A comparative study of methods to compute the free energy of an ordered assembly by molecular simulation

Sabry G. Moustafa, Andrew J. Schultz, and David A. Kofke

Citation: J. Chem. Phys. **139**, 084105 (2013); doi: 10.1063/1.4818990
View online: http://dx.doi.org/10.1063/1.4818990
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v139/i8
Published by the AIP Publishing LLC.

A comparative study of methods to compute the free energy of an ordered assembly by molecular simulation

Sabry G. Moustafa, Andrew J. Schultz, and David A. Kofke[a)]
Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260-4200, USA

(Received 24 April 2013; accepted 7 August 2013; published online 26 August 2013)

We present a comparative study of methods to compute the absolute free energy of a crystalline assembly of hard particles by molecular simulation. We consider all combinations of three choices defining the methodology: (1) the reference system: Einstein crystal (EC), interacting harmonic (IH), or r^{-12} soft spheres (SS); (2) the integration path: Frenkel-Ladd (FL) or penetrable ramp (PR); and (3) the free-energy method: overlap-sampling free-energy perturbation (OS) or thermodynamic integration (TI). We apply the methods to FCC hard spheres at the melting state. The study shows that, in the best cases, OS and TI are roughly equivalent in efficiency, with a slight advantage to TI. We also examine the multistate Bennett acceptance ratio method, and find that it offers no advantage for this particular application. The PR path shows advantage in general over FL, providing results of the same precision with 2–9 times less computation, depending on the choice of a common reference. The best combination for the FL path is TI+EC, which is how the FL method is usually implemented. For the PR path, the SS system (with either TI or OS) proves to be most effective; it gives equivalent precision to TI+FL+EC with about 6 times less computation (or 12 times less, if discounting the computational effort required to establish the SS reference free energy). Both the SS and IH references show great advantage in capturing finite-size effects, providing a variation in free-energy difference with system size that is about 10 times less than EC. This result further confirms previous work for soft-particle crystals, and suggests that free-energy calculations for a structured assembly be performed using a hybrid method, in which the finite-system free-energy difference is added to the extrapolated ($1/N \to 0$) absolute free energy of the reference system, to obtain a result that is nearly independent of system size. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4818990]

I. INTRODUCTION AND BACKGROUND

Thermodynamic stability of crystals is a topic of growing importance to a number of solid-state and nano-material applications of molecular modeling. This includes molecular crystals,[1,2] (especially pharmaceuticals[3]), alloys,[4,5] refractory materials,[6] clathrate hydrates,[7] colloids,[8,9] nanoparticles,[10–13] and more. Some of these systems are amenable to the application of *ab initio* methods for calculating the molecular energy, and thus promise a first-principles approach to prediction of stable crystal structures. In other cases, the "molecules" are models for nanoparticles, and the aim is to determine the structures that they will adopt via self-assembly. In any case, the crystal structure determines all other material properties, so knowledge of it for a given model is crucial to almost all materials applications.

Prediction of the stable crystal form should be based on comparison of the thermodynamic free energy for the candidate structures. However, molecular-modeling studies often employ only approximate treatments to obtain the free energy, and consequently their results can contain inaccuracies that are difficult to characterize. In many applications, the lattice enthalpy is used as an approximate free energy, which is strictly correct only in the limit of zero temperature. The neglected contribution of the entropy to the free energy can be important, and in particular is needed to capture any temperature-dependent phase transitions. Moreover, in some models of nano-particle or colloidal assembly, there is no energetic component, so the entropy (along with pressure) dominates behavior.[14] Lattice dynamics methods[15] are sometimes used to approximate the entropic contribution, but these will fail to the extent that the crystal is anharmonic. In general these approximations are made because it is computationally expensive to evaluate the free energy accurately and with good precision by molecular simulation.[16] This situation motivates the continued development and comparison of free-energy methods, as applied to crystalline phases. In previous work,[17–19] we examined methods for computing crystal free energies for van der Waals type models, which are suitable for description of molecular crystals. In the present study, we look at these methods in the context of hard potentials, which are often used to describe nanoparticle or colloidal assemblies.

When molecular simulation is applied to compute free energies, in practice what is computed is the difference in free energy of the system of interest (the target), with respect to a reference system. If the free energy of the reference is known, then the difference provides the absolute free energy of the target (though this is not always necessary, in that a given application may require knowledge of only the difference, e.g.,

a)Electronic mail: kofke@buffalo.edu

to determine the relative stability of the target and reference). Within this general approach, there is considerable flexibility in formulating a free-energy method. The variables include: (1) the choice of the reference system; (2) the thermodynamic path followed to join the reference to the target; and (3) the simulation method used to compute the free-energy change along the path.

Regarding the reference system, there are several to choose from for computing free energies of crystals.[20] On the one hand, any system with the same crystal symmetry as the target (or not[21]) can be used as a reference if its free energy is known from the literature, or as the result of a different set of calculations. On the other hand, ultimately such systems have to be related to an analytically tractable reference. In this respect, the choices fall into two categories. First are references based on non-interacting particles that are forced to adopt the crystalline symmetry of the target system, either by cell-occupancy constraints,[22] or by tethering them to lattice sites.[23] Second are harmonically interacting models, which can be solved using the methods of lattice dynamics.[15] This reference has the advantage that it by itself can provide a quantitative description of the target system at low temperature (except for discontinuous potentials, e.g., hard spheres). Hoover *et al.*[24] used this reference in the study of soft spheres, and it has been employed occasionally by others.[17,18,25,26]

The selection of the path connecting the reference to the target system is specific to their detailed nature, so we will not attempt a general discussion of this element of the free-energy calculation. We will, however, investigate some options in the present work to demonstrate the significance of this choice on the performance of the calculation.

Finally, we discuss the method used to compute the free-energy difference. These techniques can be classified generally into (1) density-of-states methods, in which the free-energy difference is related to the ensemble weight between different systems; and (2) work-based methods, in which the free energy difference is measured in terms of average work involved in transforming the system from the reference to the target.[16] In the present study, we consider only work-based methods, and two in particular: the "infinitely slow" work process represented by thermodynamic integration (TI),[27] and the "infinitely fast" process represented by free energy perturbation (FEP).[28] For TI, once the integration path is set, there is not much choice for how it is implemented—the main degrees of freedom are the step size and the quadrature method. For FEP, there are important choices to make in how it is implemented, and these issues have been described in detail elsewhere.[16,20,29] In this work we will use an overlap-sampling (OS)[30,31] implementation of FEP, as originally optimized by Bennett.[32] We will also examine the multi-state generalization of this due to Shirts and Chodera, and known as MBAR (multistate Bennett acceptance ratio).[33]

In previous work,[18,19] we demonstrated that crystal free energies could be computed very efficiently using an interacting harmonic reference, following a path from low to high temperature, and using OS coupled with harmonically targeted perturbations in temperature. In the present work, we consider methods for hard (or, more generally, discontinuous) potentials, for which the methods we developed previously cannot apply, because such models do not connect to a harmonic system at low temperature. We emphasize a comparison of the performance of the methods, aiming to determine the most efficient scheme to evaluate the absolute free energy precisely in the shortest time.

The prototype application is the system of hard spheres (HS). Despite the simplicity of the HS model, its dense phase can approximate the structure of more complex potentials because the short-range repulsion determines the crystal structure.[34] Further, the HS model can approximate the experimental behavior of the colloids.[35,36] Our particular interest in methods for this system is more in consideration of other hard-body potentials, which are of significant interest now as a means for understanding and predicting the assembly of nanoparticles of various shapes.[10–13] Although some references and paths examined here were used before for the HS model,[23,37–39] no computational efficiency and simulation details are available in the literature; accordingly we repeat all these calculations for the sake of consistency in the comparison. We expect that methods shown to work best for the HS model will also be good choices for computing free energies of other hard-body models.

This paper is arranged as follows. In Sec. II, we review the references, paths, and methods used to calculate the free energy of hard-particle crystals. Then in Sec. III we present and discuss free energy results for the HS system at its melting density. Finally, in Sec. IV we conclude the findings of this paper.

II. METHODS

We use Monte Carlo (MC) molecular simulation to examine how the efficiency of computing the free energy of face-centered-cubic (FCC) HS depends on the choice of the reference system and the path function. We consider three choices of a reference, namely, the interacting harmonic system, a system of soft spheres (r^{-n}, with inverse-power exponent $n = 12$), and the Einstein crystal. Two path functions are examined to switch the reference system to the target system, namely, Frenkel-Ladd, and a penetrable ramp path. We calculate the free energy difference using TI, and then repeat the calculations with OS, including MBAR, to show equivalence. In the following, we provide some background and briefly review these choices.

A. Free energy (FE)

Classically, the canonical partition function of a system of N spherically symmetric particles with only translational degrees of freedom contained in volume V at temperature T is[40]

$$Q(N, V, T) = \frac{1}{\Lambda^{3N} N!} \int \exp(-\beta U(\mathbf{r}_1, \ldots, \mathbf{r}_N)) \mathrm{d}1 \ldots \mathrm{d}N, \tag{1}$$

where Λ is the de Broglie wavelength, $\beta = 1/(k_B T)$, U is the system intermolecular potential energy, $\mathrm{d}i$ is an infinitesimal volume in configuration space given by $\mathrm{d}\mathbf{r}_i$, where $\mathrm{d}\mathbf{r}_i \equiv \mathrm{d}x_i \mathrm{d}y_i \mathrm{d}z_i$ and the integration is carried out over V. The

corresponding Helmholtz free energy (FE) A is given by

$$A(N, V, T) = -k_B T \ln Q(N, V, T). \tag{2}$$

Since the momentum contributions (i.e., Λ) to the FE cancel when the FE difference is measured, we ignore them and compute only the configurational partition function, i.e., the integration in Eq. (1).

B. Hard-sphere model (HS)

The intermolecular potential between two HS of the same diameter σ is defined as

$$u(r) = \begin{cases} \infty, & r < \sigma \\ 0, & r \geq \sigma \end{cases}, \tag{3}$$

where r is the distance between the sphere centers. Since the HS potential can be only zero or infinity, the Boltzmann factor of a configuration of N HS, $\exp(-\beta U_N)$, can be either one or zero, respectively; consequently, the configurational partition function is independent of temperature. Therefore, the full thermodynamic behavior of the HS model can be represented by just a single isotherm. We focus on the FCC solid at the melting density ($\rho\sigma^3 = 1.0409$) as it is the most stable structure.[23,37,41]

C. Free energy difference methods

The FE difference $\Delta A_{0\to1}$ is measured via molecular simulation to give the target FE from knowledge of the FE of the reference

$$A_1 = A_0 + \Delta A_{0\to1}, \tag{4}$$

where "0" and "1" refer to the reference and the target systems, respectively. The difference $\Delta A_{0\to1}$ is evaluated using either thermodynamic integration, or free-energy perturbation.

1. Thermodynamic integration (TI)

In TI, $\Delta A_{0\to1}$ is given by

$$\Delta A_{0\to1} = \int_{\lambda=0}^{1} \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda, \tag{5}$$

where λ is a scalar parameter that switches the system energy from the reference state, $U(0)$, to the target state, $U(1)$, through a specific path (see Sec. II D), and the integrand represents the ensemble average of the first derivative of the potential energy at a given λ.

It is interesting to consider the relation between the accuracy, precision, and computational effort involved in a TI calculation. The textbook application of numerical integration assumes that the integrand function is evaluated with a fixed computational cost, and is free of noise. Thus, the focus of the analysis of these methods is on the accuracy of the result, and how much it is improved by decreasing the integration step, increasing the number of function evaluations, and thus increasing the total computational cost. The underlying assumption is that the computational cost necessarily increases in proportion to the number of function evaluations. However, for TI performed in the context of molecular simulation, this is not the case. The cost to evaluate the integrand by molecular simulation can be adjusted by trading off precision: consider the evaluation of a general integral $F = \int_a^b f(x)dx$ via a Newton-Cotes formula of the form $\frac{b-a}{n}\sum_{i=1}^{n} a_i f(x_i)$ (which in this form does not exclude unequal spacings of the x_i). If each function evaluation is given independently with variance σ_i^2, then the variance of the estimate of the integral will be $\sigma_F^2 = (\frac{b-a}{n})^2 \sum_{i=1}^{n} a_i^2 \sigma_i^2$. The integrand variance σ_i^2 will vary with the number m_i of independent samples used to measure it in a molecular simulation, according to $\sigma_i^2 = \hat{\sigma}_i^2/m_i$, where $\hat{\sigma}_i$ is independent of m_i. The total computational cost is $M = \sum m_i$. For a given n and M, the optimal (minimum σ_F^2) allocation of samples is $m_i = M a_i \hat{\sigma}_i / \sum_k a_k \hat{\sigma}_k$. Putting this together, we have for the variance of the integral

$$\begin{aligned} \sigma_F^2 &= \left(\frac{b-a}{n}\right)^2 \sum_i a_i^2 \frac{\hat{\sigma}_i^2}{m_i} \\ &= \left(\frac{b-a}{n}\right)^2 \sum_i a_i^2 \hat{\sigma}_i^2 \frac{\sum_k a_k \hat{\sigma}_k}{M a_i \hat{\sigma}_i} \\ &= \left(\frac{b-a}{n}\right)^2 \frac{(\sum_i a_i \hat{\sigma}_i)^2}{M} \\ &\approx \frac{\Sigma^2}{M}, \end{aligned} \tag{6}$$

where we define $\Sigma \equiv \int_a^b \hat{\sigma}(x)dx$, which is independent of M and n. The key result here is that the precision of the estimate of the integral depends only on the total amount of sampling M applied across all the quadrature points, and not on the number of quadrature points per se. This result holds also if one chooses not to optimize the m_i, and uses just a constant amount of sampling $m = M/n$ for each quadrature point. A significant consideration omitted here is the sampling required to equilibrate each new state condition; this is difficult to characterize in general, but it will be a lower-order effect, assuming that the amount of sampling, m_i, applied to each state point is not exceedingly small.

The conclusion to take from this analysis is that for TI by molecular simulation, the number of quadrature points should be increased as needed to ensure that the accuracy of the quadrature is as good as its precision, i.e., that errors due to the finite integration step do not exceed the confidence limits of the integral. Adding quadrature points while keeping the total computational effort fixed can be done without significantly affecting the precision of the calculated integral.

2. Overlap-sampling free-energy perturbation (OS)

In FEP, while simulating a system at state 0, perturbations are periodically performed into the state labeled 1, and averages are taken of the form $\langle \exp(-\Delta(\beta U)) \rangle_0$, where $\Delta(\beta U)$ is the difference in U/k_BT accompanying the perturbation. The FE is given by the logarithm of this average, and is highly prone to bias. To obtain a good result, the simulated system must sample not only the configurations important to

system 0, but also those important to system 1. This can happen only if the important system-1 configurations are a subset of those important to system 0.[16,29,42] Often they are not, and the method fails. As a remedy, staging is introduced, to gradually transition from 0 to 1. In this context the path of the transition becomes an issue, as it does in TI. Further, unless 1-important configurations are already a subset of the 0-important configuration, the subset requirement can be met only via another staging technique, known as overlap sampling. Here, an intermediate is defined between stages i and $i+1$, such that its configurations are in the region of overlap (i.e., the intersection) of those of i and $i+1$. Then the OS is conducted, in separate simulations, from each stage (i and $i+1$) into the overlap region. This approach is quite robust, not that difficult to implement, and gives good results. It was originally introduced and optimized by Bennett.[32] The working equations are

$$\Delta A_{i\to i+1} = -k_B T \ln \frac{\langle \gamma_{OS}/e_{i+1}\rangle_{i+1}}{\langle \gamma_{OS}/e_i\rangle_i}, \tag{7}$$

where e_i is the Boltzmann factor of state i, $\exp(-\beta_i U(\lambda_i))$, and γ_{OS} is the overlap weight function, given by

$$\gamma_{OS} \equiv \frac{e_i e_{i+1}}{e_{i+1} + \alpha e_i}. \tag{8}$$

Here, α is an adjustable parameter, which for optimum performance, according to Bennett,[32] is given by $\alpha_{opt} = \exp(-\beta\Delta A)$. Since ΔA is not known *a priori*, it has to be estimated self-consistently. This can be accomplished by recording averages for multiple values of α in a single run, or by writing all the observed perturbation energies to file and solving for self-consistency in the post-simulation analysis.

If multiple states are indeed used to traverse the path from the reference to the target, there may be some advantage to performing overlap sampling between all pairs of simulated states, rather than just between states i and $i+1$. This will be helpful to the extent that the different intermediate states (say, i and $i+2$) have overlapping configurations, and moreover to the extent that the overlap region of one pair differs from that of other pairs (thereby providing new information about their relation). The scheme to do this is known as MBAR, and was presented and optimized by Shirts and Chodera.[33] With it, the relative FE of state i is given by

$$\Delta(\beta A)_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{e_i(x_{jn})}{\sum_{k=1}^{K} N_k e_k(x_{jn})/\alpha_k}, \tag{9}$$

where K is the number of states, N_j is the number of uncorrelated equilibrium samples of state j, $e_i(x_{jn})$ is the snapshot n of the Boltzmann factor as evaluated at state i for configurations sampled at state j, and α's are constants to be determined. Shirts and Chodera[33] showed that the optimum estimation of α's, after applying Bennett's criteria to all states, is when $\ln\alpha_k = -\Delta(\beta A)_k$. Therefore, the MBAR method has to be solved using self-consistent iteration methods, and this is best done via post-simulation analysis on appropriate data that is written to file during the simulation.

D. Thermodynamic paths

The path between the reference and the target systems affects the efficiency of the FE estimation. In general, one wants a path that transforms the important configurations smoothly from those of the reference to those of the target. For our HS study, two paths are considered, namely, Frenkel-Ladd (FL) and penetrable-ramp (PR) paths.

1. Frenkel-Ladd (FL)

In the FL path, the HS interactions are always in effect, and the reference potential is added gradually to the HS potential through a linear coupling parameter, λ. The reference potential is designed to pull the spheres away from each other and toward their lattice sites, such that for some large λ overlaps between spheres are negligible. The FL-path potential for a general potential is given by

$$U(\lambda) = U_{HS} + \lambda U_{ref}. \tag{10}$$

At large values of λ, the HS interactions can be neglected (for some systems—not considered here—an expansion of the lattice to zero density also is required to effect the decoupling). Therefore, at some λ_{max}, the system potential can be approximated to be $\lambda_{max}U_{ref}$, which has known free energy. The temperature of the system does not play a role by itself; rather, the $\beta\lambda$ combination is the important simulation parameter. So, without loss of generality, we set $T = 1$ when working with the Frenkel-Ladd path. The λ parameter varies such that it increases exponentially along the path[27]

$$t_i = \ln(\lambda_i + c), \tag{11}$$

where c is an adjustable parameter (chosen such that the uncertainty is minimum) and t_i is spaced equally, with $dt_i = 0.1$, in the index i, and ranges from $\ln c$ to a maximum obtained when the spheres are observed to no longer interact.

2. Penetrable ramp (PR)

This path was introduced by Almarza[38] to evaluate the HS free energy. In this path, unlike the FL path, the HS behavior is developed slowly through a ramp function as follows:

$$U(\lambda) = (1-\lambda)U_{\text{ref}} + \frac{\lambda}{1-\lambda}U_{\text{ramp}}, \tag{12}$$

where λ is the path parameter (such that $U(0) = U_{\text{ref}}$ and $U(1) = U_{HS}$) and U_{ramp} is defined as

$$\begin{aligned} U_{\text{ramp}} &= \sum_{ij}\sum u_{ij}, \\ u_{ij} &= \begin{cases} W_0(\sigma - r_{ij}), & r_{ij} < \sigma \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \tag{13}$$

where W_0 is the ramp potential strength, which can be adjusted to have a smooth transition path with the least simulation uncertainty; values of W_0 used in this work are given with the results in Sec. III. Again, unlike FL, the temperature (in addition to λ and W_0) by itself plays a role in the PR path. The temperature of the reference system is selected such that

the mean-square-displacement (MSD) of the spheres from their lattice sites is equal between the reference and the target systems.

We note that although the PR potential diverges for $r < \sigma$ in the limit $\lambda \to 1$, the integrand of Eq. (5) remains well behaved, and is given by

$$\left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda \to 1} = -\langle U_{\text{ref}} \rangle_{HS} + \frac{3}{(\beta^2 W_0)}(Z_{HS} - 1), \quad (14)$$

where Z_{HS} is the HS compressibility factor, and $\langle \dots \rangle_{HS}$ represents an average taken while sampling the HS system. The integrand at $\lambda \to 0$ is straightforward to evaluate, and also causes no complications.

E. Reference systems

1. Einstein crystal (EC)

In this model, each molecule is interacting, independently, with its respective equilibrium lattice site through a harmonic potential, all characterized by the same spring constant, k_E. Consequently there is no collective behavior exhibited by this model. The potential energy of an EC of N molecules is given by

$$U_E = k_E \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_{i0})^2, \quad (15)$$

where $\mathbf{r}_i$ is the instantaneous location of molecule i and $\mathbf{r}_{i0}$ is its respective lattice position. The corresponding Helmholtz FE of this system is given, analytically, by[37]

$$\frac{\beta A_E}{N} = \frac{3(N-1)}{2N} \ln\left(\frac{\beta k_E}{\pi}\right) - \frac{3}{2N} \ln N + \frac{1}{N} \ln \rho, \quad (16)$$

where the last two terms represent the center-of-mass (COM) motion contribution.

2. Interacting harmonic crystal (IH)

In the IH model each molecule is interacting, harmonically, with all other molecules through identical spring constants k_H. In our case, we include interactions with only the 12 first-nearest neighbors. The harmonic potential energy of this system is given by

$$U_H = \frac{1}{2} k_H \sum_{i=1}^{N} \sum_{j \in \{nbrs\}_i} \left(\mathbf{r}_{ij}^2 - \mathbf{r}_{ij,0}^2\right), \quad (17)$$

where $\mathbf{r}_{\mathbf{ij,0}}$ is the separation vector for i and j for the minimum-energy (perfect-lattice) configuration. The corresponding harmonic FE is given by[17]

$$\frac{\beta A_H}{N} = \frac{1}{2N} \sum_{i}^{3(N-1)} \ln\left(\frac{\beta \lambda_i}{2\pi}\right) - \frac{3}{2N} \ln N + \frac{1}{N} \ln \rho, \quad (18)$$

where λ_i are the eigenvalues of the Hessian and ρ is the system density. The summation runs over the $3(N-1)$ nonzero normal modes (the remaining three modes are zero and correspond to the COM motion) while the last two terms represent the correction for the CM motion. Unlike the EC system, collective behavior of molecules is expected due to the intermolecular interactions.

3. Soft-sphere crystal (SS)

The classical soft-sphere (SS) model is represented by an intermolecular inverse-power[43] potential. In our case, the inverse-power parameter, n, is taken to be 12

$$U_{SS}(r) = \epsilon \left(\frac{\sigma}{r}\right)^{12}, \quad (19)$$

where ϵ and σ are the SS model energy and size parameters. The motivation for using the SS model is that it is much more like the HS system than are the EC or IH models. Both the HS and SS potentials are purely repulsive, and both exhibit collective motions. Given these similarities, one might expect it to be comparatively easier to determine the difference in FE between the HS system and a SS reference. A comparative disadvantage of the SS system is that its FE is not analytically tractable, and can be determined only through application of molecular simulation. However, the non-trivial properties of the SS model depend on a single dimensionless state parameter, $\rho\sigma^3(\beta\epsilon)^{1/4}$, rather than separately on the two groups formed from the density and temperature, respectively. Thus, it is not difficult to establish its properties for all conditions of interest. This has been done, and the results have been fitted to give the free SS FE in a convenient mathematical representation.[18] Nevertheless, we perform these calculations again in the present work, for two reasons. First, we want to include the time needed for these calculations when comparing the computational effort required for the various FE methods considered here. Second, in the present work we use a SS reference in which only the first- and second-nearest neighbor interactions are included; the results given in Ref. 18 are extrapolated to an infinite system, and have no interaction cutoff. We employed harmonically targeted temperature perturbation (HTTP) with a low-temperature IH reference to compute the SS FE as a function of $\rho\sigma^3(\beta\epsilon)^{1/4}$. Details are as described in Ref. 18.

F. Simulation details

Standard canonical-ensemble (NVT) Monte Carlo (MC) simulations[27] were used to compute the averages specified in Sec. II C. All simulations were performed for the FCC crystal at the density of the target system ($\rho\sigma^3 = 1.0409$). For the HS and SS systems, we include interactions between only those particles that are first- or second-nearest neighbors on the perfect lattice (second neighbors are separated by $r_{cut} = (4/\rho)^{1/3}\sigma = 1.5663\sigma$ for the density studied here); only first-neighbor interactions are considered in the IH reference.

For each simulation (i.e., each point on the thermodynamic path), a total of 10^8 proposed MC trial steps were conducted, beyond an initial $50N$ equilibration trials performed before collecting any data. Contributions to the ensemble averages were taken after every N trials. The samples were subaveraged into 100 blocks that were used to provide error estimates, with the method of Kolafa[47] applied to account for any correlation between adjacent block averages. Thus, for

example, for the $N = 500$ system for which we report the most detailed results, we collected $10^8/500 = 200\,000$ data samples, and for the error analysis computed 100 block subaverages, each formed from 2000 samples taken across 10^6 MC trials. Reported confidence limits represent one standard deviation of the mean.

In all calculations the center of mass was fixed to avoid divergence associated with drifting the whole system; this was achieved by moving two HS at a time in each proposed trial move. A fixed cubic box of edge $L = (N/\rho)^{1/3}\sigma$ was used, in which N particles were positioned, at the beginning of simulation, on the equilibrium FCC sites corresponding to the box size. In all calculations, standard periodic boundary conditions were employed.

The step size of the multistage TI or OS depends on the path. For the PR path, we found that 50 steps of size $d\lambda = 0.02$ (where λ varies from 0 to 1) for all references is reasonable (we tried $d\lambda = 0.01$, but we got similar results). For the FL path, the number of path steps was selected as the fewest needed to obtain a result of accuracy that is greater than the precision of the calculated free energy; the required number is approximately 50, and the values used are given in Sec. III. The FL steps give a consistent increment in t from $\ln c$ to $\ln(c + \lambda_m)$, and the values of c and λ_m employed here are also given in Sec. III. The reference temperature of the FL path was set to unity, while the temperature for the PR path was optimized using the MSD criteria described above (which differs based on the reference system, as shown in Sec. III A).

We screened the choices by first performing MC simulations for a system of 500 HS to identify the most efficient reference and path combination; then calculations were performed on the selected combination for different system sizes ($N = 256$, 864, 1372, 2048, 2916, and 4000) to evaluate the HS FE in the thermodynamic limit ($N \to \infty$) through extrapolation.

The FE of SS reference was computed in a separate set of MC simulations, using the HTTP method.[18] Starting from $k_BT/\epsilon = 0.01$ to $k_BT/\epsilon = 0.55$ with step of 0.01, a polynomial fitting was done on this temperature range so that at any desired temperature the FE can be evaluated.

III. RESULTS AND DISCUSSION

A. Reference temperature for PR path

For the PR path, we set the reference system temperature based on equating the MSD of both the HS and reference systems at the same conditions. The MSD for the EC and IH references is known analytically and varies linearly with T;[44] on the other hand, there is no analytical form for the SS MSD; but it should behave harmonically at low temperature. The MSD versus temperature for each of the three references is presented in Fig. 1. The temperature corresponding to the intersection of the MSD of HS and each reference system is $kT/\epsilon = 0.0117$, 0.1150, and 0.4100 for EC, IH, and SS references, respectively, and these determine the temperatures used for the isothermal paths taken for the corresponding FE calculations. To check the suitability of the MSD criterion we performed FE calculations (not shown here) also at reference temperatures below and above these temperatures, and we found that the simulation error is minimized at the MSD-based temperatures.

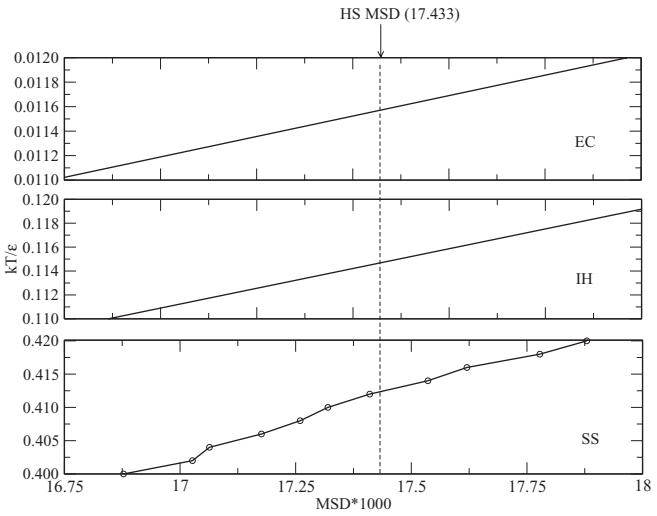

FIG. 1. *MSD* vs. T for different reference systems (EC, IH, and SS, as indicated) at $\rho\sigma^3 = 1.0409$ and $N = 500$. The dotted vertical line represents the MSD of the HS system at the same conditions.

B. Free energy results for finite-size systems

Table I summarizes free energies computed using the assorted combinations of reference system, thermodynamic path, and FE method for the $N = 500$ system. The excess FE per particle with respect to the ideal gas ($\beta A_{ex}/N = \beta(A - A_{IG})/N$; where $\beta A_{IG}/N = \ln\rho - 1 + \ln(2\pi N)/2N$) is modified by adding $\ln N/N$ to remove the logarithmic leading dependences on N, so that $\beta A_{ex}/N + \ln N/N$ scales as N^{-1}.[37]

To show that the FE computed by the 12 combinations of methods are mutually consistent, we compute the statistic,

$$c \equiv \sqrt{\left(\frac{1}{M}\sum_{i=1}^{M}\frac{(A_i - A_{avg})^2}{\sigma_i^2 + \sigma_{avg}^2}\right)}, \tag{20}$$

where M is the number of data sets (12 in our case), A_i and σ_i are the FE and uncertainty, respectively, resulting from by a specific combination of choices, and A_{avg} and σ_{avg}^2 are

$$A_{avg} = \frac{\sum_{i=1}^{M}\left(A_i/\sigma_i^2\right)}{\sum_{i=1}^{M}\left(1/\sigma_i^2\right)} \tag{21}$$

and

$$\sigma_{avg}^2 = \left[\sum_{i=1}^{M}\left(1/\sigma_i^2\right)\right]^{-1}. \tag{22}$$

For the results to be mutually consistent, c should be of order of 1.0. Using the data given in Table I, $c = 0.94$, indicating that the variation in the FE estimates from the different methods is consistent with the stochastic error estimates.

We notice that the uncertainty (as scaled to 50 simulations; see the Error row) obtained from both TI and OS are very similar when compared for the same reference and path.

TABLE I. Excess free energies of FCC HS solid at density $\rho\sigma^3 = 1.0409$, based on simulation of $N = 500$ spheres. For both TI and OS, the FE results are given for the PR and FL paths, using EC, IH, and SS references. The digit(s) in parentheses is the uncertainty of the last digit(s) in the reported value, represented by the 68% confidence limit. The number of MC trials is 10^8 for each simulation, and the total number of simulations for each free-energy calculation is given in the row labels "Path steps." This free-energy uncertainties are each scaled to a common length of 50 path steps for the purpose of comparison (multiplying by (path steps/50)$^{1/2}$), and tabulated with higher precision in the row labeled "Error." Temperature (k_BT/ϵ) and path parameters (dimensionless) are given as indicated.

Path	PR			FL		
Reference	EC	IH	SS	EC	IH	SS
			TI method			
$\beta A_{ex}/N + \ln N/N$	5.90851(16)	5.90844(12)	5.90830(6)	5.9082(2)	5.9080(3)	5.9082(2)
Error $\times 10^4$	1.6	1.2	0.61	2.6	2.4	2.2
Temperature	0.0117	0.115	0.410	1.0	1.0	1.0
Path steps	50	50	50	55	41	50
Path parameters	$W_0 = 0.8$	$W_0 = 8.0$	$W_0 = 60.0$	$\ln c = 2.0$	$\ln c = 1.0$	$\ln c = 1.0$
	$d\lambda = 0.02$	$d\lambda = 0.02$	$d\lambda = 0.02$	$\lambda_m = 1628.6$	$\lambda_m = 145.7$	$\lambda_m = 21.79$
			OS method			
$\beta A_{ex}/N + \ln N/N$	5.90852(18)	5.90831(14)	5.90825(5)	5.9086(3)	5.9079(3)	5.9080(3)
Error $\times 10^4$	1.8	1.4	0.46	3.2	2.9	2.5
Temperature	0.0117	0.115	0.410	1.0	1.0	1.0
Path steps	50	50	50	55	41	47
Path parameters	$W_0 = 2.0$	$W_0 = 20.0$	$W_0 = 80.0$	$\ln c = 2.0$	$\ln c = 1.0$	$\ln c = 0.2$
	$d\lambda = 0.02$	$d\lambda = 0.02$	$d\lambda = 0.02$	$\lambda_m = 1800.7$	$\lambda_m = 178.5$	$\lambda_m = 19.70$

More important, we find that the combination of the SS reference and the PR path exhibits the least uncertainty (at least 2 times smaller) of all the combinations. This should not be surprising, as the SS behavior is expected to be closer to the HS due to its harder repulsion (compared to IH) and its collective motion (compared to EC). On the other hand, the uncertainty associated with the FL path is independent of the reference system, suggesting that the total uncertainty in this case is dominated by the phasing in of the HS potential along the path, more than the choice of the reference.

To check if the multistate MBAR method has advantage over the pairwise OS method we conducted a separate set of simulations (apart from the results given in Table I) for both methods. It is most convenient to base the analysis on uncorrelated data, rather than attempting block averaging and compensating any residual correlation explicitly (as we do for Table I results), so we generated data for the analysis using a sampling interval of 10^5 steps; we find that the correlation between adjacent samples is statistically zero. Both methods used the same data set as an input, enabling a direct comparison. The raw data from the MC simulations were processed for MBAR using a Python script developed by Shirts and Chodera.[33]

For this comparison, we consider a cubic box of 500 HS, using SS as the reference and the PR path. Using 10^9 MC steps (i.e., 10^4 uncorrelated samples), we obtained excess free energies of 5.90839(12) from OS and 5.90841(12) from MBAR. This result indicates that adding extra perturbations (i.e., MBAR) does not improve the uncertainty of HS FE calculations, in this application. The value measured in the perturbation from the i to $i + 1$ state is strongly correlated with that from perturbations from i into subsequent states (i.e., $i + 2$, $i + 3$, etc.), and this significantly diminishes the information contributed by these extra perturbations. We observe, in particular, that the correlation between the $i \to i + 1$ and $i \to i + 2$ perturbations to be of the order of 0.9 for all states on the perturbation path. The performance observed here is not inconsistent with other studies,[48] which have found that the relative performance of MBAR and OS depends on the nature of the free-energy calculation, as well as implementation details.

Given the observed equivalence of OS and TI, we choose TI for all of the subsequent analysis, due to its simplicity to implement and post-process. Further, we will focus on using the SS reference and PR path combination, as they show the lowest uncertainty in estimating the solid HS FE.

C. Finite-size effects: Thermodynamic limit results

The integrand for the PR+SS+TI calculation (see Eq. (5)) is shown for different system sizes in Fig. 2. The results are very precise, with error bars that are not visible on the scale of the figure. The curve is not straight, but it does vary smoothly at all values of λ. The integrand is almost completely independent of N—results for the seven system sizes would not be distinguishable on the scale of the full graph. We use the inset figure to uncover the small finite-size effects. This shows the integrand relative to the value measured for the largest system. For clarity, only three values of N are plotted, and it is seen that for only the smallest size does the difference from the largest simulated system exceed the error bars. The small finite-size effects can be made visible also by examining the integral itself, as shown in Fig. 3. Here, we present the FE per particle, relative to the SS reference, at different system sizes (plotted as $1/N$). The slope is 0.40(2), which is more than ten times smaller than the slope of the full FE as reported by Polson *et al.*,[37] who used EC as a reference, or our own results using EC or size-dependent IH and SS references (Fig. 4). This result suggests that the finite-size effect is

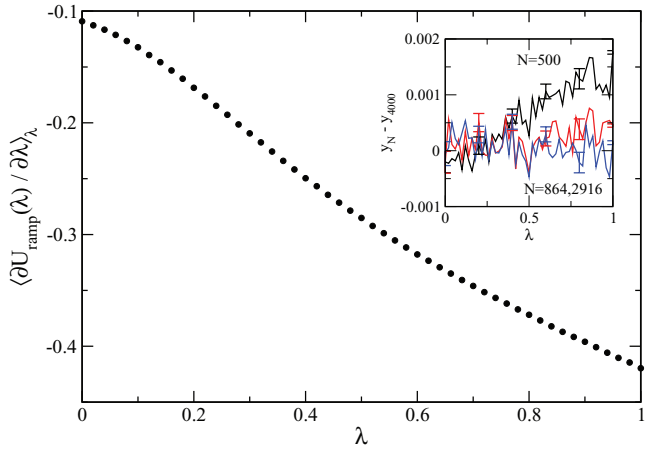

FIG. 2. Integrand of Eq. (5), $\partial A/\partial\lambda$, for the SS reference and the PR path at density $\rho\sigma^3 = 1.0409$. Results are plotted for $N = 4000$, but results taken for all other N would not be discernible on the larger figure. (Inset) For three system sizes ($N = 256$ (black), 864 (red), and 2916 (blue)), the difference at each λ with the corresponding value for the largest system ($N = 4000$).

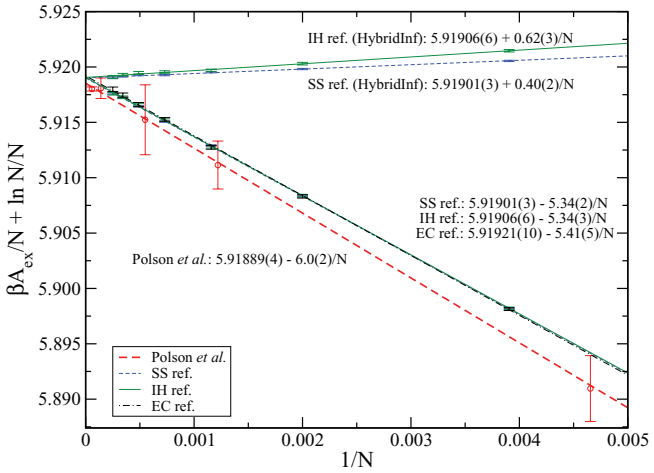

FIG. 4. Finite-size effect of the excess FE at the melting state ($\rho\sigma^3 = 1.0409$). Lines for the EC, IH, and SS references were computed by adding the FE of the corresponding size-dependent reference to the (size-dependent) FE differences computed via TI on the PR path. Results by Polson *et al.*[37] give similar results for the TI+FL+EC method. Lines labeled HybridInf are computed using a size-dependent TI+PR integral with the FE for an infinite-system reference.

dominated by the SS reference contribution, which is absent in Fig. 3. This point is made in Fig. 4, where we examine the finite-size effects in terms of the full excess FE, highlighting the role of the finite-size effects brought by the reference. We present results obtained using TI+PR with SS, IH, and EC references, as well as results reported previously by Polson *et al.*[37] using the TI+FL+EC combination. We consider results obtained both by size-dependent as well as infinite-system IH or SS references.

The nature of the finite-size effects displayed in Fig. 4 for these HS calculations is consistent with finite-size effects observed for SS (for different values of the inverse-power exponent),[17,18] and for a model of molecular nitrogen (N_2).[19] If the reference system has some semblance to the target, and in particular if it is formed from interacting particles and thus exhibits cooperative behavior (e.g., IH and SS systems in the present case), then the finite-size effects it exhibits for the FE will roughly mirror the finite-size effects in the target. Consequently, the FE difference—the difficult calculation that is the focus of this work—will be largely system-size independent.

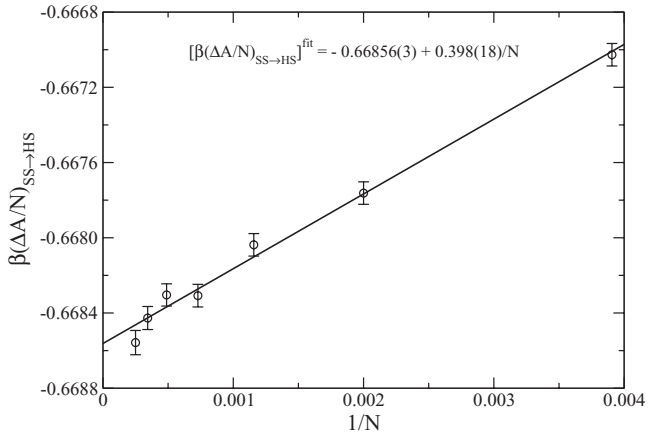

FIG. 3. Finite-size effect of the FE difference between SS and HS systems at the melting condition ($\rho\sigma^3 = 1.0409$).

Tan *et al.*[17] proposed a hybrid approach to exploit this behavior. In this hybrid method, the free energy of the reference is extrapolated to the infinite system, and this (rather than a size-dependent reference value) is added to the finite-system FE difference to obtain the target-system FE. The absolute FE computed this way will have a system-size dependence with slope given by that of the FE difference, which is nearly flat (Fig. 3). The advantage in doing this is that the FE of the reference is more easily determined as a function of system size, compared to the calculation of the FE difference, which involves TI or OS and molecular simulation. Thus, we can obtain a good estimate of the target-system FE in the thermodynamic limit while conducting small-system molecular simulations. Figure 4 includes results (labeled HybridInf) that demonstrate the idea for both SS and IH references. As shown there, although the finite-size effect of the FE difference using IH reference is quite small, its slope is a little larger than that of the SS reference. We did not apply this technique to the EC reference because it has no finite-size variation, so its "hybrid" implementation would not differ from the curve already given in the figure.

We notice in Fig. 4 that while the results from the size-dependent EC, IH, and SS references are mutually indistinguishable, they are considerably apart from the results of Polson *et al.*[37] The difference is likely to be due to the use of different box shapes for the MC simulations in the respective studies. Still, the values extrapolated to $1/N \to 0$ differ. The confidence limit given on the extrapolated value reported by Polson *et al.*[37] is much tighter than almost all of the error bars on their size-dependent data, and the high precision in this extrapolation seems to owe to the high precision of the result from their largest system simulated. This point notwithstanding, their data are not inconsistent with an extrapolation that agrees with the values we report.

The extrapolated excess FE at the thermodynamic limit, based on our calculations using SS, IH, and EC references, is 5.91901(3), 5.91906(6), and 5.91921(10), respectively; these results are close to values reported previously: 5.924(15),[22] 5.9226(10),[23] 5.91889(4),[37] 5.91930(11),[38] 5.9194(5),[46] 5.9194(2),[45] and 5.9188(6)[39] (a small adjustment is applied to the literature values, where needed, to bring them to the density examined here, $\rho\sigma^3 = 1.0409$).

D. Computation efficiency

While the SS reference with PR path showed the least uncertainty with respect to number of MC cycles, it is worthwhile also to compare the efficiency of computations with respect to the CPU time. For this purpose, we estimate the amount of CPU time t required by each method to produce a result to the same precision as the TI+PR+SS method (our base method for comparison). The comparison is slightly complicated by the need to compute the SS FE for methods that use it as a reference. One might be interested in scenarios in which the SS FE either is taken as given, or, in the interest of fair comparison of total computation effort, the time required for its calculation, t_{ref}, is included. We will present both comparisons. Regardless, we always include the contribution of the SS reference FE to the total stochastic error, according to $\sigma_{tot}^2 = \sqrt{\sigma_{\text{ref}}^2 + \sigma_\Delta^2}$, where σ_Δ is the error reported in Table I (for the actual number of path steps used); for EC and IH, effectively we take $\sigma_{\text{ref}} = 0$, $t_{\text{ref}} = 0$. We take σ_{ref} as fixed, and assume that adjustments of CPU time affect only σ_Δ as needed to make σ_{tot} equal to that for the base method (7.3×10^{-5}). CPU time is estimated assuming that $t\sigma_\Delta^2$ is an invariant with t.

Where needed, we compute the SS FE using the HTTP method.[18] For the case of PR path, we obtained the SS FE ($\beta A/N$) at $T = 0.41$ with uncertainty $\sigma_{\text{ref}} = 2.80 \times 10^{-5}$ using 7.9 CPU hours. For the FL path, the FE is required at a lower temperature, $T = 0.0459$ ($=1/\lambda_{\max}$; where $\lambda_{\max} = 21.79$, see Table I), so only 1.6 CPU hours were needed to yield a result with uncertainty $\sigma_{\text{ref}} = 1.90 \times 10^{-5}$. An additional 7.8 CPU hours were required to perform the TI calculation following the PR path to obtain the TI+PR+SS base-case FE reported in Table I. All simulations were performed on Intel Core2 Quad Q9550 Yorkfield 2.83 GHz processor, and all results are reported for the $N = 500$ system.

TABLE II. Relative CPU time for the results reported in Table I. Tabulated are the CPU times required to get the same total error as resulted from the TI+PR+SS base case, divided by the base-case CPU time $t_{\text{base}} = 15.7$ h. Values in parentheses for methods using a SS reference show relative timings that exclude the contribution t_{ref} needed to obtain the SS reference FE. Values are reported to two digits of precision, but the second digit is not highly significant.

Path	PR			FL		
Reference	EC	IH	SS	EC	IH	SS
			TI method			
t/t_{base}	2.3	1.4	1.0 (0.5)	6.0	5.9	5.2 (5.1)
			OS method			
t/t_{base}	3.5	2.1	0.9 (0.4)	10.	10.	7.5 (7.4)

Table II presents the comparison. All times are divided by the CPU time for the TI+PR+SS base case, which includes t_{ref}. Even when including computational effort of the SS reference, the PR path and SS reference are still the most efficient way to get the HS FE, requiring the least computational time for a given precision. These results reinforce the conclusions taken from Table I. We also notice that the time required for TI method is marginally lower than that of the OS method at the same path and reference combination, except for the PR+SS case. Also, the time needed for the PR path is always shorter than that of the FL path (whatever reference used). The advantage of the PR+SS methods is even more striking if one allows the SS FE as a given—in such a case this reference/path combination requires 10-20 times less computation than approaches based on the FL path.

IV. CONCLUSIONS

This work presents a comparison of the efficiency of molecular simulation methods for measuring absolute free energies of hard-particle models. The comparison is based on the HS crystal at the melting condition, but it is reasonable to expect the conclusions are general to hard-body models, or at least to rigid, convex bodies. The study employs various combinations of references (Einstein, interacting harmonic, and r^{-12} soft-sphere crystals), switching paths (Frenkel-Ladd and penetrable ramp), and free-energy methods (thermodynamic integration and overlap-sampling free-energy perturbation). It shows that if the SS used as a reference, with its FE is calculated using HTTP method, then, using TI or OS along a PR path is the most efficient route (with respect to both MC steps and CPU time) to get the crystalline HS free energy at the melting state. Coming after SS is the IH and then EC as references to the HS solid, using the PR path and TI (which is marginally better than OS for these cases). In terms of accuracy, the results from the various methods are mutually consistent, and values extrapolated to the thermodynamic limit are in good agreement with literature data.

When attempting to obtain very accurate free energies by evaluating the difference along a thermodynamic path, the choice between TI and OS is a matter of convenience and preference. For high-accuracy applications, both perform about equally well in producing a value to a given precision, all else being the same. If high accuracy is not a priority, TI can provide a rough result with as few as two simulations, while OS for the same system could fail catastrophically. However, it still not a good practice to apply TI without some means to gauge its accuracy, particularly by examining how the result changes with number of quadrature points.

The results of this study for a hard potential reinforces the observations regarding system-size dependence made previously for soft-interaction models. The difference in FE between a target system and a reference formed from interacting particles is nearly independent of system size. The hybrid technique for handling finite-size effects is thus the method of choice: the finite-system FE difference is added to the extrapolated ($1/N \to 0$) absolute FE of the reference system. This

method provides a good and rapid estimation of the infinite-size system FE using only one simulation of a relatively small system size. A second calculation at a different system size can be performed to confirm the size independence, if needed.

The methodology examined here is readily extended to more complex hard-body potentials, such as hard tetrahedra.[11] In such an application, the hard potential defining the shape of the hard body is transformed into a softer repulsive form. The PR path with OS or TI can be used compute the FE difference between the hard and soft potentials, and HTTP can then be used to lower the temperature of the soft potential to connect it to an interacting harmonic reference. The ability to perform such calculations for small systems, relying on the reference to capture the finite-size effects, can make such calculations increasingly practical, and thereby enable a rigorous evaluation of the relative stability of different candidate structures.

ACKNOWLEDGMENTS

This research was supported by the U.S. National Science Foundation (NSF), Grant Nos. CHE-0626305 and CBET-0854340.

[1]A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, S. L. Price, P. T. A. Galek, G. M. Day, and A. J. Cruz Cabeza, Int. J. Pharm. **418**, 168 (2011).
[2]G. M. Day, Crystallogr. Rev. **17**, 3 (2011).
[3]B. Rodríguez-Spong, C. Price, A. Jayasankar, A. Matzger, and N. Rodríguez-Hornedo, Adv. Drug Delivery Rev. **56**, 241 (2004).
[4]A. Van de Walle and G. Ceder, Rev. Mod. Phys. **74**, 11 (2002).
[5]S. Curtarolo, D. Morgan, and G. Ceder, CALPHAD **29**, 163 (2005).
[6]L. G. Wang and A. Van de Walle, Phys. Chem. Chem. Phys. **14**, 1529 (2012).
[7]A. K. Sum, D. T. Wu, and K. Yasuoka, MRS Bull. **36**, 205 (2011).
[8]D. Frenkel, Science **296**, 65 (2002).
[9]D. Frenkel, Nature Mater. **5**, 85 (2006).
[10]A. Haji-Akbari, M. Engel, A. S. Keys, X. Zheng, R. G. Petschek, P. Palffy-Muhoray, and S. C. Glotzer, Nature (London) **462**, 773 (2009).
[11]A. Haji-Akbari, M. Engel, and S. C. Glotzer, J. Chem. Phys. **135**, 194101 (2011).
[12]S. C. Glotzer and M. J. Solomon, Nature Mater. **6**, 557 (2007).
[13]P. F. Damasceno, M. Engel, and S. C. Glotzer, Science **337**, 453 (2012).
[14]D. Frenkel, Physica A: Stat. Mech. Appl. **263**, 26 (1999).
[15]M. Dove, *Introduction to Lattice Dynamics* (Cambridge University Press, 1993).
[16]D. A. Kofke, Fluid Phase Equilib. **228–229**, 41 (2005).
[17]T. B. Tan, A. J. Schultz, and D. A. Kofkea, J. Chem. Phys. **132**, 214103 (2010).
[18]T. B. Tan, A. J. Schultz, and D. A. Kofke, J. Chem. Phys. **133**, 134104 (2010).
[19]T. B. Tan, A. J. Schultz, and D. A. Kofke, J. Chem. Phys. **135**, 044125 (2011).
[20]P. A. Monson and D. A. Kofke, Adv. Chem. Phys. **115**, 113 (2000).
[21]A. D. Bruce, N. B. Wilding, and G. J. Ackland, Phys. Rev. Lett. **79**, 3002 (1997).
[22]W. G. Hoover and F. H. Ree, J. Chem. Phys. **49**, 3609 (1968).
[23]D. Frenkel and A. J. C. Ladd, J. Chem. Phys. **81**, 3188 (1984).
[24]W. G. Hoover, S. G. Gray, and K. W. Johnson, J. Chem. Phys. **55**, 1128 (1971).
[25]N. Lu, C. D. Barnes, and D. A. Kofke, Fluid Phase Equilib. **194–197**, 219 (2002).
[26]S. Ryu and W. Cai, Model. Simul. Mater. Sci. Eng. **16**, 085005 (2008).
[27]D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic, San Diego, 2002).
[28]R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[29]D. A. Kofke, Mol. Phys. **102**, 405 (2004).
[30]N. Lu, D. A. Kofke, and T. B. Woolf, J. Comput. Chem. **25**, 28 (2004).
[31]N. Lu, J. K. Singh, and D. A. Kofke, J. Chem. Phys. **118**, 2977 (2003).
[32]C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[33]M. R. Shirts and J. D. Chodera, J. Chem. Phys. **129**, 124105 (2008).
[34]M. D. Rintoul and S. Torquato, J. Chem. Phys. **105**, 9258 (1996).
[35]Z. Chenga, P. M. Chaikina, W. B. Russelb, W. V. Meyerc, J. Zhub, R. B. Rogersc, and R. H. Ottewilld, Mater. Des. **22**, 529 (2001).
[36]A. D. Dinsmore, J. C. Crocker, and A. G. Yodh, Curr. Opin. Colloid Interface Sci. **3**, 5 (1998).
[37]J. M. Polson, E. Trizac, S. Pronk, and D. Frenkel, J. Chem. Phys. **112**, 5339 (2000).
[38]N. G. Almarza, J. Chem. Phys. **126**, 211103 (2007).
[39]E. de Miguel, R. G. Marguta, and E. M. del Río, J. Chem. Phys. **127**, 154512 (2007).
[40]J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic, London, 1986).
[41]L. V. Woodcock, Faraday Discuss. **106**, 325 (1997).
[42]D. A. Kofke and P. T. Cummings, Mol. Phys. **92**, 973 (1997).
[43]W. G. Hoover, M. Ross, K. W. Johnson, D. Henderson, J. A. Barker, and B. C. Brown, J. Chem. Phys. **52**, 4931 (1970).
[44]G. A. Heiser, R. C. Shukla, and E. R. Cowley, Phys. Rev. B **33**, 2158 (1986).
[45]C. Vega and E. G. Noya, J. Chem. Phys. **127**, 154113 (2007).
[46]J. Chang and S. I. Sandler, J. Chem. Phys. **118**, 8390 (2003).
[47]J. Kolafa, Mol. Phys. **59**, 1035 (1986).
[48]H. Paliwal and M. R. Shirts, J. Chem. Theory Comput. **7**, 4115 (2011).