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

L. Filion, M. Hermes, R. Ni, and M. Dijkstra

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

(Dated: June 16, 2010)

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 [J. L. Harland and W. van Megen, Phys. Rev. E 55, 3054 (1997)], Sinn et al. [C. Sinn et al., Prog. Colloid Polym. Sci. 118, 266 (2001)] and Schätzel and Ackerson [K. Schätzel and B.J. Ackerson, Phys. Rev. E, 48, 3766 (1993)] and the predicted rates for monodisperse and 5% polydisperse hard spheres of Auer and Frenkel [S. Auer and D. Frenkel, Nature 409, 1020 (2001)]. When the rates are examined in long-time diffusion units, we find agreement between all the theoretically predicted nucleation rates, however, the experimental results display a markedly different behaviour for low supersaturation. Additionally, we examined the pre-critical nuclei arising in the molecular dynamics, forward flux sampling, and umbrella sampling simulations. The structure of the nuclei appear independent of the simulation method, and in all cases, the nuclei contain on average significantly more face-centered-cubic ordered particles than hexagonal-close-packed ordered particles.

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, holography, and light and x-ray scattering. Additionally, progress in particle synthesis, solvent manipulation, and the application of external fields 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. Experimentally, the phase behaviour of such a system has been examined by Pusey and Van Megen and maps well onto the phase behaviour 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 \( \eta = 0.545 \pm 0.003 \) which is in good agreement with that predicted for hard spheres. The nucleation rates have been measured using light scattering by Harland and Van Megen, Sinn et al., Schätzel and Ackerson and predicted theoretically by Auer and Frenkel.

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, towards 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 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. 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. 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 continuously being formed and melting 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 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

\[
\Delta G(R) = 4\pi \gamma R^2 - \frac{4}{3} \pi |\Delta \mu| \rho_s R^3
\]

where \( |\Delta \mu| \) is the difference in chemical potential between
the fluid and solid phases, \( \rho_s \) is the density of the solid,
and \( \gamma \) is the surface tension 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 \( \Delta G^* = 16\pi \gamma^3 / 3\rho_s^2 |\Delta \mu|^2 \). Note
that there is no system size dependence in CNT.

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\(^{13}\) 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 towards particular regions in configuration
space. The success of the method is expected to depend
largely on the choice of order parameter and biasing poten-
tial. 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 (eg. the nucleation rate) for rare events can be
determined when brute force simulations are difficult or
even not possible. In FFS, a reaction coordinate \( Q \) (simi-
lar 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 \( \Phi_{AB\lambda_0} \)
of trajectories crossing the A state boundary, typically
denoted \( \lambda_0 \), and the probability \( P(\lambda_B|\lambda_0) \) that a tra-
jectory which has crossed this boundary will reach state
B before returning to state A. Thus the transition rate
constant is written as

\[
k_{AB} = \Phi_{AB\lambda_0}P(\lambda_B|\lambda_0).
\]

Forward flux sampling facilitates the calculation of prob-
ability \( P(\lambda_B|\lambda_0) \) by breaking it up into a set of probabili-
ties between sequential values of the reaction coordinate.
Little information regarding the details of the nucleation
process is required in advance, and the choice of reac-
tion coordinate is expected to be less important than the
order parameter in US. Additionally, unlike US, FFS uti-
izes dynamical simulations 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 tech-
nique to study dynamical processes such as nucleation.
Unfortunately, however, available computational time often
limits the types of systems which can be effectively
studied by these dynamical techniques. Brownian dy-
namics 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 mixtures over a range of volume fractions.
The main difference between the two simulation methods
regards how they treat the short-time motion of the par-
ticles. 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\(^{19}\).

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 corre-
sponding pressures and supersaturations shown in Table
I. We compare these theoretical nucleation rates with the
rates measured experimentally by Refs.\(^{11,13}\).

This paper is organized as follows: in section II we
describe and examine the order parameter used to dis-
tinguish between solid- and fluid-like particles through-
out this paper, in section III we calculate essentially the
“exact” nucleation rates using MD simulations, in sec-
tions IV and V we calculate the nucleation rates of hard
spheres using US and FFS respectively, and discuss dif-
ficulties in the application of these techniques, in section
VI we summarize the theoretical results and compare the
predicted nucleation rates with the measured experimen-
tal rates of Harland and Van Megen\(^{11}\), Sinn et al.\(^{12}\) and
Schätzle and Ackerson\(^{13}\) and section VII contains our con-
clusions.

| \( \eta \) | \( \beta \sigma^3 \) | \( \Delta \mu \) |
|---|---|---|
| 0.5214 | 15.0 | 0.34 |
| 0.5284 | 16.0 | 0.44 |
| 0.5316 | 16.4 | 0.48 |
| 0.5348 | 16.9 | 0.53 |
| 0.5352 | 17.0 | 0.54 |
| 0.5381 | 17.5 | 0.58 |
| 0.5414 | 18.0 | 0.63 |
| 0.5478 | 19.1 | 0.74 |
| 0.5572 | 20.8 | 0.90 |

TABLE I: Packing fraction \( \eta = \pi \sigma^3 N/6V \) , reduced pressure \( \beta \sigma^3 \) and chemical potential difference between the fluid and solid phases \( \Delta \mu \) of the state points studied in this pa-
per. The chemical potential difference was determined using thermodynamic integration\(^{12}\) and the equations of state for the
fluid and solid are from Refs.\(^ {18,19}\) respectively.
II. 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. For this study we have chosen to use the local bond-order parameter introduced by Ten Wolde et al.\textsuperscript{[22]} 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.\textsuperscript{[3]}

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

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

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 $r_{ij} = r_j - r_i$ with $r_i$ 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 $\xi_c$ and where

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

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

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

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

The parameters contained in this algorithm include the neighbour cutoff $r_c$, the dot-product cutoff $d_c$, the critical value for the number of solid-like neighbours $\xi_c$, 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.

To investigate the effect of the choice of $\xi_c$, 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 $r_c = 1.4$, $d_c = 0.7$ and $\xi_c = 5, 7$ and 9. As shown in Fig. 1 for all values of $\xi_c$ that we examined the order parameter appears to consistently identify the particles belonging to the bulk fluid and solid regions. For comparison we also show a typical configuration of the RHCP crystal in coexistence with the fluid phase. The solid-like particles as defined by the order parameter are labelled according to the number of solid-like neighbours while the fluid-like particles are denoted by dots. The main difference between these order parameters relates to distinguishing between fluid- 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 $\xi_c$ is increased. We note that the dips in the density profile correspond to HCP stacked layers which are more pronounced for higher values of $\xi_c$.

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 labelled according to three different crystallinity criteria: the red particle have between $\xi = 5$ and 6 crystalline bonds, the green particles have between $\xi = 7$ and 8 crystalline bonds and the blue particles have $\xi \geq 9$ or more crystalline bonds. The fluid-like particles ($\xi < 5$) are denoted by dots. Bottom: The density profile of particles with a minimum number of neighbours $\xi$ as labelled. Note that the dips in the density profile correspond to HCP stacked layers.
III. 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. 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

\[ k = \frac{1}{\langle t \rangle V} \]  

(6)

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.

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 \( \sigma \sqrt{m/k_B T} \). The order parameter was measured every 10 time units and when the largest cluster exceeded the critical size by 100 percent we estimated the time \( \tau_{\text{nucl}} \) at which the critical nucleus was formed using stored previous configurations. We performed up to 20 runs for every density and averaged the nucleation times.

The results are shown in Table I. The nucleation times shown here are for a system of \( 2.0 \cdot 10^4 \) particles and in MD time units. To compare with other data we convert the MD time units to units of \( \sigma^2/(6D_l) \) with \( D_l \) 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. who used polydisperse particles to prevent crystallization. For \( \eta < 0.54 \), we find good agreement between our data for \( D_L \) and this fit.

| Volume fraction | Average nucleation time | Rate |
|-----------------|-------------------------|------|
| \( \eta \)     | \( t \sqrt{k_BT/\langle m\sigma^2 \rangle} \) | \( k\sigma^2/(6D_l) \) |
| 0.5316          | \( 1.0 \cdot 10^9 \)    | 5.10^{-9} |
| 0.5348          | \( 1.7 \cdot 10^4 \)    | 3.6.10^{-7} |
| 0.5381          | \( 1.4 \cdot 10^3 \)    | 5.3.10^{-6} |
| 0.5414          | \( 2.0 \cdot 10^2 \)    | 4.3.10^{-5} |
| 0.5478          | 42                      | 3.0.10^{-4} |
| 0.5572          | 10                      | 2.4.10^{-3} |

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 \( \sigma^2/(6D_l) \).

IV. 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. This method has been applied frequently to study rare events, such as nucleation, and specifically has been applied in the past to study the nucleation of hard spheres. In general, umbrella sampling is used to examine parts of configurational space which are unaccessible by traditional schemes, eg. 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 is 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

\[ U_{\text{bias}}(\mathbf{r}^N) = \frac{\lambda}{2}(n(\mathbf{r}^N) - n_C)^2 \]  

(7)

where \( \lambda \) 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 \( \lambda \) carefully, the simulation will fluctuate around the part of configurational space with \( n(\mathbf{r}^N) \) in the vicinity of \( n_C \). The expectation value of an observable \( A \) is then given by

\[ \langle A \rangle = \frac{\langle A/W(\mathbf{r}^N) \rangle_{\text{bias}}}{\langle 1/W(\mathbf{r}^N) \rangle_{\text{bias}}} \]  

(8)

where

\[ W(x) = e^{-\beta U_{\text{bias}}(x)} \]  

(9)

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

\[ \beta \Delta G(n) = \text{constant} - \ln(P(n)). \quad (10) \]

Many more details on this method are given elsewhere.\textsuperscript{17,25}

FIG. 2: Gibbs free energy barriers \( \beta \Delta G(n) \) as a function of cluster-size \( n \) as obtained from umbrella sampling simulations at a reduced pressure of \( \beta \rho \sigma^3 = 17 \) for varying critical number of solid-like neighbours \( \xi_c \) as labelled. For \( \xi_c = 5, 7, 9 \), the neighbour cutoff is \( r_c = 1.4 \) and for \( \xi_c = 6, 8, 10 \), \( r_c = 1.3 \). In all cases the dot product cutoff is \( d_c = 0.7 \).

FIG. 3: Classical nucleation theory fits (thick lines) to the Gibbs free energy barriers obtained from umbrella sampling simulations at a reduced pressure of \( \beta \rho \sigma^3 = 17 \) for varying \( \xi_c \) as labelled. Note that the CNT radius \( R_{\text{CNT}} \) is related to the radius \( R(\xi_c) \) measured by umbrella sampling by \( R(\xi_c) = R_{\text{CNT}} + \alpha(\xi_c) \), where \( \alpha(\xi_c) \) is a constant. The fit parameters are given in Table IV A. We have shifted the barriers for \( \xi_c = 6 - 9 \) by 5, 10, 15, 20 \( k_B T \) respectively for clarity.

For a pressure of \( \beta \rho \sigma^3 = 17 \), corresponding to a supersaturation of \( \beta |\Delta \mu| = 0.54 \), we examine the effect of one of the order parameter variables, namely \( \xi_c \), 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 \( \xi_c \) within error bars. In general, for larger values of \( \xi_c \) more particles are identified as fluid as compared with smaller values of \( \xi_c \). This is consistent with the differences between these order parameters as demonstrated in Fig. 1.

Taking the previous discussion on order parameters into consideration, we fit the barriers corresponding to \( \xi_c = 5, 6, 7, 8 \) and 9 using CNT where we assume there exists a CNT radius \( R_{\text{CNT}} \) which differs from the radius \( R(\xi_c) \) measured by the order parameter. We assume that the difference \( (\alpha) \) is a constant for each value of the critical number of solid-like neighbours \( \xi_c \) which corrects for the different ways the various order parameters identify the particles at the fluid-solid interface:

\[ R(\xi_c) = R_{\text{CNT}} + \alpha(\xi_c). \quad (11) \]

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

\[ n(\xi_c) = \frac{4 \pi R(\xi_c)^3 \rho_s}{3}. \quad (12) \]

Fitting all barriers simultaneously for the surface tension, and the various \( \alpha(\xi_c) \), we obtain the fits displayed in Fig. 3. From the various values of \( \alpha \), the associated critical CNT radius \( R_{\text{CNT}} \) can be determined.
We find \( R_{\text{CNT}}^* = 2.49\sigma \). Additionally, we find a surface tension of \( \beta \gamma \sigma^2 = 0.76 \) which roughly agrees with the results of Auer and Frenkel who obtained surface tensions of \( \beta \gamma \sigma^2 = 0.699, 0.738 \) and 0.748 for pressures \( \beta \rho a^3 \) = 15, 16 and 17 respectively. However, recent calculations by Davidchack et al.\(^{22}\) of the surface tension 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 surface tension is expected to be an average over the crystal planes. Thus our result for the surface tension and that of Ref. 4 appear to be an overestimate.

There have been a number of papers discussing possible corrections to CNT (eg. Refs. 27-29). Recent work on the 2d Ising model, a system where both the surface tension 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 \( \Delta G \) which we define as \( c \).\(^{22}\) 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 \( \xi \), we obtain the fits displayed in Fig. 1. The US barriers with this proposed form for the Gibbs free energy barrier is in much better agreement with recent calculations of Davidchack et al.\(^{22}\) than the surface tension we calculate using classical nucleation theory directly. 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.

Using both expressions for the Gibbs free energy barrier, namely CNT and ACNT, we were unable to fit the barrier corresponding to \( \beta \rho a^3 = 17 \) and \( n_0 \) = 10 simultaneously with the other predicted barriers for the same pressure. We speculate that our difficulty in fitting the barrier at \( \xi_e \) = 10 stems from an “over-biasing” of the system. Specifically, by using \( \xi_e \) = 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.

In conclusion, with the exception of \( \xi_e = 10 \), the value of \( \xi_e \) 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 (section IV B), we also calculated the Gibbs free energy \( \Delta G(n) \) for reduced pressures \( \beta \rho a^3 = 15 \) 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 \( (G(n)) \) by:

\[
\frac{1}{k} = Ae^{-\beta\Delta G(n^*)}
\]

where

\[
A \approx \rho f_{n^*} \sqrt{\frac{\beta \Delta G''(n^*)}{2\pi}}.
\]

\( n^* \) is the number of particles in the critical nucleus, \( \rho \) is the number density of the supersaturated fluid, \( f_{n^*} \) is the rate particles are attached to the critical cluster, and \( G'' \) is the second derivative of the Gibbs free energy barrier. Auer and Frenkel\(^{22}\) 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:

\[
f_{n^*} = \frac{1}{2} \langle n^2(t) \rangle - \langle n(t) \rangle^2.
\]

The mean square deviation of the cluster size can then be calculated by either employing a kinetic MC 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 umbrella sampling (US) nucleation rates, although to calculate the nucleation rates both US simulations and dynamical simulations (KMC or MD) are necessary.

The mean square deviation, or variance, in the cluster size appearing in Eq. 15 has both a short- and long-time behaviour. At short times, fluctuations are due to particles performing Brownian motion around their average positions while the long-time behaviour is caused by rearrangements of particles required for the barrier crossings.

| CNT   | 0.54 | 0.76 | 2.49 | -0.425 | -0.231 | -0.000 | 0.139 | 0.380 |
|-------|------|------|------|--------|--------|--------|-------|-------|
| ACNT  | 0.54 | 0.61 | 2.01 | -0.961 | -0.765 | -0.551 | -0.402 | -0.148 |

**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.
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\textsuperscript{20} 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 is an underestimation of the diffusion constant, and similarly, in this case, an underestimate of the attachment rate. 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/τ\textsubscript{L} where τ\textsubscript{L} is the time that it takes a particle on average to diffuse over a distance equal to its diameter i.e. \( \tau = \sigma^2/(6D) \).

![Fig. 6: The mean squared deviation (MSD) of the cluster size (\( \Delta n^2(t) \)) as 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. 15).](image)

**TABLE IV:** Nucleation rates \( k \) in units of \( D_0/c^3 \), with \( D_0 \) the short time diffusion coefficient as a function of reduced pressure \( (\beta \sigma^3) \) as predicted by umbrella sampling. \( G''(n^*) \) is the second order derivative of the Gibbs free energy at the critical nucleus size \( n^* \).

| \( \beta \sigma^3 \), \( \xi^* \), \( n^* \) | \( \beta \Delta G(n^*) \) | \( \beta \Delta G''(n^*) \) | \( f_{eq}/D_0 \) | \( k\sigma^3/D_0 \) |
|---|---|---|---|---|
| 15 | 8 | 212.4 ± 0.2 | −9.6 · 10\textsuperscript{-3} | 661.4 | 4.35 · 10\textsuperscript{-16} |
| 16 | 8 | 112.7 ± 0.6 | −1.6 · 10\textsuperscript{-3} | 429.1 | 7.80 · 10\textsuperscript{-12} |
| 17 | 6 | 102.0 ± 0.3 | −1.2 · 10\textsuperscript{-3} | 712.9 | 3.08 · 10\textsuperscript{-8} |
| 18 | 7 | 72.0 ± 0.4 | −2.0 · 10\textsuperscript{-3} | 469.8 | 1.77 · 10\textsuperscript{-8} |
| 19 | 10 | 30.0 ± 0.7 | −9.4 · 10\textsuperscript{-3} | 316.1 | 4.49 · 10\textsuperscript{-8} |

V. FORWARD FLUX SAMPLING

A. Method

The forward flux sampling method was introduced by Allen et al\textsuperscript{[3]} in 2005 to study rare events and has since been applied to a wide variety of systems. Two review articles (Refs. \textsuperscript{30,31}) 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. \textsuperscript{14} with \( \xi = 8 \), \( r_c = 1.3 \), and \( d_c = 0.7 \). In general, the reaction coordinate is used to divide phase space by a sequence of interfaces \( (\lambda_0, \ldots, \lambda_{n-1}) \).
\(\lambda_1, \ldots, \lambda_N\) associated with increasing values \(n(p^N)\) 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

\[
k_{AB} = \Phi_{AB0} P(\lambda_N|\lambda_0)
= \Phi_{AB0} \prod_{i=0}^{N-1} P(\lambda_{i+1}|\lambda_i)
\]

(16)

(17)

where \(\Phi_{AB0}\) is the steady-state flux of trajectories leaving the \(A\) state and crossing the interface \(\lambda_0\) in a volume \(V\), and \(P(\lambda_{i+1}|\lambda_i)\) is the probability that a configuration starting at interface \(\lambda_i\) will reach interface \(\lambda_{i+1}\) 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. 3 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 behaviour, short-time fluctuations related to surface fluctuations of the cluster, and a longer time cluster growth (Fig. 7). Thus, if we try to measure the flux \(\Phi_{AB}\) 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 properly. 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 pre-critical 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 pre-critical 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- and fluid-like 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 pre-critical 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 slightly novel way. We regroup the elements of the rate calculation such that

\[
k_{AB} = \tilde{\Phi}_{AB1} \prod_{i=1}^{N-1} P(\lambda_{i+1}|\lambda_i).
\]

(18)

where

\[
\tilde{\Phi}_{AB1} = \Phi_{AB0} P(\lambda_1|\lambda_0).
\]

(19)

We note that if \(\lambda_1\) is chosen such it is a relatively rare event for trajectories starting in \(A\) to reach \(\lambda_1\), then

\[
\tilde{\Phi}_{AB1} \approx \frac{1}{\langle t_{AA1} \rangle V}
\]

(20)

where \(\langle t_{AA1} \rangle\) is the average time it takes a trajectory in \(A\) to reach \(\lambda_1\). The approximation made here, in contrast to normal FFS simulations, is that the time the system spends with an order parameter greater than \(\lambda_1\) 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 \((\lambda_0)\) anywhere under \(\lambda_1\).

We choose to use \(\lambda_0 = 1\) to minimize the effect of fluctuations, as seen in Fig. 7 on the probability to reach the following interface. Here we assume that any crystalline order in a system with an order parameter of 1 likely does not arise from 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 \( t_{AAi} \), stopping whenever the trajectory first hits interface \( \lambda_1 \).

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 stepsize \( \langle \Delta_{\text{KMC}} \rangle \) per attempt to move each particle, and the frequency with which the order parameter (reaction coordinate) is measured \( \Delta t_{\text{ord}} \). 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 \( \Delta t_{\text{ord}} \) introduces an additional error into our measurement of the rate.

To examine the effects of i) the approximation associated with our method for calculating \( \Phi_{AAi} \), 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 Appendix A. 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 at a reduced pressure of \( \beta p \sigma^3 = 1000 \). The simulations were stopped when the packing fraction associated with the pressure of interest was reached. This initial configuration was then relaxed using an NPT simulation at the correct pressure \( \beta p \sigma^3 = 15, 16, 17 \). 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 \( n(r^N) = \lambda_1 \). 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 \( \lambda_1 \), and following these configurations until they either reached \( \lambda_2 \) 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 \( P(\lambda_2|\lambda_1) \). The remaining \( P(\lambda_{i+1}|\lambda_i) \)'s are calculated similarly.

To study the effect of the two KMC parameters, namely \( \Delta_{\text{KMC}} \) and \( \Delta t_{\text{ord}} \), on the nucleation rates, we have 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 \( \Delta t_{\text{ord}} \) and the maximum displacement \( \Delta_{\text{KMC}} \) 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 \( \Delta t_{\text{ord}} = 5 \) MC cycles and \( \Delta_{\text{KMC}} = 0.2 \sigma \).

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].

### VI. SUMMARY AND DISCUSSION

#### A. Nucleation Rates

In this section we examine hard-sphere nucleation rates predicted using US simulations, MD simulations and FFS simulations together with the experimental results of Harland and Van Megen,4 Simm et al.2 and Schätzle and Ackerson3 and the US simulations of monodisperse and 5% polydisperse hard-spheres mixtures examined by Auer and Frenkel. The experimental volume fractions have been scaled to yield the coexistence densities of monodisperse hard spheres. Similarly, we scale the polydisperse results of Auer and Frenkel with the coexistence densities determined in Ref. 32. Inspired by the recent work of Pusey et al.,18 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, since hydrodynamics are included in the long-time diffusion units, if we present the nucleation rates in terms of the long-time diffusion coefficient, our predicted nucleation rates should be in agreement with the experiments. 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 long-time diffusion coefficient \( D_L \) using

\[
\frac{D_L(\eta)}{D_0} = \left( 1 - \frac{\eta}{0.58} \right)^{\delta}. \tag{21}
\]

Harland and Van Megen claim that \( \delta = 2.6 \) gives a good fit to their system and Simm et al. use \( \delta = 2.58 \). Since the system Schätzle and Ackerson examine is very similar to the other two, we use \( \delta = 2.6 \) to convert their nucleation rates to long-time units. 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 long time units are shown in Fig. 8.

In Ref. 10, Pusey et al. showed that the nucleation rates for various polydispersities (0 to 6%) of hard-sphere
mixtures 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 [4] 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.

However, on the experimental side, the nucleation rates of Harland and Van Megen [1] are approximately one to two orders of magnitude below the experiments of Sinn et al. [2] and Schäztel and Ackerson [3]. This is unexpected due to the similarity between the experimental systems. In our opinion, the main difference between these experiments is the polydispersity of the particle mixtures: 5% in the case of Harland and Van Megen [1], 2.5% in the case of Sinn et al. [2] and <5% for Schäztel and Ackerson [3].

However, as demonstrated by Pusey et al. [10] and now also in
TABLE VI: Nucleation rates predicted using forward flux sampling in short-time diffusion coefficient units ($D_{	ext{th}}$). The probabilities $P(\lambda_{\text{B}}|\lambda_{1})$, number of steps between the order parameter measurements $\Delta_{\text{ord}}$, and kinetic MC stepsize are as in Tables VII, VIII, and IX. At each interface, $C_i$ copied of each successful configuration were used.

| i  | $\lambda_1$ | $C_{i-1}$ | $P(\lambda_{1}|\lambda_{i-1})$ | trial 2 | $C_{i-1}$ | $P(\lambda_{i}|\lambda_{i-1})$ |
|----|-------------|----------|-------------------------------|---------|----------|-------------------------------|
| 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 VII: Probabilities $P(\lambda_{i+1}|\lambda_i)$ for the interfaces used in calculating the nucleation rate for pressure $\beta\rho\sigma^3 = 17$ with step size $\Delta_{\text{KMC}} = 0.1\sigma$ and measuring the order parameter every $\Delta t_{\text{ord}} = 5$ MC cycles.

| i  | $\lambda_1$ | $C_{i-1}$ | $P(\lambda_{i}|\lambda_{i-1})$ | trial 2 | $C_{i-1}$ | $P(\lambda_{i}|\lambda_{i-1})$ |
|----|-------------|----------|-------------------------------|---------|----------|-------------------------------|
| 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    |                               |

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

| i  | $\lambda_1$ | $C_{i-1}$ | $P(\lambda_{i}|\lambda_{i-1})$ |
|----|-------------|----------|-------------------------------|
| 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                         |

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

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 analysed the critical clusters obtained when three different crystalline order parameters were used to bias the US simulations, namely, $\xi_\text{c} = 5, 7$ 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 ($\xi_\text{c} = 5, 7$ 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.

If 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.

To further examine whether the choice of method in-
fluenced the resulting clusters, 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 Figure 11).
There is no indication that the clusters in any of the sim-
ulation methods differed substantially.

Additionally, we examined whether the simulation
technique influenced the type of pre-critical nuclei that
formed in the simulations, ie. face-centered-cubic (FCC),
and hexagonal-close-packed (HCP). To do this we used
the order parameter introduced by Ref. 34 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 approx-
imately 80% FCC-like particles. This was unexpected
as the free energy difference between the bulk FCC and
HCP phases is about 0.001$k_B T$ per particle at melting
and hence a random stacking of hexagonal layers in the
nuclei would be expected. We speculate that this pre-
dominance of FCC stacking in the nuclei arises from sur-
face effects.

VII. CONCLUSIONS

In conclusion, we have examined crystal nucleation of
hard spheres with molecular dynamics, umbrella sam-
ping and forward flux sampling simulations. We find
that the nucleation rates predicted by all three methods
agree over the large range in volume fractions we exam-
inied. Additionally, in agreement with the recent work
of Pusey et al., we find that by measuring the nuclea-
tion 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 also agree. On examining the critical clusters,
we do not find a difference in the nuclei formed using the

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 analyzed 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.
three simulation techniques.

We have also compared our nucleation rates with previous experimental data, specifically, the nucleation rates predicted by Harland and Van Megen, Sinn et al, and Schätzl and Ackerson. The nucleation rates measured by these three experiments, in contrast to what would be expected, differ by about one order of magnitude. In general, the experimental systems are similar enough that one would have expected agreement in the rate once the rates were scaled to the coexistence densities of hard spheres. Additionally, 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. We speculate here that this difference may be due to difficulties in distinguishing between separate nuclei domains in the experiments, or measurement error in the experiments.

VIII. ACKNOWLEDGEMENTS

We would like to thank Frank Smallenburg, Matthieu Marechal, Eduardo Sanz and Chantal Valeriani for many useful discussions. We acknowledge financial support from the NWO-VICI grant and the high potential programme from Utrecht University.
Appendix A: FFS in the presence of measurement error

As mentioned in Section V 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 Section V in this paper we have applied FFS in a slightly novel manner. In this appendix, we introduce a simple model to examine the effect 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

$$\beta U(x) = 8x^2 - 2x^3.$$  \hspace{3cm} (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 $1.5 \cdot 10^{-12}$. In all the calculations in this section, we set the KMC stepsize equal to $\Delta t_{KMC} = 0.1$.

Secondly we explore the effect on the nucleation rate

| $\Delta t_{ord}$ | 1     | 2     | 5     | 10    | 50    |
|------------------|-------|-------|-------|-------|-------|
|                  | $1.2723 \cdot 10^{-12}$ | $1.0589 \cdot 10^{-12}$ | $1.8075 \cdot 10^{-12}$ | $1.5455 \cdot 10^{-12}$ | $1.3835 \cdot 10^{-12}$ |
|                  | $1.3780 \cdot 10^{-12}$ | $1.7217 \cdot 10^{-12}$ | $1.3314 \cdot 10^{-12}$ | $1.4461 \cdot 10^{-12}$ | $1.0666 \cdot 10^{-12}$ |
|                  | $1.2364 \cdot 10^{-12}$ | $1.2924 \cdot 10^{-12}$ | $1.4847 \cdot 10^{-12}$ | $1.1482 \cdot 10^{-12}$ | $1.6134 \cdot 10^{-12}$ |
|                  | $1.6942 \cdot 10^{-12}$ | $1.6422 \cdot 10^{-12}$ | $1.9482 \cdot 10^{-12}$ | $1.4383 \cdot 10^{-12}$ | $1.7550 \cdot 10^{-12}$ |
|                  | $1.2662 \cdot 10^{-12}$ | $1.2340 \cdot 10^{-12}$ | $1.5692 \cdot 10^{-12}$ | $1.6060 \cdot 10^{-12}$ | $1.2908 \cdot 10^{-12}$ |
|                  | $1.6918 \cdot 10^{-12}$ | $1.3530 \cdot 10^{-12}$ | $1.6238 \cdot 10^{-12}$ | $1.6244 \cdot 10^{-12}$ | $1.4012 \cdot 10^{-12}$ |
|                  | $1.6464 \cdot 10^{-12}$ | $1.1788 \cdot 10^{-12}$ | $1.6928 \cdot 10^{-12}$ | $1.0191 \cdot 10^{-12}$ | $1.3403 \cdot 10^{-12}$ |
|                  | $1.6809 \cdot 10^{-12}$ | $1.5860 \cdot 10^{-12}$ | $1.1903 \cdot 10^{-12}$ | $1.6227 \cdot 10^{-12}$ | $1.0582 \cdot 10^{-12}$ |
|                  | $1.4602 \cdot 10^{-12}$ | $1.7018 \cdot 10^{-12}$ | $1.3191 \cdot 10^{-12}$ | $1.3850 \cdot 10^{-12}$ | $2.3732 \cdot 10^{-12}$ |
|                  | $1.7459 \cdot 10^{-12}$ | $1.9154 \cdot 10^{-12}$ | $1.5638 \cdot 10^{-12}$ | $1.2378 \cdot 10^{-12}$ | $1.2692 \cdot 10^{-12}$ |

TABLE X: Nucleation rates for the one-dimensional potential given by Eq. A1 and shown in Fig. 13 for $\Delta t_{ord}$ as indicated. For each $\Delta t_{ord}$, 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 $C_l = 10$ 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. 20.

| $\sigma_{Gauss}$ | 0.02 | 0.04 | 0.06 | 0.08 | 0.1 |
|------------------|------|------|------|------|-----|
|                  | $1.8623 \cdot 10^{-12}$ | $1.7281 \cdot 10^{-12}$ | $1.2630 \cdot 10^{-12}$ | $1.0634 \cdot 10^{-12}$ | $1.9158 \cdot 10^{-12}$ |
|                  | $1.7627 \cdot 10^{-12}$ | $1.6900 \cdot 10^{-12}$ | $1.6402 \cdot 10^{-12}$ | $1.5655 \cdot 10^{-12}$ | $1.8785 \cdot 10^{-12}$ |
|                  | $9.9796 \cdot 10^{-13}$ | $1.6305 \cdot 10^{-12}$ | $1.5799 \cdot 10^{-12}$ | $1.6936 \cdot 10^{-12}$ | $1.4937 \cdot 10^{-12}$ |
|                  | $1.3743 \cdot 10^{-12}$ | $1.2261 \cdot 10^{-12}$ | $1.8305 \cdot 10^{-12}$ | $1.7733 \cdot 10^{-12}$ | $1.1142 \cdot 10^{-12}$ |
|                  | $1.6917 \cdot 10^{-12}$ | $1.8054 \cdot 10^{-12}$ | $1.6191 \cdot 10^{-12}$ | $1.8941 \cdot 10^{-12}$ | $1.0402 \cdot 10^{-12}$ |
|                  | $1.8482 \cdot 10^{-12}$ | $1.3337 \cdot 10^{-12}$ | $1.3283 \cdot 10^{-12}$ | $1.4039 \cdot 10^{-12}$ | $7.0735 \cdot 10^{-13}$ |
|                  | $1.5289 \cdot 10^{-12}$ | $8.6859 \cdot 10^{-13}$ | $1.3129 \cdot 10^{-12}$ | $2.7115 \cdot 10^{-12}$ | $2.4711 \cdot 10^{-12}$ |
|                  | $1.8918 \cdot 10^{-12}$ | $1.4325 \cdot 10^{-12}$ | $1.3203 \cdot 10^{-12}$ | $1.3792 \cdot 10^{-12}$ | $1.6288 \cdot 10^{-12}$ |
|                  | $1.3144 \cdot 10^{-12}$ | $1.2283 \cdot 10^{-12}$ | $1.0459 \cdot 10^{-12}$ | $1.7194 \cdot 10^{-12}$ | $1.3764 \cdot 10^{-12}$ |
|                  | $1.6654 \cdot 10^{-12}$ | $1.1236 \cdot 10^{-12}$ | $1.2572 \cdot 10^{-12}$ | $1.9631 \cdot 10^{-12}$ | $1.8976 \cdot 10^{-12}$ |

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 $\sigma_{Gauss}$ is as indicated. For each $\sigma_{Gauss}$, 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 $C_l = 10$ 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. 20.
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 XI. The average nucleation rates predicted for all values of $\Delta t_{\text{ord}}$ 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 Section V predicts the correct nucleation rates.

Finally, we examine the effect that the 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

$$x_m = x_{\text{true}} + \delta$$  \hspace{1cm} (A2)

where $x_m$ is the value of the order parameter used in the FFS simulation, $x_{\text{true}}$ is the true value of the order parameter, and $\delta$ is taken from a Gaussian distribution with a mean of 0 and a standard deviation $\sigma_{\text{Gauss}}$. In Table XI we demonstrate the effect on the predicted nucleation rate for various choices of $\sigma_{\text{Gauss}}$. The resulting nucleation rates are in good agreement with the spontaneous results. For larger $\sigma_{\text{Gauss}}$, 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. 20 as well as the effect of measurement error in the order parameter and the measurement frequency $\Delta t_{\text{ord}}$ 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. J. L. Harland and W. van Megen, Phys. Rev. E 55, 3054 (1997).
2. C. Sinn, A. Heymann, A. Stipp, and T. Palberg, Prog. Colloid Polym. Sci. 118, 266 (2001).
3. K. Schätzeli and B. J. Ackerson, Phys. Rev. E 48, 3766 (1993).
4. S. Auer and D. Frenkel, Nature 409, 1020 (2001).
5. A. D. Dinsmore et al., Appl. Optics 40, 4152 (2001).
6. S. Lee et al., Opt. Express 15, 18275 (2007).
7. S. C. Glotzer and M. J. Solomon, Nat. Mater. 6, 557 (2007).
8. A. Yethiraj and A. van Blaaderen, Nature 421, 513 (2003).
9. P. N. Pusey and W. van Megen, Nature 320, 340 (1986).
10. W. G. Hoover, J. Chem. Phys. 49, 3609 (1968).
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 et al., Philos. T. Roy. Soc. A 367, 4993 (2009).
17. D. Frenkel and B. Smit, Understanding Molecular Simulation: From Algorithms to Applications (Academic Press, San Diego, 1996).
18. R. J. Speedy, J. Phys.: Condensed. Matter 9, 8591 (1997).
19. R. J. Speedy, J. Phys.: Condensed. Matter 10, 4387 (1998).
20. P. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, Faraday Discuss. 104, 93 (1996).
21. P. R. ten Wolde, Ph.D. thesis, University of Amsterdam, 1998.
22. B. J. Alder and T. E. Wainwright, J. Chem. Phys. 31, 459 (1959).
23. E. Zaccarelli et al., Phys. Rev. Lett. 103, 135704 (2009).
24. P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, J. Chem. Phys. 104, 9932 (1996).
25. S. Auer and D. Frenkel, J. Chem. Phys. 120, 3015 (2004).
26. R. L. Davidson, J. R. Morris, and B. B. Laird, J. Chem. Phys. 125, 094710 (2006).
27. S. Ryu and W. Cai, Phys. Rev. E 81, 030601 (2010).
28. I. J. Ford, Phys. Rev. E 56, 5615 (1997).
29. S. Auer, Ph.D. thesis, University of Amsterdam, 2002.
30. R. J. Allen, C. Valeriani, and P. R. ten Wolde, J. Phys.: Cond. Matter 21, 463102 (2009).
31. F. A. Escobedo, E. E. Borrero, and J. C. Araque, J. Phys.: Cond. Matter 21, 333101 (2009).
32. M. Fasolo and P. Sollich, Phys. Rev. E 70, 041410 (2004).
33. B. O’Malley, and I. Snook, Phys. Rev. Lett. 90, 085701
(2003).
34 W. Lechner and C. Dellago, J. Chem. Phys. 129, 114707 (2008).
35 P. G. Bolhuis, D. Frenkel, S. Mau, and D. A. Huse, Nature 388, 235 (1997).
36 S. Pronk and D. Frenkel, J. Chem. Phys. 110, 4589 (1999).
37 While it does appear that Eq. 18 is completely independent of $\lambda_0$, this is not strictly correct as $\lambda_0$ 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 for the fluid.