

## Monte Carlo Sampling of a Markov Web

Georgios C. Boulougouris and Daan Frenkel\*

FOM Institute for Atomic and Molecular Physics, Kruislaan 407,  
1098 SJ Amsterdam, The Netherlands

Received November 3, 2004

**Abstract:** The efficiency of Markov-Chain Monte Carlo simulations can be enhanced by exploiting information about trial moves that would normally be rejected. The original presentation of this approach was limited to a specific MC sampling scheme. Here we present a general derivation of a method to improve the sampling efficiency of Monte Carlo simulations by collecting information about the microstates that can be linked directly to the sampled point via an independent Markov transition matrix. As an illustration, we show that our approach greatly enhances the efficiency of a scheme to compute the density of states of a square-well fluid.

Markov-Chain Monte Carlo (MCMC) methods have been used extensively for numerical evaluation of thermodynamic properties of molecular systems.<sup>1–10</sup> The core of any MCMC program is an algorithm that generates a Markov chain of configurations. By a judicious choice of the transition probability from one point in the chain to the next, the overall probability of visiting each microstate can be made proportional to its statistical weight  $\rho$  (e.g. the Boltzmann weight, in the case of thermal systems).

MCMC algorithms traditionally construct a Markov-Chain using two steps:<sup>2</sup> a. Starting from the current (old) state (o) a trial move is attempted to a new state (n) according to a trial probability  $\alpha$  (for example, displacement, rotation or regrowth of a particle<sup>4</sup>). b. The trial state is then accepted or rejected according to an acceptance rule that ensures detailed balance (or, at least, the less strict balance<sup>3</sup>) between sampled microstates.

$$\pi_{ij} = \alpha_{ij} P_{\text{acc}}(ij) \quad \forall i \neq j, \quad \pi_{ii} = 1 - \sum_j \pi_{ij} \quad (1)$$

$$\sum_j \pi_{ij} = 1, \quad \forall i \quad (2)$$

Here, and in what follows, we use the term “balance” to describe algorithms that leave the equilibrium distribution invariant. That is when we apply one Monte Carlo step to an equilibrium distribution of initial states, the average flux

into any given state  $i$  is exactly balanced by the total flux out of that state. As a consequence, the probability density is not changed by such the Monte Carlo algorithm. The stronger “detailed balance” condition states that, for any pair of states  $i$  and  $j$  in an equilibrium ensemble, the average flux from  $i$  to  $j$  is equal and opposite to the flux from  $j$  to  $i$ . Clearly, the latter condition can only be satisfied if the algorithm satisfies microscopic reversibility: i.e., if  $j$  can be reached from  $i$ , then there is a finite probability to carry out the reverse move from  $j$  to  $i$ .

In eq 1  $\pi_{ij}$  is the  $ij$ th element of the Markovian transition matrix. For  $i \neq j$ ,  $\pi_{ij}$  is given by the product of two terms: the probability of attempting a trial move from state  $i$  to state  $j$  and the probability of accepting  $j$  as the new state. The transition matrix is a stochastic matrix therefore its elements must obey eq 2.

In a Monte Carlo simulation, thermodynamic properties are evaluated as the expectation values of the corresponding instantaneous properties. When a trial move is not accepted, the instantaneous property of the old state has to be recounted in the calculated average, and no information about the rejected state is included in the computation of averages. Let us now consider MC algorithms that satisfy detailed balance. The condition of detailed balance is satisfied whenever eq 3 holds for any pair of old states (o) and new states (n).

$$\rho_o \pi_{on} = \rho_o \alpha_{on} P_{\text{acc}}(on) = \rho_n \alpha_{no} P_{\text{acc}}(no) = \rho_n \pi_{no} \quad (3)$$

Equations 1–3 do not uniquely define  $P_{\text{acc}}(on)$ : there is therefore a relative freedom in the choice of the functional

\* Corresponding author phone: +31-20-6081234; fax: +31-20-6684106; e-mail: frenkel@amolf.nl.

form. The most popular choice is the Metropolis rule.<sup>5</sup> An alternative (and usually less efficient) choice is the “symmetric” rule proposed by Barker.<sup>6</sup>

In special cases it is possible to choose the trial probability  $\alpha_{ij}$  in such a way that all trial moves are accepted.<sup>7–10</sup> Usually, however,  $P_{acc}(on) \leq 1$ , and there is a probability  $1 - P_{acc}(on)$  that the trial move will be rejected. In that case the new state is rejected, and all information about it is discarded. Recently, one of us<sup>11</sup> showed that information about the rejected states can be included in the calculation of equilibrium properties. However, the method discussed in ref 11 was restricted to one specific Monte Carlo scheme.<sup>6</sup> Below we present a more general derivation that allows the evaluation of equilibrium properties by combining importance sampling and local sampling of microstates using any combination of valid MCMC schemes. To make this explicit, we consider two sets of transition probabilities (that may or may not be the same): the first describes a conventional sampling of microstates while the second describes the local sampling of a group of microstates that are connected to the individual microstates of the first Markov chain. We use the term “Markovian web” to denote the set of microstates thus connected.

We assume that the Markov chain that connects the states of the Markovian web satisfies detailed balance. Then eq 3 applies for any given pair of connected microstates  $\{o, n\}$ . From eqs 2 and 3 the balance condition may be derived:<sup>2</sup>

$$\sum_m \rho_m \tau_{mn}^{web} = \sum_m \rho_n \tau_{nm}^{web} = \rho_n \sum_m \tau_{nm}^{web} = \rho_n \quad (4)$$

Let us now consider the sampling of an equilibrium property  $A$  of the system:

$$\langle A \rangle_\rho = \sum_n A_n \rho_n / \sum_n \rho_n \quad (5)$$

We now substitute  $\rho_n$  from eq 4. By changing the order of summation (integration in the continuum case) over all microstates we get

$$\begin{aligned} \frac{\sum_n A_n \rho_n}{\sum_n \rho_n} &= \frac{\sum_n \sum_m A_n \rho_m \tau_{nm}^{web}}{\sum_n \sum_m \rho_m \tau_{nm}^{web}} = \frac{\sum_m \rho_m \sum_n A_n \tau_{nm}^{web}}{\sum_m \rho_m \sum_n \tau_{nm}^{web}} \\ &= \frac{\sum_m \rho_m \sum_n A_n \tau_{nm}^{web}}{\sum_m \rho_m} \leftrightarrow \langle A \rangle_\rho = \langle \sum_n \tau_{nm}^{web} A_n \rangle_{\rho_m} \quad (6) \end{aligned}$$

where the last step required the use of eq 2. Equation 6 is exact and is in principle sufficient to describe how the expectation value of a property can be evaluated by combining importance sampling and integration over the local states of a Markovian web; i.e. for every state  $\{m\}$  sampled via importance sampling an integration is performed over all  $\{n\}$  states for which  $\tau_{mn}^{web} \neq 0$ . Equation 6 is quite general, and it may be implemented in many ways. In the present paper, we focus on an application to Monte Carlo

sampling of the equilibrium properties of a simple off-lattice system. The transition matrix of the Markov chain that is used to generate the Markovian web can be separated in two terms as in the case of MCMC discussed above (eq 1). The sum in eq 6 may be broken into two terms using eq (1):

$$\sum_n A_n \tau_{mn}^{web} = \sum_{n \neq m} A_n \tau_{mn}^{web} + A_m (1 - \sum_{n \neq m} \tau_{mn}^{web}) \quad (7)$$

Substituting (7) into (6) and using the fact that  $\sum_{n \neq m} \alpha_{mn} = 1$ , a new expression for the average properties is derived:

$$\langle A \rangle_\rho = \langle \sum_{n \neq m} \alpha_{mn} [A_n P_{acc}(mn) + A_m (1 - P_{acc}(mn))] \rangle_{\rho_m} \quad (8)$$

The summation in eq 8 may be performed in many ways. One possible approach is to express the sum as a weighted average over the trial probability  $\alpha_{ij}$ .

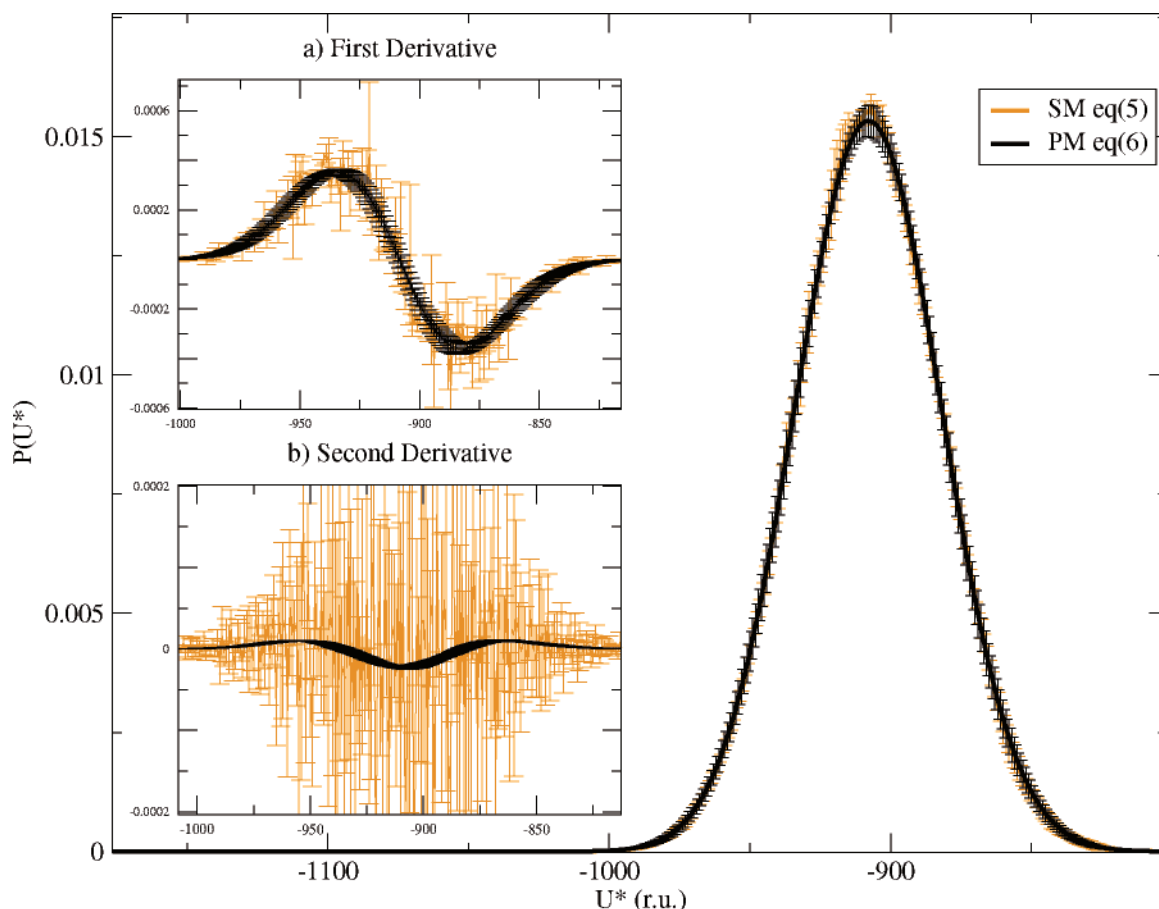
$$\langle A \rangle_\rho = \langle \langle A_n P_{acc}(mn) + A_m (1 - P_{acc}(mn)) \rangle_{\alpha_{mn}} \rangle_{\rho_m} \quad (9)$$

Equation 9 implies that the sum required for eq 6 can be computed as a weighted average, where the weight is the trial probability  $\alpha_{ij}$ . In practice, this means that we perform not only an importance sampling of the states  $n$  with corresponding weights  $\rho_m$  but also we generate a finite subset of the trial states  $n$ , with weights  $\alpha_{mn}$ .

We stress that eq 6 is general, whereas eq 9 represents one of the many possible implementations of the general case.

An important aspect of the proposed method is that the transition matrix of the Markovian web is completely independent from the Markovian matrix that is used to generate the importance sampling (the importance sampling could even have been generated by a completely different method, e.g. constant-temperature Molecular Dynamics). In other words: the two transition matrices may differ in their trial probability or in their acceptance probability (e.g. use of Metropolis or the symmetric rule) or in both.

It is advantageous to consider situations where the number of trial states is very large. Such a situation arises for instance when the system under consideration can be decomposed into noninteracting subsystems. An example is a system consisting of particles with intermolecular interactions that can be truncated beyond a finite cutoff distance  $r_c$ . If we divide the system into cells with diameter  $d \geq r_c$ , then nonadjacent cells will not interact. This means that the probability of acceptance of a trial move in one such cell does not depend on the possible outcome of a trial move inside all other nonadjacent cells. This makes it possible to construct “parallel” trial moves that attempt to move particles in a set of  $N$  noninteracting cells. In normal MCMC, the result of such a parallel move is that the system ends up in one specific final state (out of the  $2^N$  possible states). With the present algorithm we can do much better. By generating  $M$  trial states in each of the subsystems we can evaluate the term  $\sum_n A_n \tau_{mn}^{web}$  by summing all  $(M + 1)^N$  possible combinations of trial states for the subsystems. The number of trial microstates can be very large. Hence, special care should be taken to compute averages over trial states efficiently. Since the subsystems are independent, one can calculate the



**Figure 1.** Main figure: Comparison of  $P(U)$ , the probability density of the potential energy, for a system of 2048 SW particles at  $\rho^* = 0.542$  as obtained from conventional Monte Carlo sampling of serial moves (SM) using eq 5 and from parallel displacement moves of 256 particles, using eq 6 (PM). Insets: a) The first derivative of the probability density with respect to the potential energy. b) The second derivative of the probability density with respect to the potential energy. Same color coding.

probability of finding the total system in a specific macrostate (e.g. the collection of microstates with the same potential energy) as the convolution of the probabilities of the subsystems. It should be noted that the use of the convolution or even the independence between the subsystems is not a necessary requirement, but it is expected to be crucial in the case of a large number of possible trial states.

In some cases, the efficiency of the scheme may be enhanced by iterating the steps that led to the derivation of eq 6. In that case, we construct a Markov web inside a Markov web:

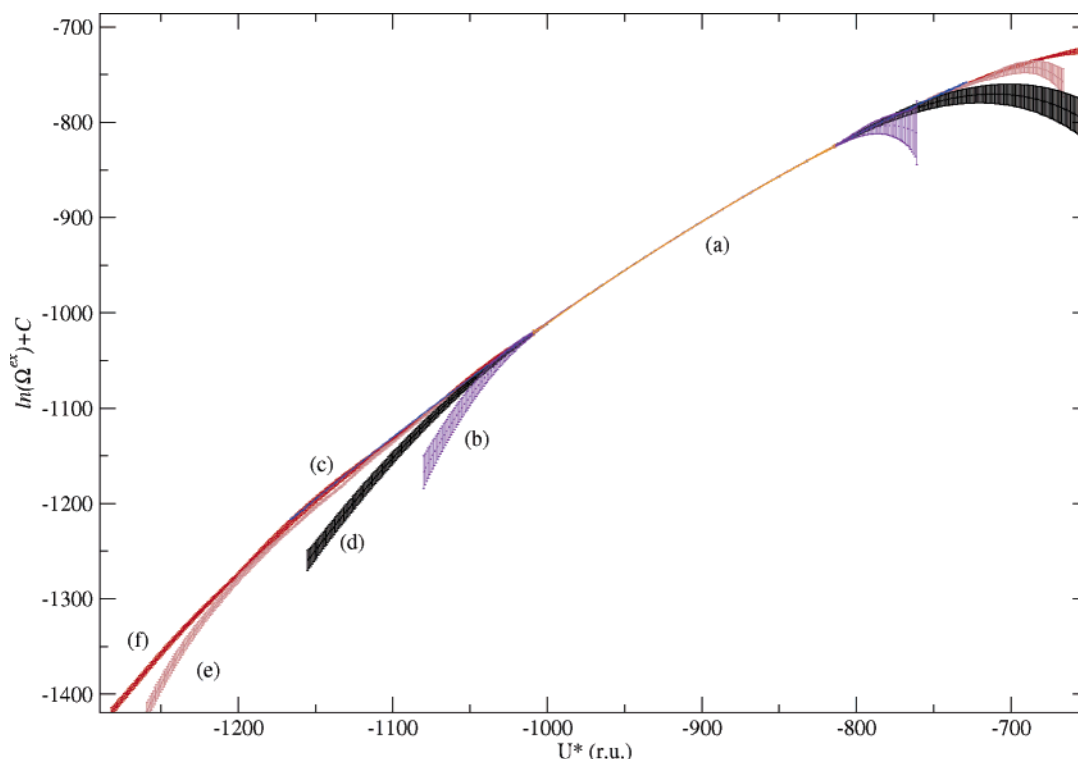
$$\langle A \rangle_\rho = \frac{\sum_k \rho_k A_k}{\sum_k \rho_k} = \frac{\sum_i \rho_i \sum_j \pi_{ij}^{(1)} \sum_k \pi_{jk}^{(2)} A_k}{\sum_i \rho_i} \quad (10)$$

In our simulations, we use such a nested Markov web because we combine a set of parallel single-particle trial displacements with a set of parallel tempering moves where we attempt to swap (again in a parallel move) systems with different temperatures.<sup>12</sup>

To illustrate the power of the proposed approach, we have applied the above algorithm to the calculation of the energy histogram of a square well (SW) system. Although the SW

potential is one of the simplest intermolecular potentials, the model exhibits glassy behavior for systems with short ranged interactions at low temperatures. This makes Monte Carlo sampling extremely difficult. Recently<sup>13</sup> we proposed an algorithm that was designed to overcome these sampling problems, allowing us to sample configuration space much more efficiently than using traditional methods. The method is an extension of the approach used in ref 14 and is based on the inclusion–exclusion theorem. It explicitly computes the accessible volumes in which a test particle would experience the square-well attraction of zero, one, two, etc. other particles. The technical details of this method are described in ref 13. What matters for the present discussion is that this algorithm, rather than considering a single trial displacement of a particle, computes the Boltzmann weights of a large number of trial positions. In a conventional algorithm, all trial states but one would be discarded. Here, we consider this algorithm precisely because it generates many trial states for a single particle move. In addition, as the square-well interactions that we consider are short ranged, we can perform  $N$  independent trial moves in parallel.

In Figure 1 we present a comparison between the traditional averaging, and the proposed methodology for the calculation of the energy histogram, and its numerical derivatives with respect to the energy, in a SW fluid with a



**Figure 2.** The logarithm of the excess density of states  $\Omega^{ex}(U^*)$  for 2048 SW particles at  $\rho^* = 0.542$  as obtained from a single run at  $T^* = 1$ : (a) (orange) averaging using eq 5, (b) (indigo) from parallel integration over 64 particle displacements using eq 6; (c) (blue) from multiple histogram reweighting of 5 systems with neighboring temperatures (see text); (d) from parallel integration over 256 particle displacements using eq 6. Curves (e) (brown) and (f) (red) were obtained by combining the parallel particle trial displacements with a set of parallel tempering moves between all 5 systems using eq 10: (e) and (f) correspond to 64 and 256 parallel particle displacements, respectively. The potential energy is expressed in units of the depth of the square well.

short-ranged attraction with a well width of  $\lambda\sigma = 1.025\sigma$ . Single-particle trial displacements were performed in a volume of  $(0.40\sigma)^3$ . The system contains 2048 SW particles at reduced density  $\rho^* = 0.542$ . The algorithm of ref 13 is used for the single-particle displacements. In our parallel calculations reported in Figure 1 we used 256 parallel displacements. Figure 1 shows that the use of the parallel algorithm enables us to achieve better statistics compared to the conventional algorithm, since it enables us to collect information for  $v^m$  points for every important sampled point, where  $v$  is the test volume in which we attempt the displacement and  $m$  is the number of parallel moves attempted. This is evident from the insets of Figure 1 where the numerical derivatives of the energy histogram are evaluated. These derivatives are related to thermodynamic derivatives, and the error in their evaluation is dramatically reduced by the parallel averaging scheme.

Figure 2 shows the excess density of states  $\Omega^{ex}(U)$  for a system of 2048 SW particles with  $\lambda = 1.025$ , at reduced density  $\rho^* = 0.542$ . The excess density of states was computed from  $P(U)$ , the histogram of the potential energy of the system, using

$$\ln[\Omega^{ex}(U)] = \ln[P(U)e^{+\beta U}] + C \quad (11)$$

For the sake of comparison, the orange curve *a* shows the results of a single run at  $T^* = 1.0$ , using conventional averaging (eq 5). In addition, we used the multiple-histogram method to obtain  $\Omega^{ex}(U)$  over a wider energy range (curve

*c*). This curve was obtained by combining the histograms obtained in 5 simulations at reduced temperatures  $T^* = \{0.8, 0.9, 1, 1.1, 1.2\}$ . The remaining curves were obtained at  $T^* = 1.0$ , using different versions of the Markovian-web algorithm: Curve *b* (yellow) was obtained using eq 6. Curves *b* (indigo) and *d* (black) are also based on eq 6, but in this case the trial moves consisted of 64 (*b*) and 256 (*d*) parallel trial displacements. The curves *e* and *f* were obtained using a “nested” web (eq 10). To obtain these curves, we combined the parallel displacement moves with parallel tempering trial moves in which we attempted all possible swaps of the temperatures of five simulations at reduced temperatures  $T^* = \{0.8, 0.9, 1, 1.1, 1.2\}$ .<sup>12</sup> Curve *e* corresponds to 64 parallel particle displacement and curve *f* to 256. In the figure, we have only included the sampling due to trial moves that originate from the system at  $T^* = 1.0$ . The figure shows that the present algorithm greatly improves the sampling of the low- and high-energy wings of the density of states. But, of course, at some point local sampling – no matter how good – is inadequate, and the estimated density of states starts to deviate from the more accurate curve obtained (at a higher cost) from a multiple-histogram calculation.

In this paper we argued that by using the basic properties of Markov chains it is possible to construct a sampling scheme where for each point of the phase space sampled via a given Markov-Chain or other equivalent means of importance sampling (e.g. Molecular Dynamics for the case of molecular systems), a subsampling is performed of the

points connected to the sampled point via an independent Markov web. By Markov web we mean the set of all points in the configuration space that are linked to the instantaneous configuration of the system via an independent Markovian transition matrix. By making use of the condition of detailed balance (or balance) of the transition matrix, we can sample these points according to their Boltzmann weight, even though the generation of the trial states is correlated to the old state. Using this approach, the sampling efficiency of the Monte Carlo simulation can be greatly enhanced.

If the intermolecular interactions are short ranged, it is straightforward to generate  $2^n$  trial states by constructing parallel trial moves of  $n$  noninteracting particles. However, the potential advantage of the method is not limited to systems with short-ranged interactions. What is essential is the existence of a large number of potential final states. For instance, Liu and Luijten<sup>10</sup> have developed a rejection-free cluster algorithm that works for systems with long-ranged interactions. In the conventional version of this algorithm, only one particular cluster state would be sampled. We expect that the present algorithm should make it possible to sample all possible cluster states. However, we have not yet attempted to do so.

We stress that the present algorithm improves the local sampling of configuration space, but not the rate at which the systems moves through configuration space. In case that diffusion in configuration space is slow (e.g. in glassy systems), the present algorithm must be combined with existing schemes, such as parallel tempering, that help the system escape from local minima in configuration space. Interestingly, the parallel-tempering algorithm itself can also be viewed as a scheme that generates many trial states yet accepts only one.

The present algorithm differs fundamentally from existing schemes that use information about virtual moves to construct information about the density of macrostates. Starting with the Widom particle-insertion scheme to compute chemical potentials (see e.g. ref 2), many schemes have been developed that employ virtual moves to construct a density of macrostates (see e.g. refs 15–17). However, in the present scheme, we use virtual moves to collect information about the properties of individual microstates. In the example that we showed, we used the present scheme to compute a density of macrostates, and, for this specific application, existing schemes might also be used. However, the difference is that the present method allows us to compute many distributions of macrostates at the same time, because the sampling considers microstates. Furthermore the proposed scheme can be used to evaluate directly the ensemble average of any desired property, whereas most of the older methods are restricted to thermodynamic properties related to the density of macrostates.

**Acknowledgment.** G.B. would like to thank I. Coluzza and S. Tindemans for helpful discussions. This work is part of the research program of the “Stichting voor Fundamenteel Onderzoek der Materie (FOM)”, which is financially supported by the “Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)”. The support by the European Community through a Marie Curie individual Fellowship with contract No. HPMF-CT-2002-02089, is gratefully acknowledged.

## References

- (1) Spyriouni, T.; Economou, I. G.; Theodorou, D. N. *J. Am. Chem. Soc.* **1999**, *121*, 3407–3413. Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. *Phys. Rev. E* **2002**, *66*, 056703. Harmandaris et al. *Macromolecules* **2003**, *36*, 1376–1387.
- (2) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: New York, 2002. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.
- (3) Manousiouthakis, V. I.; Deem, M. W. *J. Chem. Phys.* **1999**, *110*, 2753–2756.
- (4) Karayiannis, N. C.; Mavrantzas, V. G.; Theodorou, D. N. *Phys. Rev. Lett.* **2002**, *88*, 105503. Zervopoulou, E.; Mavrantzas, V. G.; Theodorou, D. N. *J. Chem. Phys.* **2001**, *115*, 2860–2875.
- (5) Metropolis et al. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (6) Barker, A. A. *Aust. J. Phys.* **1965**, *18*, 119–133.
- (7) Bortz, A. B.; Kalos, M. H.; Lebowitz, J. L. *J. Comput. Phys.* **1975**, *17*, 10–18.
- (8) Schulz, B. J.; Binder, K.; Müller, M. *Int. J. Mod. Phys. C* **2002**, *13*, 477–494.
- (9) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1987**, *58*, 86–88.
- (10) Liu, J.; Luijten, E. *Phys. Rev. Lett.* **2004**, *92*, 035504.
- (11) Frenkel, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17571–17575.
- (12) Coluzza, I.; Frenkel, D. Submitted for publication.
- (13) Boulougouris, G. C.; Frenkel, D. Submitted for publication.
- (14) Miller, M.; Frenkel, D. *Phys. Rev. Lett.* **2003**, *90*, 135702.
- (15) Wang, F. G.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, 056101. Wang, F. G.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050. Dang, L. X.; Chang, T. M.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **2002**, *117*, 3522–3523.
- (16) Fitzgerald, M.; Picard, R. R.; Silver, R. N. *Europhys. Lett.* **1999**, *46*, 282, 6269. Fitzgerald, M.; Picard, R. R.; Silver, R. N. *J. Stat. Phys.* **2000**, *98*, 321. Errington, J. R. *J. Chem. Phys.* **2003**, *118*, 22, 9915–9925. Errington, J. R. *Phys. Rev. E* **2003**, *67*, 012102.
- (17) Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **2003**, *119*, 9406–9411.