A metadynamics based approach to sampling crystallisation events

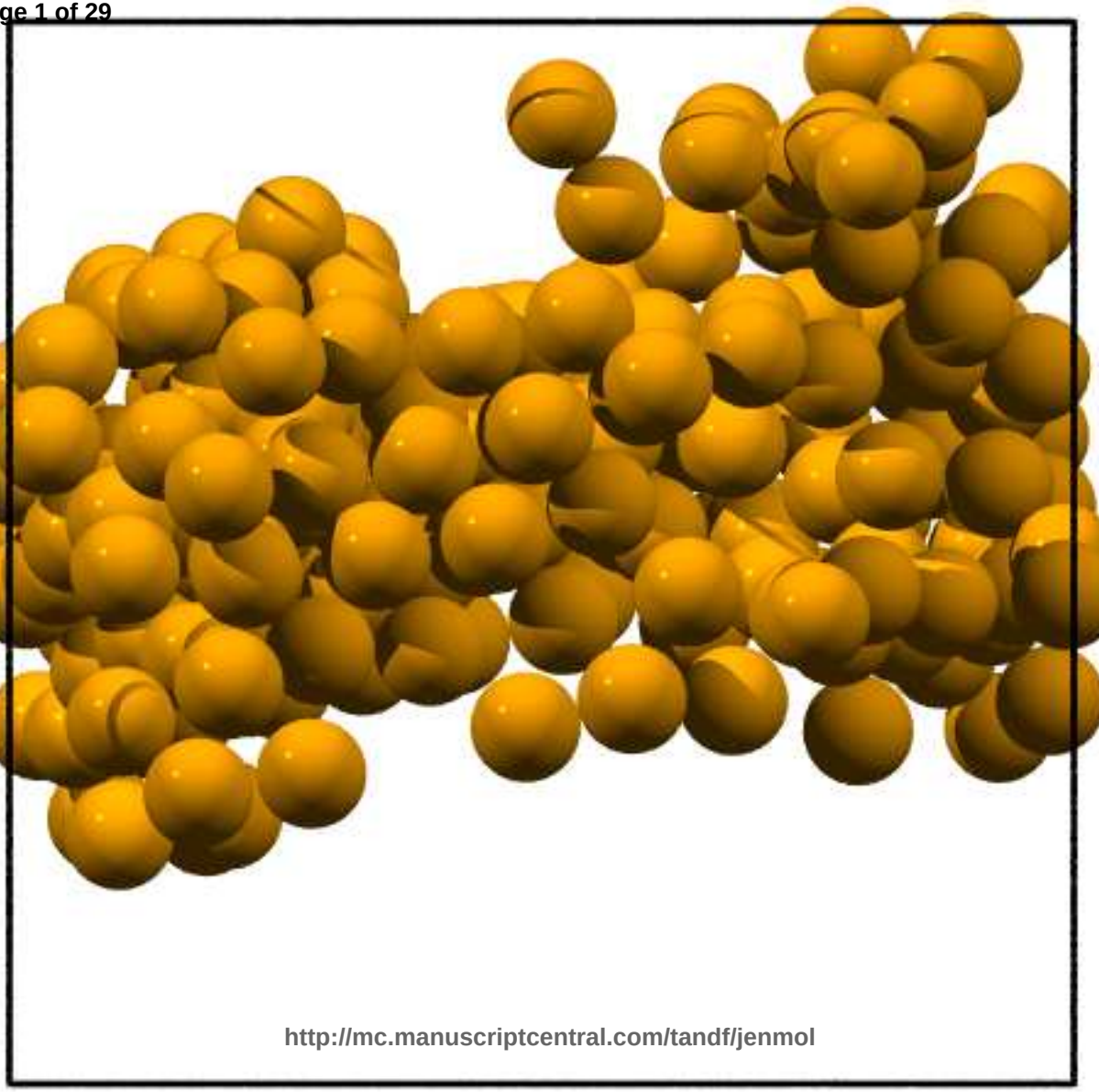
David Quigley, Mark Rodger

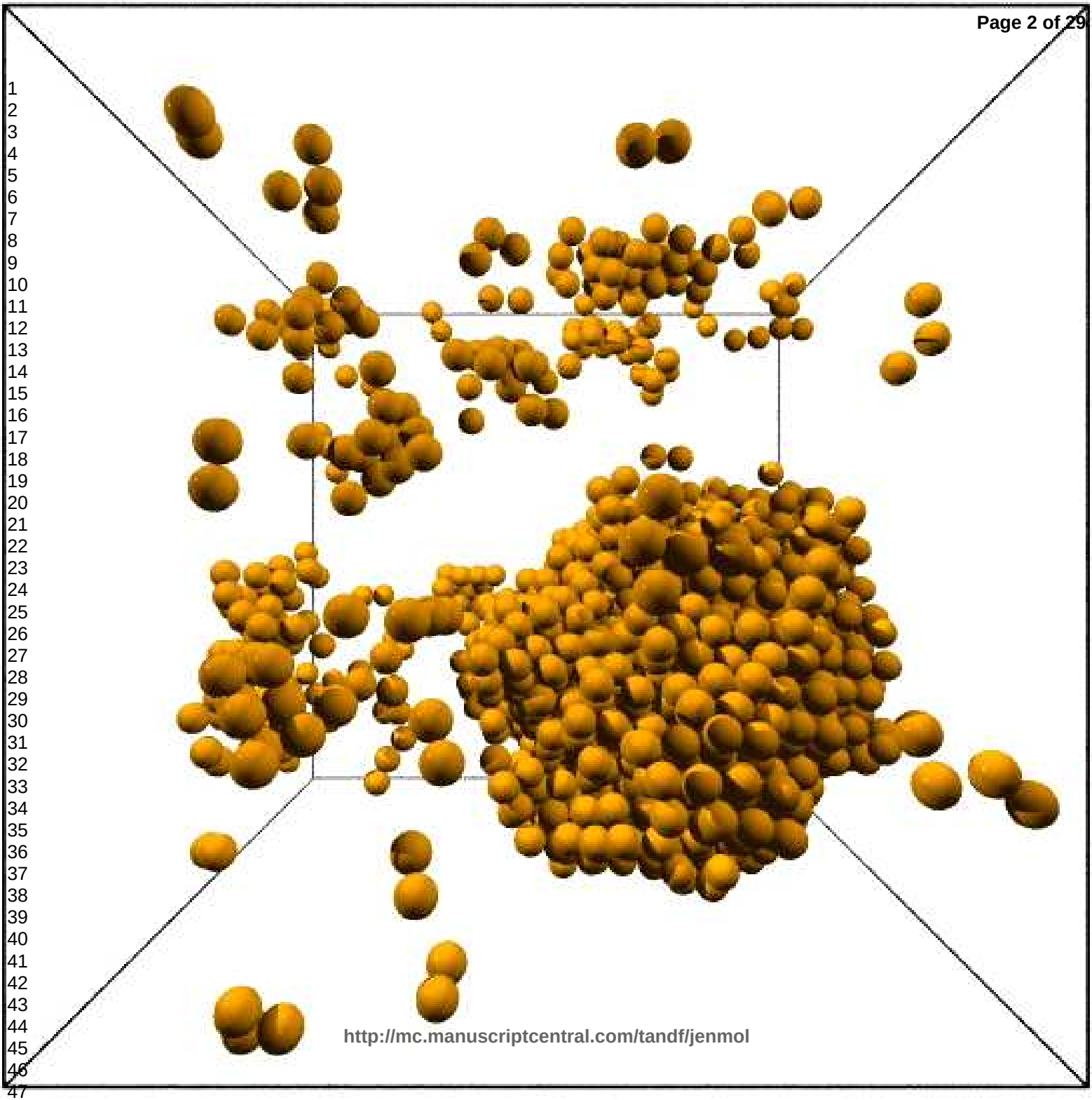
Molecular Simulation
Journal of
Experimental Nanoscience

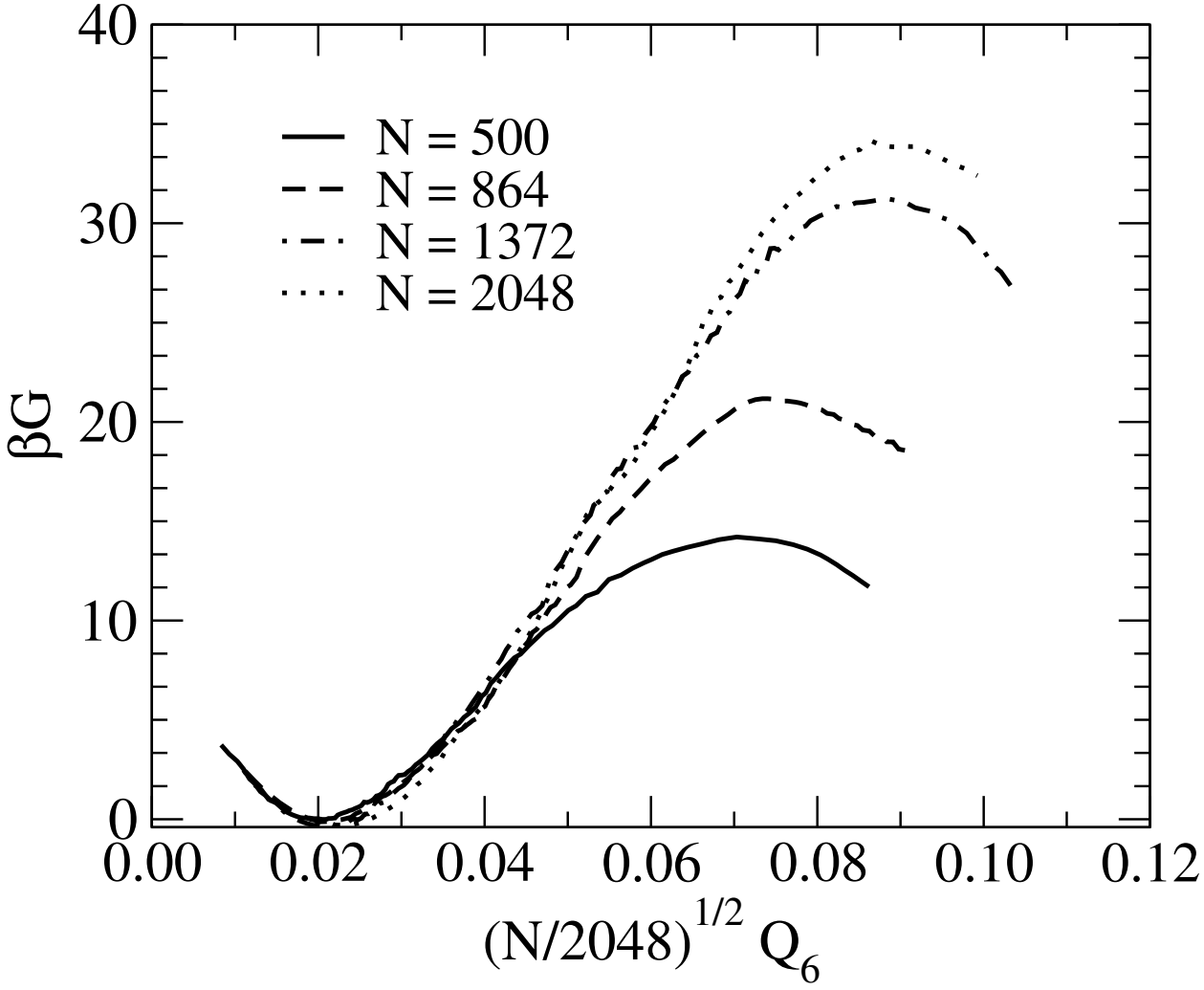
Taylor & Francis
Taylor & Francis Group

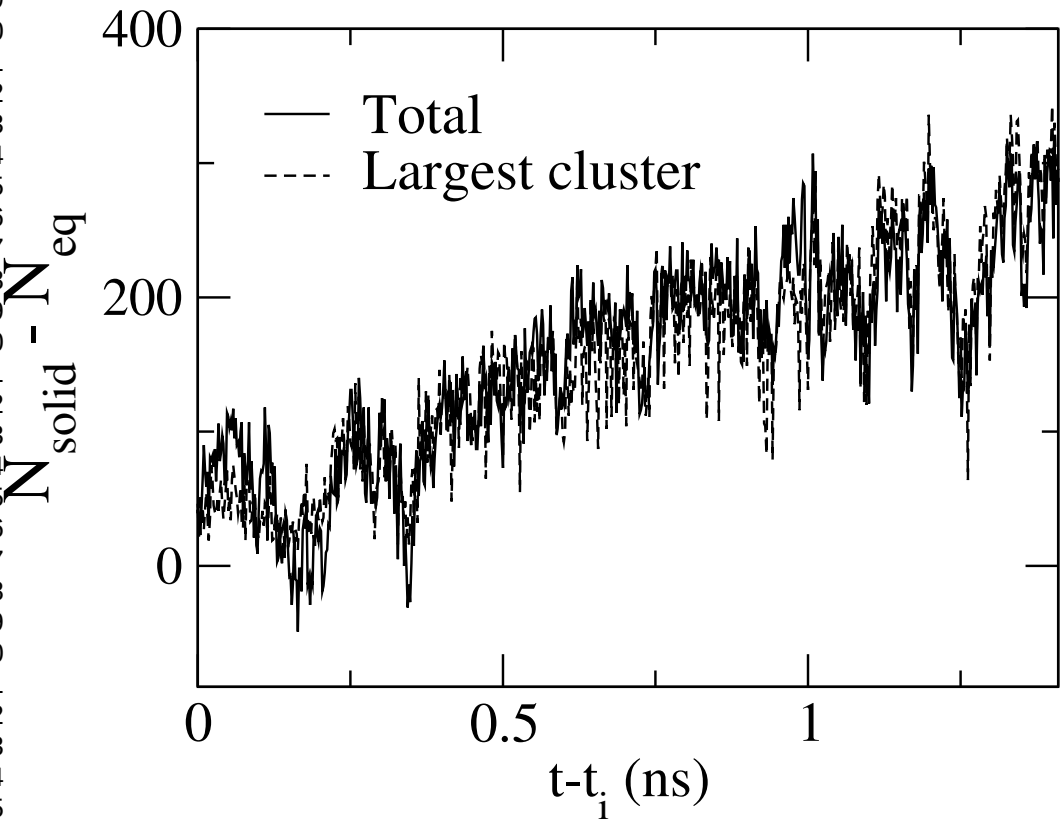
A metadynamics based approach to sampling crystallisation events

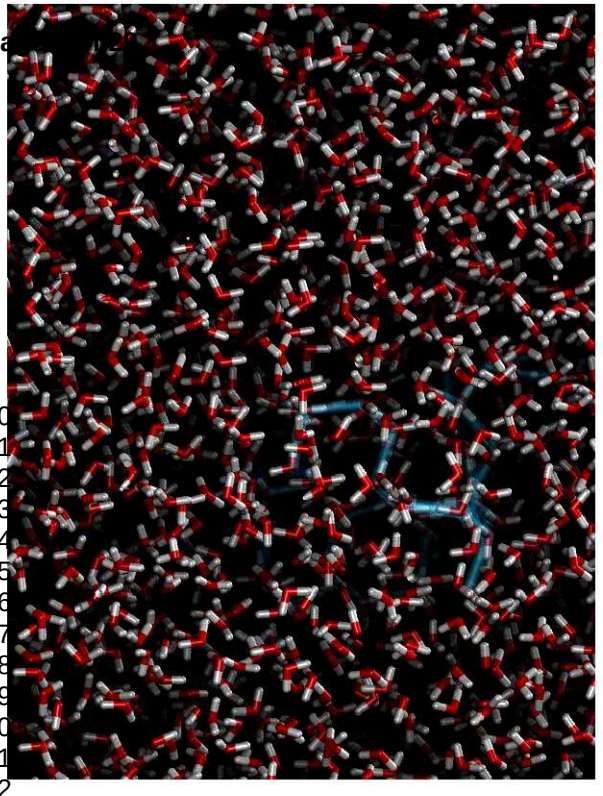
Journal:	*Molecular Simulation/Journal of Experimental Nanoscience*
Manuscript ID:	GMOS-2008-0230
Journal:	Molecular Simulation
Date Submitted by the Author:	14-Nov-2008
Complete List of Authors:	Quigley, David; University of Warwick, Chemistry and Centre for Scientific Computing Rodger, Mark; University of Warwick, Chemistry and Centre for Scientific Computing
Keywords:	Crystallisation, Freezing, Long Timescale, Lennard-Jones, Ice



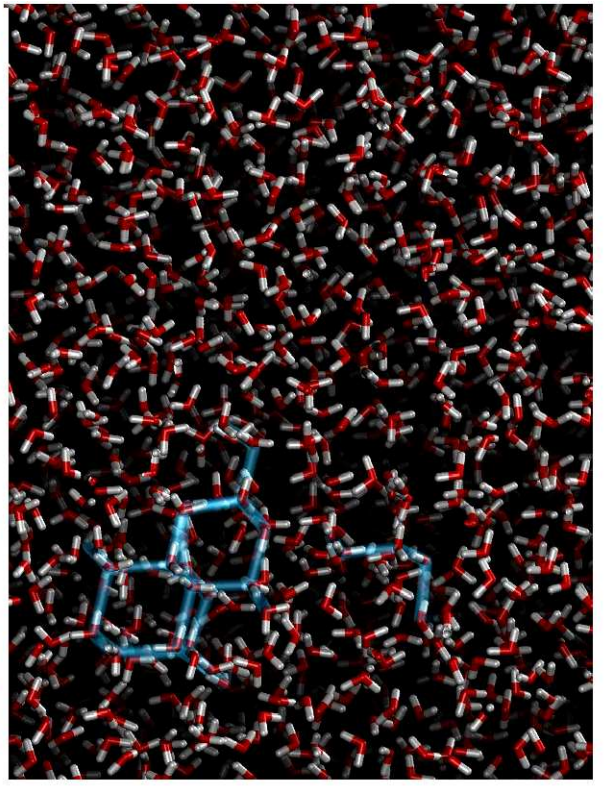



40
30
20
10
0
βG
N = 500
N = 864
N = 1372
N = 2048
0.00
0.02
0.04
0.06
0.08
0.10
0.12
(N/2048)^1/2 Q6

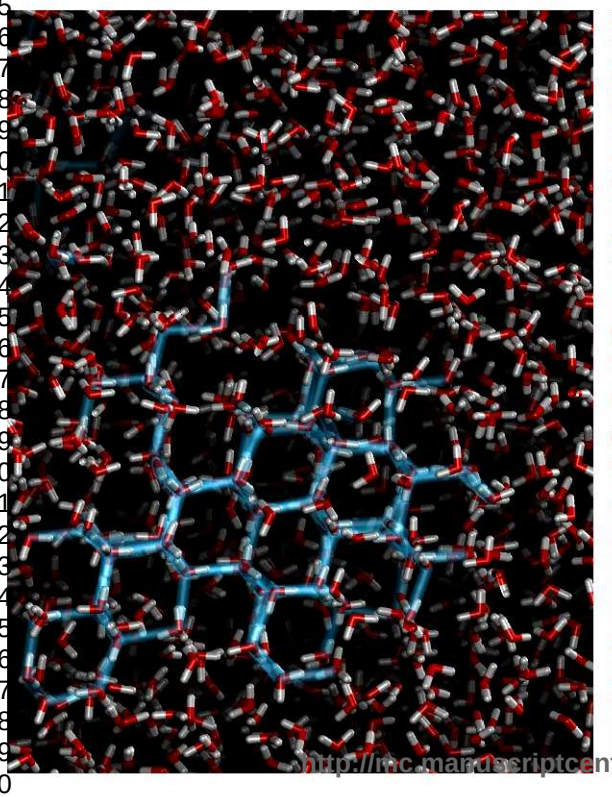

400
200
0
0
0.5
1
Total
Largest cluster
N solid - N eq
t-t i (ns)



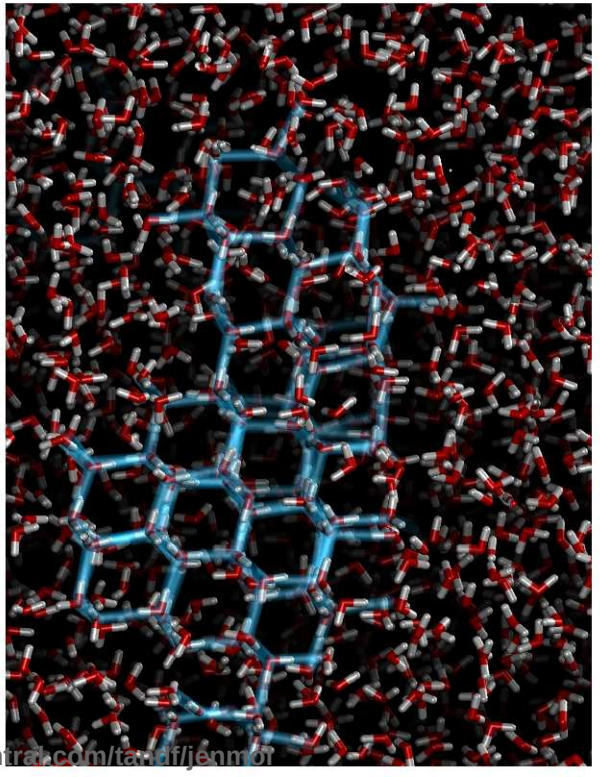
t=0.5ns



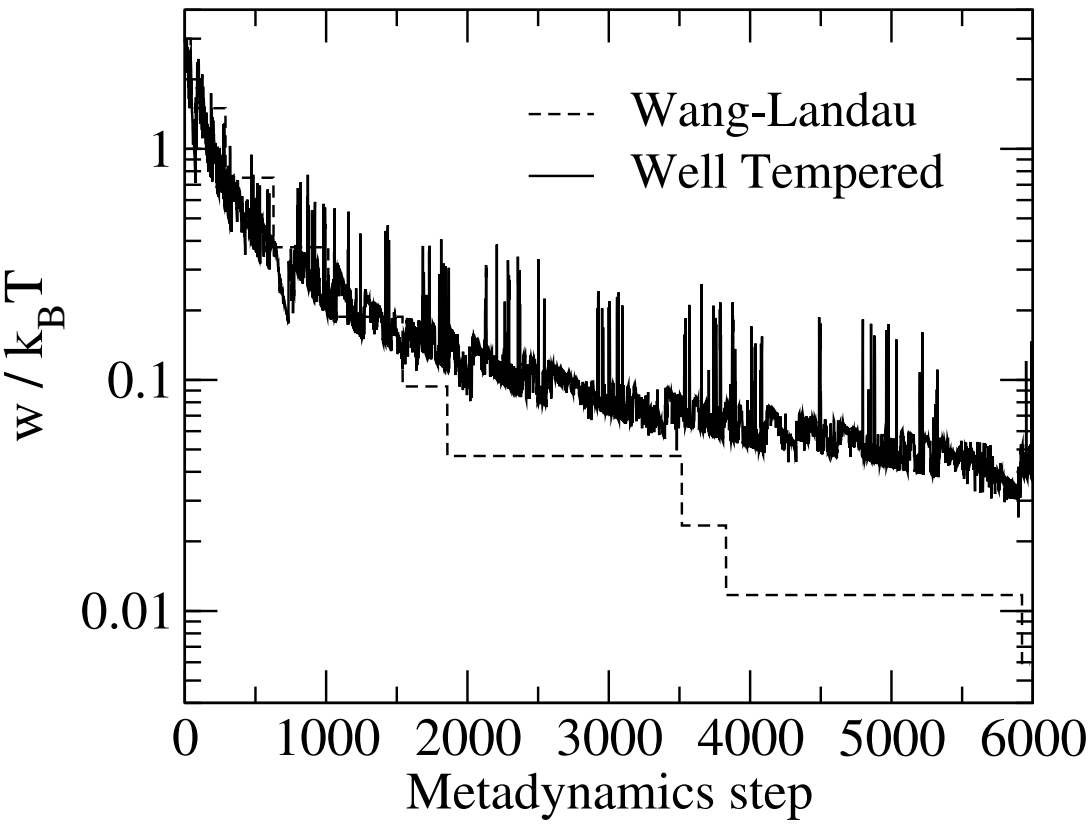
t=0.75ns

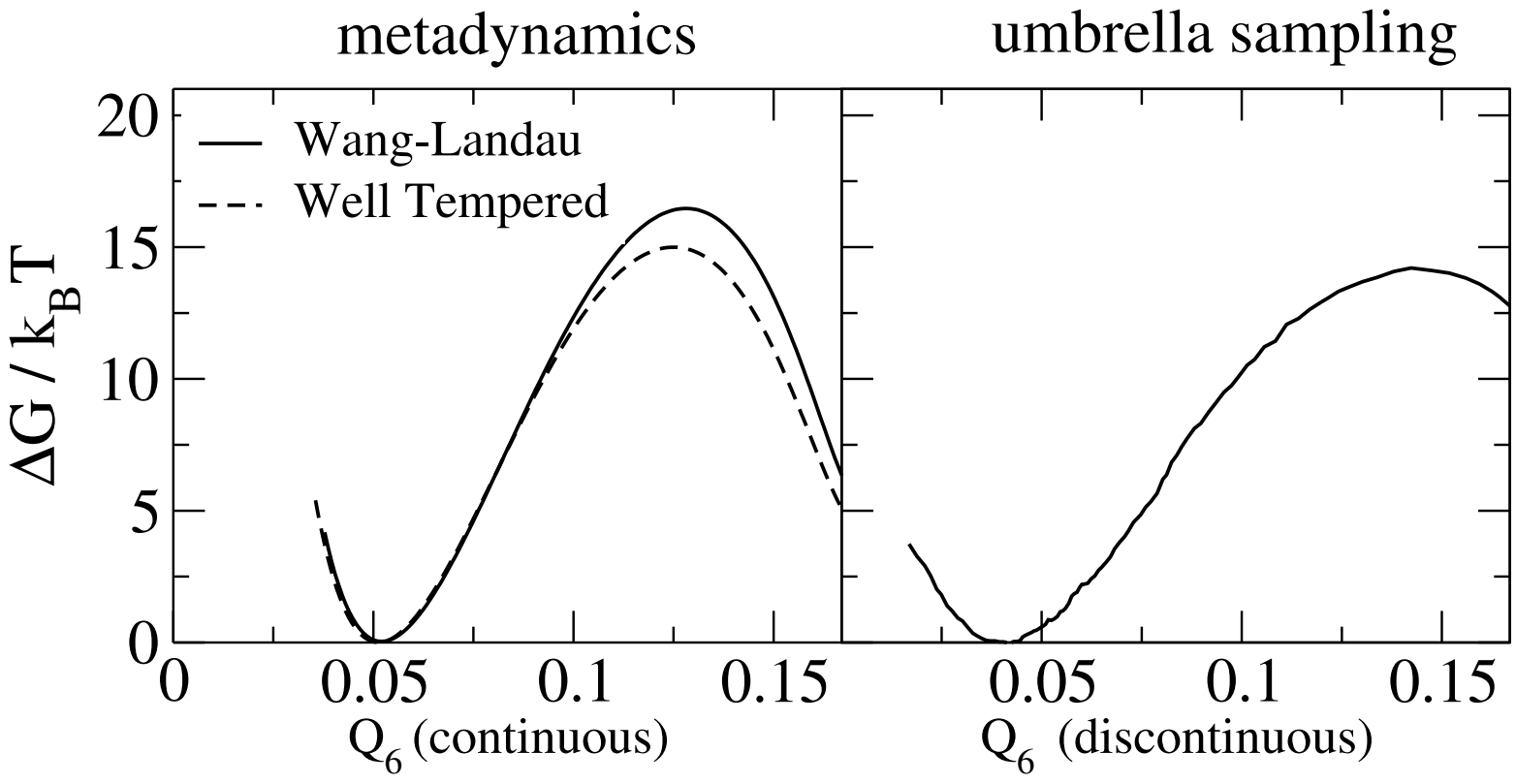


t=1.25ns



t=1.5ns


Wang-Landau
Well Tempered
1
0.1
0.01
w / k_B T
0
1000
2000
3000
4000
5000
6000
Metadynamics step


metadynamics
umbrella sampling
Wang-Landau
Well Tempered
ΔG / k_B T
Q_6 (continuous)
Q_6 (discontinuous)
0
5
10
15
20
0.05
0.1
0.15

A metadynamics based approach to sampling crystallisation events

D. Quigley and P. M. Rodger

Dept. of Chemistry and Centre for Scientific Computing,

University of Warwick,

Gibbet Hill Road, Coventry,

United Kingdom, CV4 7AL.

*

(Dated: November 14, 2008)

Abstract

We discuss the practicalities of applying the metadynamics method to sampling crystallisation events in molecular systems. Suitable choices for collective coordinates are presented along with criteria for their parameterisation. Issues arising from finite-size effects are discussed with particular reference to the generation of multiple clusters when biasing global order parameters. We also consider the applicability of two methods for enhancing the accuracy of the reconstructed free energy landscape. The discussion is illustrated with example data from freezing in the Lennard-Jones and ice-water systems.

*Electronic address: D.Quigley@warwick.ac.uk

I. INTRODUCTION

A range of problems in manufacturing, chemical processing, geology and biology involve crystallisation events. Molecular simulation of these processes is not, however, feasible using conventional methods. Often the process of interest takes place where the free energy preference for crystallisation is small (e.g. weak subcooling of a melt) and the energy barrier to forming a critical nucleus from which a bulk crystal can grow is substantial. The chance of such an event occurring within the computer time available for conventional molecular simulations methods is therefore vanishingly small. Routine simulation of crystallisation therefore requires specialist methods.

The method of Frenkel and coworkers [1, 2] has been applied to crystallisation in a variety of systems. In this approach, umbrella sampling [3] Monte-Carlo (MC) is used to reconstruct the free energy as a function of a suitably chosen reaction coordinate, or order parameter, typically the Q_6 order parameter of Steinhardt [4]. While successful, this approach can be problematic. For example, sampling the transition state requires a biased sampling of configurations corresponding to the critical nucleus; while the system is restrained at the corresponding order parameter, the nucleus may anneal to a lower energy structure, potentially preventing selection of metastable polymorphs. From a practical standpoint, implementations of umbrella sampling are generally restricted to various bespoke codes that are not usually employed outside of the author's own research group.

The increasingly popular metadynamics [5] method has recently been adapted for crystallisation studies [6–8]. This has some advantages over the umbrella sampling method, particularly in the case of multidimensional free energy landscapes. In particular crystallisation is not restricted to occur on an adiabatic free energy surface and so polymorph selection is more readily accessible. Metadynamics also has the advantage of naturally fitting into the framework of a molecular dynamics code, facilitating its use with a range of standard force-fields and simulation algorithms. We have recently developed an implementation of the metadynamics method with emphasis on crystallisation. This is based on versions 2.19 (replicated data, functional parallelism) and 3.09.3 (distributed data with parallel domain decomposition) of the versatile DL_POLY [9, 10] simulation package and is therefore potentially useful or adaptable for the study of crystallisation in a variety of molecular systems.

In this paper we describe our implementation. The metadynamics method is reviewed

in section II. In section III we describe the various order parameters which our implementation uses for collective coordinates. Finite size effects are discussed in section IV while computational details and parameterisation of the scheme are considered in section V. Two methods for improving the accuracy of free energy estimates are discussed in section VI.

II. METADYNAMICS

Here we briefly review metadynamics (metaD) as applied in our implementation. For a more detailed description and justification of the method we refer the reader to refs [11] and [12] and in particular to chapter 10 of ref [13]. We employ the 'direct' variant of the metadynamics scheme, although we note that 'discrete' metadynamics using the density as a collective coordinate has been applied to crystallisation by Prestipino and Giaquinta [14].

Metadynamics is used to map a free energy landscape as a function of M collective variables, or order parameters, which we represent as the vector $\mathbf{s}\left(\mathbf{r}^N\right)$. A history dependent bias potential $V\left[\mathbf{s}\left(\mathbf{r}^N\right)\right]$ is used to drive the system away from previously visited values of the collective variables. The bias ultimately pushes the system over free energy barriers and into previously unexplored local minima. In the 'direct' variant of metadynamics [11] the history dependent potential augments the Hamiltonian of the system without being coupled via an extended system variable.

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + U\left(\mathbf{r}^N\right) + V\left[\mathbf{s}\left(\mathbf{r}^N\right), t\right] \tag{1}$$

The force on each particle is then modified by the bias potential directly

$$\mathbf{f}_i = -\nabla_{r_i} U\left(\mathbf{r}^N\right) - \sum_{j=1}^{M} \frac{\partial V}{\partial s_j} \nabla_{r_i} s_j\left(\mathbf{r}^N\right) \tag{2}$$

This bias potential may be 'grown' by adding a Gaussian of height w and width δh, centred on the current collective variables, at periodic intervals of time τ_G,

$$V\left[\mathbf{s}\left(\mathbf{r}^N\right), t\right] = w \sum_{k=1}^{N_G} \exp\left[\frac{-\left|\mathbf{s}\left(k\tau_G\right) - \mathbf{s}\left(t\right)\right|^2}{2\delta h^2}\right] \tag{3}$$

where the integer k runs over all $N_G = \mathrm{int}\left(t/\tau_G\right)$ previously deposited Gaussians. Provided the deposition rate w/τ_G is sufficiently slow, the motion of the collective variables $\mathbf{s}$ is adiabatically separated from that of the molecular dynamics. In the limit of long simulations

times the bias potential counter balances the underlying free energy landscape so that the free energy surface can easily be recovered as

$$F_G(\mathbf{s}) = -\lim_{t\to\infty} V\left[\mathbf{s}\left(\mathbf{r}^N\right), t\right]. \tag{4}$$

The accuracy of this estimate is discussed in detail by Laio et al. [11] and is dependent on the deposition rate and the diffusion constant associated with movement of the collective variables. This error is typically of order w. Further improvement in accuracy are discussed in section VI.

III. ORDER PARAMETERS

.

To generate crystallisation events with metadynamics we must first define a suitable set of collective variables. As with any metadynamics simulation the choice of collective variables is key. In the case of crystallisation our collective variables are order parameters which distinguish between the disordered (amorphous or liquid) state and one or more crystal structures. For example the Q_6 function of Steinhardt et al. [4] has been extensively used in biased Monte-Carlo studies of Lennard-Jones and soft-sphere systems. As a single order parameter, this has the advantage of not separating the metastable bcc polymorph from the close-packed fcc structure known to be energetically preferable, and so may be used to bias crystallisation without biasing polymorph selection. This does however carry the disadvantage of making it difficult to probe the factors that determine this polymorph selection.

In the general case it is desirable to use multiple order parameters so that as few constraints as possible are placed on the reaction pathway. This is the approach we adopt. One may wish to distinguish between different crystalline forms and resolve separate pathways from the liquid to each of the available polymorphs, or indeed between polymorphs. This cannot be accomplished with a single collective variable where all such paths would be superimposed. It can also be expected that increasing the number of collective variables (provided that they are mutually independent) will improve the resulting description of the crystallisation pathway and free energy landscape. Unfortunately a significant drawback of the metadynamics method is the exponential increase in simulation time required to fully

explore landscapes of increasing dimension. In practice this restricts the number of tractable order parameters to a handful, although various methods are available to alleviate this cost [15, 16].

In order to facilitate the selection of an appropriate set of order parameters we have developed the following protocol. Having identified a set of candidate order parameters we first plot the equilibrium distribution of these from standard molecular dynamics (MD) simulations in the disordered state and any accessible crystalline polymorphs. Sets for which these distributions overlap are rejected and alternate sets are investigated until a suitable set with which to describe the known states is found. We stress that this criterion ensures only that the relevant structures are realisable and distinct within the collective variables space. Any routes found between these states are not guaranteed to be the dominant pathways of the unbiased system, but they will at least lead to upper bounds for the corresponding free energy barriers.

The following order parameters are available within our implementation. These have been successfully used in various combinations to crystallice ice 1 from supercooled liquid water [7], and calcium carbonate nanoparticles in water [8]. Other order parameter functions can easily be incorporated into the program provided routines for evaluating the corresponding gradient and stress tensor contributions are available.

A. Steinhardt order parameters

We define a smoothly varying version of the Steinhardt order parameter to ensure continuity of energy and force as required within a MD-based code. This can be computed for all combinations of atom type α with atom type β as

$$Q_l^{\alpha\beta} = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^{l} \left| \frac{1}{N_c N_\alpha} \bar{Q}_{lm}^{\alpha\beta} \right|^2 \right]^{1/2}, \tag{5}$$

where

$$\bar{Q}_{lm}^{\alpha\beta} = \sum_{b=1}^{N_b} f_c\left(r_b\right) Y_{lm}\left(\theta_b, \phi_b\right) \tag{6}$$

The index b runs over all N_b vectors connecting atoms of type α to those of type β. The spherical harmonic is computed on the polar angles of each vector measured with respect to an arbitrary choice of reference axis. Contributions are restricted to short range by the

tapering function

$$f_c(r) = \begin{cases} 1 & \text{if } r \leq r_1; \\ \frac{1}{2}\left\{\cos\left[\frac{(r-r_1)}{r_2-r_1}\pi\right]+1\right\} & \text{if } r_1 < r \leq r_2; \\ 0 & \text{if } r > r_2. \end{cases}, \tag{7}$$

which decays smoothly from one to zero between r_1 and r_2. Choices for r_1 and r_2 are discussed in section V. Approximate correspondence between this order parameter and the discontinuous version typically employed in MC studies is achieved by setting the constants N_α and N_c to the number of atoms of type α included in the sum, and the average number of β atoms within r_2 of each α respectively. Note that the order parameter is not scale invariant: increasing the system size, and hence N_α, introduces a shift in the order parameter. Comparison of values between different system sizes should be made by scaling Q_l by $\sqrt{N_\alpha/N_\alpha^{ref}}$.

Forces arising from biasing a Steinhardt order parameter can easily be decomposed into pairwise additive contributions from each of the spherical harmonics. For the vector $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, the resulting force is

$$\mathbf{f}_{ij} = -\hat{\mathbf{r}}_{ij}\frac{\partial V}{\partial Q_l^{\alpha\beta}}\frac{1}{Q_l^{\alpha\beta}}\frac{4\pi}{2l+1}\left(\frac{1}{N_c N_\alpha}\right)^2 \sum_{m=-l}^{l}\left\{\Re\bar{Q}_{lm}^{\alpha\beta}\frac{d}{dr_{ij}}\left[f_c(r_{ij})\,\Re Y_{ml}(\theta_{ij},\phi_{ij})\right] + \Im\bar{Q}_{lm}^{\alpha\beta}\frac{d}{dr_{ij}}\left[f_c(r_{ij})\,\Im Y_{ml}(\theta_{ij},\phi_{ij})\right]\right\}. \tag{8}$$

The real ($\Re$) and imaginary ($\Im$) parts of equation 6 are stored separately during computation of the order parameters and re-used as above when a second pass though all relevant pairs is made to compute forces. Crystallisation often involves a substantial density change and hence it is essential that bulk metadynamics simulations are performed at constant pressure. When the forces are decomposed in this fashion the required contributions to the stress tensor $\underline{\underline{\sigma}}$ from each pair are trivially computed as

$$\sigma_{ab} \rightarrow \sigma_{ab} - f_{ij}^a r_{ij}^b \tag{9}$$

in the usual way.

Computation of the Steinhardt order parameters is accelerated by making use of the existing DL_POLY neighbour list. This assumes that contributions from excluded atom pairs are not required, and that r_2 is less than the cut-off radius for non-bonded interactions.

B. Tetrahedral order parameters

As has been discussed by Radhakrishnan and Trout [17, 18], the tetrahedral parameter of Chau and Hardwick [19] provides a means of including second-nearest neighbour effects into measurements of order. We define a continuous version of this order parameter for angles involving atoms of the same species α

$$\zeta_\alpha = \frac{1}{N_c N_\alpha} \sum_{i=1}^{N_\alpha} \sum_{j \neq i}^{N_\alpha} \sum_{k>j}^{N_\alpha} f_c\left(r_{ij}\right) f_c\left(r_{ik}\right) \left(\cos\theta_{jik} + 1/3\right)^2 . \tag{10}$$

Here the indices i, j and k run over all atoms of species α while N_c and N_α retain their previous meanings. The value of ζ is maximised for perfect tetrahedral networks. It should therefore be considered for crystallisation of solids in which particular species are arranged in this fashion. A set of these order parameters $\{\zeta_\alpha\}$ can be also be defined with each member biased as an independent order parameter.

As with any such three body interaction, the bias forces arising from each triplet i, j, k can be decomposed into that between i and j and between i and k.

$$\begin{aligned}
\mathbf{f}_{ij} &= -\frac{\partial V}{\partial \zeta_\alpha} \left\{ \frac{2}{r_{ij}} \left(\cos\theta_{jik} + 1/3\right) f_c\left(r_{ij}\right) f_c\left(r_{ik}\right) \left(\hat{\mathbf{r}}_{ik} - \hat{\mathbf{r}}_{ij} \cos\theta_{jik}\right) + \left(\cos\theta_{jik} + 1/3\right)^2 \frac{df_c\left(r_{ij}\right)}{r_{ij}} f_c\left(r_{ik}\right) \hat{\mathbf{r}}_{ij} \right\} \\
\mathbf{f}_{ik} &= -\frac{\partial V}{\partial \zeta_\alpha} \left\{ \frac{2}{r_{ik}} \left(\cos\theta_{jik} + 1/3\right) f_c\left(r_{ij}\right) f_c\left(r_{ik}\right) \left(\hat{\mathbf{r}}_{ij} - \hat{\mathbf{r}}_{ik} \cos\theta_{jik}\right) + \left(\cos\theta_{jik} + 1/3\right)^2 \frac{df_c\left(r_{ik}\right)}{r_{ik}} f_c\left(r_{ij}\right) \hat{\mathbf{r}}_{ik} \right\}
\end{aligned} \tag{11}$$

The contribution to the stress tensor arising from biasing these order parameters can then be computed as for other three body interactions as

$$\sigma_{ab} \rightarrow \sigma_{ab} - f_{ik}^a r_{ik}^b - f_{ij}^a r_{ij}^b \tag{12}$$

Again the DL_POLY neighbour lists are used in computing these order parameters. In the parallel decomposition of DL_POLY 3 each processor must compute these forces for all atoms within its own subdomain, including those arising on atom k from triplets in which the central atom i and its neighbour j are in a neighbouring domain. It is possible for values of r_2 greater than half the simulation cut-off that j will not lie in the halo of atomic positions stored on the current subdomain. For simplicity our implementation restricts choices of r_2 such that this cannot occur.

C. Potential Energy

The use of the potential energy as a biasing order parameter was first demonstrated by Donadio et al. [20] and has since been employed in a study of Lennard-Jones nucleation [6], and by the present authors in the freezing of water [7] and crystallisation of calcium carbonate nanoparticles [8].

The potential energy U is a smooth and continuous function of all coordinates and is likely to take distinctly different values across multiple crystal structures and disordered states. It is therefore an ideal order parameter. Compared with the above order parameters it is functionally very complex and expensive to compute, although this is of no consequence since both its value and gradient are already computed at every molecular dynamics step. Forces and stress tensor contributions resulting from biasing this order parameter are also trivial to compute,

$$\begin{aligned} \mathbf{f}_i &\rightarrow \left(1 + \frac{\partial V}{\partial U}\right) \mathbf{f}_i \\ \underline{\underline{\sigma}} &\rightarrow \left(1 + \frac{\partial V}{\partial U}\right) \underline{\underline{\sigma}} \end{aligned} \tag{13}$$

Defining a *local* potential energy can be advantageous and requires somewhat more thought. For example, when studying crystallisation of a mineral nanoparticle in a large volume of water, biasing of the global potential energy will generally result in uninteresting re-arrangement of the water solvent. For such applications our implementation defines a local potential energy comprising contributions from specific atom types only, namely those involved in computation of any other order parameter (e.g. Steinhardt or tetrahedral). Any bond stretching, angle bending or long range Van der Waals interactions which include at least one atom involved in another order parameter are included. In our implementation we have chosen to include only the real space pairwise contributions to the electrostatic energy, as the reciprocal space part of the Ewald summation cannot easily be separated into local contributions. Use of the local potential energy requires separate accumulation of the force and stress contributions arising from the relevant interactions which replace $\mathbf{f}_i$ and $\underline{\underline{\sigma}}$ in equation 13 above.

Due to the essentially 'free' nature of the potential energy order parameter, it is recommended that it be tested for suitability in all applications. However we have avoided its use

in some of the examples which follow to maintain consistency with previous results.

IV. FINITE SIZE EFFECTS

In the case of a bulk first-order phase transition, crystallisation proceeds via a process of nucleation and growth. Computation of reliable free energy barriers to nucleation by metadynamics or any other method requires some consideration of finite size effects. It must be stressed that the size of the critical nucleus, which must be correctly sampled to measure the free energy barrier, increases closer to coexistence. Any attempts to compute barriers from simulations too small to accommodate the relevant critical nucleus size will be inaccurate.

A. Constant density effects

As already stated, for crystallisation events which involve a density change, simulations which sample the isothermal-isobaric ensemble are essential. The consequences of simulating at constant density have recently been investigated by Wedekind et al. [21]. If generating crystallistion events at constant density, any increase (or decrease) in density of the crystallised region must be compensated by an unphysical change of density in the surrounding medium. The severity of this change will be determined by the overall volume of the system and will ultimately become significant in all systems as the crystallised region becomes comparable in size to the simulation cell.

The consequences of this density change can be understood in terms of classical nucleation theory. Here the rate at which new material 'attaches' to a growing crystallite is known from kinetic theory to be a function of the density in the surrounding medium. If this density is continually changing as the crystal grows then the attachment rate will be substantially altered leading to unphysical growth mechanisms. We have previously suggested [7] that this effect is responsible for the two-stage growth process observed by Matsumoto et al. [22] in their 'brute-force' simulations of ice growth.

B. Periodic boundary effects

For small systems, the critical nucleus can often span the periodic domain. The surface area of this nucleus, and hence the interfacial free energy is minimised by forming an infinite cylinder rather than the expected closed spherical nucleus, leading to an underestimate of the free energy barrier as discussed in ref [7]. For systems above a threshold size, the surface area (per simulation cell) of a sphere with the same volume becomes smaller and the expected picture from nucleation theory is recovered. We should emphasise that the critical nucleus is not always spherical and indeed is not unique. Its actual shape will depend on anisotropy in the interfacial free energy as well as the rate at which new material attaches to the growing cluster. However, it is, in principle possible, to estimate this threshold system size and ensure that the simulated system is sufficiently large.

In figure 2 we present free energy barriers as a function of Q_6 for a series of system sizes in order to demonstrate this dependence. Each barrier is computed from a series of N Lennard-Jones particles at a reduced temperature of 0.92 and reduced pressure of 5.76. An example of a cylindrical domain spanning crystallite belonging to the transition state ensemble is shown in figure 1. It should be stressed that the freezing curve of the Lennard-Jones fluid is highly sensitive to truncation of the pair potential. We truncate at $r_c = 3.5\sigma$ to ensure consistency with the coexistence curve of Agrawal and Kofke [23]. The degree of sub-cooling at this temperature is approximately 17 %, however these results cannot be directly compared to other studies which employ shorter r_c [24–26]. Each barrier is computed using a combination of coarse metadynamics ($\delta h = 0.006$, $w = 7.5\, k_B T$) and umbrella sampling Monte-Carlo initialised using configurations from the metadynamics trajectory. The umbrella sampling procedure employed 15 overlapping windows across the range of Q_6. Simulations of $120,000$ MC sweeps were conducted in each window. We stress again that the size of the critical cluster increases dramatically as coexistence is approached. The influence of this finite size effect will therefore be much greater at weaker supercoolings.

Using an estimate of the critical nuclear size from Trudu et al. [6] we calculate the threshold system size for spherical critical nucleus formation as approximately $1,200$ particles under these conditions. Lack of convergence beyond this size is a consequence of generating *multiple* crystallites and is discussed in the following section.

C. Multiple crystallite generation

In a system where the penalty associated with forming an order-disorder interface is high (e.g. liquids under moderate supercooling) the probability of multiple large nuclei forming within a small volume is expected to be negligible. However it has long been recognised [24] that biasing a *global* order parameter such as Q_6 can lead to the growth of multiple crystalline regions within the simulated system. This is inconsistent with the classical nucleation theory of nucleus formation and warrants some discussion.

The origins of the effect have been analysed in detail by Tenwolde et al. [24] in the context of freezing. At a particular solid fraction χ the translational entropy gain of forming multiple solid clusters competes with the energy cost of forming multiple interfaces. Below a particular solid fraction χ_c this entropic gain dominates. Therefore any bias imposed to promote global order will initially result in growth of several crystallites. Tenwolde et al. showed that the value of χ_c varies with the inverse fourth power of the system volume and hence for small systems may artificially exceed the solid fraction corresponding to a single critical nucleus in a given simulation. This results in a transition state ensemble containing a critical cluster plus multiple smaller crystallites. A typical distribution of these solid clusters is illustrated in figure 1. Metadynamics (or umbrella sampling) will therefore *overestimate* the free energy barrier. This error is often associated with the choice of a global order parameter but is in fact demonstrated to be a finite size effect (which can be artificially *circumvented* with the use of local order parameters - see below) by the analysis of Tenwolde et al.. In principle it is possible to remove the effect altogether by simulating sufficiently large systems (i.e. several orders of magniture larger that those in figure 2) although this has not been explicity demonstrated to our knowledge.

Further examination of the expression for χ_c reveals that the decrease with respect to order-disorder interfacial free energy, γ, is exponential and can easily dominate over the entropic gain of forming multiple clusters. If γ is large this finite-size effect will therefore be very much reduced. For this reason we suggest that problems associated with biasing global order are greatly exaggerated by the Lennard-Jones system. In contrast, both our [7] and other work [17, 27] on nucleation of ice 1, show formation of a *single* nucleus when biasing global order parameters. To clarify this issue we have simulated nucleation trajectories for the freezing of liquid water to ice 1 in considerably larger systems than those previously

reported. These simulations follow the procedure given in ref [7] and are parameterised according to the criteria detailed in section V. In figure 3 we plot the total number of solid particles and the number of particles within the largest cluster during the first formation of a critical nucleus. The system size is 2496 molecules. In this particular case the biased order parameters are Q_4^{OO}, Q_6^{OO}, ζ_O and the potential energy. The cluster itself is shown in figure 4. To within the noise inherent in the identification process, the increase in solid fraction is entirely due to growth of the largest cluster and we do not see formation of multiple crystallites. Analysis of still larger simulations is underway to confirm this.

For the purposes of *nanoparticle* crystallisation we stress that the volume accessible to ordered regions within such a particle is exactly reproduced within the simulation and hence any propensity to form multiple crystalline regions within such a particle would be entirely physical, although we have yet to observe any examples of such multiple nucleation in our simulations.

Nethertheless for simulations of bulk crystallisation, one may wish to *ensure* multiple cluster generation is prevented by biasing *local* order parameters. This circumvents the finite size effect in χ_c. Two methods have been used in the literature to enforce this locality:

1. As an alternative to biasing order parameters, the size of the largest solid cluster is biased [28].

2. The calculation of order parameters is restricted to a subset of the system. This subset can be defined as a geometrical region (usually a sphere of fixed radius) [29, 30] or the N_{sub} nearest neighbours of a randomly selected particle [6].

The first approach has proven very effective for spherical particles. However, an unambiguous method for identifying solid clusters in the general case is lacking and this method effectively precludes growth from clusters by aggregation which could conceivably provide a lower free energy route to formation of the precritical nucleus. The effect of enhanced cluster entropy might therefore be overcompensated. Within an MD simulation, this method might also create anomalies if natural fluctuations change the identity of the largest cluster. While discontinuities in the forces could be avoided by smoothing, this would introduce unphysical non-local forces into the simulation.

Provided one is able to choose a suitable subset, and that the size of this subset is just sufficient to favour critical nucleus formation, the second method can be effective in com-

puting free energy barriers. The problem of critical nuclei coexisting with smaller clusters is removed; however multiple smaller clusters are still generated (with artificially high probability) within the subset when close to the liquid regime. We note that the definition of a suitable subset becomes more difficult in heterogeneous nucleation or when chemical additives are present. In these cases all choices are no longer equivalent.

To highlight the difference between these approaches we draw attention to the region of positive curvature in the free energy profile at low solid fraction seen in figure 1 of ref [29]. This profile was computed by biasing a spherical subset of molecules. This positive curvature is also visible in figure 2 of this paper and other studies using global order parameters [25, 31]. No such region is observed in studies which bias the size of the largest cluster [28]. Within a classical nucleation theory (CNT) description, it is possible to determine the criteria required for this positive curvature. The free-energy G of forming a solid volume V is

$$G\left(V\right) = \gamma A\left(V\right) - \Delta\mu V \tag{14}$$

$A\left(V\right)$ is the total surface area of the solid region(s), γ is the free-energy penalty per unit area of forming the solid-liquid interface and $\Delta\mu$ is the free-energy reduction per unit volume of solid. Within this description a positive curvature in G can only be obtained for one of the following reasons:

1. Negative value of γ.
2. Solid region(s) for which A/V increases with volume.
3. Solid region(s) for which A decreases with V.

Possibilities 1 and 2 can be discounted as unphysical. The third possibility can however be realised by the aggregation of multiple growing crystallites. We can therefore see that biasing order within a subset merely removes the problem of critical nuclei coexisting with smaller clusters, with any artificially high entropic favourability of forming multiple *precritical* clusters unaffected. It this context it is interesting to note that in a study of Lennard-Jones nucleation using transition path sampling (TPS) techniques [26] (which do not bias an order parameter) no such region of positive curvature was observed. However, results for the barrier height are in agreement with simulations which biased a global Q_6 at the same temperature and pressure [31].

Of the two local biasing schemes, the subset method is most easily incorporated within our implementation. Other methods of avoiding the multiple cluster problem are also under investigation.

V. PARAMETERISATION

In this section we discuss the approach taken in choosing the various parameters required for efficient operation of metadynamics for sampling crystallisation.

A. Tapering function

Before computing equilibrium order parameters, the tapering function (equation 7) must be parameterised by defining r_1 and r_2. These two radii must be chosen carefully. Poor choices can result in forces dominated by 'bond'-stretching [38] when the system is biased. To minimise this effect we define two criteria. Firstly, the range $r_1 \rightarrow r_2$ should correspond as closely as possible to minima in the relevant pair correlation function for the disordered and ordered states. This ensures any spurious bond stretching is eliminated from the equilibrium states and for most cases restricts computation of order parameters to nearest neighbours only. As discussed by Radhakrishnan and Trout [18] the influence of second nearest neighbour effects is implicitly included in a tetrahedral order parameter, whereas the potential energy order parameter included contributions from all neighbours upto the simulation cutoff radius. The second criteria maximises $r_2 - r_1$ within the limits of the first. This minimises any residual spurious forces

It should be noted that this parameterisation should be performed based on equilibrium simulations at the temperature and pressure at which the crystallisation is to be sampled. Application to higher pressures in particular may shift neighbour distances significantly and warrant a revision of these parameters.

B. Preconditioning

The number of Gaussians needed to fill the initial basin can be considerably reduced by a simple preconditioning of the problem. Each order parameter is scaled by a constant factor

such that the initial basin of attraction is approximately spherical in the collective variable space. This becomes particularly important when employing many collective variables, or when fluctuations in the collective variables differ by several order of magnitude.

The scaling constants are identified by monitoring the order parameters during an equilibrium simulation in the disordered state. The distribution of samples for each order parameter is then approximated by a Gaussian distribution of width σ. Each collective variable is then scaled by a constant $1/\sigma$ throughout the metadynamics simulation.

C. Gaussian deposition rate and width

As discussed in some detail in ref [11] the accuracy of a metadynamics calculation is determined by the properties of the history dependent bias potential. In particular the deposition rate w/τ_g and width δ_h of the Gaussian augmentations should be chosen carefully. Narrow Gaussians can be expected to reconstruct a more detailed free energy landscape, but require longer simulation times to saturate the available minima. Wider Gaussians will fill basins rapidly, but may simultaneously increase the effective height of barriers leading to overfilling and a reduction in efficiency.

A natural choice when using collective variables scaled as above is to set $\delta h = 1$. This then corresponds to Gaussians with a width in each direction matching the equilibrium fluctuations. If, however, the curvature of the initial basin is small (as if often the case for liquids), this choice quickly becomes inefficient as the system explores higher energies and wider fluctuations. Choices of $\delta h = 2$-3 offer a reasonable compromise, performing an efficient exploration of space at the expense of some lost detail at the basin minimum (which is easily explored with equilibrium simulations). We have experimented with a scheme using adaptive Gaussian widths based on fluctuations during the previous metadynamics step. The necessary averaging required a large interval between depositions which offset any gain in efficiency. We note that in this limit the metadynamics method effectively reduces to the adaptive umbrella sampling scheme described by Mezei [32].

Choosing an optimal deposition rate can be difficult without an *a-priori* estimate of the energy scales involved in the problem. Typical energy barriers to crystallisation can be estimated from classical nucleation theory provided data is available on bulk and interfacial free energies. These can range from a few tens to several hundreds of k_BT in magnitude.

If no such data is available one must resort to performing coarse simulations with large Gaussians to obtain an initial estimate before repeating with a more appropriate choice. In our simulations we employ Gaussian with $w \approx 1-5\%$ of the largest expected barrier height. Having made this choice we identify an appropriate deposition interval by monitoring the evolution of the collective variables after deposition of test Gaussians at random positions near equilibrium. The time taken to stabilise at new values is then used as the deposition interval. In our simulations this typically corresponds to a few hundred simulation timesteps.

VI. ACCURACY AND CONVERGENCE

A common criticism of the metadynamics method is lack of accuracy. If the choice of collective variables is not optimal, or if the deposition rate is large, considerable overfilling of basins can occur before the simulation finds an escape pathway. If the simulation is terminated before this overfilling is compensated an undesirable bias is introduced into the reconstructed free energy landscape. The simulation should therefore only be terminated when rapid diffusion between basins is identified. In this limit the underlying free energy landscape is exactly (to within the error computed in ref [11]) compenstated by the bias potential and the estimate is unbiased.

In the case of crystallisation the transition pathway involves complex rearrangement over many degrees of freedom. The time associated with transitions can be very long even in the limit of a perfectly flat landscape. Identifying an optimal time to terminate the simulation can therefore be difficult. In refs [7] and [8] we have erred on the side of caution and terminated the simulation only after establishing that the proceeding several nanoseconds of metadynamics produced only a shift in the free energy landcape to within an accuracy suitable for the energy scale of the problem. For example in ref [7], the simulations were terminated only after observing that both the height of the nucleation barrier and the relative depth of the two minima changed by no more than $2k_BT$ (approximately 5% of the barrier height) over 15 ns. If further accuracy is required then it may be prudent to use configurations from the metaD trajectory to initialise a one dimensional umbrella sampling along the reaction pathway identified by metadynamics. Such an approach has been adopted for biomolecular simulations by Ensing et al. [33]. Similarly candidate metaD pathways can be used to seed TPS calculations which might be particularly powerfull where pathways

obtained on the adiabatic free energy surface are not the most probable members of the transition path ensemble.

Recently various simple methods have been suggested to improve the accuracy of metadynamics itself. We have implemented two of these in the context of crystallisation and discuss their merits below.

A. The Wang-Landau recursion approach

This approach has been proposed by Min et al. [34]. Here the metadynamics simulation proceeds as normal, initially using a large Gaussian height w_1. A histogram is accumulated which tracks the number of visits to each square of a grid in the collective variable space. One this histogram is approximately flat (the authors suggest a criterion that all values are $> 80\,\%$ of the mean) the height of all subsequent Gaussians is updated as $w \to w/2$. The histogram is reset and the process is repeated. At the j step the accuracy of the recovered free energy landscape is increased to order $w_1/2j$ and will ultimately reach any desired accuracy provided the flatness criterion is appropriate. The scheme is therefore similar in spirit to the Wang-Landau sampling method [35].

As with any Wang-Landau based method, a suitable domain for the histograms must be defined which covers all areas of physical interest. In the case of a multidimensional landscape of order parameters this can be be somewhat difficult. In particular a square grid spanning the accessible range of all order parameters will include *combinations* which are not physically realisable. Furthermore the accessible regions are not known *a-priori*, making definition of a suitable grid problematic. Our implemention of the recursion approach therefore uses a different criterion for updating the Gaussian heights if more than one collective variable is used. We periodically monitor the evolution of order parameters and identify the time t_c after which several barrier crossing events have occurred. The Gaussian heights are then updated at multiples of t_c and we carefully check that further crossings occur within these intervals.

This approach has been taken in a recent study of protein controlled nanoparticle crystallisation [36]. In this case the large energy scale of the problem ensures that only one or two reductions in height are required before the resulting landscape does not change significantly.

B. Well-tempered metadynamics

An alternative and even simpler method to improve the convergence of metadynamics calculations has been proposed by Barducci et al. [37]. Rather than controlling the height of Gaussians based on multidimensional spatial criteria, the well-tempered metadynamics method uses a maximum energy criterion. A threshold energy V_{max} is defined which lies above the height of the largest anticipated barrier. In the case of crystallisation this can be estimated from nucleation theory or from a coarse metadynamics simulation. At each step, the Gaussian deposited is of height

$$w = \exp\left[-V_{aug}\left(\mathbf{s}\right)/V_{max}\right]. \tag{15}$$

In this manner the augmentations become smaller as the threshold is approached. The risk of overfilling is therefore substantially reduced. Once all basins accessible below the threshold level have been filled, further simulation will act to smooth any roughness in the reconstructed landscape without large shifts in the energy.

C. Comparison

To demonstrate the accuracy of these methods we have recomputed the free energy barrier to nucleation for the 500 particle Lennard-Jones fluid as a function of the Q_6 order parameter. Even in this trivially simple system, previous metadynamics calculations of the barrier (figure 2) required refinement with umbrella sampling to achieve a useful accuracy. For both the Wang-Landau recursion and well tempered schemes we employ an initial Gaussian height of $w = 3\,k_BT$ and a width δh equal to the one standard deviation in Q_6 at equilibrium. We have deliberately chosen w to be very large in this case in order to demonstrate ultimate convergence of the two schemes, regardless of this poor initial choice. Each simulation employed a deposition interval of 4 ps.

As we are interested only in the barrier to nucleation and not in the relative free energies of the two states (which can be readily found in the literature) we add an artificial energy penalty to all states with $Q_6 > 0.2$. This takes the form of a wall which smoothly increases from 0 to $1000\,k_BT$ over an interval of 0.025. This restricts the simulation to the interesting region and removes the considerable waiting time involved in saturating the crystalline free

energy minimum. The reconstructed free energy landscape is therefore only valid for $Q_6 < 0.2$.

In the case of well-tempered metadynamics we set $V_{max} = 30\,k_BT$, i.e. sufficient to allow transitions to the crystalline state. For the Wang-Landau recursion scheme we use the stardard flatness criterion modified to disregard the histogram bin adjacent the artificial wall. Both simulations are terminated after 6000 metadynamics steps. Heights of Gaussians deposited at each step are plotted in figure 5. In this case the Wang-Landau approach is most effective in reducing the height of the Gaussians. Alternative (lower) choices of V_{max} may increase the competitiveness of the well tempered metadynamics scheme; however this choice is not easily made in advance of a simulation.

Free energy barriers computed from these two simulations are compared to the umbrella sampling result in figure 6. Note that the umbrella sampling result is not precisely comparable due to the need for a continuous definition of Q_6 in metadynamics, whereas discontinuous definitions are normally used in umbrella sampling MC studies. Each profile is adjusted such that the liquid minimum is at zero free energy. In this case, the well tempered result is liable to alter on a scale of $0.1\,k_BT$ with further simulation whereas the Wang-Landau result is considerably more converged. Both are within $1\,k_BT$ of the umbrella sampling result which we regard as acceptable agreement within the differing definitions of Q_6.

VII. CONCLUSIONS

We have developed an implementation of the metadynamics method applied to crystallisation with broad applicability. The metadynamics method is ideally suited to circumventing the long-timescale problems associated with generating crystallisation trajectories. As well as providing seed trajectories for umbrella sampling or TPS calculations it is capable of recovering the free energy landscape to arbitrary accuracy when combined with the methods discussed in section VI. Of these methods the Wang-Landau recursion scheme is most effective at improving accuracy in one dimension. The well tempered metadynamics scheme is likely to be most effective in higher dimensions.

As with all methods for sampling crystallisation in the bulk, metadynamics suffers from finite size effects which can lead to unreliable estimates for the free energy barrier to nucleation, and unphysical growth mechanisms. Our implementation of the scheme within

DL_POLY 2 and 3 has potential to simulate systems sufficiently large to overcome these effects and is applicable to a range of materials. We have also demonstrated that problems associated with the generation of multiple crystallites do not need to be circumvented with the use of local order parameters in the case of large solid-liquid interfacial free energy.

VIII. ACKNOWLEDGEMENTS

Computing facilities were provided by the Centre for Scientific Computing of the University of Warwick and by the HECToR service. Early access to this service is gratefully acknowledged. Funding was provided by the EPSRC under grants GR/S80127 and EP/F054785 (HECToR Capability Challenge). The authors would also like to thank Dr Martyn Foster and Dr Ilian Todorov for assistance with the DL_POLY 3 code on HECToR. We also thank Prof. Mike Allen for useful discussions.

[1] J. S. VanDuijneveldt and D. Frenkel, J. Chem. Phys **96**, 4655 (1992).

[2] R. M. Lynden-Bell, J. S. Vanduijneveldt, and D. Frenkel, Mol. Phys. **80**, 801 (1993).

[3] G. Torrie and J. Valleau, Chem. Phys. Lett. **28**, 578 (1974).

[4] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Phys. Rev. B. **28**, 784 (1983).

[5] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 562 (2002).

[6] F. Trudu, D. Donadio, and M. Parrinello, Phys. Rev. Lett. **97**, 105701 (2006).

[7] D. Quigley and P. M. Rodger, J. Chem. Phys. **128**, 154518 (2008).

[8] D. Quigley and P. M. Rodger, J. Chem. Phys. **128**, 221101 (2008).

[9] DL_POLY is a molecular dynamics simulation package written by W. Smith, T.R. Forester and I.T. Todorov and has been obtained from STFC's Daresbury Laboratory via the website http://www.ccp5.ac.uk/DL_POLY.

[10] W. Smith, C. W. Yong, and P. M. Rodger, Molec .Sim. p. 385 (2002).

[11] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, J. Phys. Chem. B. **109**, 6714 (2005).

[12] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. **99**, 6714 (2005).

[13] A. Laio and M. Parrinello, *Computer Simulations in Condensed Matter Systems: From Mate-*

rials to Chemical Biology Volume 1 (Springer Berlin / Heidelberg, 2006), vol. 703 of *Lecture Notes in Physics*, chap. 10, pp. 315–417.

[14] S. Prestipino and P. V. Giaquinta, J. Chem. Phys. **128**, 114707 (2008).

[15] S. Piana and A. Laio, J. Phys. Chem. B **111**, 4553 (2007).

[16] D. Branduardi, F. L. Gervasio, and P. M., J. Chem. Phys. **126**, 054103 (2007).

[17] R. Radhakrishnan and B. L. Trout, Phys. Rev. Lett. **90**, 1786 (2003).

[18] R. Radhakrishnan and B. L. Trout, J. Chem. Phys. **117**, 158301 (2002).

[19] P. L. Chau and A. J. Hardwick, Mol. Phys. **93**, 511 (1998).

[20] D. Donadio, P. Raiteri, and M. Parrinello, J. Phys. Chem. B **109**, 5421 (2005).

[21] J. Wedekind, D. Reguera, and R. Strey, J. Chem. Phys. **125**, 214505 (2006).

[22] M. Matsumoto, S. Saito, and I. Ohmine, Nature **416**, 409 (2002).

[23] R. Agrawal and D. A. Kofke, Mol. Phys. **85**, 43 (1995).

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

[25] M. Chopra, M. M., and J. de Pablo, J. Chem. Phys. **124**, 134102 (2006).

[26] D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, Phys. Rev. Lett. **94**, 235703 (2005).

[27] R. Radhakrishnan and B. L. Trout, J. Am. Chem. Soc. **125**, 7743 (2003).

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

[29] J. M. Leyssale, J. Delhommelle, and C. Millot, J. Chem. Phys. **122**, 104510 (2005).

[30] J. M. Leyssale, J. Delhommelle, and C. Millot, J. Chem. Phys. **122**, 184518 (2005).

[31] P. R. Tenwolde, M. J. Ruizmontero, and D. Frenkel, Phys. Rev. Lett. **75**, 2714 (1995).

[32] M. Mezei, J. Comp. Phys. **68**, 237 (1985).

[33] B. Ensing, A. Laio, M. Parrinello, and M. L. Klein, J. Phys. Chem. B **109**, 6676 (2005).

[34] D. Min, Y. Lui, and W. Carbone, I. Yang, J. Chem. Phys. **126**, 194104 (2007).

[35] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[36] D. Quigley, C. L. Freeman, P. M. Rodger, and J. H. Harding, in preparation.

[37] A. Barducci, G. Bussi, and M. Parrinello, Phys. Rev. Let. **100**, 020603 (2008).

[38] In this context we mean a purely geometrical connection vector rather than chemical bond.

FIG. 1: (a) Approximately cylindrical domain-spanning cluster identified as belonging to the transition state ensemble in a simulation of Lennard-Jones freezing at 17 % supercooling with 864 particles. (b) Coexistence of the largest solid crystallite with multiple smaller crystallites in a similar simulation of 10976 Lennard-Jones particles. In both cases particles are identified as solid according to the criteria described in ref [1]. Liquid-like particles are hidden.

FIG. 2: Free energy barriers to formation of a critical solid nucleus in the supercooled Lennard-Jones fluid at $T^* = 0.92$, $P^* = 5.76$.

FIG. 3: Total number of solid water molecules, and number within the largest cluster as a function of time minus the induction time (t_i) to the first nucleation event in a metadynamics simulation of 2496 TIP4P water molecules freezing at 180 K. Each value is plotted minus the average equilibrium (unbiased) background value upto the formation of a critical cluster. The total number of molecules in the simulation is 2496.

FIG. 4: Growth of a critical nucleus in a metadynamics simulation of 2496 TIP4P water molecules freezing at 180 K. Hydrogen bonds connecting molecules identified as solid are highlighted. Solid molecules are identified by the same method used in ref [7]. Times correspond to the x-axis in figure 3.

FIG. 5: Heights of Gaussians deposited during 6000 steps of metadynamics, simulating nucleation in a 500 particle Lennard-Jones system. Heights of Gaussians are controlled by the two schemes described in section VI.

FIG. 6: Free energy barriers to nucleation of the Lennard-Jones solid in a system of 500 particles at 17 % supercooling.