UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Reaction coordinates for the crystal nucleation of colloidal suspensions extracted from the reweighted path ensemble

Lechner, W.; Dellago, C.; Bolhuis, P.G.

DOI
10.1063/1.3651367

Publication date
2011

Document Version
Final published version

Published in
Journal of Chemical Physics

Link to publication

Citation for published version (APA):
Lechner, W., Dellago, C., & Bolhuis, P. G. (2011). Reaction coordinates for the crystal nucleation of colloidal suspensions extracted from the reweighted path ensemble. *Journal of Chemical Physics*, *135*(15). https://doi.org/10.1063/1.3651367

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Sep 2026

Reaction coordinates for the crystal nucleation of colloidal suspensions extracted from the reweighted path ensemble

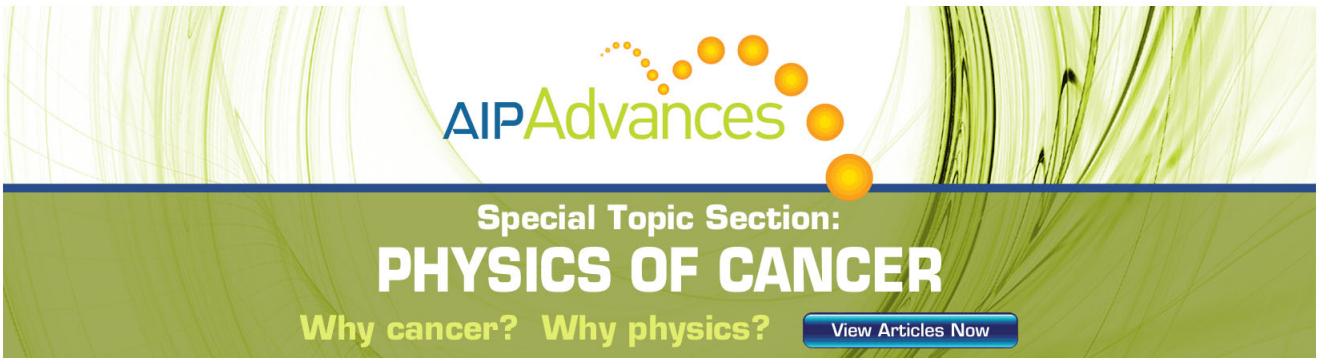
Wolfgang Lechner, Christoph Dellago, and Peter G. Bolhuis

Citation: J. Chem. Phys. **135**, 154110 (2011); doi: 10.1063/1.3651367
View online: http://dx.doi.org/10.1063/1.3651367
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v135/i15
Published by the American Institute of Physics.

Additional information on J. Chem. Phys.

Journal Homepage: http://jcp.aip.org/
Journal Information: http://jcp.aip.org/about/about_the_journal
Top downloads: http://jcp.aip.org/features/most_downloaded
Information for Authors: http://jcp.aip.org/authors

ADVERTISEMENT



Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

Reaction coordinates for the crystal nucleation of colloidal suspensions extracted from the reweighted path ensemble

Wolfgang Lechner,[1,a)] Christoph Dellago,[2] and Peter G. Bolhuis[1]
[1]*Van 't Hoff Institute for Molecular Sciences, PO Box 94157, 1090 GD Amsterdam, The Netherlands*
[2]*Faculty of Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria*

(Received 25 July 2011; accepted 22 September 2011; published online 20 October 2011)

We study the mechanisms of the homogeneous crystal nucleation from the supercooled liquid to the crystal phase in the Gaussian core model for colloidal suspensions with the aim to find optimal reaction coordinates. We introduce a set of novel collective variables based on the local structure of particles. By applying likelihood maximization of the committor function for the reweighted path ensemble constructed by replica exchange transition interface sampling, we select the optimal reaction coordinates from the set of collective variables. We find that the size of the cloud of prestructured particles surrounding the crystalline nucleus enhances the description of the transition. Further, we show that the rearrangement of the inner core of the nucleus according to Ostwald's step rule is a separate process, independent of the growth of the nucleus. © *2011 American Institute of Physics*. [doi:10.1063/1.3651367]

I. INTRODUCTION

Homogeneous nucleation from the liquid to the crystalline phase is usually described within the framework of classical nucleation theory (CNT). This theory predicts that at moderate undercooling the free energy of a crystalline nucleus growing in the supercooled fluid consists of two terms, a negative volume term, taking into account the particles in the nucleus that are already in the more stable solid phase, and a positive surface term, which comes from the surface free energy penalty of the interface between solid and liquid. Underlying assumptions of CNT are: (i) that the cluster has a spherical shape, (ii) that the nucleation process is independent of the particular crystal structure, (iii) that only the largest cluster in the system is important for the transition, and (iv) that the surface tension is independent of the cluster size. The Gibbs free energy difference between the supercooled liquid with and without a spherical crystal nucleus with radius r is then given by

$$\Delta G(r) = \frac{4}{3}\rho_s \pi r^3 \Delta\mu + 4\pi r^2 \gamma. \tag{1}$$

Here, $\Delta\mu$ is the chemical potential difference between the meta-stable liquid and the solid phase, ρ_s is the density of the solid phase, and γ is the solid-liquid surface tension. This free energy exhibits a maximum at $r^* = -2\gamma/\rho_s\Delta\mu$, and only if a solid nucleus becomes large enough to reach r^*, the nucleus will continue to grow spontaneously towards a bulk crystal. At coexistence, $\Delta\mu$ vanishes and therefore r^* becomes infinitely large and spontaneous freezing will not be observed. On the other hand, when the system is undercooled far below coexistence, the free energy barrier becomes small compared to k_BT and spinodal decomposition is observed.

The recent considerable interest in the nucleation of colloidal particles stems from the novel experimental techniques, such as optical tweezers and confocal microscopy, that allow one to study homogeneous nucleation in real time and three-dimensional space[1,2] and, on the other hand, the progress in computational techniques[3–7] that allow for accurate sampling of rare transitions. These advances also provide exciting new challenges that are currently extensively studied. First, the transition rates predicted by computer simulations based on the classical theory are, for some cases, orders of magnitudes off from the experimental measured ones.[8,9] Second, the predicted transition states cannot be fully described by the size of the nucleus only.[7] This discrepancy is revealed by investigation of the distribution of the commitment probability (also known as the committor distribution) at the top of the free energy barrier that should be peaked at $p_B = 0.5$ but is found to be rather flat. In this paper, we focus on the latter challenge and show which of the assumptions made by the classical theory may oversimplify the description of the nucleation process in the context of the novel insights into theory and experiment.

As a consequence of assumption (i), that the nucleus has a spherical shape, the surface and volume term are not independent variables, but are both described by a single coordinate, the radius r. Usually, in computer simulations (see, e.g., Ref. 10) the size of the nucleus is measured in terms of N, the number of connected solid particles that form the cluster. The CNT then predicts that the free energy barrier is due to the contributions of the volume, which scales with N and the surface $S \sim N^{2/3}$. In this paper, we show that the crystallization transition can be described more accurately by including the size of a prestructured cloud of particles surrounding the nucleus as an additional parameter. Defining this cloud requires novel order parameters that allow one to distinguish between crystalline cluster and prestructured surrounding, which will be introduced below.

a)Author to whom correspondence should be addressed. Electronic mail: w.lechner@uibk.ac.at.

 © 2011 American Institute of Physics

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

Assumption (ii), that the particular crystal structure of the nucleus does not enter the free energy was challenged by several recent studies. In some regions of the phase diagram of colloidal systems with strong short-ranged attractions, freezing is observed to be a two step process rather than the simple nucleation mechanism postulated by CNT.[12] This behavior can also be observed for the vapor-solid transition in the Lennard-Jones model.[13] Other researchers investigated the dependence of the nucleation on the underlying crystal structure. Ostwald's step rule[14] predicts that the system will not necessarily freeze into the thermodynamically most stable phase, but the one that is most easily accessible. Based on Landau theory, Alexander and McTague[15] argued that this first step would always be a crystal with body centered cubic structure (BCC). The general applicability of this theory is still debated. Klein[16] showed, based on the same formalism, that close to the spinodal, the BCC crystal would not be stable. Computer simulations indicated that the freezing of a Lennard-Jones system indeed exhibits solid clusters with a BCC ordering at the surface and a face centered cubic (FCC) structure in the core.[17] In contrast, experiments on hard-sphere colloids[1] exhibit random hexagonal close packed clusters rather than the predicted BCC crystals. Here, we study the crystal structure of the nucleus with novel averaged local bond order parameters showing that the structural rearrangement of the core is an independent event from the nucleation, thus providing a confirmation of the Ostwald step rule.

The applicability of assumption (iii) is well established. The rate of nucleation is determined by the probability to find a critical nucleus in the system volume. For large clusters, CNT is accurate while for small clusters the free energy has to be renormalized with the distribution of all cluster sizes in the system (see Ref. 17). The possibility of a size dependent surface tension, ruled out by assumption (iv), has been studied, e.g., in Ref. 18 and will not be addressed in this work.

The remainder of this paper is structured as follows. The model system we used will be introduced in Sec. II. A crucial part of the analysis concerning the distinction between different crystal structures will be discussed in Sec. III. In Sec. IV, we introduce our approach to test the quality of reaction coordinates based on the reweighted path ensemble with a nonlinear mapping onto a string. Finally, the results are presented in Sec. V, followed by a discussion on the prestructured cloud, recrossings at the transition state, and Ostwalds step rule in Sec. VI. We end with concluding remarks in Sec. VII.

II. GAUSSIAN CORE MODEL

As a general model for soft colloidal particles in solution (e.g., polymer chains), we used the GCM.[19,20] In this model, particles interact via a pair potential of Gaussian shape,

$$v(r) = \varepsilon e^{-r^2/\xi^2}. \tag{2}$$

Here, r is the particle-particle distance, ε and ξ determine the strength and the range of the potential, respectively, and k_B is the Boltzmann constant. In the following, we will use reduced units. Denoting the true temperature, density and pressure by T^*, ρ^*, and P^*, respectively, the reduced temperature is $T = k_B T^*/\varepsilon$, the reduced density is $\rho = \rho^*/\xi^3$, and the reduced pressure is $P = P^*\xi^3/\varepsilon$. Although the pair interaction of the system is rather simple, the GCM exhibits a feature-rich phase diagram (see Refs. 21–23). A particularly interesting feature of the GCM phase diagram is the phenomena called *re-entrant melting* (see Fig. 1), which refers to the following process. Starting at low densities and low temperatures, the system's dynamics is essentially that of a hard sphere system. Increasing the density will cause the system to freeze but, due to the lack of a hard core, the system will melt again after further compression. Another anomalous feature of the GCM is that the solid phase actually consists of two distinct regions: one at low densities, where the FCC crystal structure is more stable, and one at high densities, where the BCC structure becomes thermodynamically stable.

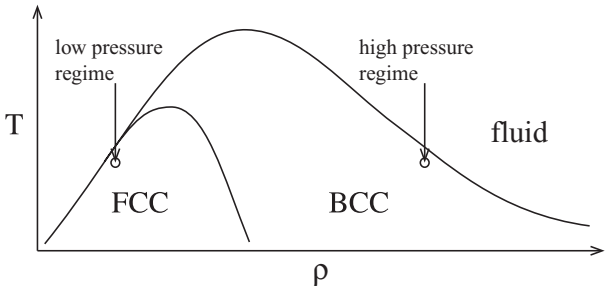


FIG. 1. Schematic phase diagram of the Gaussian Core model (GCM) in the temperature density plane taken from Refs. 21–23. The solid lines indicate (rather narrow) first order coexistence regions. At high density the fluid becomes stable again, a phenomenon known as re-entrant melting.

We studied the crystal nucleation in both the high and in the low density regime. At a pressure of $P = 0.011$, which corresponds to a density $\rho \approx 0.12$ and moderate undercooling at $T < 0.0030$, the most stable phase is a FCC crystal. In the high pressure regime of $P = 1.0$ ($\rho \approx 0.6$) and temperatures below $T < 0.0018$, the stable phase is a BCC crystal. With this model, it is thus possible to study the nucleation transition from a liquid to a BCC crystal and to a FCC crystal in the same system simply by adjusting the pressure. This allows us to study to what extent the underlying thermodynamically stable phase influences the structure of the crystal nucleus. To answer this question, we select optimal reaction coordinates to represent the nucleation from a set of collective variables that provide structural information about the nucleus. In Sec. III these collective variables are introduced.

III. POSSIBLE REACTION COORDINATES

The most widely used method to distinguish between solid and liquidlike particles, introduced by ten Wolde *et al.*,[17] is based on Steinhardt's bond order parameters.[24] The method basically consists of two steps. First, for each particle in the system the vector $q^i_{lm} = 1/n_{nn}(i) \sum_{j=1}^{n_{nn}(i)} Y_{lm}(\mathbf{r}_{ij})$ is calculated (with l typically $l = 6$), where Y_{lm} are the spherical harmonics, and $n_{nn}(i)$ is the number of nearest neighbors of particle i within a distance r_{nn}.[24] This vector holds information on the relative positions of the nearest neighbors of particle i. Second, for each pair of neighboring particles i and j the

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

normalized scalar product of $s_{i,j} = \langle q^i_{lm} q^{j*}_{lm} \rangle$ is evaluated. If the bond orders of the nearest neighbors of particle i and j are strongly correlated the number $s_{i,j}$ will be close to 1, if they are uncorrelated the scalar product will vanish. Connections between particles can, therefore, be defined as liquidlike, if $s_{i,j}$ is smaller than a certain threshold, usually $s_{i,j} < 0.5$, otherwise as solidlike bonds. As a last step, particles with more than six to eight solid connections are defined as solidlike particles. The strength of this method is that the definition of the solidity of particles is independent of the reference frame and also independent of the particular crystal structure. In this method, *solidity* means strongly correlated relative positions of the nearest neighbors.

Additional information about the particular local structure around a particle can be (in principle) calculated from the *local* bond order parameters, again based on q_{lm}.[24] For a choice of $l = 4$, the vector q_{4m} is sensitive to cubic symmetries and for a choice of $l = 6$ q_{6m} is sensitive to hexagonal symmetries. The self-scalar products $q^i_l = (4\pi/(2l+1) \sum_{l=-m}^{m} |q^i_{lm}|^2)^{1/2}$ can then be used to distinguish between different structures. Unfortunately, in the interesting region for nucleation – close to coexistence – these order parameters exhibit large fluctuations.

In fact, even for perfect crystals the distributions of the local bond order parameters overlap to a large extent, such that it is impossible to distinguish structures from these parameters alone. Frenkel and co-workers[10,11,25] developed a method that allows one to distinguish between FCC, BCC, and liquidlike clusters based on the histograms of q_4 and q_6 of the perfect crystals at moderate undercooling. The histogram of the q_4 and q_6 values of all particles in the cluster are considered as a superposition of the reference histograms of the perfect crystals. The coefficients are then the ratios of the particular crystal structures in the cluster. This allows one to calculate the structural composition of the cluster as well as the radial dependence thereof.

An averaged version of the local bond order resolves the local structure of individual particles better and allows for a distinction of crystal structures even in the regime close to coexistence.[26] Averaging the vectors q_{4m} and q_{6m} over the nearest neighbors of particle i and the particle i itself, and subsequently taking the self-scalar product of these vectors leads to an averaged version $\bar{q}_4$ and $\bar{q}_6$ of the bond order parameters. The probability histograms of these averaged local bond order parameters show minimal overlap and, hence can distinguish FCC, BCC, HCP, and liquidlike particles. The consequence of averaging over the nearest neighbors is that the structure of the second neighbor shell is taken into account. Therefore, this method has the same spatial resolution as the solid-liquid distinction of ten Wolde *et al.*[17]

In this work, we employed the averaged version of the local bond order parameters to distinguish between crystal structures as well as to define a novel interpretation of solidity.[7] The interpretation is based on the two-dimensional (2D) histograms of the parameters $\bar{q}_4$ and $\bar{q}_6$ for a perfect FCC, BCC, and HCP crystal as well as for an undercooled liquid. These reference histograms, evaluated for the system at the pressure and temperature at which the nucleation was studied, show almost no overlap.

As the reference histograms of the perfect structures give the probabilities to find a $\bar{q}_4$–$\bar{q}_6$ pair in the particular perfect crystal, particle i is defined to be of the structure that has the largest of all four probabilities at the given instantaneous $\bar{q}_4$–$\bar{q}_6$ for particle i. If all probabilities are smaller than $p < 10^{-5}$, the structure is defined as *undefined structure*. Finally, we define a particle to be *solidlike*, if the probability of finding the local structure in the liquid phase vanishes. This is an alternative view of local solidity: while the method of ten Wolde and Frenkel defines solidity as strong correlation between the local structures of the neighbors of a particle, our method defines solidity by a vanishing probability to find a particular structure in the undercooled liquid.

The information about the crystal structure of each individual particle in the system opens a wide range of possible collective variables. Besides using the ten Wolde-Frenkel definition of solidity to determine the number of particles in the largest cluster, N_{tf}, we can now also employ our new solidity definition to obtain a novel cluster size parameter N_{ld}. Moreover, to study the influence of the structure of the cluster, we define the following additional collective variables: the fraction of BCC particles in the cluster nBCC, that of FCC particles nFCC, that of HCP particles nHCP and that of undefined structures nUND. We also compute the ratio of crystal structures of the whole system: sBCC is the fraction of BCC particles in the system, sFCC, sHCP, sUND, sLIQ, and sSOL are the fraction of FCC, HCP, undefined structures, liquidlike particles and solid particles, respectively, where $s\mathrm{SOL} + s\mathrm{LIQ} = 1$ and $s\mathrm{BCC} + s\mathrm{FCC} + s\mathrm{HCP} + s\mathrm{UND} = s\mathrm{SOL}$. Including the shape of the cluster as a reaction coordinate might improve the classical nucleation theory. Therefore, we consider the number of *skin* particles N_s and the number of skin links N_{sl}. Skin particles are defined as liquidlike particles that have at least one nearest neighbor that is part of the largest cluster. The link number of a liquidlike skin particle is given by the number of its nearest neighbors that are part of the cluster. Summing all links of all skin particles yields N_{sl}. Moroni *et al.*[27] found evidence for possible enhancement of the reaction coordinate by taking the parameter $Q_{6,cl}$ into account, the average of the local bond order parameter q_6 over all particles in the largest cluster. We added $Q_{4,cl}$ and $Q_{6,cl}$ to our set of test reaction coordinates. All 15 collective variables were tested as possible candidates for enhancing the reaction coordinate.

IV. METHODS

The challenge is to extract information about the reaction coordinate without making *a priori* assumptions about the transition. Our approach requires two steps. First, we construct the reweighted path ensemble (RPE; Ref. 28) from a replica exchange transition interface sampling (RETIS) simulation[29,30] as described in Subsection IV A. Second, we present ways to extract and analyze a number of properties from the RPE in a *post production* step that does not require any additional sampling. These analysis methods will be addressed in Subsection IV B.

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

A. Simulation

For the degree of undercooling used in this study, the free energy barrier separating the metastable liquid state from the solid state is of the order $\Delta F \approx 10 - 20 k_B T$. The transition rates are so low that one cannot sample the barrier with straightforward methods. There are several solutions to this sampling problem, for instance, umbrella sampling, constraint MD, Wang-Landau sampling, to name just a few.[31] These methods are powerful tools to calculate of free energy profiles along a predefined collective variable. However, finding the optimal collective variable that can act as a model for the true reaction coordinate is rather the question that we address than the input. A method that does not require any *a priori* knowledge about the reaction coordinate is transition path sampling (TPS).[3,4] We employ an enhanced path sampling method called RETIS.[5,29,30] While in a TPS simulation only paths that leave the state A (the liquid state) and enter the state B (the solid state) are sampled, a TIS simulation includes all possible trajectories. In TIS, interfaces are introduced between the two stable states and paths are accepted or rejected with the condition that they cross a given interface. For enhanced sampling, a number of TIS simulations are run in parallel in a replica exchange scheme. The resulting paths can then be reweighted in accordance with the correct Boltzmann weights leading to a full description of the transition.[28] This RPE is an estimate for an infinitely long trajectory in a system with two metastable states A and B separated by a free energy barrier.

In the following we will summarize these methods.

1. Path ensembles

A path or trajectory of length L is defined as a sequence of phase points $\mathbf{x}^L = \{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_L\}$, where phase point $\mathbf{x}$ contains the positions $\mathbf{r}$ and momenta $\mathbf{p}$ of all particles in the system and two subsequent phase points are separated by a time step Δt. The equilibrium probability to find a path is

$$\mathcal{P}[\mathbf{x}^L] = \frac{\pi[\mathbf{x}^L]}{\mathcal{Z}}, \tag{3}$$

where $\pi[\mathbf{x}^L] = \rho(\mathbf{x}_0)\prod_{i=1}^{N} p(\mathbf{x}_i \to \mathbf{x}_{i+1})$ is the product of the distribution of initial points $\rho(\mathbf{x}_0)$ and the short time Markovian transition probabilities to go from one phase point to the next. $\mathcal{Z} = \int \mathcal{D}\mathbf{x}^L \pi(\mathbf{x}^L)$ is the normalization factor, where $\int \mathcal{D}\mathbf{x}^L$ denotes the integral over paths. Transition path sampling allows one to sample a subset of these paths, namely, all paths that start in state A and end in state B. Their probability is

$$\mathcal{P}_{AB}[\mathbf{x}^L] = \frac{h_{AB}[\mathbf{x}^L]\pi[\mathbf{x}^L]}{\mathcal{Z}_{AB}}. \tag{4}$$

Here, $h_{AB}[\mathbf{x}^L]$ is the indicator function which is unity if the trajectory starts in A and ends in B and zero otherwise. The transition path ensemble has been proven to be extremely useful for finding dominant reaction paths. A drawback of this ensemble is, however, that it does not hold direct information on the free energy or rate constant. The calculation of free energy also requires the missing paths that start in A and go back to A, as well as paths that start and end in B. TIS samples such paths by introducing an ordered set of n non-intersecting *interfaces*, defined by a parameter λ_i, between the stable states, and accepting or rejecting paths depending on whether or not they cross the interface λ_i. Here, the interfaces are fully described by the cluster size N_{ld}. We refer to trajectories that leave region A as belonging to the *forward* and those that leave region B as belonging to the *backward* process. The individual forward path probability for each interface is

$$\mathcal{P}_{A\lambda_i}[\mathbf{x}^L] = \frac{h_{A\lambda_i}[\mathbf{x}^L]\pi[\mathbf{x}^L]}{\mathcal{Z}_{A\lambda_i}}, \tag{5}$$

where the function $h_{A\lambda_i}[\mathbf{x}^L] = 1$ if the path starts in A and crosses interface λ_i and is zero otherwise. The path probability for backward paths that start in B is defined accordingly,

$$\mathcal{P}_{B\lambda_i}[\mathbf{x}^L] = \frac{h_{B\lambda_i}[\mathbf{x}^L]\pi[\mathbf{x}^L]}{\mathcal{Z}_{B\lambda_i}}. \tag{6}$$

2. Replica exchange transition interface sampling

In a RETIS simulation, the path distributions of Eqs. (5) and (6) are sampled in parallel with replica exchange methods developed to enhance sampling.[29,30] We introduce n_f forward and n_b backward replicas with their according interfaces. The total number of interfaces is thus $n = n_f + n_b$ (note that this definition does not include the stable state boundaries as a first and last interface, in contrast to, e.g., Refs. 28 and 30). The simulation consists of four possible moves, chosen with equal probabilities at each step.

(i) The shooting move selects, for each of the $n_f + n_b$ replicas, randomly one configuration from the current trajectory as a new shooting point. The velocities of this phase point are slightly changed and the equations of motions are integrated forward and backward in time, until the trajectory either hits region A or region B. This trial trajectory is accepted or rejected according to the acceptance rules of TIS. Trajectories belonging to a *forward* process are accepted when they start in A and cross their respective interface, i.e., the maximum N_{ld} reached along the path is larger than λ_i. Analogously, trajectories belonging to the *backward* process have to start in B and cross the interface λ_i.

(ii) The swapping move allows replicas to exchange their interfaces. Each move consisted of $(n_f - 1)n_f/2$ swapping attempts between replicas with forward trajectories and $(n_b - 1)n_b/2$ attempts between replicas with backward trajectories. Therefore, each possible swap is attempted in average once in each move. A swap between two forward trajectories is accepted when both swapped trajectories fulfill the TIS criterion.

(iii) An additional swapping move allows for exchange between forward and backward trajectories. A forward trajectory ending in B and a backward trajectory ending in A can be swapped. All velocities in all path points along the trajectories are reversed. This move is attempted $n_f \times n_b$ times per step with random pairs.

(iv) At the first interface, an additional replica samples trajectories that start at the interface λ_0, visit A and recross

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

the first interface.[29] This additional replica is used to calculate the flux through the first interface and to feed independent new configurations into the replica exchange simulation. Note that this is only possible for state A, the undercooled liquid, as the B state will continue to grow until the system is fully crystallized.

The trajectories from the path sampling simulation are not sampled from their natural distribution as they are biased by the acceptance procedure that rejects all paths that do not cross the given interface. Nevertheless, the paths can be reweighted such that the ensemble of configurations from the reweighted paths are drawn from the Boltzmann distribution.

3. Reweighted path ensemble

The individual paths from each interface can be combined to an estimate of the full path ensemble $\mathcal{P}[\mathbf{x}^L]$ by reweighting the individual paths with their appropriate weight,

$$\mathcal{P}[\mathbf{x}^L] = c_A \sum_{j=0}^{n} \mathcal{P}_{A\lambda_j}[\mathbf{x}^L] W_A[\mathbf{x}^L] + c_B \sum_{j=0}^{n} \mathcal{P}_{B\lambda_j}[\mathbf{x}^L] W_B[\mathbf{x}^L]. \tag{7}$$

Here, the function $W^A[\mathbf{x}^L] = \sum_{i=0}^{n-1} \bar{w}_i^A \theta(\lambda_{\max}[\mathbf{x}^L] - \lambda_i)\theta \times (\lambda_{i+1} - \lambda_{\max}[\mathbf{x}^L])$ selects, using the Heaviside step functions θ, the correct interface weight $\bar{w}_i^A$ for each path $\mathbf{x}^L$ starting from A based on the maximum value $\lambda_{\max}$ of λ along the path. Similarly, $W^B[\mathbf{x}^L] = \sum_{i=0}^{n-1} \bar{w}_i^B \theta(\lambda_{\min}[\mathbf{x}^L] - \lambda_i)\theta(\lambda_{i+1} - \lambda_{\min}[\mathbf{x}^L])$ selects the weights $\bar{w}_i^B$ for paths from B based on the minimum λ along the path. Note that the weights $\bar{w}_i^A = 0$ for replicas belonging to the backward process, and vice verse $\bar{w}_i^B = 0$ for the forward replicas.

As demonstrated in Ref. 28, the weights for each interface $\bar{w}_i^A$ and $\bar{w}_i^B$ can be calculated from a weighted histogram analysis method (WHAM) procedure of the forward and backward crossing probability histograms. The crossing probability that a path reaches a value of λ, given that it crossed interface i, is defined for the forward and backward process, respectively, as

$$P_A(\lambda|\lambda_i) = \langle \theta(\lambda_{\max}[\mathbf{x}^L] - \lambda) h_A(x_0) \rangle_{\lambda_i}, \quad P_B(\lambda|\lambda_i) = \langle \theta(\lambda_{\min}[\mathbf{x}^L] - \lambda) h_B(x_0) \rangle_{\lambda_i}, \tag{8}$$

where the angular brackets denote a (RE)TIS path ensemble average for interface λ_i. The crossing probability $P_{A,B}(\lambda_i|\lambda_i)$ is unity per definition, because each path has to cross the interface to be accepted. For larger values of λ, the crossing probability decreases. In principle, the crossing probability over the whole range of λ can be calculated from the first interface. This, however, is not possible in practice because the sampling at the transition state is poor. The WHAM is used to construct an estimated crossing probability by weighting of the individual crossing probabilities at each interface. Here, the crossing probabilities were cut off at values lower than 1%.

Substituting the resulting weights $\bar{w}_i^A$ and $\bar{w}_i^B$ into Eq. (7) yields the reweighted path ensemble up to the two constants c_A and c_B that are calculated from the statistics of the A-to-B paths at the interface λ_i,[28]

$$c_A = \frac{m_A(\lambda_i)}{m_{AB}(\lambda_i) P_A(\lambda_i|\lambda_0)}, \quad c_B = \frac{m_B(\lambda_i)}{m_{BA}(\lambda_i) P_B(\lambda_i|\lambda_n)}. \tag{9}$$

Here, $m_{AB}(\lambda_i)$ is the number of A-to-B paths found at interface i and $m_A(\lambda_i)$ the unnormalized histogram at λ_i, i.e., the total number of paths found at the interface. The constants c_A and c_B fix all relative weights in the RPE. For a complete path ensemble, including the stable states, the paths from the additional first interface have to be added as described in Ref. 28.

In the following subsection we will describe all the *post production* steps, which can all be viewed as projections of the RPE.

B. Analysis

The reweighted path ensemble is the starting point for several *post production* analyses, in which we take advantage of the fact that the RPE holds information about the statics (free energy) but also about the dynamics of the system (committor function).[28] In the following, we will summarize how the free energy profile and a committor estimation along arbitrary collective variables can be calculated. Further, we show how we can extract the information about the reaction coordinate with a likelihood maximization scheme. We emphasize that these calculations do not require any additional sampling.

1. Free energy profiles

The configurations from the paths in the RPE have the correct Boltzmann weight and, therefore, the free energy of any arbitrary set of collective variables, $\mathbf{q}(\mathbf{x}) = \{q_1(\mathbf{x}), q_2(\mathbf{x}), ..., q_m(\mathbf{x})\}$, follows from the weighted configurations by projecting them onto the desired variables without any additional sampling. The configurational density is

$$\rho(\mathbf{q}) = \frac{1}{Z} \int \mathcal{D}\mathbf{x}^L \mathcal{P}[\mathbf{x}^L] \sum_{k=0}^{L} \delta(\mathbf{q}(\mathbf{x}_k) - \mathbf{q}). \tag{10}$$

Here, Z is the normalization constant. The sum in the integral runs over all configurations k in all paths. The free energy as a function of $\mathbf{q}$ is

$$\Delta F(\mathbf{q}) = -k_B T \ln \rho(\mathbf{q}). \tag{11}$$

2. Committor function

The methods used here to find optimal reaction coordinates are based on the committor $p_B(\mathbf{x})$, which is defined as the probability that a given configuration with randomized momenta results in a trajectory that reaches B before A. In practice, the exact committor function cannot be calculated because it would take an infinite number of trajectories for each configuration in the system. However, in the projected space of collective variables $\mathbf{q}$ the RPE can

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

be used to estimate the averaged committor function $p_B(\mathbf{q})$ efficiently,[28]

$$p_B(\mathbf{q}) = \frac{\int \mathcal{D}\mathbf{x}^L \mathcal{P}[\mathbf{x}^L] h_B[\mathbf{x}^L] \sum_{k=0}^{L} \delta(\mathbf{q}(\mathbf{x}_k) - \mathbf{q})}{\int \mathcal{D}\mathbf{x}^L \mathcal{P}[\mathbf{x}^L] \sum_{k=0}^{L} \delta(\mathbf{q}(\mathbf{x}_k) - \mathbf{q})}. \quad (12)$$

Here, $h_B[\mathbf{x}^L]$ is the indicator function which is unity if the trajectory reaches B before it reaches A, and otherwise zero. Again, the projection can be done in any arbitrary coordinate $\mathbf{q}$. This projection of the committor function itself does not give us direct insight into the importance of a reaction coordinate, but is the basis of a maximum likelihood estimation summarized in the next subsection.

3. Maximum likelihood estimation

The use of the maximum likelihood estimation to evaluate the quality of reaction coordinates was introduced by Peters and co-workers.[32,33] The method requires the definition of an estimator for the committor function which is used as a model for the real committor,

$$\hat{p}_B(\mathbf{q}) = \frac{1}{2}(1 + \tanh[r(\mathbf{q})]), \quad (13)$$

and the logarithmic likelihood for $\hat{p}_B$,

$$\ln L = \sum_{\mathbf{x}\to A} \ln[1 - \hat{p}_B(\mathbf{q})] + \sum_{\mathbf{x}\to B} \ln \hat{p}_B(\mathbf{q}). \quad (14)$$

Here, $\hat{p}_A = 1 - \hat{p}_B$ and $r(\mathbf{q})$ is a projection of the n-dimensional vector $\mathbf{q}$ onto a single dimension. In the original work by Peters and Trout, this projection is a linear combination of the collective variables $r(\mathbf{q}) = c_0 + \sum_i c_i q_i$. The likelihood is then maximized with respect to the coefficients c_i and the optimal reaction coordinate is the one with the largest maximum likelihood after this optimization. The data points used for the maximization are the shooting points taken from an aimless shooting simulation.[32] By design, these shooting points are located close to the transition state and; therefore, the linear projection is justified.

We lift the restriction to be close to the TS by including information about the entire transition, using all the weighted configurations from the RPE paths as underlying data.[34] We view each configuration and the path it belongs to as a single realization of a trial calculation of the committor. These configurations are distributed according to their Boltzmann weight and; therefore, the distribution of the configurations is peaked around the stable states. However, the likelihood is peaked around the states close to the transition state. It turns out that these two factors cancel each other to a large degree and the contributions to the estimate result from the whole range in a more or less uniform way (see the Appendix in Ref. 34).

Using all configurations from the RPE requires a nonlinear description of the transition. This is achieved by projecting the configuration onto a piecewise linear string living in a low-dimensional collective variable space.[35] This string describes the progression σ of the transition, where $\sigma = 0$ corresponds to state A and $\sigma = 1$ to state B, respectively. The string is constructed from piecewise linear connected string points $\mathbf{S}^M = \{\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_M\}$ in the order parameter space $\mathbf{q}$. Each string point represents a value of σ ranging from $\sigma(\mathbf{s}_0) = 0$ in stable state A to $\sigma(\mathbf{s}_M) = 1$ in state B. Each configuration from the RPE is projected onto the string leading to a value of $\sigma(\mathbf{x})$ for each configuration that can be interpreted as the progress of the nucleation along the string. An additional function $f(\sigma)$ maps σ to a number $r(\mathbf{x})$ defined in the interval $[-\infty, \infty]$ which is used as the parameter for the maximum likelihood method. Thus, the logarithmic likelihood is

$$\ln L = \sum_{\mathbf{x}\to\mathbf{A}} W(\mathbf{x}) \ln[1 - \hat{p}_B(r(\mathbf{x}))] + \sum_{\mathbf{x}\to\mathbf{B}} W(\mathbf{x}) \ln \hat{p}_B(r(\mathbf{x})), \quad (15)$$

where the sums run over all configurations $\mathbf{x}$ in the RPE with their according weight $W(\mathbf{x})$, and the projection r is a function of the string

$$r(\mathbf{x}) = f(\sigma(\mathbf{S}^M(\mathbf{q}(\mathbf{x})))). \quad (16)$$

Here, $\mathbf{S}^M(\mathbf{q})$ is a general annotation for the mapping of the configurations onto the string. In practice, this can be achieved by a Voronoi construction or a slightly more sophisticated method that also takes the curvature of the string into account.[34] Equation (15) is maximized as a function of the positions of the string points $\mathbf{S}$ and as a function of the mapping $f(\sigma)$. The maximum likelihood maximization based on the nonlinear string is optimized in a steepest descent scheme[34] consisting of 3 Monte Carlo moves.

(i) The string move, where a new trial string is generated by displacing each string point by a small amount. Equation (15) is then evaluated for the trial string and accepted when the log-likelihood increased. We restrict the first string point to region A and the last one to B. The states A and B are only defined by the variable N_{ld}. Therefore, the first and last points are only allowed to move perpendicularly to N_{ld}. A second constraint on the movement of the string points is the equidistance condition required for the projection method used here.[34] Equidistance of the string points is achieved by moving the string points randomly and then rearranging the points to be equidistant again.[36] Due to the geometry of the string, configurations at the ends of the string can be located outside the range of σ. To avoid inaccuracies at the string ends, two additional virtual points are added that serve as extensions for $\sigma < 0$ and $\sigma > 1$.

(ii) The mapping $f(\sigma)$ is assumed to be a piecewise monotonically increasing linear function defined by M interpolation nodes (M is the number of string points). A new trial function $f'(\sigma)$ is generated by displacing the nodes with the constraint that $f'(\sigma)$ remains a monotonic function. Again the likelihood with the trial function $f'(\sigma)$ is evaluated and accepted if the likelihood increases and otherwise rejected.

(iii) The relative scaling of the collective variables is the equivalent to the metric tensor $\mathbf{M}$ in the string method with vanishing off-diagonal elements (see Refs. 37 and 38). In a third step, the metric tensor is optimized with respect to the log-likelihood. Here, the metric tensor is a 2×2 matrix with the elements $M_{00} = 1$, $M_{10} = M_{01} = 0$, and $M_{11} = s$. In this optimization step, a trial metric tensor is generated by varying s. Then, all

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

configurations are mapped onto the string in the scaled variable space and the likelihood is evaluated leaving the string and the mapping function constant. If the trial likelihood is larger than the initial one, the new scaling is accepted.

The log-likelihood from data with different dimensions cannot be compared in a straightforward way because the likelihood naturally increases with every additional dimension.[39] Peters and Trout used the Bayesian information criterion that allows for a comparison of models with different dimensions. In the nonlinear reaction coordinate analysis, a similar criterion was introduced[34] to find the optimal number of string beads M, where one maximizes

$$\ln \mathcal{L}(M) = \max_{\mathbf{S}^M, f^M} \ln L(M) - \mathrm{BIC}(M, N_d), \tag{17}$$

with

$$\mathrm{BIC}(M, N_d) = \frac{k(M)}{2} \ln N_d. \tag{18}$$

Here, $N_d = \sum_x W(x)$ is the total number of data points in the RPE, and $k(M)$ is the number of degrees of freedom in the string optimization. Taking into account that the number of interpolation nodes is equivalent to the number of string points and that the first and last string points are only allowed to move perpendicularly to N_{ld}, the number of degrees of freedom becomes $k(M) = 3M - 2$ for $M \geq 2$.

With the above method, the number of configurations that have to be taken into account becomes large (in the order of $N_d \propto 10^8$). The mapping onto the string requires several mathematical operations and the optimization steps of the string might become computationally expensive. Therefore, it is useful to work with histograms of points rather than the points itself. This can be achieved by rewriting the logarithmic likelihood as follows:

$$\begin{aligned}
\ln L &= \sum_{\mathbf{x}\to B} W(\mathbf{x}) \ln \hat{p}_B(r(\mathbf{x})) + \sum_{\mathbf{x}\to A} W(\mathbf{x}) \ln \hat{p}_A(r(\mathbf{x}) \\
&= \sum_{\mathbf{q}} N_B(\mathbf{q}) \ln \hat{p}_B(r(\mathbf{q})) + \sum_{\mathbf{q}} N_A(\mathbf{q}) \ln \hat{p}_A(r(\mathbf{q})) \\
&= \sum_{\mathbf{q}} N_d \rho(\mathbf{q}) p_B(\mathbf{q}) \ln \hat{p}_B(r(\mathbf{q})) \\
&\quad + \sum_{\mathbf{q}} N_d \rho(\mathbf{q}) p_A(\mathbf{q}) \ln \hat{p}_A(r(\mathbf{q})).
\end{aligned} \tag{19}$$

In the second line, we have replaced the two sums over all $\mathbf{x}$ from trajectories that end in A and B, respectively, with sums over the $\mathbf{q}$-space. Instead of summing up individual points we now sum over two histograms $N_A(\mathbf{q})$ and $N_B(\mathbf{q})$. $N_A(\mathbf{q})$ is generated from all configurations that are part of a trajectory that ends in A, while trajectories that end in B contribute only to $N_B(\mathbf{q})$. The last line serves to illustrate the meaning of N_A and N_B. The probability to find a configuration with order parameters $\mathbf{q}$ in the RPE is the equilibrium probability $\rho(\mathbf{q})$. The probability that this configuration leads to the stable state B is the committor p_B (for state A it is p_A). In other words, the histograms N_A and N_B are simply the product of the equilibrium distribution and committor function and can be defined generally for configurations $\mathbf{r}$,

$$\frac{N_A(\mathbf{r})}{N_d} = \rho(\mathbf{r}) p_A(\mathbf{r}), \quad \frac{N_B(\mathbf{r})}{N_d} = \rho(\mathbf{r}) p_B(\mathbf{r}). \tag{20}$$

V. RESULTS

The simulations are done at two different pressure regimes: at pressure $P = 0.011$ and $T = 0.003$, the low pressure regime where the most stable phase is FCC, and a pressure $P = 1.0$ and $T = 0.0018$, where BCC is more stable. The number of particles is $N = 10976$ in both cases. The potential is cut off at $r_c = 4.0$, which basically means that the potential was calculated up to machine precision. The nearest neighbor criterion is $r_{nn} \leq 1.7$ in the high pressure simulations and $r_{nn} \leq 2.7$ in the low pressure simulation. For the calculation of the forces and the potential energy the cell list method is used with $N_{cells} = 216$ cells in the high pressure simulation and $N_{cells} = 729$ cells at low pressure. The high friction limit was sampled using the velocity Verlet algorithm in combination with the Andersen thermostat with velocity reinitialization frequency $\nu = 0.01$ in combination with the Andersen barostat. We define region A as $N_{ld} \leq 5$ and region B as $N_{ld} \geq 400$. Note that we assume that when the cluster has reached region B, it will be entirely committed to the crystal phase. Between the stable states we introduce eight non-intersecting interfaces at the positions $N_{ld} = 10, 20, 40, 60, 80, 100, 160, 200$ for the forward process and eight interfaces at $N_{ld} = 80, 100, 160, 200, 240, 280, 320, 360$ for the backward process. In both simulations, we have generated 7000 trajectories for each of the 16 interfaces.

A. Reweighted path ensemble

The RPE is constructed from the RETIS paths according to Eq. (7). Weights of the individual interfaces and the constants c_A and c_B are calculated from the WHAM analysis of the crossing probabilities as described above. Figure 2 depicts the weighted crossing probability histograms from the forward and from the backward part of the simulation for the low pressure simulation and the high pressure simulation, respectively. From these crossing probabilities, the nucleation rate constant follows by multiplication with the first interface flux ϕ_{A,λ_0}.[5] The resulting rates are $k_{\mathrm{high}} = 6.32 \times 10^{-7}$ ($\phi_{A,\lambda_0} = 0.018$) and $k_{\mathrm{low}} = 1.53 \times 10^{-6}$ ($\phi_{A,\lambda_0} = 0.092$) for the high and low pressure regime, respectively.

B. Free energy profiles

The paths in the RPE have the correct statistical weights and; therefore, the configurations taken from these reweighted paths are drawn from the natural Boltzmann distribution. Therefore, a projection of these configurations onto a set of collective variables leads directly to the free energy as a function of these order parameters (see Eqs. (10) and (11)). The order parameters can be chosen arbitrarily, which means that from the weighted paths one can deduce the free energy in any parameter space without doing any further sampling.

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

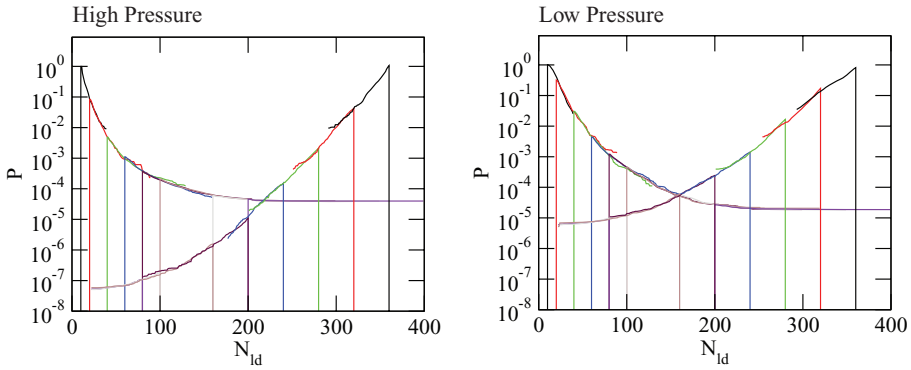


FIG. 2. Reweighted crossing probability histograms as a result of the WHAM analysis of the crossing probabilities in the high pressure regime (left) and low pressure regime (right).

Figure 3 depicts the free energy profile as a function of N_{ld} and N_{tf}, respectively, calculated from the RPE. An additional unbiased simulation allows for a normalization of the free energy with respect to the metastable liquid, where all cluster sizes are taken into account.[8] While, in principle, this additional sampling is not needed as the information about state A is included in the RPE, here, for practical reasons we did not store full trajectories but only stored the 15 order parameters described above which does not include the information about all clusters in the system. N_{ld} is an important projection because it serves as order parameter to define the interfaces and the stable states and it is the sole reaction coordinate suggested by CNT. In both the high and the low pressure regime, the free energy profile exhibits a single peak without any intermediate minima. A common choice for the reaction coordinate is N_{tf} (e.g., Ref. 17), which we depict for completeness in Fig. 3 on the right. Note that the according free energy barriers as function of N_{tf} and N_{ld}, respectively, are almost equally high. We come back to this point in Sec. VI.

C. Optimal reaction coordinate

A main goal of this study is to find reaction coordinates that enhance the description of the transition compared to the CNT. The CNT predicts that the size of the largest cluster is the only relevant reaction coordinate to describe nucleation. However, Moroni *et al.*[27] have shown that for a Lennard-Jones system, the reaction can be described more accurately by including structural information about the cluster.

We maximize the log-likelihood of the committor as a function of a nonlinear mapping of pairs of candidates for reaction coordinates. The aim is to find a reaction coordinate that enhances the classical description with the single reaction coordinate N_{ld}. We do this by comparing pairs of reaction coordinates (N_{ld}, q_i), where q_i is one from the following list of parameters: nBCC, nFCC, nHCP, nUND, sBCC, sFCC, sHCP, sUND, sLIQ, sSOL, N_{tf}, N_s, N_{sl}, $Q_{6,\,cl}$, and $Q_{4,\,cl}$ (as defined in Sec. III). This means we are comparing 15 pairs of possible reaction coordinates.

Figure 4 depicts the optimized strings and mapping functions, respectively, from the high and low pressure simulation after 20 000 optimization steps. Each step consisted of a string displacement, a move in the mapping function, and a scaling move. The mapping function $f(\sigma)$ is already well described with two points. Note, that the predicted transition state is almost identical for all studied numbers of interpolation nodes. Both at low and at high pressure, the string as well as the mapping function converge to a similar result for different numbers of string points. In particular, the mapping functions fall together at the transition state. Note that in this method the geometric interpretation of the transition state is the collection of all configurations at σ^* with $r(\sigma^*) = 0$.

For each of the 15 combinations of order parameters (N_{ld}, q_i) we have optimized the string, the mapping function and the scaling. For convenient comparison of the results,

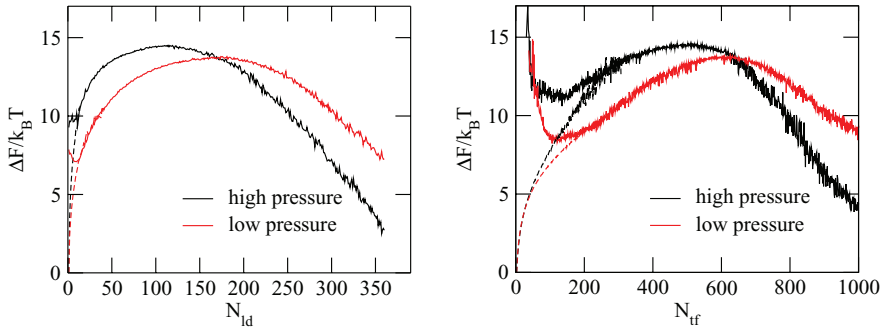


FIG. 3. (Left) Free energy profile from the projection of the RPE (solid) as a function of N_{ld} from the high pressure regime (black) and low pressure regime (red). (Right) Free energy profiles as a function of N_{tf}. The relative normalization between N_{tf} and N_{ld} free energy barrier is given by the RPE as the configurations are weighted with the correct Boltzmann weight and therefore also its projections. For completeness we have added the histogram of all cluster sizes (dashed).[8,18]

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

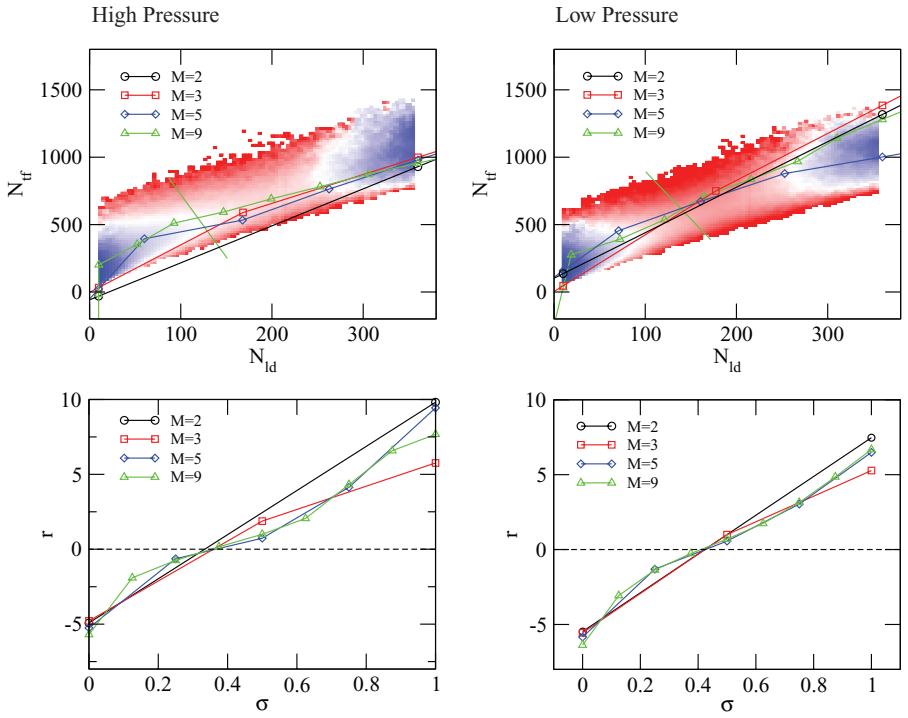


FIG. 4. (Top left) Optimized strings with $M = 2, 3, 5, 9$ string points in the order parameter space spanned by N_{ld} and N_{tf} from the high pressure simulations. The color map in the background represents the free energy calculated from the RPE as a function of N_{ld} and N_{tf}. The estimated transition state for the $M = 9$ string is indicated as a green line, parametrized as $N_{tf} = 1588-8.5\ N_{ld}$. (Bottom left) The mapping function $r = f(\sigma)$ belonging to the strings above. The transition state is estimated as configurations perpendicular to the string (in the scaled variables) at $r = 0$, in the high pressure regime at $\sigma = 0.4$. (Top right) The optimized strings in the low pressure regime. Here, the transition state is parametrized by $N_{tf} = 1601-7.04\ N_{ld}$. (Bottom right) Optimized mapping function with a transition state at $\sigma = 0.43$ in the low pressure regime.

we introduce a relative information gain $g = (\ln L(N_{ld}, q_i) - \ln L(N_{ld}))/b$ with $b = 1/2\ln L(N_d)$. As a result, only a $g \geq 1$ leads to an information gain due to an additional collective variable q_i that satisfies the Bayesian information criterion. The results are listed in Table I. From the full set of 15 collective variables only three combinations [(N_{ld}, N_{tf}), (N_{ld}, N_s), (N_{ld}, N_{sl})] satisfy this criterion and exhibit a $g \geq 1$ (or close to 1). These three best combinations are the same in the low and high pressure regimes. In both regimes the combination (N_{ld}, N_{tf}) is the optimal choice. A comparison of the information of different numbers of string points shows that for the nucleation a linear description is sufficient. The combination of (N_{ld}, N_{tf}) is the optimal reaction coordinate independent of the number of string points used in this study. This indicates that the method is robust with respect to M. Adding a third component to the reaction coordinate does not improve the likelihood significantly.

TABLE I. Relative gain (g) in information compared to the single reaction coordinate N_{ld} with the error $\Delta = \langle (g - \langle g \rangle)^2 \rangle^{0.5}$ calculated from six independent optimization runs. s is the optimized relative scaling between N_{ld} and the respective collective variables.

	High pressure				Low pressure			
Rank	OP	g	Δ	s	OP	g	Δ	s
1	N_{tf}	2.35	0.46	0.32	N_{tf}	2.23	0.24	0.29
2	N_s	1.01	0.45	0.77	N_s	1.54	0.36	1.61
3	N_{sl}	0.95	0.33	0.28	N_{sl}	1.39	0.06	0.46

The results lead to a novel interpretation of surface and volume term in the CNT presented in Ref. 7. The particles in the core identified as crystalline particles, N_{ld}, are surrounded by a cloud of particles that are highly correlated in their structure but not ordered in a crystal structure. N_{tf} represents the sum of the core and the surface particles. In Ref. 7, we have shown that the free energy, one would expect from this interpretation is in excellent agreement with the CNT. However, the two-dimensional reaction coordinate (N_{ld}, N_{tf}) is required for an accurate description of the nucleation pathway as shown in the next subsection.

D. Committor analysis

For a perfect reaction coordinate each configuration located at the transition state should exhibit a commitment probability of $p_B = 0.5$. The deviation from a peaked distribution of the p_B values is, therefore, a useful measure for the quality of a reaction coordinate.[40] A bad choice would lead to a flat distribution or, in the worst case, to a minimum at the putative transition state. Figure 5 depicts the committor distribution of 100 independent configurations located at the transition state using N_{tf} alone and the transition state predicted by the string method which is a combination of N_{tf} and N_{ld}, and parametrized by $N_{tf} = 1588-8.5\ N_{ld}$ and $N_{tf} = 1601-7.04\ N_{ld}$ for low and high pressure, respectively. This analysis shows that the committor distribution is peaked around $p_B = 0.5$ when we consider both of the order parameters. Taking only into account, the often used N_{tf} reaction

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

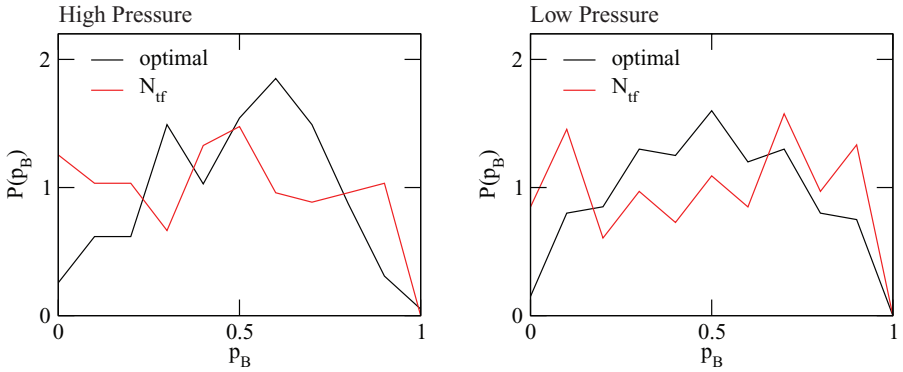


FIG. 5. Commitment probability profile from $N = 100$ individual configurations taken from the predicted transition state at high pressure (left) and low pressure (right). The committor was evaluated for 100 trajectories each with velocities randomly assigned according to the Boltzmann distribution. The transition state defined with N_{tf} as reaction coordinate (red) consists of 100 configurations with clusters of sizes $N_{tf} = 560$–580 in the low pressure and $N_{tf} = 490$–510 in the high pressure regime, respectively. This ensemble of configurations leads to an almost flat histogram indicating a rather sub-optimal choice for a reaction coordinate in both cases. The committor distribution of configurations taken from $r(\sigma) = 0$ are plotted in black. The means and standard deviations (Ref. 41) for the optimal reaction coordinate and the N_{tf} reaction coordinate in brackets are: $\mu_{h,\text{high}} = 0.49\,(0.47)$ and $\sigma_{h,\text{high}} = 0.22\,(0.55)$ and $\mu_{h,\text{low}} = 0.49\,(0.50)$ and $\sigma_{h,\text{low}} = 0.24\,(0.30)$.

coordinate the distribution becomes flat. Note, that a perfect reaction coordinate would lead to a binomial committor distribution with a standard deviation of $\sigma_p = 0.05$ for $N = 100$ trial trajectories.[41] Thus, while the optimized reaction coordinate is much better than the N_{tf} coordinate alone, it is still far from perfect.

E. Structural composition of the nucleus

The thermodynamic stable phase at high pressure is a BCC structured crystal, at low pressure it is a FCC crystal. This raises the question whether or not, and to what extent, the underlying stable phase influences the formation of the nucleus. This is particularly interesting in the light of Ostwald's step rule, which indicates that the nucleus can form an intermediate structure first and later rearranges into the stable phase. It is therefore useful to study the structural composition of the nucleus in both regimes at different stages of the nucleation. Figure 6 (top left) depicts the structural composition of 200 independent critical nuclei taken from the high pressure simulation. The HCP structure is the dominant structure. Only in the core the onset of a BCC is clearly visible. For post-critical clusters, shown in Fig. 6 (bottom left), the BCC structure in the core is becoming larger and about equally probably as the HCP structure. This indicates that the HCP-BCC transformation occurs in clusters larger than the critical size in a separate transition. This is a clear indication for the applicability of Ostwalds step rule in the high pressure regime, where first a HCP cluster forms which then transforms its core to a BCC crystal. A typical post critical cluster is shown in Fig. 7 as illustration. The core of the nucleus is already in the BCC phase surrounded by an intermediate layer of HCP particles. The nucleus is embedded in the prestructured surface cloud.

The composition of critical nuclei in the low pressure regime are shown in Fig. 6 (right). Again, the HCP structure is dominant but in the low pressure regime, it stays dominant also for large clusters. Here, the core does not rearrange to the most stable FCC phase.

VI. DISCUSSION

A. Free energy and rates

It is instructive to view the free energy barriers (Fig. 3) in the context of the crossing probabilities, the flux and the rates. The plateau of the forward crossing probability (see Fig. 2), the probability that a trajectory starting at the first interface reaches B, is $P_{AB}^{\text{high}} = 3.4 \times 10^{-5}$ and $P_{AB}^{\text{low}} = 1.67 \times 10^{-5}$ in the high and low pressure regime, respectively. At first sight, this may seem surprising as the free energy barrier from the high pressure regime is the larger one and; therefore, one might expect the crossing probability to be lower. However, the barrier height cannot directly be compared with the crossing probability, but is related to the rate which is the product of flux and crossing probability. Here, the flux in the low pressure regime is larger by a factor of 5 which results in a total fraction of the rates $k_{\text{high}}/k_{\text{low}} = 0.42$ which in turn results exactly in the difference of 0.9 k_BT between the heights of the free energy barriers. Another way to see this connection is the following: the first interface is defined as $N_{ld} = 5$ for both regimes. The crossing probability depends on the difference between the height of the barrier and the free energy at the position of first interface. This difference is smaller in the high pressure regime leading to a higher crossing probability.

While the flux and the crossing probabilities are depending on the position of the first interface, the rate and the free energy barrier are independent of this choice. Indeed, the flux compensates exactly for the crossing probabilities to give the correct rates according to the absolute free energy differences. The practical consequence of this is that even though the rates and the free energy barriers are independent of the choice of the first interface, it does result in a trade-off of accuracy in the calculation of flux and crossing probability.

B. Novel interpretation of CNT

The results of the likelihood maximization indicate that the optimal choice of reaction coordinates is a combination

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

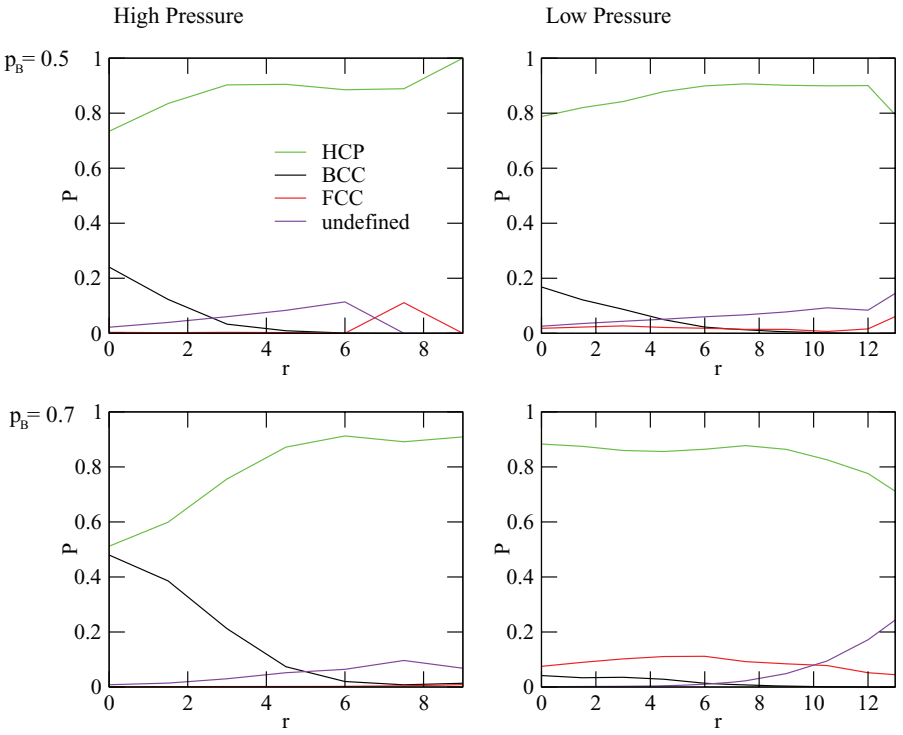


FIG. 6. (Top left) Structural composition as a function of the distance to the center of mass averaged over 200 independent clusters of critical size in the high pressure regime. The HCP structure dominates, however, the formation of BCC core in the center becomes evident. (Bottom left) Averaged structural composition for post critical clusters in the high pressure regime. Half of the clusters have a BCC structured core. This core is surrounded by a HCP layer. (Top right) In the low pressure regime HCP is dominant for small clusters as well as for large clusters (bottom right).

of N_{tf} and N_{ld}. These two order parameters are based on two different definitions of solidity. For the definition of N_{tf}, one considers correlations between the local structures around neighboring particles. The definition of solidity for the parameter N_{ld} requires that the local structure around a particle has a vanishing probability in the supercooled liquid phase. Both definitions have exactly the same spatial resolution because both hold information of q_{lm} evaluated for the nearest neighbors and their second shell neighbors. We find that the crystalline core (the N_{ld} cluster) is embedded in a cloud of particles that are solidlike in the sense that their local structure is highly correlated but liquidlike in the sense that the probability to find the structure in the supercooled liquid is larger than to find it in a solid phase.

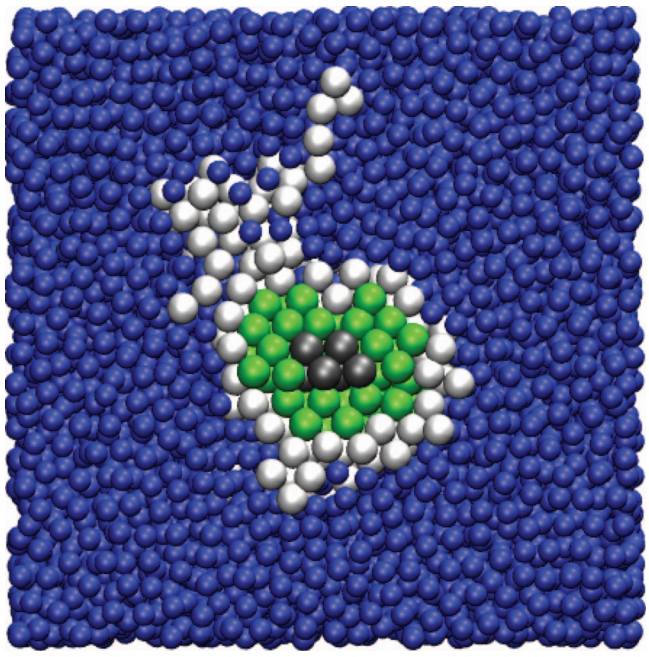


FIG. 7. Typical post critical cluster in the high pressure regime. The nucleus consists of several layers. The innermost core (black) has rearranged to a BCC structured cluster and is surrounded by a layer of HCP particles (green). The nucleus is embedded in the prestructured surface cloud (white). The undercooled liquid is indicated by blue particles.

This cloud of prestructured particles that surrounds the crystalline core of the nucleus leads to an interpretation of the surface that is consistent with the CNT.[7] To see this, we consider a general form of CNT where surface term and volume term are independent, $\Delta G = \Delta\mu N_{\text{core}} + \epsilon_s N_{\text{surface}}$, where ϵ_s is the surface energy and $\Delta\mu$ is the chemical potential difference. Here, we interpret the crystalline core as the volume term $N_{\text{core}} = N_{ld}$ and the surrounding cloud as the surface $N_{\text{surface}} = N_{tf} - N_{ld}$. The resulting free energy has the form $\Delta G = c_1 + c_2 N_{ld} + c_3 N_{tf}$, where c_i are constants. In the special case that $N_{tf} \propto N_{ld}^{(2/3)}$, one recovers the free energy barrier predicted by the CNT (see Eq. (1)). For this system, we find that $N_{tf} \propto N_{ld}^{0.68}$ in the high pressure regime and $N_{tf} \propto N_{ld}^{0.60}$ in the low pressure regime, which is indeed consistent with CNT. Note that another possible definition of the surface does not exhibit this feature. The number of *skin* particles defined as the number of particles that are liquidlike and have a particle that is part of the cluster as nearest neighbor scales as $N_s \propto N_{ld}^{0.37}$. This small exponent is a result of the roughness of the clusters. In contrast, the cloud particles

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

constitute a rather thick layer, which smooths the rough cluster surface and recovers the compact scaling.

The appearance of the surface cloud in our simulations is also not simply a geometrical artifact from a different sensitivity of the two methods for N_{ld} and N_{tf} with respect to interfaces. We note that both methods have the same spatial resolution and include the same information about the structure without any directional bias in a radial manner. Nevertheless, one could argue that one of the two methods might be more sensitive to interfaces, therefore, considering a layer of particles as liquidlike that the other criterium considers solidlike. We have checked this possibility by performing a simulation at the high pressure conditions of an idealized system consisting of a slab of a perfect FCC crystal immersed in liquidlike particles, originally at randomized positions. In this idealized scenario, both N_{ld} and N_{tf} identify the same layer as interface layer. As the system equilibrates, the liquid particles start ordering at the interface leading to an additional layer similar to the surface cloud. We note that the understanding of this interface may be further improved by using order parameters with variable spatial resolution (see, e.g., Ref. 42).

C. Variational transition state theory

Variational transition state theory[44] states that the optimal dividing surface (in other words the best reaction coordinate as the dividing surface is often defined in terms of the reaction coordinate) shows the smallest number of recrossing trajectories. The ratio of reactive trajectories and the number of crossing points on the dividing surface, also known as the transmission coefficient, is thus a good measure for the quality of the RC. The higher the transmission coefficient, the better the choice of RC. The transmission coefficient is also the dynamical correction factor that transforms the estimate of transition state theory (TST) into the exact reaction rate constant. We can compute the transmission coefficient directly from the RPE.[43] Focusing on the high pressure case, we determined the dividing surface as the top of the free energy barrier, for the N_{ld}, N_{tf} and the optimized RC, yielding, respectively, $N_{ld} = 120$, $N_{tf} = 550$, and $N_{tf} = 1588 - 8.5\, N_{ld}$ (see Figs. 3 and 4). The transmission coefficients are, respectively, $\kappa_{N_{ld}} = 0.059$, $\kappa_{N_{tf}} = 0.038$, and $\kappa_{opt} = 0.062$. Indeed the optimized RC has the highest transmission coefficient, hence the lowest number of recrossings, followed closely by the N_{ld} parameter. The N_{tf} order parameter has a significantly higher number of recrossings and is thus a less accurate RC. These findings agree with the results of the likelihood maximization. Interestingly, the free energy barrier heights for N_{ld} and N_{tf} shown in Fig. 3 do not seem to differ. However, the barrier height should be multiplied by the average order parameter velocity at the top of the barrier, in order to get the TST rate constant.[43] This correction is dependent on the order parameter and is expected to be much larger for the N_{tf} than for N_{ld}, thus compensating for the difference in transmission coefficient and eventually leading to an identical true rate, which by necessity is independent from the choice of reaction coordinate.

D. Structure of the growing nucleus

Next, we will take a closer look at the structural properties of the prestructured surface cloud and discuss its relation to previous studies on the nucleation of colloids. First, we analyze the structure of the prestructured surface cloud. Figure 8 depicts the 2D-histograms of the order parameters $\bar{q}_4$ and $\bar{q}_6$ for different crystal structures and a projection of the histograms onto the $\bar{q}_6$ axis. The histogram of the supercooled liquid, BCC, FCC, and HCP structures were used as reference to determine N_{ld} in the analysis.[7] Here, we have added the $\bar{q}_4$–$\bar{q}_6$ histogram of the prestructured surface cloud. The configurations are taken from 200 post-critical clusters from the RETIS simulation in the high pressure regime. The histogram of the prestructured surface cloud clearly separates from the supercooled liquid phase. In particular, the histogram of the prestructured surface cloud is distributed around a larger $\bar{q}_6$ value with respect to the supercooled liquid, while the parameter $\bar{q}_4$ is identical. Note that the prestructured cloud region is close to the HCP region, thus explaining why this structure is the first to form out of the supercooled liquid.

As a next step, we revisit the supercooled liquid phase. The histogram of the supercooled liquid (Fig. 8, blue) consists of all possible local configurations in the meta-stable liquid phase where, due to spontaneous fluctuations, highly correlated regions form continuously. In the definition of N_{tf}, these are crystalline clusters, while in the definition of N_{ld} they are part of the supercooled liquid! We add the histogram of the particles that are part of spontaneously formed N_{tf} clusters in the supercooled liquid (Fig. 8, orange). The structural properties of spontaneously formed N_{tf} clusters are the same as those of the prestructured surface cloud in terms of $\bar{q}_4$ and $\bar{q}_6$, except that the prestructured surface cloud has a slightly larger average $\bar{q}_6$ than the spontaneously formed N_{tf} clusters. Therefore, we identified the highly correlated particles in the supercooled liquid phase as the large-$\bar{q}_6$ part of the supercooled liquid histogram and find that their structure is the same as that of the prestructured surface cloud.

This suggests an initial phase of the nucleation, where first a large-$\bar{q}_6$ region forms spontaneously that enhances the formation of a truly crystalline cluster. The large-$\bar{q}_6$ region surrounds the crystalline nucleus, growing with it. Therefore, our conclusions are similar to the mechanism that was recently predicted by Kawasaki and Tanaka[45] for colloidal hard spheres. They find that the nucleus forms in a large-$\bar{q}_6$ region that is similar to a HCP structure. A similar conclusion was drawn by Schilling *et al.*,[46] who found that a highly correlated nucleus is surrounded by a less structured layer for crystal growth at a larger undercooling. Finally, the results are also in agreement with the conclusions by Moroni *et al.*[27] The authors find that the reaction coordinate can be enhanced by taking into account N_{tf} and $Q_{6,\,cl}$. This is in agreement with our results, as $Q_{6,\,cl}$ and N_{ld} both hold the same information about the crystallinity of the cluster, except that N_{ld} allows for the analysis of the structure on a particle-level, while $Q_{6,\,cl}$ is averaged over the whole cluster. Recently, Beckham and Peters[47] confirmed this importance of $Q_{6,\,cl}$ as a reaction coordinate using transition path sampling in combination with a likelihood maximization at the transition state. They find that

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

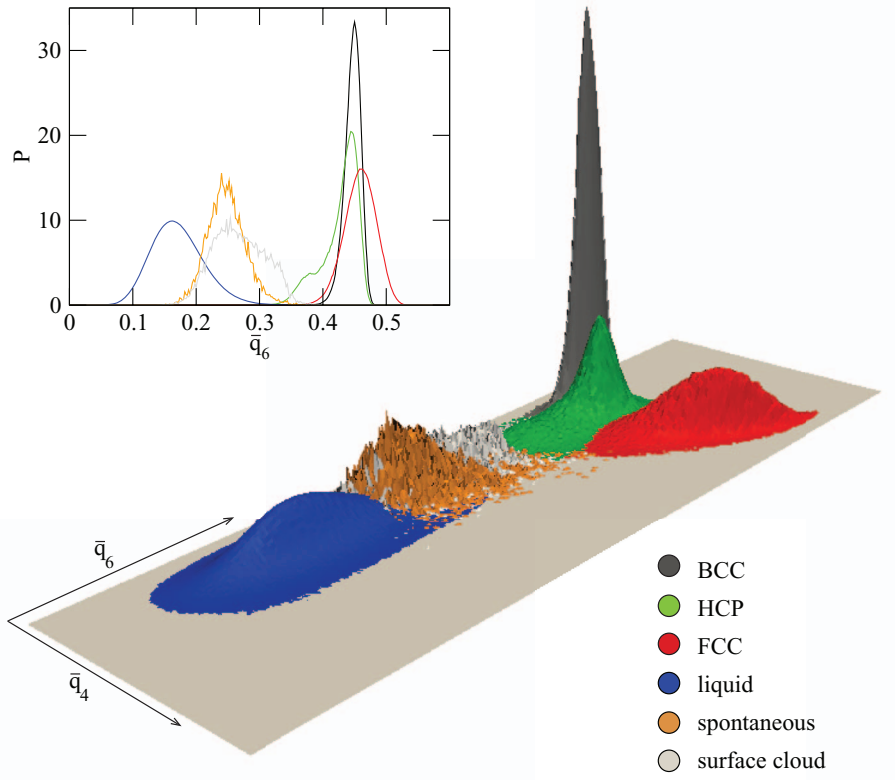


FIG. 8. Distributions of the $\bar{q}_4$ and $\bar{q}_6$ order parameters for different phases in the high pressure regime. The histograms of the perfect crystal structures and the supercooled liquid serve as reference for the structural analysis.[7] Particles that are part of the prestructured surface cloud (gray) and clusters that form spontaneously (orange) in the supercooled liquid have a similar structure and can be clearly distinguished from the liquid and the crystal structures. (Top left) Projection of the distributions onto the $\bar{q}_6$ coordinate.

the product of cluster size and crystallinity $Q_{6,\,cl}$ improves the description considerably.

The results further shed light on the applicability of Ostwald's step rule. We find that the core consists almost only of HCP and BCC, while FCC particles and undefined structures are rare and located close to the surface. Considering the dominance of the HCP structure in the nucleus, the GCM shows a similar nucleation pathway as a recent study found for C_{60} nano-particles[48] and colloidal particles.[1] The study of the composition of the crystal structures of the nuclei at the transition state in comparison to post critical nuclei suggests a two step mechanism of the nucleation. First, a predominately HCP cluster forms, in both regimes with different underlying stable crystal structure. In the high pressure regime, the core then transforms into a BCC-HCP mixture. In the low pressure regime, the portion of FCC also increases but HCP remains the dominant structure even for clusters larger than $N_{ld} > 360$. This is an indication for Ostwald's step rule which implies that the nucleation does not necessarily lead directly to the most stable phase but rather to the one which is most easily accessible from the meta-stable phase. In the low pressure regime, the second step might not take place because HCP and FCC are too close in terms of free energy.

VII. CONCLUSIONS

In summary, we applied a novel method to distinguish crystal structures and establish the degree of solidity of particles together with a likelihood maximization of the committor function. We identified a new interpretation of surface and volume in the CNT and observed Ostwald's step rule with *atomic* resolution. Structural analysis revealed that the crystal nucleus has an *onion* composition. The innermost core consists of particles in the stable crystalline phase, surrounded by a layer of HCP ordered particles. Around the latter we find the prestructured surface cloud,[7] which is surrounded by the undercooled liquid. Finally, we mention that our methodology is generally applicable not only to crystallization in colloidal suspensions, but to all crystallization phenomena, and indeed to other rare events, from chemical reactions to biomolecular isomerization.

ACKNOWLEDGMENTS

This work is part of the research program of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organization for Scientific Research (NWO) and financially supported by the Austrian Science Foundation (FWF) within the SFB ViCoM (F 41).

[1] J. Zhu, M. Li, R. Rogers, W. Meyer, R. H. Ottewill, STS-37 Space Shuttle Crew, W. B. Russel, and P. M. Chaikin, Nature (London) **387**, 883 (1997).
[2] U. Gasser, E. R. Weeks, A. Schofield, P. N. Pusey, and D. A. Weitz, Science **292**, 258 (2001).
[3] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[4] C. Dellago, P. G. Bolhuis, and P. L. Geissler, Adv. Chem. Phys. **123**, 1 (2002).

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

[5]T. S. van Erp, D. Moroni, and P. G. Bolhuis, J. Chem. Phys. **118**, 7762 (2003).
[6]R. J. Allen, D. Frenkel, and P. R. ten Wolde, J. Chem. Phys. **124**, 194111 (2006).
[7]W. Lechner, C. Dellago, and P. G. Bolhuis, Phys. Rev. Lett. **106**, 085701 (2011).
[8]S. Auer and D. Frenkel, Adv. Polym. Sci. **173**, 149 (2005).
[9]L. Filion, M. Hermes, R. Ni, and M. Dijkstra, J. Chem. Phys. **133**, 244115 (2010).
[10]P. R. ten Wolde and D. Frenkel, Phys. Chem. Chem. Phys. **1**, 2191 (1999).
[11]I. Volkov, M. Cieplak, J. Koplik, and J. R. Banavar, Phys. Rev. E **66**, 061401 (2002).
[12]V. L. Anderson and H. N. W. Lekkerkerker, Nature (London) **416**, 811 (2002).
[13]J. A. van Meel, A. J. Page, R. P. Sear, and D. Frenkel, J. Chem. Phys. **129**, 20450 (2008).
[14]W. Ostwald, Z. Phys. Chem. **22**, 289 (1897).
[15]S. Alexander and J. McTague, Phys. Rev. Lett. **41**, 702 (1978).
[16]W. Klein, Phys. Rev. E **64**, 056110 (2001).
[17]P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, J. Chem. Phys. **110**, 1591 (1999).
[18]P. R. ten Wolde and D. Frenkel, J. Chem. Phys. **109**, 9901 (1998).
[19]P. J. Flory and W. R. Krigbaum, J. Chem. Phys. **18**, 1086 (1950).
[20]F. H. Stillinger, J. Chem. Phys. **65**, 3968 (1976).
[21]S. Prestipino, F. Saija, and P. V. Giaquinta, Phys. Rev. E **71**, 050102(R) (2005).
[22]S. Prestipino, F. Saija, and P. V. Giaquinta, J. Chem. Phys. **123**, 144110 (2005).
[23]P. Mausbach, A. Ahmed, and R. J. Sadus, J. Chem. Phys. **131**, 184507 (2009).
[24]P. L. Steinhardt, D. R. Nelson, and M. Ronchetti, Phys. Rev. B **28**, 784 (1983).
[25]P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, J. Chem. Phys. **104**, 9932 (1996).
[26]W. Lechner and C. Dellago, J. Chem. Phys. **129**, 114707 (2008).
[27]D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, Phys. Rev. Lett. **94**, 235703 (2005).
[28]J. Rogal, W. Lechner, J. Juraszek, B. Ensing, and P. G. Bolhuis, J. Chem. Phys. **133**, 174109 (2010).
[29]T. S. van Erp, Phys. Rev. Lett. **98**, 268301 (2007).
[30]P. G. Bolhuis, J. Chem. Phys. **129**, 144108 (2008).
[31]D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, San Diego, 2002).
[32]B. Peters and B. L. Trout, J. Chem. Phys. **125**, 054108 (2006).
[33]B. Peters, G. T. Beckham, and B. L. Trout, J. Chem. Phys. **127**, 034109 (2007).
[34]W. Lechner, J. Rogal, J. Juraszek, B. Ensing, and P. G. Bolhuis, J. Chem. Phys. **133**, 174110 (2010).
[35]W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).
[36]L. Maragliano and E. Vanden-Eijnden, Chem. Phys. Lett. **44**, 182 (2007).
[37]L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, J. Chem. Phys. **125**, 024106 (2006).
[38]E. Vanden-Eijnden and M. Venturoli, J. Chem. Phys. **130**, 194103 (2009).
[39]G. Schwartz, Ann. Stat. **6**, 461 (1978).
[40]P. L. Geissler, C. Dellago, and D. Chandler, J. Phys. Chem. B **103**, 3706 (1999).
[41]B. Peters, J. Chem. Phys. **125**, 241101 (2006).
[42]R. Shetty, F. A. Escobedo, D. Choudhary, and P. Clancy, J. Chem. Phys. **117**, 4000 (2002).
[43]P. G. Bolhuis and W. Lechner, "On the relation between projections of the reweighted path ensemble," J. Stat. Phys. (in press).
[44]J. Horiuti, Bull. Chem. Soc. Jpn. **13**, 210 (1938).
[45]T. Kawasaki and H. Tanaka, Proc. Natl. Acad. Sci. U.S.A. **107**, 14036 (2010).
[46]T. Schilling, H. J. Schöpe, M. Oettel, G. Opletal, and I. Snook, Phys. Rev. Lett. **105**, 025701 (2010).
[47]G. T. Beckham and B. Peters, J. Phys. Chem. Lett. **2**, 1133 (2011).
[48]K. N. Ngale, C. Desgranges, and J. Delhommelle, J. Chem. Phys. **131**, 244515 (2009).

Downloaded 13 Jun 2012 to 145.18.109.227. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions