

Original citation:

Lifanov, Yuri, Vorselaars, Bart and Quigley, David (Researcher in physics). (2016) Nucleation barrier reconstruction via the seeding method in a lattice model with competing nucleation pathways. The Journal of Chemical Physics, 145 (21). 211912.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/81578>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>

A note on versions:

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Nucleation barrier reconstruction via the seeding method in a lattice model with competing nucleation pathways

Yuri Lifanov,¹ Bart Vorselaars,² and David Quigley³

¹Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, United Kingdom

²School of Mathematics and Physics, University of Lincoln, Lincolnshire LN6 7TS, United Kingdom

³Department of Physics and Centre for Scientific Computing, University of Warwick, Coventry CV4 7AL, United Kingdom

(Received 8 June 2016; accepted 22 August 2016; published online 13 September 2016)

We study a three-species analogue of the Potts lattice gas model of nucleation from solution in a regime where partially disordered solute is a viable thermodynamic phase. Using a multicanonical sampling protocol, we compute phase diagrams for the system, from which we determine a parameter regime where the partially disordered phase is metastable almost everywhere in the temperature–fugacity plane. The resulting model shows non-trivial nucleation and growth behaviour, which we examine via multidimensional free energy calculations. We consider the applicability of the model in capturing the multi-stage nucleation mechanisms of polymorphic biominerals (e.g., CaCO_3). We then quantitatively explore the kinetics of nucleation in our model using the increasingly popular “seeding” method. We compare the resulting free energy barrier heights to those obtained via explicit free energy calculations over a wide range of temperatures and fugacities, carefully considering the propagation of statistical error. We find that the ability of the “seeding” method to reproduce accurate free energy barriers is dependent on the degree of supersaturation, and severely limited by the use of a nucleation driving force $\Delta\mu$ computed for bulk phases. We discuss possible reasons for this in terms of underlying kinetic assumptions, and those of classical nucleation theory. © 2016 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). [<http://dx.doi.org/10.1063/1.4962216>]

I. INTRODUCTION

Predicting the rate at which crystalline material precipitates from a supersaturated solution remains a significant challenge to atomistic and molecular simulation due to a large disparity between the physical time scales of the process and the time scales which are accessible to computational models.^{1–3} To date, very few studies have computed quantitative absolute nucleation rates from such simulations.⁴

A particularly challenging example where nucleation kinetics remain well beyond the reach of atomistic simulation is biomineralisation. Here nucleation and growth, mediated by organic components, appears to progress via a multistage process with a high degree polymorph control.⁵ This serves as a natural example of self-assembly — a phenomenon of great technological interest.⁶ The prototypical system for study of biomineralisation and polymorph selection is calcium carbonate (CaCO_3), where some degree of controlled assembly has been demonstrated in a laboratory.^{7–9} Recent *in situ* transition electron microscopy (TEM) experiments¹⁰ have provided a striking visualisation of crystallisation of calcite — the most stable polymorph of CaCO_3 at ambient conditions — via vaterite — a metastable polymorph whose molecular structure is characterised by a degree of structural disorder.¹¹

Some important insights into parts of the CaCO_3 nucleation pathway have been gained through molecular dynamics (MD) simulations.^{12–14} However, the expense of

detailed atomistic models and the low solubility of CaCO_3 prohibit modelling of the entire multi-stage process. More quantitative studies of multi-stage processes^{15–17} are possible using simple lattice models. In common with biomineral nucleation, these show existence of amorphous precursors to the assembly of anisotropic particles. To our knowledge, however, pathways proceeding via partially disordered phases or featuring the dissolution–regrowth mechanism¹⁰ have not been previously captured.

Lattice models remain a useful tool in nucleation theory, yielding insights into nonclassical phenomena,^{15–17} multistep pathways,^{18,19} heterogeneous nucleation,^{20,21} and limitations of calculation methodology.^{22,23}

In the following contribution, we present a lattice model of nucleation from solution, where the solute may form disordered, ordered, or semi-ordered solids. We map the phase diagram of the system and show that, in the limit of slow growth, the transition from solvent rich to ordered crystalline state proceeds via the two metastable solute phases. The temperature dependence of the barrier to nucleation of these three phases indicates the existence of a parameter regime where heights of these barriers become comparable. However, we also find that the barriers to solid state transformation between the three solute phases, once nucleated, are too low to allow our model to capture the dissolution and regrowth pathway to crystallisation.

We then study the kinetics of nucleation in our model using two distinct choices of microscopic dynamics over a

broad range of parameters. We use the increasingly popular “seeding” approach,^{4,24–26} which uses information extracted from simulations to parametrise a classical nucleation theory (CNT) expression for the nucleation rate. We consider the propagation of statistical error and critically examine the approximations made in this method. We show that the reconstructed nucleation free energy barrier is limited by the use of a CNT driving force parameter $\Delta\mu$ calculated for bulk phases rather than finite-size nuclei. As a result, barriers can be inconsistent with those obtained via explicit free energy calculation.

The definition of the lattice model is given in Sec. II. In Sec. III we specify the details of the free energy calculations presented in Sec. IV, where we examine the phase behaviour and the possible nucleation pathways in the model. In Sec. V we move on to the discussion of kinetics of nucleation under the different choices of microscopic dynamics (Sec. V A), developing an error estimation approach for the “seeding” method in Sec. V B and discussing generation of appropriately structured seeds in Sec. V C, before presenting the obtained by the method results in Secs. V D and V E. Finally, in Sec. VI we summarise and discuss the results presented in Sections IV and V.

II. MODEL

Our model is an extension of that introduced by Duff and Peters.^{16,27} We study a three component system of anisotropic particles on a cubic lattice, where each particle i is characterised by species $s_i \in \{1, 2, 3\}$ and orientation $q_i \in \{1, \dots, 24\}$. We label species 1 as solvent and species 2 and 3 as solute. The nearest neighbouring particles (i, j) interact isotropically with strength $K(s_i, s_j)$, while diagonally neighbouring particles (k, l) interact anisotropically with strength $A(s_k, s_l)$, giving us the following lattice energy E :

$$E = - \sum_{(i,j)} K(s_i, s_j) - \sum_{[k,l]} \delta(q_k, q_l) A(s_k, s_l), \quad (1)$$

where the first and second summations are performed, respectively, over the unique nearest and diagonal neighbour pairs, $\delta(\cdot, \cdot)$ is the Kronecker delta function, and the lattice structure is primitive cubic with 6 nearest and 12 diagonal neighbours. We sample lattice configurations from the semigrand (μVT) ensemble, i.e., $\sum_{s=1}^3 N_s = N$, where N_s is the number of particles of species s and $N = L^3$ is the number of lattice sites, using the Metropolis algorithm with the moves: (1) Transmutation $s_i \rightarrow s'_i$, (2) Reorientation $q_i \rightarrow q'_i$. The two Monte Carlo (MC) moves are attempted with equal probability and accepted with probability,

$$\min \{1, f_{ss'} \exp[-\beta \Delta E]\}, \quad (2)$$

where $f_{ss'} = \exp[\beta(\mu_{s'} - \mu_s)]$ is the ratio of fugacities of species s' and s , $\beta = (k_B T)^{-1}$ is the inverse temperature of the system, and ΔE is the change in energy of the lattice configuration due to the proposed move. In Secs. V and VI of the article, we will refer to the MC move set defined here as transmutation-reorientation kinetics. For convenience, we define f to be the solute to solvent fugacity ratio f_{12} .

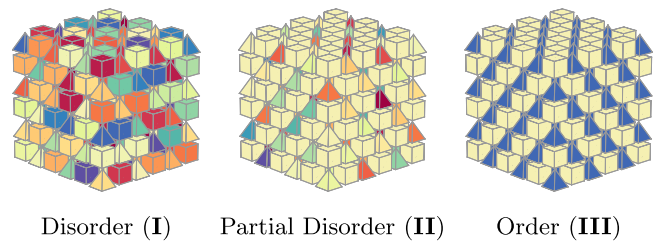


FIG. 1. Visualisations of bulk forms of the three solute phases on a cubic lattice of length $L = 6$, with solute species $s = 2$ and $s = 3$ drawn as cubes and tetrahedrons, respectively. Due to the nature of isotropic interaction, solute particles assemble into checkerboard structures. The shapes are colour coded according to the corresponding orientation label q .

We study the phase behaviour of the system as a function of temperature $k_B T$ and fugacity ratio f in the parameter regime where $\mu_2 = \mu_3$ with the following interaction strengths:

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & c/(c+1) & 0 \\ 0 & 0 & 1/(c+1) \end{bmatrix}. \quad (3)$$

For $c = 0.5$, i.e., where anisotropic interactions of species 2 are stronger than those of species 3, we observe, for $k_B T \leq 1.5$, three distinct energetically stable solute rich states (Fig. 1) and a solvent rich state for low values of f .

III. FREE ENERGY METHODS

We employ a multicanonical “flat histogram” method, refined using the Wang-Landau recursion²⁸ algorithm in conjunction with a histogram reweighting procedure.^{29,30} Our implementation of the method relies on constructing functions $\eta_{N_1}(N_1)$ and $\eta_E(E)$, which, when used as bias energies in our Metropolis acceptance criteria, allow our MC scheme to sample uniformly in N_1 and E , where N_1 is the total number of solvent particles in the system. It can be shown that uniform sampling is achieved when $\eta_{N_1}(N_1) = \ln \mathbb{P}(N_1)$ and $\eta_E(E) = \ln \mathbb{P}(E)$, where $\mathbb{P}(N_1)$ and $\mathbb{P}(E)$ are the equilibrium probability distributions of quantities N_1 and E .

We proceed by segmenting the ranges of the two quantities into overlapping intervals of equal width, ranging between 40 and 80 bins. We use bin widths of 1 and 6 for N_1 and E , respectively. Within each interval, the sampling scheme performs a biased Metropolis random walk while simultaneously refining the bias function using increments $\Delta\eta \in \{2^{-10}, 2^{-11}, \dots, 2^{-26}\}$ and generating a histogram h of the quantity of interest. Once the histogram achieves the flatness criteria $\max(h) \leq 1.1\bar{h}$ and $\min(h) \geq 0.9\bar{h}$, where \bar{h} is the mean value of the histogram across all bins, the increment $\Delta\eta$ is reduced and the histogram is discarded. After obtaining the estimates of $\eta_{N_1}(N_1)$ and $\eta_E(E)$, we sample histograms of N_1 and E with the, now fixed, bias. The maximum likelihood estimates of $\ln \mathbb{P}(N_1)$ and $\ln \mathbb{P}(E)$ are then recovered via weighted histogram analysis (WHAM).^{31–33} The obtained distributions are reweighted with respect to parameters f and β to produce approximate coordinates of

the points where two or more states, e.g., solvent rich ($N_1 \approx N$) and solute rich ($N_1 \approx 0$), are equally probable. We rerun the scheme 20 times for each coexistence point to obtain error bars.

To explore the nucleation pathways in our model, we implement an equilibrium path sampling (EPS) algorithm,^{16,34} which produces an estimate of the equilibrium probability distribution of a desired quantity without biasing the MC. We first define an order parameter (n, ν, χ) which characterises the size and degree of orientational order within the largest cluster of solute particles on the lattice, where n is the cluster size, and ν and χ are cluster orientational order parameters defined as follows:

$$\nu = \max \{p_2 P_2^{-1}, p_3 P_3^{-1}\}, \quad (4)$$

$$\chi = \begin{cases} \nu^{-1} \min \{p_2 P_2^{-1}, p_3 P_3^{-1}\} & \text{if } \nu > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where P_i and p_i are, respectively, the total number and the number of aligned diagonally neighbouring pairs of particles of species i , present in the cluster. To avoid singularities we only consider configurations where $P_i \geq 1$. We argue that, for a cluster of $n > 20$, the (ν, χ) coordinate is sufficiently representative of the orientational ordering of the cluster to classify it as amorphous (I), semi-ordered (II), or crystalline (III).

We employ the standard “geometric” definition of clustering, where a solute particle is considered to be part of a cluster if it is a nearest neighbour of at least one other solute particle. The quantity n is taken as the size of the largest such cluster on the lattice. This definition is known to be problematic at high temperatures above the surface roughening transition;³⁵ however, in this work we operate in the low temperature regime where the roughening effects are negligible.

We estimate the equilibrium joint probability distribution $\mathbb{P}(n, \nu, \chi)$ for $n \in [22, 322]$ by, once again, segmenting the range of the order parameter into overlapping windows of 4 bins wide along each axis. We use a bin width of 1 along n , and $1/16$ along ν and χ . In each window, the EPS algorithm performs path MC,³⁶ where the state of the Markov chain is some path $\vec{\sigma} = (\sigma_1, \dots, \sigma_\tau)$, with $\sigma_i \in \{(s, q)\}^N$ being a lattice configuration. In our implementation, a path is generated from a seed lattice state by propagating it a random number $m \in \{1, \dots, \tau - 1\}$ of sweeps forward and $\tau - m - 1$ sweeps backward, where a sweep is equivalent to N Metropolis MC moves and, for our purposes, the MC is time symmetric. We fix the path length $\tau = 7$. The generated path is accepted with probability 1 if at least one of the comprising states falls within the range of (n, ν, χ) , as specified by the window, and rejected otherwise. To generate a new path $\vec{\sigma}'$ from an already accepted path $\vec{\sigma}$, EPS selects a random snapshot σ_k of the old path and uses it as a seed for the new path. Upon seeding the path MC, we allow a relaxation stage which terminates after accepting 5×10^4 new configurations (up to 7 configurations per path). This is followed by a sampling stage, which aims to accept at least 10^5 new configurations.

All EPS calculations presented in this work were carried out in cubic systems with linear dimension of $L = 32$ lattice sites. For all conditions studied, this is sufficient to capture nuclei at the final stage of the nucleation pathway without self-interaction between periodic images. For computation of phase diagrams via multicanonical sampling with Wang-Landau recursion, the structural details of the mixed-phase region are unimportant allowing use of smaller system sizes $L \in \{4, 8\}$ for rapid convergence of bias energies.

IV. PHASE BEHAVIOUR AND NUCLEATION PATHWAYS

With the help of the multicanonical sampling method, as described in Sec. III, we verify that the solute-solvent, I–II and II–III phase transitions in our model are first-order, as is consistent with the current understanding of the $Q \geq 3$ 3D Potts models^{37–39} as well as previous studies of the Potts Lattice Gas (PLG) model.¹⁶ We obtain phase diagrams (Fig. 2) of our system for $c \in \{0.5, 1\}$, showing that the stability of the partially disordered solute phase is determined by the relative strength of anisotropic interactions of solute species. By considering the solute rich state of our model as an interweave of two decoupled $Q = 24$ state Potts lattices with 12 nearest neighbours (due to diagonal interactions), we show (Fig. 2(c)) that the coordinates β_\dagger of solute phase coexistence points are in close agreement with the existing mean field result,⁴⁰

$$\beta_\dagger^{-1}(Q) = RN_{\text{nbr}} \frac{Q-2}{Q-1} [2 \ln(Q-1)]^{-1}, \quad (6)$$

where R is the strength of anisotropic interaction between aligned particles and N_{nbr} is the maximum number of neighbours with which a solute particle can interact anisotropically. In our case, $N_{\text{nbr}} = 12$ due to diagonal interactions. While we observe finite widths of the stability region of the partially disordered phase for $c = 1$ on small

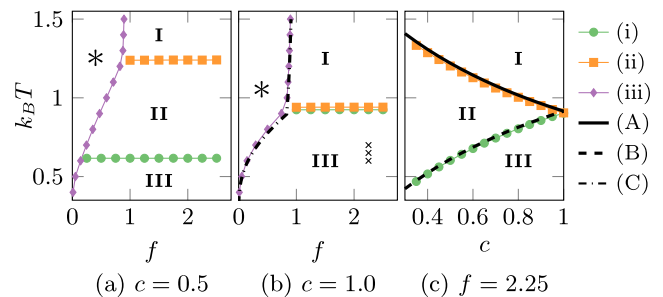


FIG. 2. Phase diagrams for different anisotropic interaction strengths, where f is the parameter controlling the solute saturation. Coexistence lines shown: (i) II and III, (ii) I and II, (iii) Solute and solvent. In (a) and (b), the regions of stability of the solvent rich state are marked by an asterisk (*). Shown in (c) is the comparison of mean field predictions (A) and (B) as given by (6), and computational results for the coordinates of the solute phase coexistence points. The equilibrium distributions of nuclei were sampled at parameter values indicated by black crosses in (b). Lines (i) and (ii) were estimated via the multicanonical approach described in Sec. III applied to the model on a cubic lattice with a linear dimension $L = 8$. Lines (iii) were estimated using the same approach but in a cubic system with $L = 4$. In (b), line (iii) is in good agreement with the mean field approximation (7) shown by (C).

lattices, we find that the proximity of the two solute phase coexistence lines increases with system size. Hence, in accordance with the mean field prediction, we expect the partially ordered phase to be metastable everywhere but the disorder–order coexistence line in the thermodynamic limit.

We derive an approximation to the solute–solvent coexistence line for $c = 1$ by considering the lattice adaptation⁴¹ of the Widom expressions^{42,43} for solute and solvent chemical potentials in the semigrand ensemble. Assuming that, at conditions of solute–solvent coexistence, the free energy change due to insertion of one solvent particle into a system with N solute particles is equal to that due to insertion of one solute particle into a system with N solvent particles, we arrive at the ideal gas like approximation for the coordinates $f_*(\beta)$ of solute–solvent coexistence points,

$$f_*(\beta) = \begin{cases} \left[Q(Q-1 + e^{\beta/2})^{-1} \right]^6, & \beta \leq \beta_*, \\ Q(Q-1 + e^{6\beta})^{-1/2}, & \beta \geq \beta_*, \end{cases} \quad (7)$$

where $(Q-1 + e^{\beta_*/2})^{12} Q^{-10} = Q-1 + e^{6\beta_*}$ determines the Q dependent value $\beta_*(Q)$ — the inverse temperature of the order–disorder transition. The estimates of solute–solvent coexistence points, obtained via the multicanonical sampling of $\ln \mathbb{P}(N_1)$ in a cubic $L = 4$ system, appear in good agreement with Eq. (7) (Fig. 2(b)). Additionally, we find that $\beta_*(Q)$ is in reasonable agreement with Eq. (6) for $Q \geq 24$, converging to $\beta_+(Q)$ as a power law for large Q .

We study the energetics of transition of our model between solvent rich and solute rich states at conditions where the system is most stable in the crystalline phase (Fig. 2(b)). Using EPS along with nearest neighbour interpolation based on Delaunay triangulation, we obtain free energy surfaces $\mathcal{F}(n, \nu, \chi) = -k_B T \ln \mathbb{P}(n, \nu, \chi)$ for $k_B T \in \{0.6, 0.65, 0.7\}$ (respectively, 33%, 28%, and 22% undercooling with respect to the order–disorder transition), at a constant value of the fugacity ratio $f = 2.25$ (167%, 160%, and 155% supersaturation for the respective temperatures). By examining the local minima $(\nu, \chi)^{\ddagger}(n) = \min_{(\nu, \chi)} \mathcal{F}(n, \nu, \chi)$, we note that the preferred orientational ordering of the solute nuclei varies with nucleus size. Thus, we argue that thermodynamic transition pathways, i.e., those limited to slow nuclei growth, starting in the solvent rich phase and leading to the solute crystal, proceed via amorphous and partially disordered precursors. Defining the critical nucleus size n^{\ddagger} as the maximum of free energy $\mathcal{F}(n) = -k_B T \ln \int_0^1 \int_0^1 d\nu d\chi \mathbb{P}(n, \nu, \chi)$, we further show that the preferred orientational ordering $(\nu, \chi)^{\ddagger}(n^{\ddagger})$ of the critical nuclei is temperature dependent (Fig. 3).

We compute the conditional probability distributions $\mathbb{P}(n|\{\nu\}, \{\chi\}) = \int_{\{\nu\}} \int_{\{\chi\}} d\nu d\chi \mathbb{P}(n, \nu, \chi)$ to obtain free energies $\mathcal{F}_{\mathbf{X}}(n) = -k_B T \ln \mathbb{P}(n|\{\nu\}_{\mathbf{X}}, \{\chi\}_{\mathbf{X}})$ of nuclei of the three solute phases, where \mathbf{X} is the phase label [disordered (I), partially ordered (II), or ordered (III)]. To do so, we define three basins in the (ν, χ) plane: (1) $\{\nu\}_{\text{I}} = [0, 0.5], \{\chi\}_{\text{I}} = [0, 1]$. (2) $\{\nu\}_{\text{II}} = (0.5, 1], \{\chi\}_{\text{II}} = [0, 0.5]$. (3) $\{\nu\}_{\text{III}} = (0.5, 1], \{\chi\}_{\text{III}} = (0.5, 1]$. Plots of the three free energy curves at different temperatures are shown in Fig. 4. Here we argue that at lower temperatures ($k_B T \leq 0.6$), the free

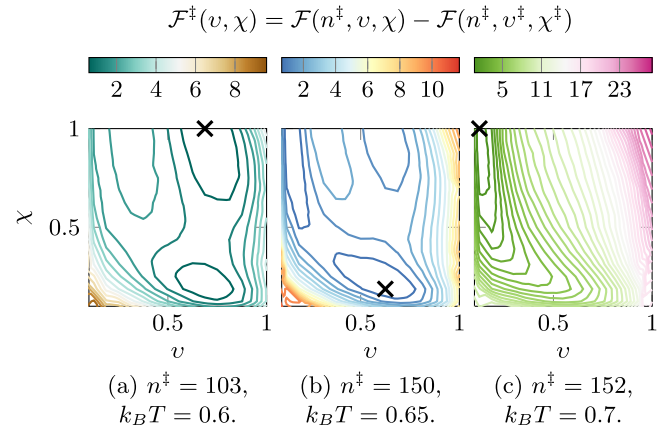


FIG. 3. Free energy surfaces $\mathcal{F}^{\ddagger}(\nu, \chi)$ for critical nuclei n^{\ddagger} at three different temperatures and constant fugacity ratio $f = 2.25$. Contour lines are drawn at intervals of $1.0k_B T$ and the coordinates $(\nu, \chi)^{\ddagger}(n^{\ddagger})$ of the local minima $\mathcal{F}^{\ddagger}(\nu^{\ddagger}, \chi^{\ddagger}) = 0$ are indicated by the crosses. The plots demonstrate the temperature dependence of preferred ordering $(\nu, \chi)^{\ddagger}(n^{\ddagger})$ of critical nuclei n^{\ddagger} . The three regimes shown are as follows: (a) critical nuclei most stable when ordered; (b) critical nuclei most stable when partially ordered; and (c) critical nuclei most stable when disordered.

energy barrier to nucleation of the ordered phase is lowest. The picture changes, however, at temperatures $k_B T \geq 0.7$, where the ordered phase has the highest free energy barrier. Hence, we anticipate a crossover regime at intermediate temperatures $0.6 < k_B T < 0.7$, where the heights of all three free energy barriers are comparable. This hints at the existence of a parameter regime where nuclei of all three solute phases can nucleate on similar time scales, with nuclei of the most stable phase growing at the expense of the others via dissolution and reprecipitation. However, as illustrated by the sparsity of contours between minima in Fig. 3, we find that barriers to solid state transformation between the solute phases are typically comparable to thermal energy at the critical nucleus size, or absent entirely, in either case vanishing completely for large n . This implies that the most probable route to reaching the thermodynamically stable phase is via transformations within the solute rich phase.

The vanishing of the barriers to solid state transformation within the larger nuclei is not surprising, since we study the model in a temperature regime far below the

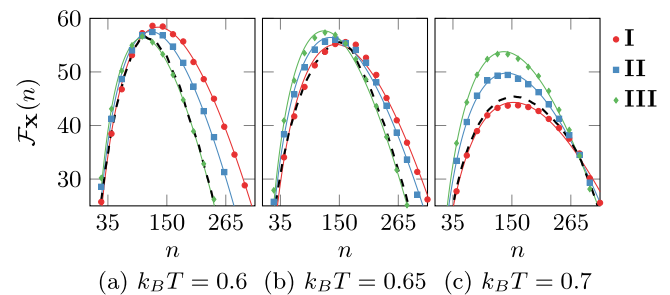


FIG. 4. Numerically obtained free energy curves $\mathcal{F}_{\mathbf{X}}(n)$ (markers) for nuclei of the three phases at $f = 2.25$, showing the corresponding CNT least squares fits (lines). For reference, shown by the dashed lines are estimates of the free energies $\mathcal{F}(n)$ obtained via a one-dimensional analogue of the EPS procedure described in Sec. III. We find both variants of the EPS approach in excellent agreement with respect to the estimates of $\mathcal{F}(n)$.

thermodynamic and kinetic (for $L \geq 32$ cubic systems) metastability limits of bulk disordered and partially disordered solute states. In addition, we are able to qualitatively reproduce the temperature dependence of the preferred orientational ordering within the small nuclei by studying the system in the solute rich state on small ($L < 10$) cubic lattices with reflecting boundary conditions, demonstrating that the relative thermodynamic stability of the three solute phases is strongly size dependent. Thus, the stability of disordered and partially disordered structures in small nuclei can be interpreted as analogous to a finite size effect.

The shapes of the three barriers in Fig. 4, with the exception of faceting effects, are well fitted by the classical nucleation theory (CNT) expression of the form $F(n) = F_0 + An^{2/3} - Bn$, where A and B are related to, respectively, the surface and volume free energy densities of the nucleus and F_0 relaxes the fit by allowing the free energy of the metastable solvent-rich phase to deviate from zero.¹⁶ The obtained least squares fits yield parameter values which are consistent with the energetics of our model—the cost of forming solute-solvent interface and the gain associated with the growth of solute domain both increase with the degree of solute ordering.

We observe a dramatic reduction in the quality of the CNT fits if setting $F_0 = 0$, i.e., fitting the commonly used expression $\Delta F(n) = F(n) - F(0)$, consistent with other reports of poor quantitative performance of the functional form $\Delta F(n) = An^{2/3} - Bn$ in representing explicitly computed free energy barriers.^{22,44,45} Fitting $\Delta F(n)$ to the *overall* nucleation barriers $\mathcal{F}(n)$ in the range $n \in [n^\ddagger - 50, n^\ddagger + 50]$, we find the barrier heights adequately captured by the CNT fits, suggesting that the standard CNT framework alone is sufficient to formulate an effective description of the energetics of nucleation in the present model, despite the evident multi-stage character of the process. This poses an interesting test case for the increasingly popular “seeding” method for nucleation barrier and rate estimation, which aims to parametrise a CNT based model of nucleation kinetics by extracting the necessary quantities from trajectories of the nucleus size coordinate. Below, we will assess the capability of the “seeding” method to formulate an effective description of the multi-stage nucleation process under different kinetic regimes, thus evaluating the sensitivity of the method to variations in kinetic growth pathways and violation of the Markov assumption on the kinetics of the nucleus size coordinate.

V. SEEDING METHOD

In the framework of CNT, the nucleation rate J is given by^{1,46}

$$J = \rho J_\ddagger^+ Z \exp[-\beta \Delta F(n^\ddagger)], \quad (8)$$

where ρ is the solute monomer density, J_\ddagger^+ is the rate of monomer attachment to the critical cluster, $F(n) = -k_B T \ln \mathbb{P}(n)$ is the free energy of nucleus of size n , $\Delta F(n) = F(n) - F(0)$, and $Z = \sqrt{-\beta F''(n^\ddagger)/2\pi}$ is the Zeldovich factor proportional to the square root of the second n derivative F'' of $F(n)$ evaluated at n^\ddagger — the size of the critical

nucleus, i.e., $n^\ddagger = \operatorname{argmax}_n \{F(n)\}$. We distinguish between the empirical value n^\ddagger of the critical nucleus, as used in Sec. IV, and the value n^\ddagger maximising the standard CNT formula for $F(n)$ since the two may not agree in general.

A closely related expression⁴⁷ to (8) can be derived by assuming the continuum approximation to time evolution of nucleus size $n(t)$ to obey the Langevin equation in the overdamped limit: $\dot{n}(t) = -\beta D_n \nabla_n F[n(t)] + \xi_n(t)$, where $\xi_n(t)$ is a random process satisfying $\langle \xi_n(t) \rangle = 0$ and $\langle \xi_n(t) \xi_n(t') \rangle = 2D_n \delta(t - t')$, with $\delta(t)$ being the Dirac delta function, and D_n —the diffusivity of the nucleus size coordinate at the top of the barrier $\Delta F(n)$ — is taken to be equal to the monomer attachment rate $J_\ddagger^+ = D_n$.

The “seeding” approach to nucleation rate calculations builds on the above equations and has been reported to yield accurate rate estimates from atomistic simulations when applied to nucleation from the melt.⁴⁸ Seeding, in this context, amounts to preparing the initial state of the system to contain a nucleus of size n_0 and observing the time evolution $n(t)$ with the initial condition $n(0) = n_0$. Taking the ensemble average on both sides of the overdamped Langevin equation yields

$$\langle \dot{n}_0 \rangle_{n_0} = \langle \dot{n}(0) \rangle_{n_0} = -\beta D_n \left. \frac{dF}{dn} \right|_{n_0}, \quad (9)$$

which, in principle, allows one to reconstruct the free energy profile $F(n)$ from the initial drift velocities $\langle \dot{n}_0 \rangle_{n_0}$. In practice, however, accurate estimation of drift velocities for the whole range of the relevant n_0 is often computationally expensive and the free energy barrier $\Delta F(n^\ddagger)$ is instead estimated by fitting the n derivative of the CNT expression,

$$\Delta F(n) = -\Delta\mu n + \gamma n^{2/3}, \quad (10)$$

where $\Delta\mu$ can be defined as the difference per particle between the free energies of solute and solvent rich states and γ is the product of the nucleus shape factor and the free energy per unit area of the nucleus surface, to a handful of estimates of $\langle \dot{n}_0 \rangle_{n_0}$ in the vicinity of n^\ddagger , where D_n is assumed to be approximately independent of n . In both cases, the mean initial drift velocities must be expressed in units of βD_n , leading to propagation of error in estimates of D_n to estimates of the barrier height, which is exponentially amplified in the rate calculation. We, therefore, pay particular attention to error analysis in this work.

A. Microscopic kinetics

In principle, the outcome of the nucleation barrier reconstruction method should be independent of the microscopic kinetics, provided that the relevant statistical ensemble of the system’s configurations is sampled appropriately and the solute chemical potentials are correctly maintained. The transmutation–reorientation (TR) MC move set, as defined in Sec. II, provides kinetics which correctly maintain the chemical potentials of all particle species at the expense of incorrectly representing mass transport due to the unphysical transmutation move. A more realistic representation of mass transport can be provided by replacing the transmutation move, in the TR move set, with nearest neighbour particle exchange move, i.e., analogous to Kawasaki dynamics in the

kinetic Ising model,^{49,50} while keeping the reorientation move and the move attempt probabilities unchanged. We will refer to this alternative move set as diffusion–reorientation (DR) kinetics. Due to the conservation of particle counts under the DR kinetics, solute depletion effects can occur during the nucleation process; however, as detailed in Sec. V C, part of the “seeding” approach is to ensure that these depletion effects are negligible.

Apart from being a more realistic model of particle transport, the DR kinetics differ from TR in two ways: (1) The Markov assumption on $n(t)$, as made by the “seeding” method, is violated more strongly due to the stronger memory effects in local particle density fluctuations.^{23,51} (2) The rates of nuclei growth are much lower, due to the diffusion limited character of the particle attachment process, allowing substantially greater time scales for structural relaxation of the growing nuclei between successive solute attachment events. By applying the “seeding” approach under both sets of kinetics, we are, therefore, able to assess the sensitivity of the barrier reconstruction method to deviations of $n(t)$ from the assumed Langevin dynamics description as well as differences in the observed nuclei growth pathways.

B. Fitting and error estimation procedures

In order to estimate the quantities $\langle \dot{n}_0 \rangle_{n_0}$ and D_n , one typically samples M independent trajectories $n_j(t) : n_j(0) = n_0$ at some number W of uniformly spaced time points $t_i = i\Delta t$, $i \in \{1, \dots, W\}$ for a range of initial cluster sizes n_0 . The estimator for the mean drift velocity $\langle \dot{n}_0 \rangle_{n_0}$ is given simply by the gradient of the least squares fit to the time series of the mean displacement $\langle \Delta n(n_0, t) \rangle = \langle n(t) - n_0 \rangle$. In accordance with CNT, one expects that trajectories $n(t) : n(0) = n^\ddagger$, starting at the critical cluster size, will yield $\langle \Delta n(n^\ddagger, t) \rangle = 0$, allowing the diffusivity D_n to be estimated via the halved gradient of the linear least squares fit to the time series of the mean squared displacement. In practice, however, deviations from the expected zero drift behaviour are common, and average squared deviation from the mean $\langle \text{SD}_n(n_0, t) \rangle = \langle [n(t) - \langle n(t) \rangle]^2 \rangle$ is used instead. Both $\langle \Delta n(n_0, t) \rangle$ and $\langle \text{SD}_n(n_0, t) \rangle$ can be expected to satisfy $\langle \Delta n(n_0, 0) \rangle = 0$ and $\langle \text{SD}_n(n_0, 0) \rangle = 0$, allowing the use of a linear model $g(t) \propto t$ with zero vertical offset for extraction of gradients. The mean drift velocity and diffusivity estimators are then given by, respectively, $v_n(n_0)$ and $\zeta_n(n^\ddagger)$, which, if writing out the averaging and least squares fitting operations explicitly, can be written as

$$v_n(n_0) = \frac{\sum_{i=1}^W t_i \sum_{j=1}^M \Delta n_j(n_0, t_i)}{M \sum_{i=1}^W t_i^2}, \quad (11)$$

$$\zeta_n(n_0) = \frac{\sum_{i=1}^W t_i \sum_{j=1}^M \text{SD}_n^{(j)}(n_0, t_i)}{2M \sum_{i=1}^W t_i^2}, \quad (12)$$

where $\Delta n_j(n_0, t)$ and $\text{SD}_n^{(j)}(n_0, t)$ are obtained from independent trajectories $n_j(t)$ with $n_j(0) = x_0$.

Assuming the CNT form of the nucleus free energy, as given by Eq. (10), we can extract the CNT critical nucleus

size n^\ddagger and the height of the free energy barrier $\Delta F(n^\ddagger)$,

$$n^\ddagger = \left(\frac{2\gamma}{3\Delta\mu} \right)^3, \quad \Delta F(n^\ddagger) = \frac{1}{2} n^\ddagger \Delta\mu, \quad (13)$$

by estimating γ via a fit derived from Eq. (9),

$$-\beta^{-1} D_n^{-1} \langle \dot{n}_0 \rangle_{n_0} = \frac{2}{3} \gamma n_0^{-1/3} - \Delta\mu. \quad (14)$$

For known $\Delta\mu$, the fitting procedure can be simplified by rearranging Eq. (14) for γ . If Brownian motion is a good approximation to $n(t)$ and Eq. (10) holds, one should find that $\gamma^\ddagger(n_0) = \gamma$, where

$$\gamma^\ddagger(n_0) = \frac{3}{2} \left[\Delta\mu - \frac{\langle \dot{n}_0 \rangle_{n_0}}{\beta D_n} \right] n_0^{1/3}, \quad (15)$$

hence the least squares estimator γ_{LS} of γ is given by an average over the range $n_0 \in [a, b]$,

$$\gamma_{\text{LS}} = k_\gamma \left[\Delta\mu \sum_{n_0=a}^b n_0^{1/3} - \frac{k_B T}{\zeta_n^*} \sum_{n_0=a}^b v_n(n_0) n_0^{1/3} \right], \quad (16)$$

where $k_\gamma = 3/(2(b-a+1))$ and $\zeta_n^* = \zeta_n(n^\ddagger)$. For normally distributed estimates $v_n(n_0)$ of $\langle \dot{n}_0 \rangle_{n_0}$ with known parameters, the parameters of the normal distribution of the sum $\sum_{n_0=a}^b v_n(n_0) n_0^{1/3}$ can be deduced via the scaling and addition properties of normal distributions. If estimates ζ_n^* of D_n also follow a normal distribution, then the quantity $\sum_{n_0=a}^b v_n(n_0) n_0^{1/3} / \zeta_n^*$ follows the distribution of the quotient of noncentral normal variates,^{52,53} for which the cumulative distribution function (CDF) is known and can be computed accurately by appropriately integrating the bivariate normal probability density function.⁵⁴ Thus, knowing the parameters of the normal distributions of $\zeta_n(n^\ddagger)$ and $v_n(n_0)$, we can compute the confidence interval on γ_{LS} as well as on $\Delta F(n^\ddagger)$.

The above error estimation approach makes a number of assumptions, in particular, that (1) estimators $v_n(n_0)$ and ζ_n^* are unbiased and normally distributed and (2) estimators $v_n(n_0)$ and ζ_n^* are independent. While (2) is straightforward to guarantee by using independent sets of trajectories for computation of $v_n(n_0)$ and ζ_n^* , we find that (1) is not guaranteed even if the dynamics of $n(t)$ exactly follow the overdamped limit of the Langevin equation. In the Appendix, we show that ζ_n^* may not follow the normal distribution for small M and can be a biased estimator of D_n if $n(t)$ is a trajectory of a 1D Brownian particle in a bistable potential. We, therefore, recommend use of $M \geq 100$ for estimation of D_n ; however, explicit verification of normality of ζ_n^* for large M may be infeasible in many applications of the “seeding” method.

C. Seed generation

We employ the “seeding” method in our lattice model to estimate the nucleation barrier heights in the range of $k_B T \in \{0.6, 0.65, 0.7\}$ and $f \in \{2, 3, \dots, 7\}$. At each parameter point, we sample $M = 10^2$ independent trajectories of the largest cluster size $n_j(t)$ at time values $t_i = i\Delta t$, $i \in \{1, \dots, W\}$ with $n_j(0) = n_0 \in [3, 498]$, $\Delta t = 10 \text{ MCS}$ and $W = 10^2$, where a unit of time represents a single MC sweep (MCS) of the whole lattice.

The initial configurations for each independent trajectory were generated by equilibrating a randomly grown compact cluster of solute particles in pure solvent. The initial compact clusters were generated iteratively by, starting with a single solute particle on the lattice, placing a solute particle into a randomly chosen neighbouring site of a randomly chosen member particle of the cluster until the desired cluster size is reached. The term “equilibration” here is used in reference to a statistical ensemble of configurations where the number of solute particles belonging to the constructed cluster is conserved. To achieve correct sampling from the said ensemble, we restricted the MC to the set of sites \mathbf{S} , which included solute sites belonging to the grown solute cluster and solvent sites in the cluster’s nearest neighbourhood. The equilibration procedure employed transmutation and reorientation moves on solute sites in \mathbf{S} as well as non-local particle swap moves on pairs of solute and solvent sites in \mathbf{S} , with all move types having equal attempt probabilities. Acceptance of a non-local particle swap move may lead to a change in \mathbf{S} , which we took into account in the calculation of the move acceptance probability to assure detailed balance. Moves leading to a change in the size of the grown cluster were rejected with probability 1.

We find that after a burn-in period of $10^3 n$ iterations of the procedure, the nuclei attain compact shapes and the corresponding distributions of internal orientational order parameter are consistent with those obtained via the three-dimensional EPS approach applied to the μVT ensemble. The agreement between the two methods of sampling distributions of cluster orientational order parameter is not surprising, since the difference between the effects of pure solvent and saturated solution on the structure of the nuclei is negligible due to the short range nature of particle interactions and the low counts of solute particles in the supersaturated solution. Upon equilibration of the nucleus, we fill the lattice uniformly with $N\rho(k_B T, f)$ solute particles, where $\rho(k_B T, f)$ is the fugacity and temperature dependent solute concentration, which we compute numerically as the mean fraction of solute particles in the system in the metastable state.

Following seeding, the trajectories $n_j(t)$ were generated under composition conserving unconstrained DR dynamics,^{49,50} i.e., equally probable reorientation and nearest neighbour particle exchange moves, as well as the composition nonconserving TR dynamics described in Sec. II. Since DR dynamics samples system configurations from the NVT ensemble, conserving the solute particle count, the resultant evolution of nucleus size can be affected by depletion or augmentation of solute—an unphysical effect in open systems.⁴ To avoid this effect in our sample of trajectories $n_j(t)$, we employ the larger cubic lattice sizes of $L = 64$. In addition, we verify that the maximum change in nucleus size, along all obtained trajectories for DR dynamics, is negligible in comparison to the total number of the available solute particles in the system.

D. Drift velocities and diffusivities

Typical realisations of $\langle \Delta n(n_0, t) \rangle = \langle n(t) - n_0 \rangle$ are shown in Fig. 5, where we see that, consistent with the intuition

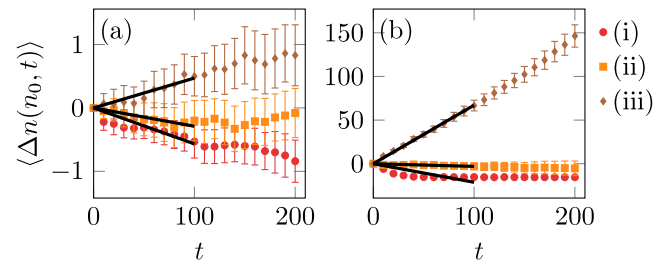


FIG. 5. Plots of mean displacements $\langle \Delta n(n_0, t) \rangle$ against time t for DR [(a) $n^* = 82$] and TR [(b) $n^* = 78$] dynamics at $k_B T = 0.7$, $f = 3$. Error bars represent the 95% confidence intervals based on 10^2 observations assuming normal distribution. The three sets of points correspond to trajectories initialised with n_0 given by (i) $n^* - 60$, (ii) n^* , and (iii) $n^* + 10^2$. Solid line segments represent the linear fits with gradients $v_n(n_0)$ used as estimates of initial drift velocity $\langle \dot{n}_0 \rangle_{n_0}$ along the cluster size coordinate.

of Sec. V A, nuclei growth rates for TR kinetics are approximately two orders of magnitude faster than those for DR. We observe roughly linear behaviour of $\langle \Delta n(n_0, t) \rangle$ on short time scales $t \leq 100 \text{ MCS}$; however, based on 10^2 observations, we expect considerable relative error margins on initial drift velocities measured under DR dynamics. Estimates $v_n(n_0)$ of initial drift velocities $\langle \dot{n}_0 \rangle_{n_0}$ along the cluster size coordinate were calculated by fitting straight lines with zero vertical offset (Eq. (11)) to time series $\langle \Delta n(n_0, t) \rangle$ over the range $t \leq 100 \text{ MCS}$. Estimates $\zeta_n(n_0)$ of cluster size dependent diffusivity D_n along n were extracted from time series $\langle \text{SD}_n(n_0, t) \rangle = \langle [n(t) - \langle n(t) \rangle]^2 \rangle$, $n(0) = n_0$ (Fig. 6(a)) in a similar fashion [Eq. (12)].

In accordance with CNT, trajectories starting at the critical cluster size n^\dagger are expected to yield $v_n(n^\dagger) \approx 0$. Computing the value n^\dagger , however, requires knowledge of the value γ_{LS} , which, in itself, requires the estimate ζ_n^* and, hence, the coordinate of the barrier top. Thus, we estimate the coordinate of the barrier top as $n^* = \text{argmax}_n \sum_{n_0=3}^n -v_n(n_0)$, which exploits the proportionality of $dF(n)/dn$ to the mean initial drift velocity in Eq. (9). Again, it is important to stress that n^* is an empirical value of the critical cluster size, which may differ from the fitted CNT value n^\dagger . We observe that within the range $n_0 \in [n^* - 20, n^* + 20]$ of initial cluster sizes n_0 , the gradient estimates $\zeta_n(n_0)$ of the mean squared displacements

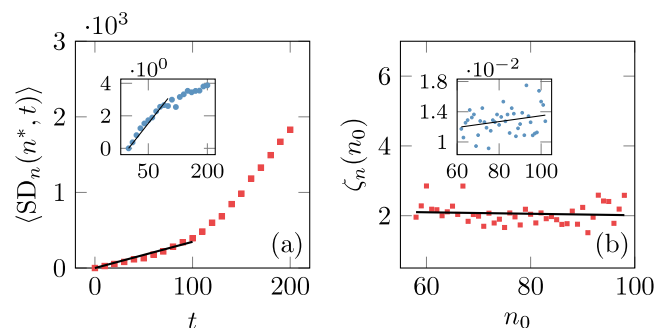


FIG. 6. Plots of $\langle \text{SD}_n(n_0, t) \rangle$ for $n_0 = n^*$ (a) and diffusion coefficient estimates ζ_n for n_0 in the range $n_0 \in [n^* - 20, n^* + 20]$ (b) obtained under TR (main panels) and DR (insets) kinetics at $k_B T = 0.7$, $f = 3$. In (a), the solid line segments represent linear fits used in estimation of cluster size diffusivity $\zeta_n(n^*)$. In (b), solid lines represent linear fits used for detrending the data for estimation of statistical uncertainty in $\zeta_n(n^*)$.

vary slowly with n_0 , as shown in Fig. 6(b), which is consistent with the intuition that diffusivity along the n coordinate is approximately constant close to the top of the free energy barrier.¹ We therefore use the data for $\zeta_n(n_0)$ in the range $n_0 \in [n^* - 20, n^* + 20]$ to estimate the statistical uncertainty in our measurement $\zeta_n(n^*)$ of the diffusivity D_n along the cluster size coordinate.

E. Choice of $\Delta\mu$

In our model the degree of saturation of the solution is dependent only on the difference between solute and solvent chemical potentials. Thus we can set $\mu_1 = 0$ without loss of generality, allowing the chemical potential $\mu_2 = \mu_3$ of solute to fully determine the chemical composition of the system. Therefore, the standard choice of the value of the CNT parameter $\Delta\mu$, in our case, would be

$$\Delta\mu_{\text{coex}}(\beta, f) = \beta^{-1}[\ln f - \ln f_*(\beta)], \quad (17)$$

with $f_*(\beta)$ given by Eq. (7), corresponding to the difference between the bulk solute chemical potentials in states of metastable supersaturation and solute-solvent coexistence.^{1,4,55} It is, however, understood that usage of the bulk value $\Delta\mu = \Delta\mu_{\text{coex}}$ as the driving force to nucleation may be unsuitable for the description of the microscopic particle attachment process.⁵⁶ An alternative value $\Delta\mu_{\text{fit}}(\beta, f)$ can be obtained via a two parameter fit of Eq. (14), allowing the appropriate value of $\Delta\mu$ to be determined by the microscopic dynamics of the system.

Due to the comparatively low computational cost of free energy calculations in our model, we are able to compute the nucleation barriers over the specified range of parameter values explicitly with high accuracy by utilizing the one-dimensional analogue of the EPS procedure, described in

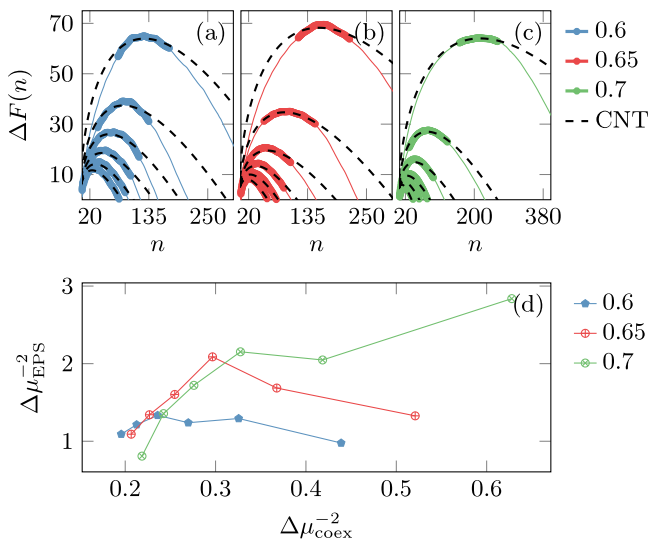


FIG. 7. Fits of Eq. (10) (dashed lines) to EPS data (solid lines) for (a) $k_B T = 0.6$, (b) $k_B T = 0.65$, and (c) $k_B T = 0.7$. The six curves in (a), (b), and (c) are sampled at the six values of $f \in \{2, 3, \dots, 7\}$ with higher barriers corresponding to lower values of f . The thicker portions of the solid lines in (a), (b), and (c) indicate the ranges of data used in the corresponding two parameter fits of Eq. (10). The corresponding fitted values $\Delta\mu_{\text{EPS}}$ are plotted in (d) as $\Delta\mu_{\text{EPS}}^{-2}$ against $\Delta\mu_{\text{coex}}^{-2}$ given by Eq. (17).

Sec. III, applied to cubic systems of length $L = 32$ sites. To assure absence of finite size effects in the obtained barrier estimates, we verify that the EPS data are reproducible in smaller systems with $L = 16$. We obtain data which are adequately described by Eq. (10) (Fig. 7), yielding the set of values $\Delta\mu_{\text{EPS}}(\beta, f)$ via a two parameter fit of Eq. (10) to the EPS estimates of $\Delta F(n)$ over a range of 100 values of n near the top of the barrier (Fig. 7(d)).

We now consider the effect of the choice of $\Delta\mu$ on the capacity of the “seeding” method to reconstruct the explicitly computed free energy barriers by carrying out the fitting procedure [Eq. (16)] for each of the three possible choices: $\Delta\mu_{\text{coex}}$, $\Delta\mu_{\text{fit}}$, and $\Delta\mu_{\text{EPS}}$. We first obtain estimates of $\Delta\mu_{\text{fit}}$ via two parameter fits of Eq. (14) to the scaled average initial drift velocities for DR and TR dynamics over a range of $n_0 \in [n^* - 20, n^* + 20]$. Close to the top of the barrier, i.e., $n \in [n^* - 20, n^* + 20]$, our data show reasonable agreement with $\gamma^\dagger(n) = \text{const.}$ [Eq. (15)] for $\Delta\mu \neq \Delta\mu_{\text{coex}}$ (Figs. 8(b) and 8(c)), while the usage of the bulk value $\Delta\mu_{\text{coex}}$ often yields a clear n dependence of $\gamma^\dagger(n)$ (Fig. 8(a)). Despite the comparable quality of the three fits of Eq. (14) (Fig. 8(d)), with occasionally exceptionally poor fits using

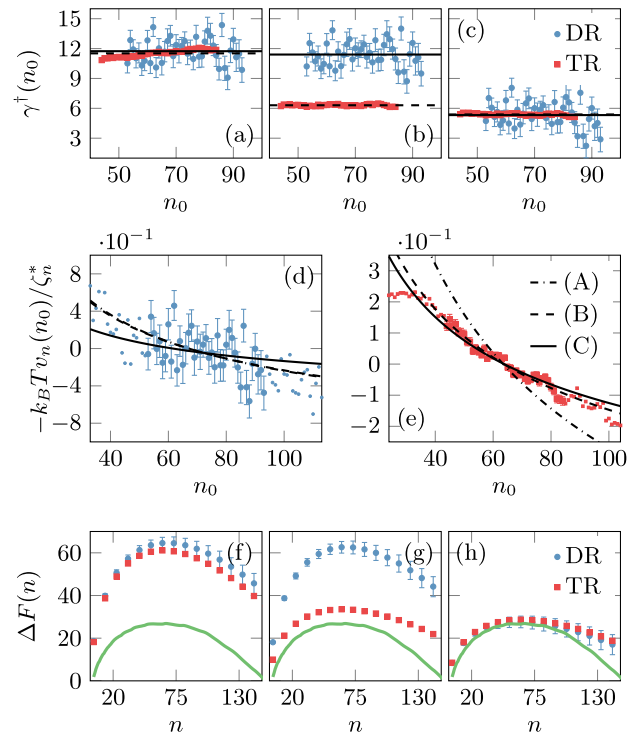


FIG. 8. Fits of CNT to scaled mean drift velocity data via Eq. (16) for $k_B T = 0.6$, $f = 4$. The shown data points were obtained under DR (circles) and TR (squares) dynamics. Error bars represent the 95% confidence intervals obtained via the CDF of noncentral normal quotients. In (a) ($\Delta\mu = \Delta\mu_{\text{coex}}$), (b) ($\Delta\mu = \Delta\mu_{\text{fit}}$), and (c) ($\Delta\mu = \Delta\mu_{\text{EPS}}$), we plot the data used for fitting of γ^\dagger via Eq. (16), with dashed and solid lines corresponding to the fits to TR and DR data, respectively, over the range of $n_0 \in [n^* - 20, n^* + 20]$. In (d) and (e), we show the corresponding fits to the estimates of $-\beta^{-1} D_n^{-1}(n_0)n_0$ [Eq. (14)] for a wider range of n_0 , with data points outside the fitting range shown as small markers with $\zeta_n^* = \zeta_n(n^*)$. The curves (A), (B), and (C), in (d) and (e), correspond, respectively, to the three choices of $\Delta\mu$ values: $\Delta\mu_{\text{coex}}$, $\Delta\mu_{\text{fit}}$, and $\Delta\mu_{\text{EPS}}$. Finally in (f) ($\Delta\mu = \Delta\mu_{\text{coex}}$), (g) ($\Delta\mu = \Delta\mu_{\text{fit}}$), and (h) ($\Delta\mu = \Delta\mu_{\text{EPS}}$), we show the resultant shapes of the CNT free energy barriers in comparison to the $\Delta F(n)$ obtained via EPS (solid line).

$\Delta\mu_{\text{coex}}$ (Fig. 8(e)), we find that barrier reconstruction via the “seeding” approach is highly sensitive to the choice of $\Delta\mu$ (Figs. 8(f)–8(h)), only yielding a consistent agreement with EPS for the two sets of dynamics if using $\Delta\mu = \Delta\mu_{\text{EPS}}$.

To illustrate the sensitivity of the “seeding” method further, we plot all barrier height estimates in Fig. 9. For $f \geq 3$, the EPS barrier height estimates fall on a straight line $\Delta F(n^\ddagger) \propto \Delta\mu_{\text{coex}}^{-2}$ (Fig. 9(a)) as is consistent with CNT.¹ Closer to coexistence ($f = 2$) for $k_B T \in \{0.65, 0.7\}$, however, the linear trend does not hold, which cannot be accounted for by statistical errors or finite size effects in our free energy calculations, and, therefore, may be due to the errors in the mean field approximation to $f_*(\beta)$ in Eq. (7). We find that usage of the bulk values of $\Delta\mu$ yields a systematic error of at least a factor of 2 in the barrier height estimates due to the “seeding” method (Fig. 9(b)). Although the approach employing a two parameter fit of Eq. (14) can yield reasonable agreement between EPS and the “seeding” method data, usage of $\Delta\mu = \Delta\mu_{\text{fit}}$ leads to barrier estimates which vary greatly with temperature, saturation, and the choice of the system’s kinetics (Fig. 9(c)). Setting $\Delta\mu = \Delta\mu_{\text{EPS}}$, on the other hand, recovers barrier height estimates which are in excellent agreement with the explicit free energy calculations for $f \geq 3$ independent

of the system’s kinetics, yet deviate by, roughly, 43% for $k_B T \in \{0.6, 0.65\}$, $f = 2$ (Fig. 9(c)). Thus, we argue that, even for a well chosen $\Delta\mu$, the “seeding” method cannot guarantee an accurate estimate of the nucleation barrier height for low supersaturations in our model. In our case, particularly, the 43% error in barrier height can lead to underestimation of the nucleation rate by up to 10 orders of magnitude when using Eq. (8).

The kinetic prefactor $\rho D_n Z$ of Eq. (8) can readily be obtained from our calculations, with estimates $\zeta_n(n^*)$ of D_n being, as to be expected, the only significant kinetics dependent contribution to the rate J for $\Delta\mu = \Delta\mu_{\text{EPS}}$. For the explicitly computed nucleation barriers, the Zeldovich factor can be estimated via a parabolic fit near n^\ddagger , which allows us to compare rate estimates due to EPS and the “seeding” method. Taking into account the statistical errors in our estimates, we arrive at the same observations analysing J as a function of $\Delta\mu_{\text{coex}}$ as outlined above, obtaining reasonable agreement with the CNT prediction $\ln J \propto \Delta\mu_{\text{coex}}^{-2}$.

VI. CONCLUSIONS

We have introduced a novel multicomponent lattice model, in which the slow growth limited solute crystallisation pathway proceeds via the metastable disordered and partially disordered solute phases. We have shown that the heights of the barriers to nucleation of the metastable phases in relation to that of the stable crystal vary with temperature, leading to a parameter regime where the free energy barriers to nucleation of all three phases are equal. Due to the low barriers to solid state transformation between the three solute phases, we argue that the present model cannot expect to favour the dissolution–regrowth pathway relevant to homogeneous nucleation of calcite from solution.¹⁰ Design of minimal models to capture this process needs to incorporate large energy barriers to direct transformation between solute-rich phases. We hope to report on such models in a future communication.

Turning our attention to nucleation kinetics and the “seeding” method, given that the nucleation barrier height is exponentiated in the CNT rate expression, the largest contribution to error in rate estimates via the “seeding” method lies in the barrier reconstruction. We have shown how to estimate statistical uncertainties in this procedure, and hence demonstrated statistically significant deviation of nucleation barrier height estimates from those obtained via explicit free energy calculations, subject to the definition of the CNT parameter $\Delta\mu$. We found that the discrepancies between the two barrier estimation methods vanish over a broad range of parameter space, with the exception of conditions of lower supersaturation, if using values of $\Delta\mu$ informed by the shape of the explicitly computed free energy barriers. At lower supersaturations, however, the discrepancies between the two methods remain significant, which cannot be explained by the errors in our calculations. A possible source for these discrepancies lies in our choice of reaction coordinate. It is known that the choice of the reaction coordinate plays a major role in quantitative treatment of nucleation.⁵⁷ While we

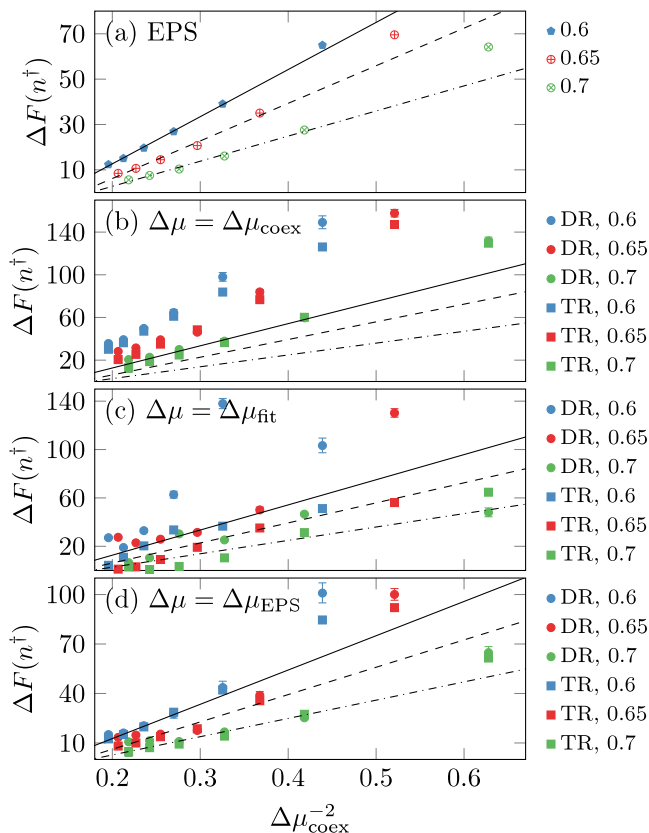


FIG. 9. Comparison of nucleation barrier height estimates obtained via EPS and the “seeding” method. In (a) we plot the EPS estimates of barrier heights against $\Delta\mu_{\text{coex}}^{-2}$, yielding good agreement with the linear trend $\Delta F(n^\ddagger) \propto \Delta\mu_{\text{coex}}^{-2}$ for $f \geq 3$ at the three values of $k_B T$. In (b)–(d), we show the barrier height estimates obtained via the “seeding” method under DR (circles) and TR (squares) dynamics, taking the $\Delta\mu$ values as, respectively, $\Delta\mu_{\text{coex}}$, $\Delta\mu_{\text{fit}}$, and $\Delta\mu_{\text{EPS}}$, in relation to the linear fits for $f \geq 3$ to the EPS data. Error bars represent the 95% confidence intervals obtained via the CDF of noncentral normal quotients.

cannot completely rule out the possibility of having chosen a suboptimal definition of n , results of the committor test⁵⁸ suggest the quality of our definition is at worst comparable to those used in off-lattice simulations.

Additional explanations for the failure of the “seeding” method to correctly reconstruct $F(n)$ at weak supersaturation lie in the underlying assumptions of CNT. The functional form of $\Delta F(n)$ remains a subject of debate.^{22,59} In this work we have not made any assumptions regarding the shape of nuclei except that surface area scales as $n^{2/3}$. The single cluster assumption underlying this statement is potentially problematic close to $n = 0$ where the entropic benefit of multiple clusters may offset the increased interfacial free energy. This trade-off is system size dependent and not corrected for in our free energy calculations.

The “seeding” method also relies on the assumption that the dynamics of $n(t)$ can be approximated by the over-damped Langevin equation. This assumption enters into the calculation of the barrier height as well as the kinetic prefactor. The validity of the Markov assumption for $n(t)$ is expected to depend on the choice of microscopic kinetics. It is known that TR dynamics yield only approximately Markovian $n(t)$, while the behaviour of $n(t)$ under the more realistic mass transport of DR dynamics completely violates the Markov assumption.²³ Despite the apparent differences between the two sets of kinetics, we observed a remarkable agreement between the two corresponding free energy barrier reconstructions, particularly at higher supersaturations, suggesting that the Markov assumption does not play a great role here. The agreement also suggests that the “seeding” method can recover consistent barrier estimates in different regimes of nuclei growth, provided that the structures of the initial seeds are correctly prepared. In future communications, we will report on direct calculations of nucleation rates via path sampling methods, with the hope to examine more closely the role of the Markov assumption in computing accurate nucleation rates.

ACKNOWLEDGMENTS

We would like to thank Stefan Grosskinsky (Centre for Complexity Science, University of Warwick), Carina Karner (Computational Physics, University of Vienna), and the authors of Ref. 16 for useful discussions and advice. We are also grateful for support from EPSRC, Grant Nos. EP/E501311/1 and EP/H00341X/1. Additionally, we gratefully acknowledge the use of facilities provided by the University of Warwick Centre for Scientific Computing and Queen Mary’s MidPlus computational facilities, supported by QMUL Research-IT, and funded by EPSRC Grant No. EP/K000128/1. We thank the anonymous referee for providing detailed and constructive feedback for this article. Data associated with this work is available via the Warwick Research Archive Portal at <http://wrap.warwick.ac.uk/81418/>.

APPENDIX: SEEDING METHOD FOR BROWNIAN MOTION

We now illustrate the “seeding” approach by computing barrier crossing rates j_x of a one-dimensional Brownian

particle in the bistable potential $U(x) = x^4 - 2x^2$. The equivalent to the CNT rate expression [Eq. (8)] for this system is the Kramers’⁶⁰ rate,

$$j_x \approx \beta D_x \exp \{-\beta \Delta U\} \sqrt{|U''(-1)| |U''(0)|} / (2\pi), \quad (\text{A1})$$

where $D_x = k_B T$ is the diffusivity of the Brownian particle, $\Delta U = U(0) - U(-1)$ is the potential barrier height, and $U''(x) = 12x^2 - 4$. Employing the BAOAB-limit method,⁶¹ we estimate the initial drifts $\langle \dot{x}_0 \rangle_{x_0}$ and D_x taking $x_0 \in \{-0.9, -0.8, \dots, 0.9\}$ at $k_B T \in \{0.15, 0.175, \dots, 0.25\}$. We employ the estimators $v_x(x_0)$ and $\zeta_x(x_0)$, computed in the same fashion as shown by Eqs. (11) and (12), taking $\Delta t = 10^{-4}$ and $W \in \{5, 10, 100\}$.

Assuming that on short time scales the coordinate $x(t)$ evolves according to $\dot{x}(t) \approx -\beta D_x \nabla_x U(x_0) + \xi_x(t)$, it is straightforward to obtain the exact distributions of $\Delta x(x_0, t)$ and $\text{SD}_x(x_0, t)$ as, respectively, normal — $\mathcal{N}[t \nabla_x U(x_0), 2D_x t]$, and gamma — $\Gamma(0.5, 4D_x t)$. While the estimator v_x is clearly normally distributed, the exact form of the distribution of ζ_x is much more complex,⁶² however we find that for $M \geq 100$ the normal distribution gives a reasonable approximation (Fig. 10(a)). This is to be expected, since the probability distribution of the sum $\sum_{j=1}^M \text{SD}_x^{(j)}(x_0, t)$ is $\Gamma(M/2, 4MD_x t)$, which is well approximated by the normal distribution for large M .

In light of the computational expense of obtaining large numbers of trajectory sets, it might be tempting to estimate the error in ζ_x by fitting least squares lines to individual realisations of $\text{SD}_x^{(j)}(x_0, t)$. While this approach will recover the correct mean value, the distribution of estimates will drastically deviate from normal (Fig. 10(b)) yielding a wide and highly asymmetric confidence bound. In fact, for low M the estimator $\zeta_x(x^\dagger)$, where $x^\dagger = 0$ is the coordinate of the barrier top, does not correspond to the maximum likelihood estimator of D_x .

We find that both $\langle \Delta x(x_0, t) \rangle$ and $\langle \text{SD}_x(x_0, t) \rangle$ are well fitted by the linear model over the chosen range of times and temperatures (Fig. 11). While the estimates of drift velocities match the gradient of the potential well (Fig. 12(a)), we observe a parabolic deviation of the diffusion coefficient estimates ζ_x from the known value $D_x = k_B T$ (Fig. 12(b)). In particular, at the top of the barrier, we expect values

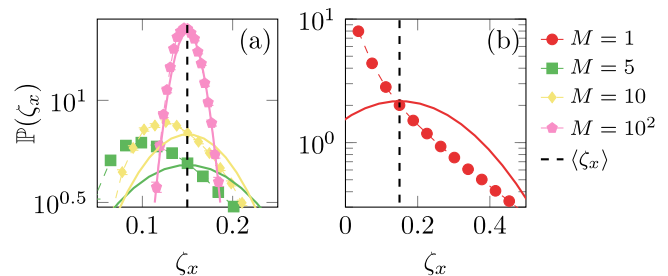


FIG. 10. Typical distributions of the estimator ζ_x at $k_B T = 0.15$, $W = 10$ and varying M (markers), and the corresponding normal distribution fits (solid lines). The deviation from normality is clear for low M ; however, it is negligible at $M = 100$. The $M = 1$ estimator corresponds to fitting linear trends to individual trajectories $\text{SD}_x^{(j)}(x_0, t)$ and the resultant distribution is highly asymmetric.

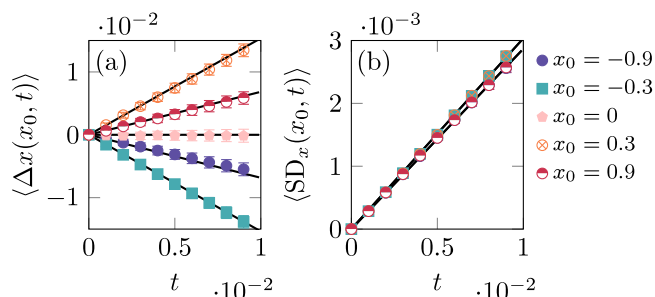


FIG. 11. Mean displacements (a) and mean squared displacements (b) for $x(t)$ against time at $k_B T = 0.15$, $M = W = 10^2$. Error bars represent the 95% confidence intervals, the width of which is typically smaller than that of the markers. The solid lines in (a) are drawn with the gradient given by $\nabla_x U(x_0)$. In (b), the solid lines represent the linear least squares fits to the data.

of ζ_x between 1% and 4% higher than $k_B T$. Factoring in the error of barrier height and D_x estimates due to increasingly high drift contributions to $\langle SD_x(x_0, t) \rangle$ away from the barrier top, we observe good agreement of Eq. (A1) with brute force estimates of inverse mean first passage times⁶³ (Fig. 12(c)). It is important to highlight that the derivation of Eq. (A1) relies on harmonic approximations of the potential $U(x)$ near the top and bottom of the barrier. These approximations, in our case, lead to small errors (Fig. 12(c)) which are cancelled by the systematic errors in estimates of the diffusion coefficient. A more accurate value of j_x can be obtained by numerically integrating the Kramers'⁶⁰ expression for the stationary current: $j_x = D_x \mathcal{P}(-1) e^{\beta U(-1)} / \int_{-1}^1 dx e^{\beta U(x)}$, out of the quasi-stationary state $\mathcal{P}(x)$ near $x = -1$, of the appropriate Fokker-Planck equation.

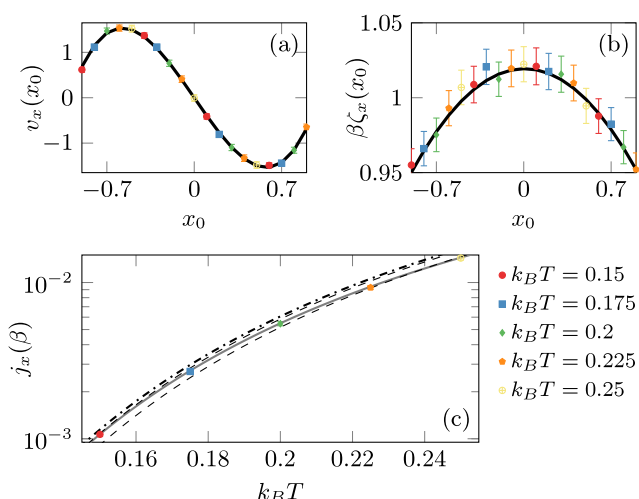


FIG. 12. Mean drift velocity (a), diffusion coefficient (b), and rate (c) estimates for the Brownian dynamics in $U(x) = x^4 - 2x^2$. In (a), the mean drifts are in excellent agreement with the gradient of the potential shown as a solid line. In (b), the estimates of the diffusion coefficient vary parabolically in the vicinity of the barrier top—the solid line shows a parabolic fit to the data to guide the eye. In (c), the points represent brute force computed inverse mean first passage time estimates, dashed lines show the range of rate estimates obtained via the procedure analogous to the “seeding” method, while the solid grey line gives the more accurate values of j_x computed via numerical integration of the Kramers’ stationary current equation. The dashed-dotted line in (c) is the plot of Eq. (A1) taking $\beta D_x = 1$.

- ¹V. Agarwal and B. Peters, “Solute precipitate nucleation: A review of theory and simulation advances,” in *Advances in Chemical Physics: Volume 155* (John Wiley & Sons, Inc., 2014), pp. 97–160.
- ²J. Anwar and D. Zahn, *Angew. Chem., Int. Ed.* **50**, 1996 (2011).
- ³F. Giberti, M. Salvalaglio, and M. Parrinello, *IUCr* **2**, 256 (2015).
- ⁴N. E. R. Zimmermann, B. Vorselaars, D. Quigley, and B. Peters, *J. Am. Chem. Soc.* **137**, 13352 (2015).
- ⁵W. E. Müller, *Molecular Biomineralization* (Springer-Verlag, 2011).
- ⁶S.-H. Yu and S. Chen, “Biomineralization: Self-assembly processes,” in *Encyclopedia of Inorganic and Bioinorganic Chemistry* (John Wiley & Sons, Ltd., 2011).
- ⁷S. Mann, B. R. Heywood, S. Rajam, and J. D. Birchall, *Nature* **334**, 692 (1988).
- ⁸J. Aizenberg, *J. Cryst. Growth* **211**, 143 (2000).
- ⁹K. Naka and Y. Chujo, *Chem. Mater.* **13**, 3245 (2001).
- ¹⁰J. D. Rodriguez-Blanco, S. Shaw, and L. G. Benning, *Nanoscale* **3**, 265 (2011).
- ¹¹R. Demichelis, P. Raiteri, J. D. Gale, and R. Dovesi, *CrystEngComm* **14**, 44 (2012).
- ¹²D. Quigley and P. M. Rodger, *J. Chem. Phys.* **128**, 221101 (2008).
- ¹³D. Quigley, C. L. Freeman, J. H. Harding, and P. M. Rodger, *J. Chem. Phys.* **134**, 044703 (2011).
- ¹⁴A. F. Wallace, L. O. Hedges, A. Fernandez-Martinez, P. Raiteri, J. D. Gale, G. A. Waychunas, S. Whitlam, J. F. Banfield, and J. J. De Yoreo, *Science* **341**, 885 (2013).
- ¹⁵S. Whitlam, *J. Chem. Phys.* **132**, 194901 (2010).
- ¹⁶N. Duff and B. Peters, *J. Chem. Phys.* **131**, 184101 (2009).
- ¹⁷L. O. Hedges and S. Whitlam, *J. Chem. Phys.* **135**, 164902 (2011).
- ¹⁸D. P. Sanders, H. Larralde, and F. Leyvraz, *Phys. Rev. B* **75**, 132101 (2007).
- ¹⁹A. Okamoto, T. Kuwatani, T. Omori, and K. Hukushima, *Phys. Rev. E* **92**, 042130 (2015).
- ²⁰A. J. Page and R. P. Sear, *Phys. Rev. Lett.* **97**, 065701 (2006).
- ²¹R. P. Sear, *J. Phys.: Condens. Matter* **24**, 052205 (2012).
- ²²S. Ryu and W. Cai, *Phys. Rev. E* **81**, 030601 (2010).
- ²³J. Kuipers and G. T. Barkema, *Phys. Rev. E* **82**, 011128 (2010).
- ²⁴B. C. Knott, V. Molinero, M. F. Doherty, and B. Peters, *J. Am. Chem. Soc.* **134**, 19544 (2012).
- ²⁵E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. F. Abascal, and C. Valeriani, *J. Am. Chem. Soc.* **135**, 15008 (2013).
- ²⁶J. R. Espinosa, C. Vega, C. Valeriani, and E. Sanz, *J. Chem. Phys.* **142**, 194709 (2015).
- ²⁷V. Agarwal and B. Peters, *J. Chem. Phys.* **140**, 084111 (2014).
- ²⁸F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- ²⁹A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
- ³⁰N. B. Wilding, *Am. J. Phys.* **69**, 1147 (2001).
- ³¹A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
- ³²S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, *J. Comput. Chem.* **13**, 1011 (1992).
- ³³F. Zhu and G. Hummer, *J. Comput. Chem.* **33**, 453 (2012).
- ³⁴R. Radhakrishnan and T. Schlick, *J. Chem. Phys.* **121**, 2436 (2004).
- ³⁵F. Schmitz, P. Virnau, and K. Binder, *Phys. Rev. E* **87**, 053302 (2013).
- ³⁶C. Dellago, P. G. Bolhuis, and P. L. Geissler, “Transition path sampling,” in *Advances in Chemical Physics* (John Wiley & Sons, Inc., 2003), pp. 1–78.
- ³⁷F. Y. Wu, *Rev. Mod. Phys.* **54**, 235 (1982).
- ³⁸A. Bazavov, B. A. Berg, and S. Dubey, *Nucl. Phys. B* **802**, 421 (2008).
- ³⁹C. Bonati and M. D’Elia, *Phys. Rev. D* **82**, 114515 (2010).
- ⁴⁰L. Mittag and M. J. Stephen, *J. Phys. A: Math., Nucl. Gen.* **7**, L109 (1974).
- ⁴¹D. Winter, P. Virnau, and K. Binder, *J. Phys.: Condens. Matter* **21**, 464118 (2009).
- ⁴²B. Widom, *J. Chem. Phys.* **39**, 2808 (1963).
- ⁴³D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, 2002).
- ⁴⁴L. Filion, M. Hermes, R. Ni, and M. Dijkstra, *J. Chem. Phys.* **133**, 244115 (2010).
- ⁴⁵S. Jungblut and C. Dellago, *J. Chem. Phys.* **134**, 104501 (2011).
- ⁴⁶H. Vehkamäki, *Classical Nucleation Theory in Multicomponent Systems* (Springer Science and Business Media, 2006).
- ⁴⁷P. Hänggi, P. Talkner, and M. Borkovec, *Rev. Mod. Phys.* **62**, 251 (1990).
- ⁴⁸J. R. Espinosa, C. Vega, C. Valeriani, and E. Sanz, *J. Chem. Phys.* **144**, 034501 (2016).
- ⁴⁹K. Kawasaki, *Phys. Rev.* **145**, 224 (1966).
- ⁵⁰R. P. Sear, *EPL* **83**, 66002 (2008).

- ⁵¹J. Kuipers and G. T. Barkema, [Phys. Rev. E](#) **79**, 062101 (2009).
- ⁵²S. Cabuk and M. D. Springer, [Commun. Stat. Theory Methods](#) **19**, 1157 (1990).
- ⁵³D. ÖKsoy and L. a Aroian, [Commun. Stat. Simul. Comput.](#) **23**, 223 (1994).
- ⁵⁴Z. Drezner and G. O. Wesolowsky, [J. Stat. Comput. Simul.](#) **35**, 101 (1990).
- ⁵⁵B. Cheng, G. A. Tribello, and M. Ceriotti, [Phys. Rev. B](#) **92**, 180102 (2015).
- ⁵⁶J. H. Harding, D. M. Duffy, M. L. Sushko, P. M. Rodger, D. Quigley, and J. A. Elliott, [Chem. Rev.](#) **108**, 4823 (2008).
- ⁵⁷G. T. Beckham and B. Peters, [J. Phys. Chem. Lett.](#) **2**, 1133 (2011).
- ⁵⁸B. Peters, [J. Chem. Phys.](#) **125**, 241101 (2006).
- ⁵⁹S. Prestipino, A. Laio, and E. Tosatti, [Phys. Rev. Lett.](#) **108**, 225701 (2012).
- ⁶⁰H. A. Kramers, [Physica](#) **7**, 284 (1940).
- ⁶¹B. Leimkuhler and C. Matthews, [AMRX](#) **2013**, 34 (2013).
- ⁶²P. G. Moschopoulos, [Ann. Inst. Stat. Math.](#) **37**, 541 (1985).
- ⁶³J. Wedekind, R. Strey, and D. Reguera, [J. Chem. Phys.](#) **126**, 134103 (2007).