eq. (A1) and shown in Fig. 13 where the order parameter is given by Eq. (A2) and σ_{Gauss} is as indicated. For each σ_{Gauss}, we performed 10 independent FFS simulations. The average rate and associated standard deviation is also as indicated. In all cases, 100 configurations were started in the fluid, and at each interface $C_i = 10$ copies of the successful configurations were used to calculate the proceeding probabilities. The interfaces were placed at $\lambda_0 = 0$, $\lambda_1 = 1.5$, $\lambda_2 = 1.7$, $\lambda_3 = 1.9$, $\lambda_4 = 2.2$, $\lambda_5 = 2.6$, $\lambda_6 = 3.3$, and $\lambda_7 = 4.0$ and the flux was calculated using Eq. (21).

| σ_{Gauss}	0.02	0.04	0.06	0.08	0.1
	1.8623×10^{-12}	1.7281×10^{-12}	1.2630×10^{-12}	1.0634×10^{-12}	1.9158×10^{-12}
	1.7627×10^{-12}	1.6090×10^{-12}	1.6402×10^{-12}	1.5655×10^{-12}	1.8785×10^{-12}
	9.9796×10^{-13}	1.6305×10^{-12}	1.5799×10^{-12}	1.6936×10^{-12}	1.4937×10^{-12}
	1.3743×10^{-12}	1.2261×10^{-12}	1.8305×10^{-12}	1.7733×10^{-12}	1.1142×10^{-12}
	1.6917×10^{-12}	1.8054×10^{-12}	1.6191×10^{-12}	1.8941×10^{-12}	1.0402×10^{-12}
	1.1842×10^{-12}	1.3337×10^{-12}	1.3283×10^{-12}	1.4039×10^{-12}	7.0735×10^{-13}
	1.5289×10^{-12}	8.6859×10^{-13}	1.3129×10^{-12}	2.7115×10^{-12}	2.4711×10^{-12}
	1.8918×10^{-12}	1.4325×10^{-12}	1.3203×10^{-12}	1.3792×10^{-12}	1.6288×10^{-12}
	1.3144×10^{-12}	1.2283×10^{-12}	1.0459×10^{-12}	1.7194×10^{-12}	1.3764×10^{-12}
	1.6654×10^{-12}	1.1236×10^{-12}	1.2572×10^{-12}	1.9631×10^{-12}	1.8976×10^{-12}
Avg. Rate	1.5×10^{-12}	1.4×10^{-12}	1.4×10^{-12}	1.7×10^{-12}	1.6×10^{-12}
Std. Error	9.5×10^{-14}	9.4×10^{-14}	7.5×10^{-14}	1.4×10^{-13}	1.6×10^{-13}
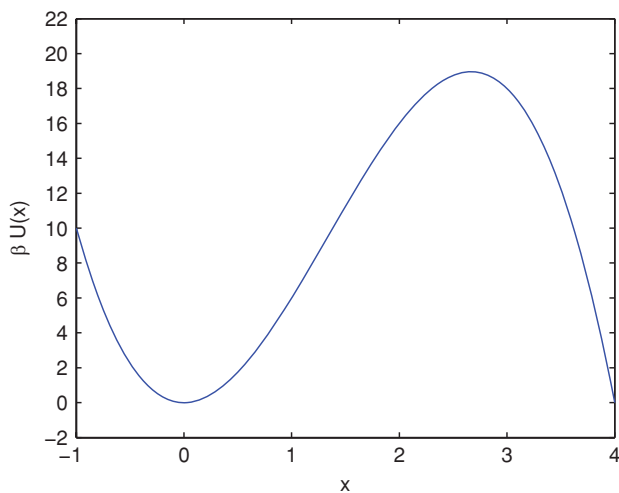
FIG. 13. Toy model potential used to study forward flux sampling in the presence of various types of measurement errors.

$\beta p \sigma^3 = 17$, however the error bars are larger for the higher pressure MD results.

In Ref. 16, Pusey *et al.* showed that the nucleation rates for various polydispersities (0%–6%) of hard spheres collapsed onto the same curve when the rates were plotted in units of the long-time diffusion coefficient. We find similar results here. Both the monodisperse and polydisperse US results of Auer and Frenkel,[10] in addition to our own US predictions of the nucleation rate, agree well within the expected measurement error. Additionally, we find that the simulation results of the US, FFS, and MD all agree. Whereas the simulation results agree well with the experimental results for the nucleation rate at high supersaturation there is still a significant difference at low supersaturations. Unfortunately, the origin of this discrepancy remains unsolved.

However, on the experimental side, the nucleation rates of Harland and van Megen[5] are approximately 1–2 orders of magnitude below the experiments of Sinn *et al.*[8] and Schätzel and Ackerson.[9] This is unexpected due to the similarity between the experimental systems. The main difference between these experiments is the size polydispersity of the particles: 5% in the case of Harland and van Megen,[5] 2.5% in the case of Sinn *et al.*,[8] and < 5% for Schätzel and Ackerson.[9] However, as demonstrated by Pusey *et al.*,[16] and now also in Fig. 8, the nucleation rate when measured in long-time diffusion coefficient units should not be effected by the polydispersity. Thus, this seems unlikely as an explanation.

B. Nuclei

To examine whether the structure and shape of the critical clusters from US simulations depended on the precise threshold values used for the crystalline order parameters, we compared and analyzed the critical clusters obtained when three different crystalline order parameters were used to bias the US simulations, namely, $\xi_c = 5, 7,$ and 9. Subsequently, we analyzed these critical clusters using the three different order parameters. In Fig. 9, two typical critical clusters from different biasing order parameters are shown on the top and bottom

rows. The nucleus of the cluster, shown in blue, was identified by all three cluster criteria ($\xi_c = 5, 7,$ and 9). The main difference between the criteria is the location of the fluid-solid interface as shown by the green and red particles. The strictest order parameter finds only the more ordered center whereas the loosest version detects the more disordered particles at the interface as well. In Fig. 10 we show some of the nuclei obtained from MD simulations. These snapshots were taken just before the nuclei grew out so they are not necessarily precisely at the top of the nucleation barrier. They appear very similar in roughness and aspect ratio to those obtained from US simulations. We note here that this is not meant to be a thorough study of the critical clusters, but rather just a rough comparison to demonstrate, to a first approximation, the clusters formed by the three simulation techniques are the same. A more thorough examination of the structure of the nuclei for high supersaturations can be found in Ref. 42.

To further examine whether the choice of method influenced the resulting clusters, particularly the presence of the biasing potential in the US simulations and the choice of reaction coordinate and interfaces in FFS, we calculated the radius of gyration tensor for each of the methods for pressure $\beta p \sigma^3 = 17$ as a function of cluster size (see Fig. 11). There is no indication that the clusters in any of the simulation methods differed substantially.

Additionally, we examined whether the simulation technique influenced the type of precritical nuclei that formed in the simulations, i.e., FCC and HCP. To do this we used the order parameter introduced by Ref. 41 which allows us to identify each particle in the cluster as either FCC-like or HCP-like. The results for a wide range in nucleus size is shown in Fig. 12. We find complete agreement between the three simulation techniques. Specifically, in all cases we find that the nucleus is composed of approximately 80% FCC-like particles. This was unexpected as the free-energy difference between the bulk FCC and HCP phases is about $0.001k_BT$ per particle at melting[43] and hence random-hexagonal-close-packing order in the nuclei would be expected.[44] Note that using our order parameter this would appear as an approximately 50% occurrence of FCC-like and HCP-like particles in the nucleus. We speculate that this predominance of FCC-like particles in the nuclei arises from surface effects.

VIII. CONCLUSIONS

In this paper, we have examined in detail three independent simulation techniques for studying nucleation processes and predicting nucleation rates, namely forward flux sampling, umbrella sampling, and molecular dynamics. We have shown that the three simulation techniques are completely consistent in their prediction of the nucleation rates for hard spheres over the large range of volume fractions studied, despite the fact that they treat the dynamics differently. Additionally, in agreement with the recent work of Pusey *et al.*,[16] we find that by measuring the nucleation rates in terms of the long-time diffusion constant and scaling to the coexistence density of monodisperse hard spheres, the 5% polydisperse results of Auer and Frenkel[10] also agree. On examining

the critical clusters, we also do not find a difference in the nuclei formed using the three simulation techniques. Hence, we conclude that the original prediction of Auer and Frenkel[10] for the nucleation rates in hard-sphere systems was indeed robust.

We have also compared our nucleation rates with previous experimental data, specifically, the nucleation rates predicted by Harland and van Megen,[5] Sinn *et al.*[8] and Schätzel and Ackerson.[9] As was found first by Auer and Frenkel,[10] while the simulation results agree well with the experimental results for high supersaturations, there is a significant difference between the simulations and experiments for smaller volume fractions. The agreement between the three theoretical methods examined in this paper, namely molecular dynamics, umbrella sampling, and forward flux sampling, seems to indicate that either there is a fundamental difference between the simulations and theory, which we are not taking into account, such as some form of collective hydrodynamics which are included in the experiments but not considered in the theory or some difficulty in interpreting the experimental data. In either case, the origin of the huge discrepancy in the theoretical and experimental nucleation rates remains a mystery.

ACKNOWLEDGMENTS

APPENDIX: FFS IN THE PRESENCE OF MEASUREMENT ERROR

As mentioned in Sec. VI of this paper, the FFS technique assumes that the reaction coordinate is known exactly at all times. However, for the hard-sphere system examined in this paper, this is not possible due to the computational time required for measuring the order parameter. In applying the FFS technique to hard spheres, two separate types of error are introduced: (i) error associated with our inability to know the value of the reaction coordinate at all times, and (ii) an error in measuring the number of particles in a cluster for a given configuration. Additionally, as discussed in Sec. VI, in this paper we have applied FFS in a slightly novel manner. In this appendix, we introduce a simple model to examine the effect of this approximation and the effect such measurement errors have on the nucleation rate predicted by forward flux sampling.

To this end, we study the transition rate for a single Brownian particle to surmount a one dimensional potential energy barrier given by

$$\beta U(x) = 8x^2 - 2x^3. \tag{A1}$$

A plot of the barrier is shown in Fig. 13. For this potential, we consider the "liquid" state to be near $x = 0$ and the "solid" phase to be near $x = 4$.

We first determine the "exact" nucleation rate using spontaneous simulations. To do this we perform a random walk

starting at $x = 0$ and determine the time it takes the random walk to surmount the barrier. The rate is then given by $R = 1/\langle t \rangle$. Performing 40 such random walks we find the nucleation rate to be 1.3×10^{-12}. In all the calculations in this section, we set the KMC step size equal to $\Delta_{KMC} = 0.025$.

Second, we explore the effect on the nucleation rate of not knowing the value of the order parameter at all times. For this purpose we have performed FFS simulations when the order parameter was measured every $\Delta t_{\text{ord}} = 1, 2, 5, 10, 50$ kinetic Monte Carlo steps. The results are shown in Table X. The average nucleation rates predicted for all values of Δt_{ord} clearly are the same within error. Similarly, the standard error associated with $\Delta t_{\text{ord}} = 1, 2, 5, 10$ are approximately the same, and is only marginally larger for $\Delta t_{\text{ord}} = 50$. Hence, we conclude that the frequency of measuring the order parameter does not significantly affect the predicted nucleation rate. Additionally, these nucleation rates agree with the nucleation rate predicted from spontaneous simulations, indicating that of applying FFS as outlined in Sec. VI predicts the correct nucleation rates.

Finally, we examine the effect that measurement error in the cluster size has on the nucleation rate. For this purpose, we apply a noise term to our order parameter such that

$$x_m = x_{\text{true}} + \delta, \tag{A2}$$

where x_m is the value of the order parameter used in the FFS simulation, x_{true} is the true value of the order parameter, and δ is taken from a Gaussian distribution with a mean of 0 and a standard deviation σ_{Gauss}. In Table XI we demonstrate the effect on the predicted nucleation rate for various choices of σ_{Gauss}. The resulting nucleation rates are in good agreement with the spontaneous results. For larger σ_{Gauss}, e.g., $\sigma_{\text{Gauss}} = 0.08$ and 0.1, the standard error in the results is slightly larger, however, the predicted nucleation rates are still correct.

In summary, we have examined the effect of the approximation described by Eq. (21), as well as the effect of measurement error in the order parameter and the measurement frequency Δt_{ord} of the order parameter. We do not find a significant effect on the predicted nucleation rates. Thus, we conclude that FFS should be robust to the types of error we are introducing when we apply the technique to hard spheres.

[1] A. D. Dinsmore, E. R. Weeks, V. Prasad, A. C. Levitt, and D. A. Weitz, Appl. Opt. **40**, 4152 (2001).
[2] S.-H. Lee, Y. Roichman, G.-R. Yi, S.-H. Kim, S.-M. Yang, A. van Blaaderen, P. van Oostrum, and D. G. Grier, Opt. Express **15**, 18275 (2007).
[3] S. C. Glotzer and M. J. Solomon, Nature Mater. **6**, 557 (2007).
[4] A. Yethiraj and A. van Blaaderen, Nature **421**, 513 (2003).
[5] J. L. Harland and W. van Megen, Phys. Rev. E **55**, 3054 (1997).
[6] P. N. Pusey and W. van Megen, Nature **320**, 340 (1986).
[7] W. G. Hoover and F. H. Ree, J. Chem. Phys. **49**, 3609 (1968).
[8] C. Sinn, A. Heymann, A. Stipp, and T. Palberg, Prog. Colloid Polym. Sci. **118**, 266 (2001).
[9] K. Schätzel and B. J. Ackerson, Phys. Rev. E **48**, 3766 (1993).
[10] S. Auer and D. Frenkel, Nature **409**, 1020 (2001).
[11] G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974).
[12] J. S. van Duijneveldt and D. Frenkel, J. Chem. Phys. **96**, 4655 (1992).
[13] R. J. Allen, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett **94**, 018104 (2005).
[14] R. J. Allen, D. Frenkel, and P. R. ten Wolde, J. Chem. Phys. **124**, 024102 (2006).
[15] R. J. Allen, D. Frenkel, and P. R. ten Wolde, J. Chem. Phys. **124**, 194111 (2006).

[16] P. N. Pusey, E. Zaccarelli, C. Valeriani, E. Sanz, W. C. K. Poon, and M. E. Cates, Philos. Trans. Roy. Soc. London, Ser. A **367**, 4993 (2009).

[17] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego, 1996).

[18] R. J. Speedy, J. Phys.: Condens. Matter **9**, 8591 (1997).

[19] R. J. Speedy, J. Phys.: Condens. Matter **10**, 4387 (1998).

[20] T. Kawasaki and H. Tanaka, Proc. Natl. Acad. Sc. U.S.A. **107**, 14036 (2010).

[21] J. T. Padding and A. A. Louis, Phys. Rev. E **77**, 011402 (2008).

[22] S. Auer, W. C. K. Poon, and D. Frenkel, Phys. Rev. E **67**, 020401 (2003).

[23] S. Auer and D. Frenkel, J. Phys.: Condens. Matter **14**, 7667 (2002).

[24] P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, Faraday Discuss. **104**, 93 (1996).

[25] P. R. ten Wolde, Ph.D. thesis, University of Amsterdam, 1998.

[26] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **31**, 459 (1959).

[27] E. Zaccarelli, C. Valeriani, E. Sanz, W. C. K. Poon, M. E. Cates, and P. N. Pusey, Phys. Rev. Lett. **103**, 135704 (2009).

[28] P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, J. Chem. Phys. **104**, 9932 (1996).

[29] S. Auer and D. Frenkel, J. Chem. Phys. **120**, 3015 (2004).

[30] R. L. Davidchack, J. R. Morris, and B. B. Laird, J. Chem. Phys. **125**, 094710 (2006).

[31] S. Ryu and W. Cai, Phys. Rev. E **81**, 030601 (2010).

[32] I. J. Ford, Phys. Rev. E **56**, 5615 (1997).

[33] B. Cichocki and K. Hinsen, Physica A **166**, 473 (1990).

[34] E. Sanz and D. Marenduzzo, J. Chem. Phys. **132**, 194102 (2010).

[35] A. Cacciuto, S. Auer, and D. Frenkel, J. Chem. Phys. **119**, 7467 (2003).

[36] S. Auer, Ph.D. thesis, University of Amsterdam, 2002.

[37] R. J. Allen, C. Valeriani, and P. R. ten Wolde, J. Phys.: Cond. Matter **21**, 463102 (2009).

[38] F. A. Escobedo, E. E. Borrero, and J. C. Araque, J. Phys.: Condens. Matter **21**, 333101 (2009).

[39] M. Fasolo and P. Sollich, Phys. Rev. E **70**, 041410 (2004).

[40] B. O'Malley and I. Snook, Phys. Rev. Lett. **90**, 085701 (2003).

[41] W. Lechner and C. Dellago, J. Chem. Phys. **129**, 114707 (2008).

[42] T. Schilling, H. J. Schöpe, M. Oettel, G. Opletal, and I. Snook, Phys. Rev. Lett. **105**, 025701 (2010).

[43] P. G. Bolhuis, D. Frenkel, S. Mau, and D. A. Huse, Nature **388**, 235 (1997).

[44] S. Pronk and D. Frenkel, J. Chem. Phys. **110**, 4589 (1999).

[45] While it does appear that Eq. (19) is completely independent of λ_0, this is not strictly correct as λ_0 creates the border for state A and state A is expected to be a metastable, equilibrated state. For the purposes of this paper, the difference is insignificant as the average time for a nucleation event is much longer than the relaxation time of the fluid.