

LA-UR- 09-00198

Approved for public release;  
distribution is unlimited.

*Title:* Optimal Sampling Efficiency in Monte Carlo Sampling with  
an Approximate Potential

*Author(s):* Joshua D. Coe, Thomas D. Sewell, and M. Sam Shaw

*Intended for:* Journal of Chemical Physics



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

## Optimal Sampling Efficiency in Monte Carlo Simulation with an Approximate Potential

Joshua D. Coe<sup>§</sup>, Thomas D. Sewell<sup>†</sup>, and M. Sam Shaw<sup>§</sup>

<sup>§</sup>*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545*

<sup>†</sup>*Department of Chemistry, University of Missouri-Columbia, MO 65211-7600*

Building on the work of Iftimie *et al.* [J. Chem. Phys. **113**, 4852 (2000)] and Gelb [J. Chem. Phys. **118**, 7747 (2003)], Boltzmann sampling of an approximate potential (the “reference” system) is used to build a Markov chain in the isothermal-isobaric ensemble. At the endpoints of the chain, the energy is evaluated at a higher level of approximation (the “full” system) and a composite move encompassing all of the intervening steps is accepted on the basis of a modified Metropolis criterion. For reference system chains of sufficient length, consecutive full energies are statistically decorrelated and thus far fewer are required to build ensemble averages with a given variance. Without modifying the original algorithm, however, the maximum reference chain length is too short to decorrelate full configurations without dramatically lowering the acceptance probability of the composite move. This difficulty stems from the fact that the reference and full potentials sample different statistical distributions. By manipulating the thermodynamic variables characterizing the reference system (pressure and temperature, in this case), we maximize the average acceptance probability of composite moves, lengthening significantly the random walk between consecutive full energy evaluations. In this manner, the number of full energy evaluations needed to precisely characterize equilibrium properties is dramatically reduced. The method is applied to a model fluid, but implications for sampling high-dimensional systems with *ab initio* or density functional theory (DFT) potentials are discussed.

## I. INTRODUCTION

Characterization of thermodynamic equilibrium using Markov Chain Monte Carlo (MC)<sup>2</sup> methods is now well-established practice<sup>1-4</sup>. Instead of building time-averages for an ensemble of trajectories, as in molecular dynamics (MD)<sup>1,2,5</sup>, configurational integrals are sampled directly at points dictated by a random walk. New points are added to the Markov chain on the basis of an acceptance criterion, most often that of Metropolis<sup>6</sup>, and the simulation is complete when variance in (thermodynamic) ensemble averages has dropped to an acceptable level. This level varies inevitably with application, but the number of steps required to achieve a target variance usually rises with the dimensionality of configuration space. For this reason, precision sampling of high-dimensional systems remains a serious challenge.

Methodological improvements in solving the electronic Schrödinger equation, coupled with steady advances in computing power, have made single-point calculation of *ab initio* (AI)<sup>7</sup> or density functional theory (DFT)<sup>8</sup> energies routine even for very large systems<sup>7</sup>. Paired with algorithms for extracting forces from wavefunctions (or densities) analytically<sup>9</sup>, these improvements lead directly to steady growth in the application of *ab initio* MD (AIMD)<sup>10</sup>. The potential energy surface (PES) in AIMD is built “on-the-fly” using quantum chemistry in place of *ad hoc* functional forms, permitting more robust and accurate sampling of phase space. Expanded use of AIMD has not been matched, however, by commensurate growth of *ab initio* (MC)<sup>2</sup>, although the use of *ab initio* potentials in (MC)<sup>2</sup> simulation has been the subject of very recent attention<sup>11,12</sup>. While MD steps are collective and deterministic, standard (MC)<sup>2</sup> steps are individual and stochastic; the computational exchange made in substituting MD for (MC)<sup>2</sup> is that of

force calculation at every time step in return for steps encompassing all particles in the system. The single-particle character of standard  $(MC)^2$  steps can be exploited to lower their cost from  $O(N^2)$  to  $O(N)$  in a system of  $N$  particles described by a pair potential, but no analogous reduction is afforded self-consistent potentials including interaction levels much higher than the pair. It remains true, however, that  $(MC)^2$  possesses inherent flexibility unavailable to MD<sup>13</sup>, such as constant temperature or pressure sampling without need of a stochastic bath, or chemical equilibrium sampling without need of a reactive potential surface<sup>14</sup>. For these reasons, it is worthwhile to explore  $(MC)^2$  algorithms harnessing the accuracy of *ab initio* quantum chemistry without requiring full system energy evaluation following every single-particle displacement.

One alternative is to build trial moves from collective displacement of several or even all particles. The acceptance probability of a collective move will be much lower than that of its constituents taken independently, however, because single-particle steps are chosen randomly and thus lack information regarding the potential. This fact is illustrated clearly in a hard-sphere fluid, where the likelihood that two particles will overlap increases monotonically with the number of particles displaced; if a collective step yields even a single overlap, its acceptance probability will vanish entirely. This will result in many wasted trial steps, a weighty consideration if each acceptance test requires significant computing time. The radius of trial moves could be dramatically reduced in order to salvage the acceptance probability, but only at the expense of slow configuration space exploration; as before, precise equilibrium averages will require many energy evaluations. In this sense, MD steps can be viewed as “directed” forms of multi-particle  $(MC)^2$  moves, in the sense that time-reversible integration of the equations of motion

guarantees energy conservation and thus unit acceptance probability of the “trial move”. No such guarantee exists when trial moves are chosen stochastically.

An alternative means of building  $(MC)^2$  steps was introduced recently by Ifimie, *et al.*<sup>15</sup>, followed by an independent treatment from Gelb<sup>16</sup>. Although several monikers have been applied<sup>17</sup>, we will refer to this procedure as Nested Markov Chain Monte Carlo ( $N(MC)^2$ ). The method is conceptually related to  $(MC)^2$  with stochastic potential switching<sup>18</sup>, multiple “time steps”<sup>19</sup>, multilevel summation<sup>20</sup>, and resolution exchange<sup>21</sup>. In  $N(MC)^2$  a series of elementary moves (in the  $NPT$  ensemble, single-particle or volume adjustments), each accepted with Boltzmann weight, is made in a “reference” system defined by an inexpensive (but less accurate) potential. At the endpoints of this sequence, the energy is evaluated again with a more accurate potential defining the “full” system. Through appropriate modification of the acceptance criterion, the reference system Markov chain is transformed into a composite trial step accepted with Boltzmann weight in the full system. As long as the reference potential captures adequately the physics of the full potential, these composite trial moves retain a reasonable probability of acceptance; the more reference steps comprising a composite move (or the less capably the reference potential captures the interactions present in the full potential), the lower its acceptance probability. The difficulty is that the reference and full potentials sample different statistical distributions, and so the number of reference steps combinable into a composite step is strongly limited by the practical need of a reasonable acceptance probability for the latter. In spite of this difficulty,  $N(MC)^2$  permits  $(MC)^2$  sampling of an accurate potential without having to evaluate it following every single-particle displacement, and in this sense represents an important step toward realistic

implementation of  $(MC)^2$  with an *ab initio* potential. Although its application already has been fairly extensive<sup>12,22</sup>, the present work attempts to improve upon the original  $N(MC)^2$  algorithm by addressing its principal weakness; namely, the potentially poor overlap of reference and full distributions.

In order to minimize the number of full energy evaluations required to achieve target variance in ensemble averages, configurations at which the full energy is evaluated should be as decorrelated (*vide infra*) as possible. Decorrelation requires separation by a large number of reference steps, a number constrained also by the acceptance probability for the composite step. By manipulating the thermodynamic variables of the reference system, we show how to maximize the overlap of reference and full distributions. This procedure maximizes also the acceptance probability for composite steps built from a fixed number of reference steps, minimizes the correlation of energies sampled in the full system, and thereby lowers considerably the number of full energy evaluations needed to sample with high precision.

Section II describes the potentials used to generate the results that follow. The next sections provide a brief overview of conventional Monte Carlo sampling (III) and basic  $N(MC)^2$  (IV). Section V contains our primary contribution, wherein we outline a means of optimizing  $N(MC)^2$  sampling efficiency. Section VI summarizes and offers some suggestions for further development.

## II. POTENTIAL MODEL

The  $N(MC)^2$  procedure evaluates the energy of a configuration using two different potentials: an approximate potential for single-particle steps, and a more accurate one for composite steps. We assume that quantities for comparison with

experiment are computed in the full system. In the present work, the purpose of which is to test and optimize the procedure, we will utilize combinations of pair potentials equal in computational expense but differing in their parameterization. In the future, a model potential will be used as reference for a full system characterized with DFT<sup>23</sup>.

The model potentials used below describe diatomic molecules of fixed bond length and with interaction sites at their atomic centers. Pairwise interaction of atomic sites  $a$  and  $b$  are modeled with the Buckingham exponential-6 potential:

$$\varphi(r_{ab}) = \frac{\epsilon}{\alpha - 6} \left( 6e^{\alpha(1-r_{ab})} - \frac{\alpha}{r_{ab}^6} \right), \quad (1)$$

where

$$r_{ab}^2 = \frac{0.25r_{ij}^2(l_i^2 + l_j^2) + (-1)^a(\mathbf{r}_{ij} \cdot \mathbf{l}_i) + (-1)^b(\mathbf{r}_{ij} \cdot \mathbf{l}_j) + (-1)^{a+b}0.5(\mathbf{l}_i \cdot \mathbf{l}_j)}{r_0^2}. \quad (2)$$

The site-site separation distance  $r_{ab}$  has been expressed in terms of the center-of-mass (COM) separation vector ( $\mathbf{r}_{ij}$ ) for interacting molecules  $i$  and  $j$ , and the individual bond vectors  $\mathbf{l}_i$  and  $\mathbf{l}_j$  (of length  $l_i$  and  $l_j$ ). The full interaction of two diatomics is then

$$\varphi_{ij} = \sum_{a=1}^2 \sum_{b=1}^2 \varphi(r_{ab}). \quad (3)$$

The potential parameters  $\alpha$ ,  $\epsilon$ , and  $r_0$  were chosen to roughly approximate compressed nitrogen fluid on its shock Hugoniot locus<sup>24</sup>. Details of the fitting procedure used to determine these parameters will be described in an upcoming publication<sup>23</sup>. The final values are

$$\begin{aligned} \epsilon &= 34.156 \text{ K} \\ r_0 &= 4.037 \text{ Angstrom} . \\ \alpha &= 12.29 \end{aligned} \quad (4)$$

We enforced a minimum allowable  $r_{ij}$  slightly greater than the  $N_2$  bond length ( $r_{ij}^{\min}=1.20$  Å) as a guarantee of smooth behavior throughout the simulations. In testing the  $N(MC)^2$  procedure,  $l_i$  and  $l_j$  were fixed at 1.10 Å in the full system but shortened in 0.05 Å increments to generate a series of reference systems. The reference potential approaches a purely spherical COM interaction as  $l_j \rightarrow 0$ , thus providing a poorer approximation to the full potential. Because bond lengths in the full system are fixed, and  $l_i=l_j$  in all reference systems, we will refer only to  $l$  (and always in the context of the reference system) in what follows.

Although each site-site interaction described by (1) is spherical, the sum of these interactions (3) for a pair of molecules is highly anisotropic. Figure 1 schematically illustrates some of the quantities appearing in (1)-(2) for a pair of molecules. The pair is drawn in three fiducial configurations<sup>25</sup> labeled **T**, **L**, and **X**, defined by the quartet of angles  $(\theta_1, \chi_1, \theta_2, \chi_2)$ .  $\theta$  ( $\chi$ ) is the angle in (out of) the plane of the paper, and angles are zeroed to the configuration in which the molecules are parallel to one another (not shown). The subscripts label the molecules. Figure 2 displays the variation in potential for each of these configurations as a function of COM separation and bond length  $l$ . The full potential ( $l=1.10$  Å) is compared with a purely spherical potential ( $l=0$ ) and one of the reference potentials ( $l=0.90$  Å) used below. The ordinate is drawn on a log scale, and it is clear that the reference potential may differ substantially both from the full potential and from that of a purely isotropic interaction.

### III. STANDARD MONTE CARLO SAMPLING

In keeping with an earlier presentation of the  $N(MC)^2$  method<sup>16</sup>, we have adopted the structure and notation of Ref. 26 to describe Monte Carlo sampling. Matrices are



indicated by bold lettering, and their individual elements by a subscripted, italicized form of the same symbol. The system is described by a state vector  $\boldsymbol{\pi}$ , each element of which defines the probability that the system is in state  $\pi_i$ . These probabilities vary as steps are added to the Markov chain, a process performed by repeated application of the transition matrix  $\mathbf{p}$  to the state vector  $\boldsymbol{\pi}$ ,

$$\boldsymbol{\pi}^T(n)\mathbf{p} = \boldsymbol{\pi}^T(n+1) \quad (5)$$

Here we have written the states as transposed column vectors and indicated the step number by  $n$ . Following Metropolis<sup>6</sup>, we demand that the  $\pi_i$  be asymptotically distributed according to their Boltzmann weights:

$$\pi_i^* \propto e^{W_i}, \quad (6)$$

where

$$\boldsymbol{\pi}^* = \lim_{n \rightarrow \infty} \boldsymbol{\pi}(1)\mathbf{p}^n. \quad (7)$$

In the isothermal-isobaric ensemble<sup>27</sup>, for which the corresponding potential is the Gibbs free energy<sup>28</sup>,

$$W_i = -\beta(U_i + PV_i) + N \ln V_i. \quad (8)$$

$U_i$  and  $V_i$  are the internal energy and volume (respectively) of state  $i$ ,  $N$  is the total number of atoms in the system, and  $\beta$  has its usual meaning as the inverse product of temperature with the Boltzmann constant,  $(k_B T)^{-1}$ . A simple step toward realization of (6) is construction of the transition matrix  $\mathbf{p}$  such that

$$\boldsymbol{\pi}^* \mathbf{p} = \boldsymbol{\pi}^*, \quad (9)$$

meaning that once reached, the limiting distribution is permanently maintained. This is known as the *balance condition*. The elements of  $\mathbf{p}$  define the probability of transition between various states, meaning that

$$p_{ij} \geq 0 \quad \forall i, j. \quad (10)$$

Conservation of probability mandates that

$$\sum_j p_{ij} = 1 \quad \forall i \quad (11)$$

as well. Equation (11) identifies  $\mathbf{p}$  as a *stochastic* matrix. The Markov chain is irreducible (or ergodic) if there exists some  $n$  such that

$$[p^n]_{ij} > 0, \quad \forall i, j, \quad (12)$$

establishing that any final state can be reached from any initial state simply by repeated application of  $\mathbf{p}$  to an (arbitrary) initial state vector  $\boldsymbol{\pi}(1)$ . If  $\mathbf{p}$  is stochastic and irreducible, the Perron-Frobenius theorem<sup>4</sup> ensures that it possesses a single left eigenvector having unit eigenvalue, and that this eigenvector represents the limiting distribution. This guarantee does not, however, ensure that the limiting distribution is the Boltzmann distribution. For this, an explicit form of  $p_{ij}$  compatible with (6) must be specified.

A helpful constraint in this regard is *microscopic reversibility*,

$$\pi_i p_{ij} = \pi_j p_{ji}. \quad (13)$$

Although (13) is unnecessarily strong<sup>1,29</sup>, its combination with (11) guarantees satisfaction of (9). We now restrict  $p_{ij}$  to the product form

$$p_{ij} = q_{ij} \alpha_{ij}, \quad (14)$$

where  $q_{ij}$  is the unbiased (or marginal) probability of making a trial step from state  $i$  to state  $j$ , and  $\alpha_{ij}$  is the (conditional) probability of accepting such a move. The average number of systems attempting this transition will be  $\pi_i q_{ij}$ , and the average number attempting the reverse transition will be  $\pi_j q_{ji}$ . Metropolis *et al.*<sup>6</sup> were the first to show that (6) will be satisfied when

$$\alpha_{ij} = \min\left(\frac{q_{ji}\pi_j}{q_{ij}\pi_i}, 1\right), \quad (15)$$

if the  $\pi_i$  are defined as  $e^{-W_i}$ . In the (very common) event that  $q_{ij} = q_{ji}$  by construction, this reduces to

$$\alpha_{ij} = \min\left(e^{W_j - W_i}, 1\right). \quad (16)$$

The choice of  $\alpha_{ij}$  given in (15) and (16) satisfies microscopic reversibility as well, so long as there is a nonzero probability of remaining in the same state,

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij} \neq 0. \quad (17)$$

The matrix elements  $q_{ij}$  represent the probability of making a trial move, such as a displacement or a volume change. For single-particle displacements limited to a sphere of cutoff radius  $r_c$ ,  $q_{ij}$  is the uniform probability of choosing a trial state  $j$  in which a single particle has been moved to a different point within the sphere; this uniformity is what permits reduction of (15) to (16). For more sophisticated move types such as the composite moves introduced below, the distribution of trial moves may *not* be uniform, in which case  $\mathbf{q}$  will assume a more complicated form. In that case, the simple

decomposition of  $p_{ij}$  assumed in (14) can be leveraged to yield the new matrix  $\alpha$  in relatively straightforward fashion. This procedure is illustrated in the next section.

#### IV. NESTED MARKOV CHAIN MONTE CARLO

The N(MC)<sup>2</sup> procedure distinguishes the full system of interest from a reference system defined by an alternate potential. In what follows, reference system quantities will be indicated with superscripted zeros. Reference and full system volumes are identical, so no attempt will be made to distinguish the two.

Consider a sequence of  $M$  elementary reference steps connecting configurations  $i$  and  $j$ . Each of these steps is accepted on the basis of the standard Metropolis criterion (15) using reference system energies. We wish to transform this sequence of steps accepted in the reference system into a *trial* step made in the *full* system. The full system  $q_{ij}$  are no longer drawn from a uniform distribution; rather, they are built from a sequence of  $M$  points accepted with Boltzmann weight in the reference system. What is the appropriate form of the new  $\alpha_{ij}$ , the acceptance probability in the full system? The full system  $q_{ij}$  are

$$q_{ij} = \prod_{k=1}^M p_{k-1,k}^{(0)} = \prod_{k=1}^M q_{k-1,k}^{(0)} \alpha_{k-1,k}^{(0)} = \prod_{k=1}^M q_{k-1,k}^{(0)} \min\left(\frac{\pi_k^{(0)} q_{k,k-1}^{(0)}}{\pi_{k-1}^{(0)} q_{k-1,k}^{(0)}}, 1\right), \quad (18)$$

meaning that

$$\frac{q_{ij}}{q_{ji}} = \frac{q_{0,1}^{(0)} \min\left(\frac{\pi_1^{(0)} q_{1,0}^{(0)}}{\pi_0^{(0)} q_{0,1}^{(0)}}, 1\right) \times q_{1,2}^{(0)} \min\left(\frac{\pi_2^{(0)} q_{2,1}^{(0)}}{\pi_1^{(0)} q_{1,2}^{(0)}}, 1\right) \times \dots \times q_{M-1,M}^{(0)} \min\left(\frac{\pi_M^{(0)} q_{M,M-1}^{(0)}}{\pi_{M-1}^{(0)} q_{M-1,M}^{(0)}}, 1\right)}{q_{M,M-1}^{(0)} \min\left(\frac{\pi_{M-1}^{(0)} q_{M-1,M}^{(0)}}{\pi_M^{(0)} q_{M,M-1}^{(0)}}, 1\right) \times \dots \times q_{2,1}^{(0)} \min\left(\frac{\pi_1^{(0)} q_{1,2}^{(0)}}{\pi_2^{(0)} q_{2,1}^{(0)}}, 1\right) \times q_{0,1}^{(0)} \min\left(\frac{\pi_0^{(0)} q_{0,1}^{(0)}}{\pi_1^{(0)} q_{1,0}^{(0)}}, 1\right)}. \quad (19)$$

Following Gelb<sup>16</sup>, note that

$$\frac{q_{k-1,k}^{(0)} \min\left(\frac{\pi_k^{(0)} q_{k,k-1}^{(0)}}{\pi_{k-1}^{(0)} q_{k-1,k}^{(0)}}, 1\right)}{q_{k,k-1}^{(0)} \min\left(\frac{\pi_{k-1}^{(0)} q_{k-1,k}^{(0)}}{\pi_k^{(0)} q_{k,k-1}^{(0)}}, 1\right)} = \frac{\pi_k^{(0)}}{\pi_{k-1}^{(0)}}, \quad (20)$$

which, in combination with reordering of factors in (19), implies that

$$\frac{q_{ij}}{q_{ji}} = \frac{\pi_1^{(0)}}{\pi_i^{(0)}} \times \frac{\pi_2^{(0)}}{\pi_1^{(0)}} \times \frac{\pi_3^{(0)}}{\pi_2^{(0)}} \times \dots \times \frac{\pi_j^{(0)}}{\pi_{M-1}^{(0)}} = \frac{\pi_j^{(0)}}{\pi_i^{(0)}}. \quad (21)$$

Substituting (21) into (15) gives

$$\alpha_{ij} = \min\left(\frac{\pi_j \pi_i^{(0)}}{\pi_i \pi_j^{(0)}}, 1\right), \quad (22)$$

the acceptance probability of composite moves required for Metropolis sampling of the full potential. In comparing  $\alpha_{ij}$  with  $\alpha_{ij}^{(0)}$ , the standard ratio of Boltzmann factors for initial and final states of the full system has been augmented by the inverse of the standard ratio in the reference system (to which Gelb refers as a ‘correction factor’). The  $\pi_k^{(0)}$  in (22) are evaluated using the reference system temperature ( $T^{(0)}$ ) and pressure ( $P^{(0)}$ ), but this in no way precludes use of a different pressure ( $P$ ) and temperature ( $T$ ) in building  $\pi_k$  for the full system. Abbreviating the difference between full and reference potentials for state  $k$  as  $W_k - W_k^{(0)} \equiv \delta W_k$ , and  $\delta W_j - \delta W_i \equiv \Delta W$ , (22) can be reexpressed as:

$$\alpha_{ij} = \begin{cases} 1 & , \Delta W \geq 0 \\ e^{\Delta W} & , \Delta W < 0 \end{cases}. \quad (23)$$

If the reference energy always were related to the full energy by a simple constant shift

$$U_n^{(0)} + c = U_n, \quad (24)$$

the product in (22) would never deviate from unity, and thus *all* composite moves would be accepted, regardless of the magnitude of  $M$ . A distribution of  $\delta W$  implies a distribution of  $\Delta W$ , the mean of which is determined by the extent to which the reference potential deviates from the full (or by the number of reference steps between full energy evaluations). Because a Dirac  $\delta(0)$  distribution of  $\Delta W$  would yield unit acceptance probability, reducing the absolute value of the first two moments of the  $\Delta W$  distribution raises the mean value of  $\alpha_j$  in (23). These moments are dictated partly by the thermodynamic state of the reference system, a fact upon which we build the optimization procedure described in the next section.

Unless otherwise indicated all full system results are for a periodic system of 100 diatomic molecules at temperature  $T=728$  K and pressure  $P=4.84$  GPa, although our methodology is in no way restricted to such extreme environments. After an equilibration period of  $O(10^4)$  reference steps<sup>30</sup>, results were collected from an additional  $O(10^7)$  reference steps and averaged over 5-10 Markov chains started from randomly chosen initial conditions.

The rate of convergence for ensemble averages depends on the statistical independence of the sampling points, in a sense now defined. The left panel of Figure 3 presents the distribution of reference energy per particle  $u^{(0)}$  as given by (1)-(4) and for  $l=1.00$  Å, calculated at a fixed number of steps  $O$  from a reference configuration  $j$ . At an offset  $O=10$  steps, the energies  $u^{(0)}(j)$  and  $u^{(0)}(j+O)$  are highly correlated and thus the distribution is narrowly peaked about  $u^{(0)}(j)$ . As the offset grows larger, the distribution widens gradually up to  $O=3000$ , at which point the distribution ceases to broaden. The right panel provides a quantitative measure of this effect through the standard deviation  $\sigma$

of the Gaussian distribution. The width of the distribution at  $O=4000$  is indistinguishable from that at  $O=3000$ . In fact, energies are correlated only slightly at  $O=500$ , but a conservative estimate of  $O_{corr}$  is made in order to clarify the benchmarks provided below. We define the correlation length  $O_{corr}$ , then, to be equal roughly to 3000 for this set of reference potentials.

Gelb's presentation of the N(MC)<sup>2</sup> method<sup>16</sup> suggested a metric for evaluating its computational efficiency (maximum speedup), but did not attempt to quantify its sampling efficiency. Because the reference and full potentials used here do not differ in computational expense, we reverse the emphasis and defer discussion of the total efficiency (some combination of sampling and computational efficiencies) for a later work<sup>23</sup>. The sampling efficiency of the method provides a measure of the rate at which it will explore the relevant space. This quantity is not determined by the acceptance probability alone, but in balancing the need to separate full energy evaluations by as many reference steps as possible (up to  $O_{corr}$ ) with maintenance of a reasonable acceptance probability for each composite move. In light of these considerations, we define the sampling efficiency  $E_s$  for a given reference potential (characterized here by  $O_{corr}$ ) and offset  $O$  as

$$E_s(O, O_{corr}) \equiv \frac{\bar{\alpha} \min(O, O_{corr})}{O_{corr}}. \quad (25)$$

$\bar{\alpha}$  is the average acceptance probability of a composite move from state  $i$  to state  $j$  when the states are separated by  $O$  reference steps, and the min function reflects the efficiency loss in increasing  $O$  beyond  $O_{corr}$ . The min function really should be replaced by one passing smoothly to  $O_{corr}$ , but (25) is sufficient for our purposes here. The possible range of  $E_s$  as defined by (25) is  $[0,1]$ , and the goal of the procedure introduced below will be to

maximize this quantity through variation of  $\bar{\alpha}$ . If  $O$  is large but  $\bar{\alpha}$  is small, then accepted composite steps will explore configuration space rapidly but much computational effort will be wasted on rejected steps; for small  $O$  and large  $\bar{\alpha}$ , composite steps will be accepted with high probability but little ground will be covered in the process.

We now examine the performance of  $N(\text{MC})^2$  for  $l$  values in the range 0.90-1.05 Å. Figure 4 illustrates the acceptance probability and Figure 5 the sampling efficiency as defined by (25) for  $O=1-3000$ . As the reference potential deviates more strongly from the full potential (*i.e.*, as  $l$  decreases), the performance of the method deteriorates rapidly, as evidenced by the downward shift in both the acceptance probability and efficiency curves. As is to be expected, the acceptance probability also falls as the magnitude of the offset rises. The inset in Figure 4 demonstrates the ability to obtain a good acceptance probability even with a poor reference potential, albeit at the cost of lowering  $O$  (meaning that a greater number of total sampling points will be required). In this context it is important to emphasize that using  $O$  values close to  $O_{\text{corr}}$  is ideal, but not at all necessary for sampling the full potential much more efficiently than with conventional  $(\text{MC})^2$ . On this point, note that the efficiency using any reference potential is minimal at  $O=1$ , which corresponds roughly to conventional  $(\text{MC})^2$ . Efficiency no longer increases monotonically with the offset as the reference potential deviates more strongly from the full potential; results for  $l=0.95$  Å and  $l=0.90$  Å exhibit maxima around  $O=250$  steps. These results will be scrutinized quantitatively below, after introducing an optimized variant of  $N(\text{MC})^2$ .

## V. OPTIMIZED $N(\text{MC})^2$ PROCEDURE



The average acceptance probability  $\bar{A}$  for composite steps connecting configurations  $i$  and  $j$  can be expressed exactly in the limit that  $i$  and  $j$  are fully decorrelated:

$$\bar{A} \equiv \lim_{O \rightarrow O_{corr}} \bar{\alpha}. \quad (26)$$

The initial states  $i$  will, by construction, possess relative weights  $e^{W_i}$  drawn from the full distribution. The final states  $j$  will, in the  $O_{corr}$  limit, be drawn randomly from the reference distribution and thus carry weights  $e^{W_j^{(0)}}$ . The acceptance probability of a composite trial step from state  $i$  to state  $j$  is  $\alpha_{ij}$ , and this quantity is averaged over the configuration and volume spaces of all decorrelated  $(i, j)$  pairs to obtain  $\bar{A}$ ,

$$\bar{A} = \frac{\iiint \alpha_{ij} e^{W_i + W_j^{(0)}} d\tau_i dV_i d\tau_j dV_j}{\iiint e^{W_i + W_j^{(0)}} d\tau_i dV_i d\tau_j dV_j} = \frac{\iiint \alpha_{ij} e^{\delta W_i} (e^{W_i^{(0)} + W_j^{(0)}}) d\tau_i dV_i d\tau_j dV_j}{\iiint e^{\delta W_i} (e^{W_i^{(0)} + W_j^{(0)}}) d\tau_i dV_i d\tau_j dV_j}. \quad (27)$$

Because composite steps are built from a sequence of elementary moves accepted with Boltzmann weight in the reference system, the terms appearing in parenthesis in (27) are implicitly taken into account when (MC)<sup>2</sup> sampling on the basis of the reference potential. Indicating a double average over initial and final states by nested brackets, (27) can be condensed as follows:

$$\bar{A} = \frac{\langle \langle \alpha_{ij} e^{\delta W_i} \rangle \rangle_0}{\langle \langle e^{\delta W_i} \rangle \rangle_0}, \quad (28)$$

where the subscripted '0' indicates that the averaging is performed entirely in the reference ensemble.  $\bar{A}$  can be built from (28) by sampling  $\delta W$  at a collection of decorrelated configurations (each separated by  $O \geq O_{corr}$  reference steps), meaning that

the sampled points will be drawn purely from the reference distribution and thus without application of (22). We refer to this reference distribution sampling as the ‘reweighting’ calculation<sup>31,32</sup>, and  $\bar{A}$  evaluated on this basis will be denoted  $\bar{A}_{rw}$ .  $\bar{A}_{rw}$  constitutes an *a priori* estimate of  $\bar{A}$ , in the sense that it provides an acceptance probability for N(MC)<sup>2</sup> composite steps (but only in the  $O_{corr}$  limit) without recourse to an actual N(MC)<sup>2</sup> simulation.

We now step through the procedure for performing an optimized N(MC)<sup>2</sup> simulation at a prescribed set of thermodynamic conditions ( $P=P', T=T'$ ). The reference (full) system weights  $W_k^{(0)}$  ( $W_k$ ) appearing in (27) depend on the reference (full) system temperature and pressure through (8), meaning that  $\bar{A} \equiv f(P^{(0)}, T^{(0)}, P, T)$ . Hereafter, the variable dependencies of  $\bar{A}$  will be listed in this order. From a single set of reference configurations collected in the reweighting calculation at  $(P^{(0)}, T^{(0)})$ , a family of  $\bar{A}_{rw}$  differing only in the values assigned to  $(P, T)$  can be constructed from (28). Because it is the thermodynamic state of the full system *only* that we wish to match with experiment,  $(P^{(0)}, T^{(0)})$  can be treated separately from  $(P, T)$  and the latter varied as free parameters in order to maximize  $\bar{A}_{rw}$  for a given set of configurations. Previously<sup>31</sup> we applied a similar idea to the thermodynamics of fluid N<sub>2</sub> as described by DFT, but strictly in the context of reweighting configurations already sampled using traditional (MC)<sup>2</sup>. The reference system parameters can be varied to yield maximal  $\bar{A}_{rw} = \bar{A}_{max}^{(0)}$  at optimal  $(P^{(0)} = P_{opt}^{(0)}, T^{(0)} = T_{opt}^{(0)})$ :

$$\begin{aligned} \bar{A}_{max}^{(0)}(P_{opt}^{(0)}, T_{opt}^{(0)}, P', T') = \\ \max \left\{ \bar{A}_{rw}(x, y, P, T) : P_1^{(0)} \leq x \leq P_2^{(0)}, T_1^{(0)} \leq y \leq T_2^{(0)} : P = P', T = T' \right\}, \end{aligned} \quad (29)$$

where the reference system  $(x,y)=(P^{(0)},T^{(0)})$  have been scanned over a given domain. This approach is permissible, but requires iteratively resampling  $\delta W$  (which includes evaluation of the full potential). Alternatively one can satisfy

$$\bar{A}_{\max}(P',T',P_{opt},T_{opt}) = \max \left\{ \bar{A}_{rw}(P^{(0)},T^{(0)},x,y) : P^{(0)} = P', T^{(0)} = T' : P_1 \leq x \leq P_2, T_1 \leq y \leq T_2 \right\}, \quad (30)$$

using the same set of reference configurations for each  $(x,y)$  pair. In general,  $\bar{A}_{\max}^{(0)}$  is a function of  $(x,y)=(P,T)$  and  $\bar{A}_{\max}$  is a function of  $(x,y)=(P^{(0)},T^{(0)})$ ; in (29) and (30), we have specified an actual *value* for these functions at designated values of  $(P^{(0)},T^{(0)},P,T)$ . Upon solution of (30) there are at least two different ways of returning the full system to the thermodynamic state of interest at  $(P=P',T=T')$ . The first is to collect reweighting samples at multiple  $(P^{(0)},T^{(0)})$ , solve (30) at each thermodynamic state to yield a set of corresponding  $(P_{opt},T_{opt})$  pairs, calculate ensemble averages at each new pair using the optimized N(MC)<sup>2</sup> procedure, then interpolate between those averages to obtain approximate values at the  $(P=P',T=T')$  combinations desired. We will take this approach in a future publication<sup>23</sup> where N(MC)<sup>2</sup> will be used to characterize the shock Hugoniot locus of N<sub>2</sub> over a wide range of thermodynamic conditions. Here we assume a simpler approach, more suitable for use of N(MC)<sup>2</sup> at an isolated thermodynamic state. After solving (30) for  $(P_{opt},T_{opt})$ , we linearly extrapolate back to the original, desired  $(P',T')$  and apply the same transformation to the reference variables, yielding approximate

$(P_{opt}^{(0)},T_{opt}^{(0)})$ :

$$P_{opt}^{(0)} \approx P^{(0)} + (P' - P_{opt}), \quad T_{opt}^{(0)} \approx T^{(0)} + (T' - T_{opt}). \quad (31)$$

Optimized N(MC)<sup>2</sup> is then performed at  $(P_{opt}^{(0)}, T_{opt}^{(0)}, P', T')$ . Concrete examples of this procedure are shown in Table 1. Beginning at  $T=T^{(0)}=T'=728$  K and  $P=P^{(0)}=P'=4.84$  GPa, unoptimized N(MC)<sup>2</sup> simulations were carried out with all four reference system bond lengths  $l$  and the resultant  $\bar{A}_{MC}(P', T', P', T')$  recorded.  $\bar{A}_{MC}$  values represent an *a posteriori* estimate of  $\bar{A}$ , calculated simply as the number of accepted composite steps divided by the total number of composite trial steps in a N(MC)<sup>2</sup> simulation. The reference distribution of  $\delta W$  was then sampled at  $O(10^4)$  points, from which  $\bar{A}_{rw}(P', T', P', T')$  was built using (28).  $P$  and  $T$  were varied with (30) to yield  $\bar{A}_{max}$  and  $(P_{opt}, T_{opt})$ , then (31) was used to generate  $(P_{opt}^{(0)}, T_{opt}^{(0)})$ . Finally, N(MC)<sup>2</sup> simulations using the two optimized sets  $(P', T', P_{opt}, T_{opt})$  and  $(P_{opt}^{(0)}, T_{opt}^{(0)}, P', T')$  were performed to yield the corresponding  $\bar{A}_{MC}$ . Note that optimized reference system variables were obtained by solution of (31), not (29); thus, no *a priori* estimate of  $\bar{A}_{MC}(P_{opt}^{(0)}, T_{opt}^{(0)}, P', T')$  is available. Numbers in parenthesis indicate statistical uncertainty in the final digit recorded. Discrepancies of greater than one  $\sigma$  between theoretical and computed values most likely reflect use of incompletely decorrelated samples.

We found the surface describing  $\bar{A}_{rw}$  as a function of  $P$  and  $T$  (or  $P^{(0)}$  and  $T^{(0)}$ ) to be generally smooth, and the scanned range of pressure and temperature values can be squeezed iteratively in combination with finer meshes until a maximum is located; this approach proceeds with little difficulty. The left panel in Figure 6 presents a contour plot of predicted acceptance probabilities scanned over a range 750-850 K and 4.75-5.75 GPa, for the reference potential  $l=1.00$  Å. Contour values were obtained in the process of

solving (30), so the temperature and pressure being varied are that of the full system. The arrow points in the direction of uniformly increasing contour values, and the vertical dotted line marks the temperature at which  $\bar{A}_{rw}$  is maximal. A trio of acceptance probability “isotherms” drawn from the contour plot is depicted in the right panel; these curves scan  $\bar{A}_{rw}$  over a range of pressures at fixed temperature. The overall maximum is clearly identifiable at  $T=790$  K,  $P=5.35$  GPa.

Solution of (29) or (30) requires sampling at enough points to provide a reliable estimate of  $\bar{A}_{rw}$  from (28). If such estimates require a large number of sampling points, then the sampling efficiency gained by optimizing  $N(\text{MC})^2$  will be lost in the overhead of performing the optimization itself. It is reasonable, then, to ask how many sampling points  $N_{rw}$  are required to predict stable values of  $(P_{\text{opt}}, T_{\text{opt}})$  or  $(P_{\text{opt}}^{(0)}, T_{\text{opt}}^{(0)})$ . The convergence of  $(P_{\text{opt}}, T_{\text{opt}})$  with respect to  $N_{rw}$  is illustrated in Figure 7, where it appears to be faster for reference potentials closer to the full potential; while  $(P_{\text{opt}}, T_{\text{opt}})$  for  $l=1.05$  Å are converged at  $N_{rw}=1000$ ,  $(P_{\text{opt}}, T_{\text{opt}})$  for  $l=0.90$  Å are clearly unconverged even for  $N_{rw}=5000$ . We hasten to note, however, that convergence of the acceptance probability is much more important than convergence of the thermodynamic parameters. If  $\bar{A}_{\text{opt}}$  exhibits a broad, flat peak when expressed as a function of  $P$  and  $T$ , then strict convergence of the latter two is not necessary to ensure a dramatically improved acceptance probability. Figure 8 confirms that this is indeed the case: the  $\bar{A}_{\text{opt}}$  for all four reference potentials stabilize at at roughly 1000 steps to the  $(P_{\text{opt}}, T_{\text{opt}})$  shown in Table 1.

Having obtained solutions to (30) and extrapolated back to  $(P_{\text{opt}}^{(0)}, T_{\text{opt}}^{(0)})$  with (31), we now examine the performance of optimized  $N(\text{MC})^2$  using the new set of

thermodynamic reference variables. Optimal acceptance probabilities and efficiencies are shown in Figures 9 and 10, and should be compared directly to those of Figures 4 and 5, respectively (note that the ordinate scales in Figures 5 and 10 differ). Improvements in acceptance probability as a percentage of the unoptimized values for  $O=100-3000$  are shown in Table 2. Improvement is significant for all potentials at all values of  $O$ , but the marginal gain increases as  $O$  grows larger and (in general) as the reference potential deviates more strongly from the full potential. The greatest performance improvements are for  $l=0.95$  Å, possibly indicating that already at  $l=0.90$  Å the physics embodied by the reference potential starts to deviate too strongly from that of the full potential for the optimization procedure to be fully effective.

The distribution  $\delta U \equiv U - U^{(0)}$  of potential energy differences sampled by trial states in  $N(\text{MC})^2$  interpolates between the distribution found by Metropolis sampling on the basis of the reference or full potential alone: for small  $O$ , the distribution of trial  $\delta U$  is roughly that encountered in sampling on the basis of  $U$  only; in the  $O_{corr}$  limit, the  $\delta U$  distribution corresponds exactly to that found in sampling strictly on the basis of  $U^{(0)}$  (in this case, the trial state loses its “memory” of the initial state). Because trial states are assigned Boltzmann weight in the full system, the acceptance probability of  $N(\text{MC})^2$  steps is reflected indirectly in the overlap of the  $\delta U$  distributions for trial and accepted states (both distributions are a function of volume). Figure 11 illustrates  $\delta u$  (where  $u \equiv (U - U^{(0)})/N$ ) vs.  $v$  ( $v \equiv V/N$ ) for  $l=1.05$  Å and  $O=1000$ . Unoptimized  $N(\text{MC})^2$  results are shown in the left panel, optimized results in the right. The overlap of the trial (black points) and accepted (red points) distributions increases significantly upon optimization, where we have indicated the values of  $(P_0, T_0, P, T)$  used in the underlying simulations. Figure 12

illustrates the same data, but for  $l=0.90$  Å and  $O=1000$ . Again, optimization increases the overlap considerably. The density of accepted points is much lower than in Figure 11, in keeping with the acceptance probabilities listed in Table 2. The system exhibits also a strong tendency to become “trapped” at certain volumes, indicated by the broken, vertical collections of trial points. As stated in IV, the acceptance probability for  $N(\text{MC})^2$  steps remains good even with a poor reference potential if one is willing to employ a smaller  $O$ . Figure 13 shows the same distributions as in Figure 12, but for  $O=250$  instead of  $O=1000$ . Not only are the overlaps between trial and accepted distributions greater before and after optimization, but the volume trapping noted in connection with Figure 12 is absent almost entirely. There is considerably more skew in the trial distribution for smaller offsets; this reflects correlation between initial and final points in the composite trial step. Although this combination of reference potential and  $O$  yields high overlap and thus a good acceptance probability, many more sampling points will be required to minimize the variance in ensemble averages.

From the sets of trial and accepted points shown in Figures 11-12, one can build the  $\Delta W$  distributions appearing in (23), thus establishing a direct link between trial distributions and acceptance probabilities. Figure 14 depicts the  $\Delta W$  distributions built from Figures 11 (left panel) and 12 (right panel). As expected, the distributions for the better reference potential ( $l=1.05$  Å) are narrower and their means lie closer to zero. Optimization substantially lowers the absolute value of  $\langle\langle\Delta W\rangle\rangle$  and  $\sigma(\Delta W)$  for both values of  $l$  (we have indicated the double averaging over initial and final points by nested brackets), and both of these factors contribute to a higher average acceptance probability for composite moves. This exercise provides an alternate means of evaluating the

performance of the optimization procedure, in that the first two moments of the  $\Delta W$  distribution are closely related to the acceptance probability. Table 3 provides an overview of these moments (before and after optimization) for all of the potentials surveyed.

## VI. SUMMARY

Nested Markov Chain Monte Carlo (N(MC)<sup>2</sup>) allows (MC)<sup>2</sup> sampling with a potential of given accuracy (the full potential) without needing to evaluate it following every elementary step. By stringing together a sequence of single-particle moves accepted on the basis of a more approximate potential (the reference system), trial steps in the full system are made to encompass multiple particles. The acceptance probability of this composite step can be maximized in at least two different ways. The first, described above, involves manipulation of the thermodynamic conditions characterizing the reference system such that the variance of the  $\delta U$  vs.  $V$  distribution<sup>33</sup> is minimal. A second means of optimization, not explored here, is direct, iterative modification of the reference potential to conform with ‘targets’ (such as average energies or volumes) computed with the full potential. Changes could be made adaptively to the reference potential functional form or its parameterization, or (in the case of an *ab initio* reference potential) the basis set or level of convergence. An iterative strategy similar to that used in optimal pulse-shaping<sup>34</sup> or empirical structure refinement<sup>35</sup> could be employed in combination with the thermodynamic optimization described above to yield an even more efficient route to accurate equilibrium sampling.

Clearly there are limits to the range of reference potentials suitable to any given full potential. Qualitative differences in the nature of intermolecular forces, such as the



complete absence of attractions, cannot be salvaged using the optimized N(MC)<sup>2</sup> procedure; hard spheres will never be a good reference for *ab initio* water. Once a pair of potentials is different enough that the overlap of their  $\delta U$  vs.  $V$  distributions is nearly zero, collecting statistics to evaluate (27) becomes quite difficult. These difficulties were apparent even in the  $l=0.90$  Å case above, in spite of the fact that our reference and full potentials have the same functional form. Bennett's methods<sup>36</sup> for estimating  $\Delta W$  surely are useful in this context, but it is unlikely that even optimal performance will be acceptable if such techniques are required simply to estimate  $\bar{A}$  (*i.e.*, thermodynamic optimization can only move the distribution so far).

## VII. ACKNOWLEDGEMENTS

JDC thanks the Office of the Director at Los Alamos National Laboratory (LANL) for support in the form of a Director's Postdoctoral Fellowship. MSS is supported by the LANL High Explosives Project of the National Nuclear Security Administration (NNSA) Advanced Strategic Computing Program (HE-ASC). TDS is supported by the LANL Laboratory Directed Research and Development (LDRD) Program and by the Army Research Office. LANL is operated by Los Alamos National Security L.L.C. under the auspices of the NNSA and the United States Department of Energy, under Contract No. DE-AC52-06NA25396.

- 1 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. (Oxford  
University Press, Oxford, UK, 1987).
- 2 D. Frenkel and B. Smit, *Understanding Molecular Simulation: From  
Algorithms to Applications*. (Academic Press, New York NY, 2002).
- 3 D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in  
Statistical Physics*. (Cambridge University Press, Cambridge, UK, 2000).
- 4 W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed.  
(John Wiley & Sons, New York, 1957).
- 5 B. J. Alder and T. E. Wainwright, *J. Chem. Phys.* 31, 459 (1959).
- 6 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E.  
Teller, *J. Chem. Phys.* 21, 1087 (1953).
- 7 S. Goedecker, *Rev. Mod. Phys.* 71 (4), 1085 (1999).
- 8 W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional  
Theory*, 2nd ed. (Wiley-VCH, 2001); R. G. Parr and W. Yang, *Density-  
Functional Theory of Atoms and Molecules*. (Oxford University Press, New  
York, NY, 1989).
- 9 Y. Yamaguchi, J. D. Goddard, Y. Osamura, and H. F. Schaefer, *A New  
Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio  
Molecular Electronic Structure Theory*. (Oxford University Press, 1994).
- 10 R. Car and M. Parrinello, *Phys. Rev. Lett.* 55 (22), 2471 (1985); M. E.  
Tuckerman, *J. Phys.: Condens. Matter* 14, R1297 (2002).
- 11 L. D. Crosby and T. L. Windus, *J. Phys. Chem. A* 113 (3), 607 (2009).
- 12 A. Nakayama, N. Seki, and T. Taketsugu, *J. Chem. Phys.* 130 (2), 024107  
(2009).
- 13 Of course the converse is also true, the most obvious manifestation of which  
is MC's limitation to equilibrium contexts.
- 14 M. S. Shaw, *J. Chem. Phys.* 94 (11), 7550 (1991); J. K. Johnson, A. Z.  
Panagiotopoulos, and K. E. Gubbins, *Mol. Phys.* 81, 717 (1994); J. K.  
Johnson, *Adv. Chem. Phys.* 105 (1998); C. H. Turner, J. K. Brennan, M.  
Lisal, W. R. Smith, J. K. Johnson, and K. E. Gubbins, *Mol. Sim.* 34, 119  
(2008).
- 15 R. Iftimie, D. Salahub, D. Wei, and J. Schofield, *J. Chem. Phys.* 113 (2), 4852  
(2000).
- 16 L. D. Gelb, *J. Chem. Phys.* 118, 7747 (2003).
- 17 These include two-surface Monte Carlo (TSMC) and the molecular  
mechanics based importance function (MMBIF) method. We eschew the first  
because it is too evocative of electronically excited states, and the second  
because it inaccurately characterizes the method's generality.
- 18 C. H. Mak, *J. Chem. Phys.* 122 (21), 214110 (2005).
- 19 K. Bernacki, B. Hetenyi, and B. J. Berne, *J. Chem. Phys.* 121 (1), 44 (2004);  
B. Hetenyi, K. Bernacki, and B. J. Berne, *J. Chem. Phys.* 117 (18), 8203  
(2002).

20 A. Brandt, V. Ilyin, N. Makedonska, and I. Suwan, *J. Mol. Liquids* 127 (1-3),  
37 (2006).

21 E. Lyman, F. M. Ytreberg, and D. M. Zuckerman, *Phys. Rev. Lett.* 96 (2),  
028105 (2006); E. Lyman and D. M. Zuckerman, *J. Chem. Theory Comput.* 2  
22 (3), 656 (2006).

P. Bandyopadhyay, *J. Chem. Phys.* 122 (9), 091102 (2005); P.  
Bandyopadhyay, *J. Chem. Phys.* 128 (13), 134103 (2008); P. Bandyopadhyay,  
*Theor. Chem. Acc.* 120 (1-3), 307 (2008); L. D. Gelb and T. N. Carnahan,  
*Chem. Phys. Lett.* 417 (4-6), 283 (2006); R. Iftimie, D. Salahub, and J.  
Schofield, *J. Chem. Phys.* 119 (21), 11285 (2003); R. Iftimie and J. Schofield,  
*J. Chem. Phys.* 114 (15), 6763 (2001); R. Iftimie and J. Schofield, *Int. J.*  
*Quant. Chem.* 91 (3), 404 (2002); I.-F. W. Kuo, C. J. Mundy, M. J. McGrath,  
J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein,  
F. Mohamed, M. Krack, and M. Parrinello, *J. Phys. Chem.* 108B (34), 12990  
(2004); D. R. Matussek, S. Osborne, and A. St-Amant, *J. Chem. Phys.* 128  
(15), 154110 (2008); M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, and C. J.  
Mundy, *Mol. Phys.* 104 (22-24), 3619 (2006); M. J. McGrath, J. I. Siepmann,  
I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M.  
Krack, *ChemPhysChem* 6 (9), 1894 (2004); M. J. McGrath, J. I. Siepmann,  
I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, M. Sprik, J. Hutter, F.  
Mohamed, M. Krack, and M. Parrinello, *Comp. Phys. Comm.* 169 (1-3), 289  
(2005); J. Michel, R. D. Taylor, and J. W. Essex, *J. Chem. Theo. Comp.* 2 (3),  
732 (2006); J. Michel, M. L. Verdonik, and J. W. Essex, *J. Med. Chem.* 49  
(25), 7427 (2006).

23 J. D. Coe, T. D. Sewell, and M. S. Shaw, in preparation.

24 Y. B. Zel'dovitch and Y. P. Raizer, *Physics of Shock Waves and High-*  
*Temperature Hydrodynamic Phenomena.* (Dover, Mineola, NY, 2002).

25 P. J. Hay, R. T. Pack, and R. L. Martin, *J. Chem. Phys.* 81 (3), 1360 (1984); J.  
D. Johnson, M. S. Shaw, and B. L. Holian, *J. Chem. Phys.* 80 (3), 1279 (1983).

26 W. K. Hastings, *Biometrika* 57 (1), 97 (1970).

27 W. W. Wood, *J. Chem. Phys.* 52 (2), 729 (1970).

28 The  $N \ln V_i$  term arises from the use of scaled coordinates.

29 V. I. Manousiouthakis and M. W. Deem, *J. Chem. Phys.* 110 (6), 2753 (1998).

30 Due to the absence of strongly attractive forces, as well as the high  
temperature and pressure of the system, the equilibration periods required  
were much shorter than for, say, water at STP.

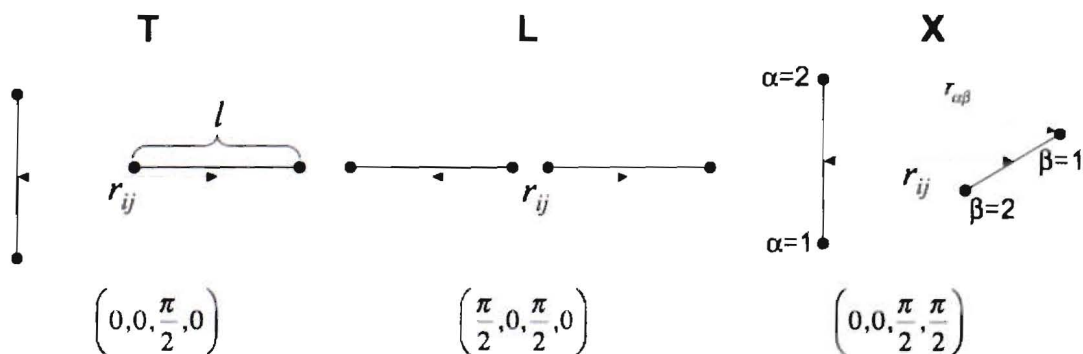
31 M. S. Shaw and C. Tymczak, in *Shock Compression of Condensed Matter*  
edited by M. D. Furnish (American Institute of Physics, 2005), Vol. 845, pp.  
179.

32 I. R. McDonald and K. Singer, *J. Chem. Phys.* 47 (11), 4766 (1967).

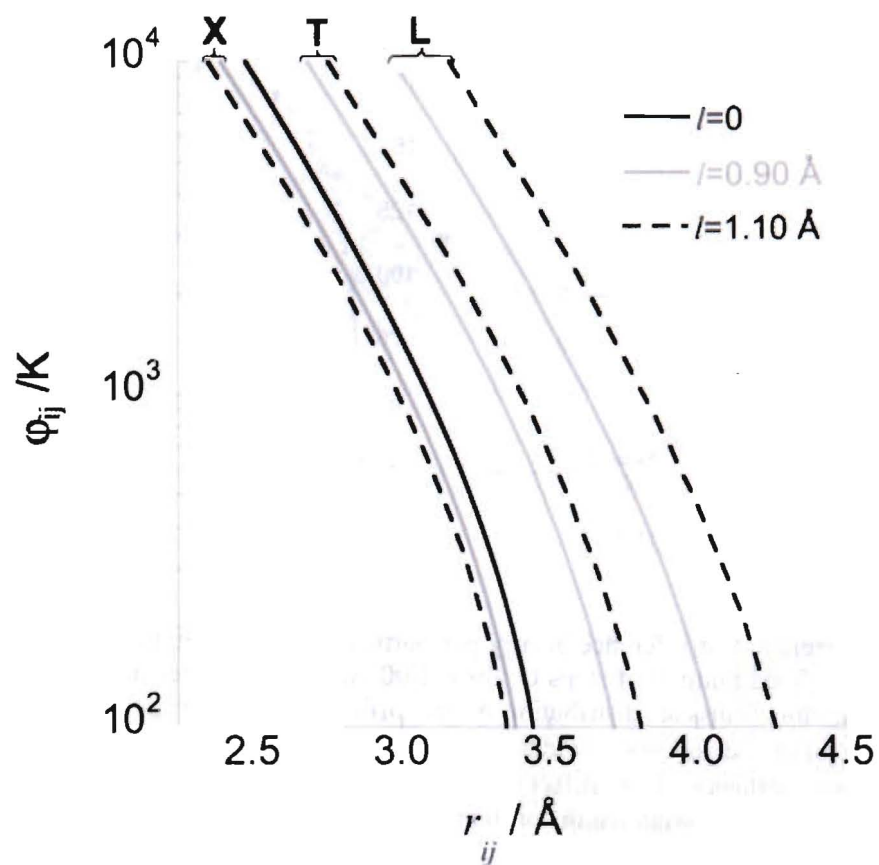
33 This is the case only for the isothermal-isobaric ensemble. In the canonical  
ensemble, for instance, it would be the  $\delta U$  vs.  $P$  distribution.

34 E. C. Carroll, J. L. White, A. C. Florean, P. H. Bucksbaum, and R. J.  
Sension, *J. Phys. Chem. A* 112 (30), 6811 (2008); H. Rabitz, *Science* 299  
(5606), 525 (2003).

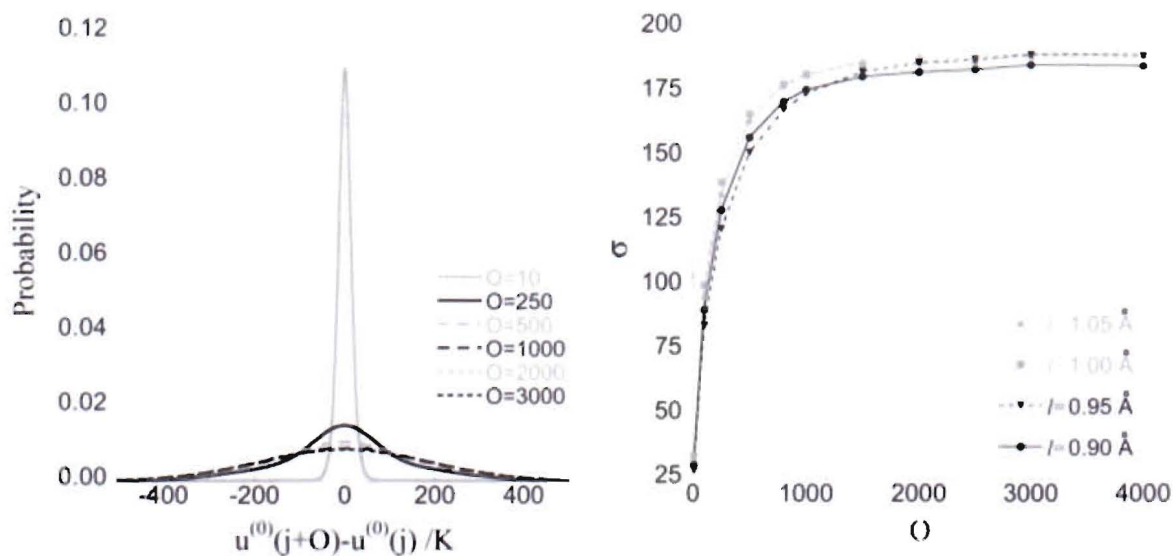
- <sup>35</sup> **A. K. Soper, Chem. Phys. 202, 295 (1996); A. K. Soper, Mol. Phys. 99 (17), 1503 (2001).**
- <sup>36</sup> **C. H. Bennett, J. Comp. Phys. 22 (2), 245 (1976).**



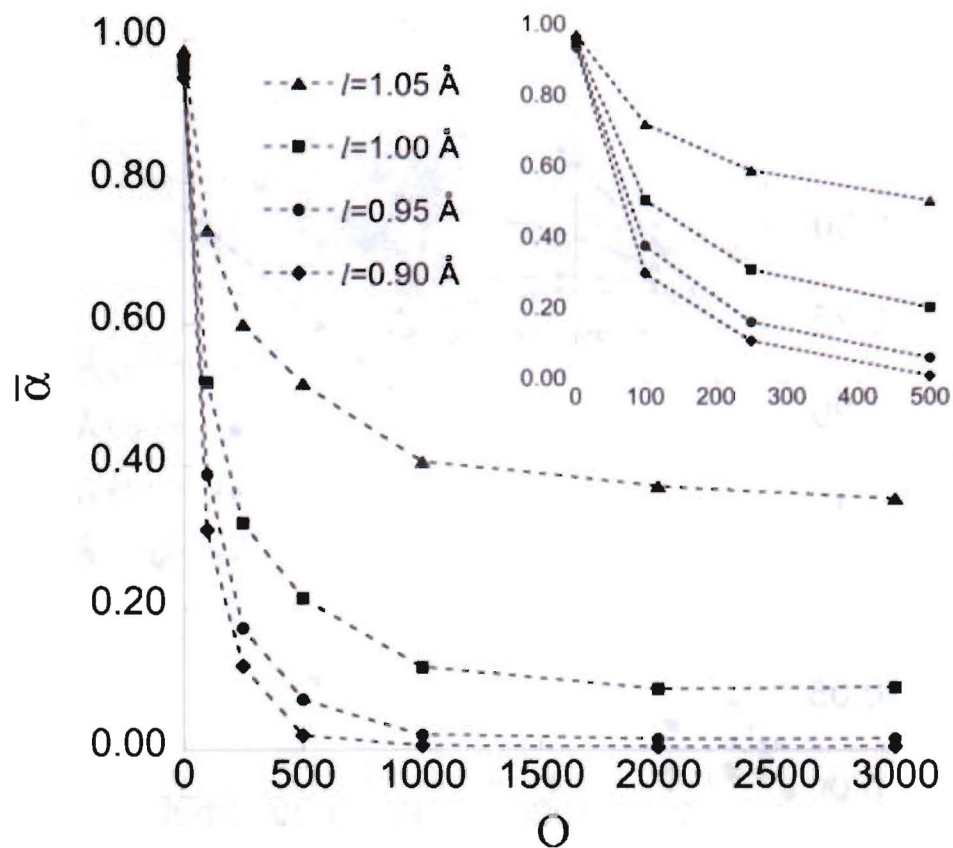
**Figure 1.** Diatomic pair configurations used to illustrate the anisotropy of the potential defined by (1)-(4) and its variation with intramolecular bondlength  $l$  in Figure 2. Bond lengths are fixed at 1.10 Å in the full system, but varied from 0.90-1.05 Å in the reference systems. Configuration types **T**, **L**, and **X** are characterized by four angles,  $(\theta_1, \chi_1, \theta_2, \chi_2)$ , where  $\theta$  and  $\chi$  represent rotation of molecules 1 and 2 in and out of the plane of the page. Diatomic bond vectors in the **T** and **L** configurations therefore are coplanar whereas those in the **X** configuration are perpendicular. Clockwise rotations are positive, and all angle values are zeroed to those of the parallel configuration (not shown).



**Figure 2.** Illustration of the pair potential defined by (1)-(4) as a function of the center-of-mass-separation  $r_{ij}$ . The full potential ( $l=1.10 \text{ \AA}$ ) is compared to a spherical potential ( $l=0$ , where all configurations are equivalent), as well as to the reference potential with the shortest bond length ( $l=0.90 \text{ \AA}$ ). Variations in potential  $\phi_{ij}$  are plotted on a log scale, revealing its highly anisotropic character. See Figure 1 for an explanation of the **X**, **T**, and **L** configurations.

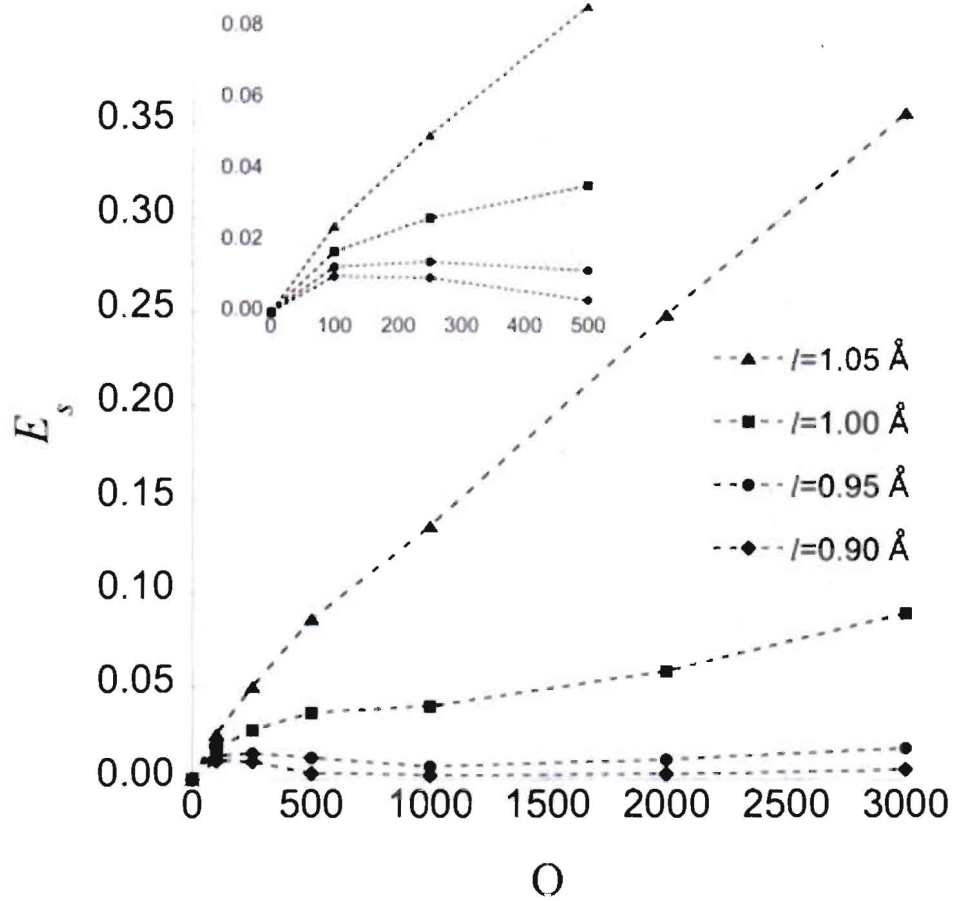


**Figure 3.** Correlation of reference energy per particle  $u^{(0)}$  at configurations  $j$  with those separated by a fixed number of steps  $O$ , for  $l=1.00 \text{ \AA}$ . The left panel shows the uniform broadening of the Gaussian distribution as the offset rises. The right panel illustrates convergence of the distribution width, as characterized by standard deviation  $\sigma$ , for several different values of  $l$ . At offset  $O_{corr}=3000$  steps, the energies are approximately decorrelated and the Gaussian width constant.

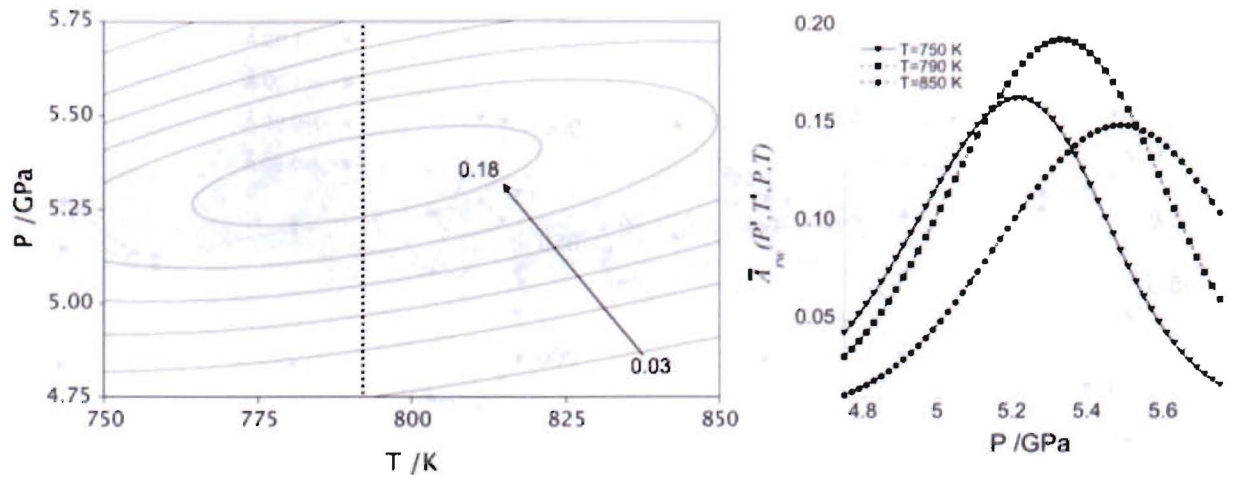


**Figure 4.** Acceptance probabilities obtained using the  $N(\text{MC})^2$  procedure with the potential defined by (1)-(4). Bondlength  $l$  is held fixed at  $1.10 \text{ \AA}$  in the full system, but varied from  $1.05$ - $0.90 \text{ \AA}$  in the reference systems. Acceptance probabilities diminish rapidly as the offset  $O$  between full energy evaluations increases or as the similarity of the reference and full potentials lessens. For all points, statistical errors in  $\bar{\alpha}$  are smaller than the symbol size. The inset enlarges the behaviour at small  $O$ , showing that better acceptance probabilities can be obtained even with the poorest reference potential by shortening the reference system Markov chain.

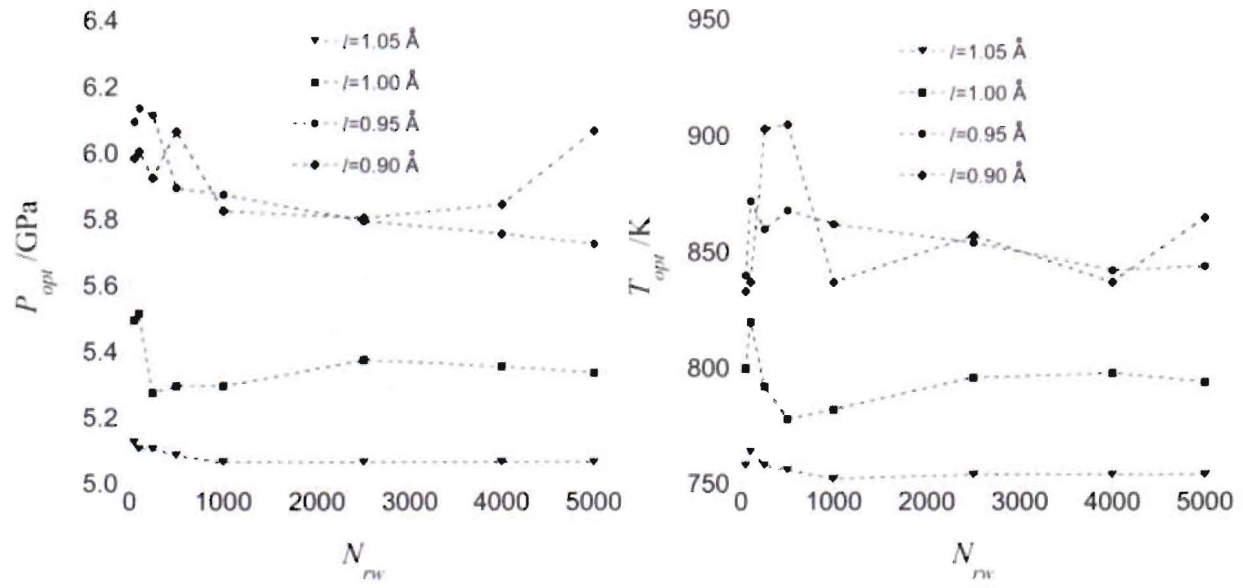




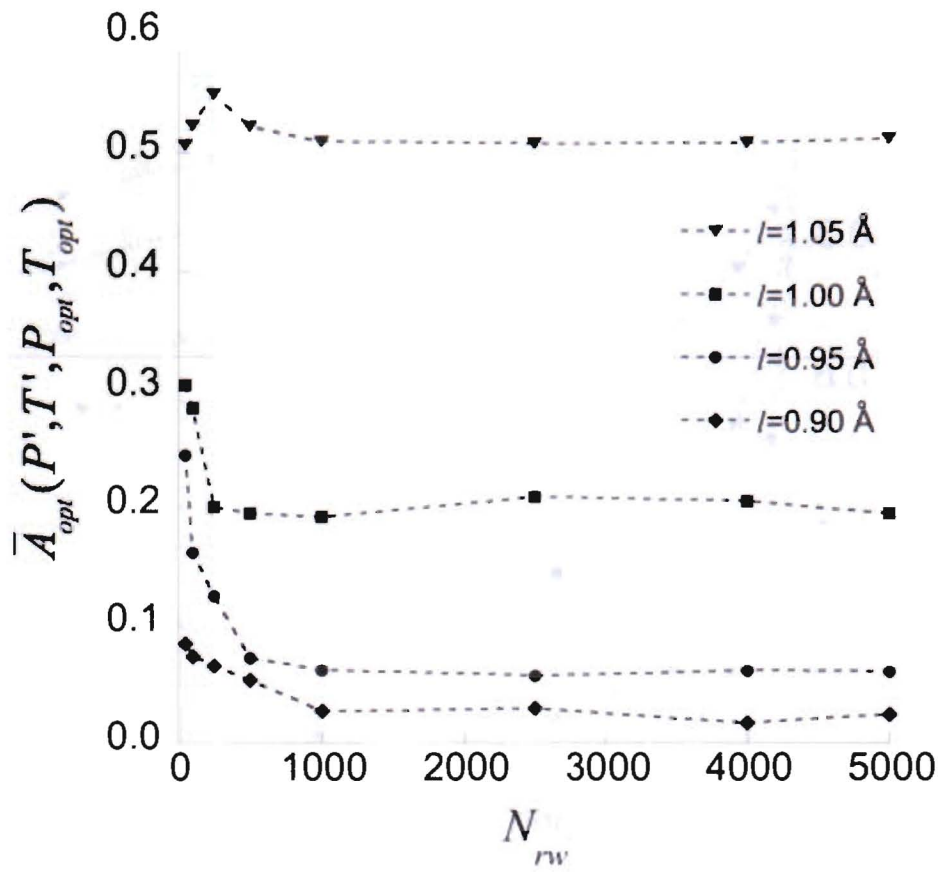
**Figure 5.** Sampling efficiencies  $E_s$  corresponding to the acceptance probabilities shown in Figure 4 and as defined by (25) in the text. Local maxima occur at small  $O$  as the quality of the reference potential deteriorates, but note efficiencies are minimal at  $O=1$ , corresponding roughly to that of standard Monte Carlo. The inset enlarges the behaviour at small  $O$ .



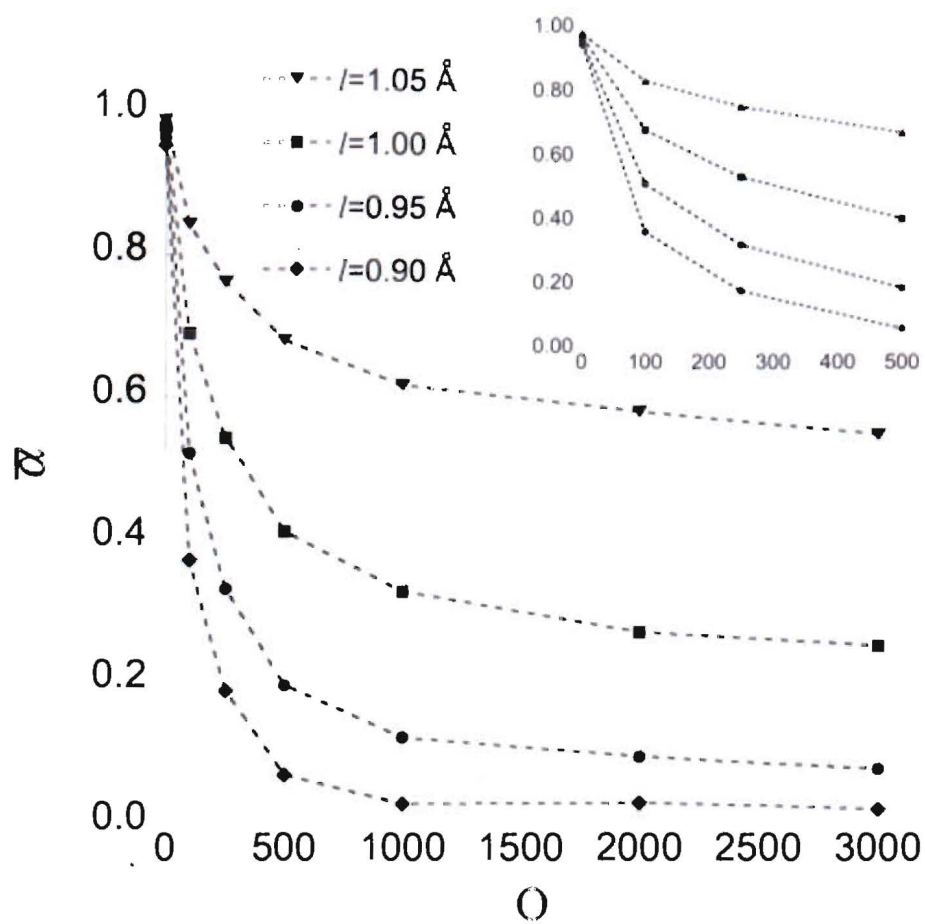
**Figure 6.** Maximization of the acceptance probability  $\bar{A}_{rw}(P', T', P, T)$  by variation of full system pressure and temperature in (28). The contours in the left panel show acceptance probabilities for pressures of 4.75-5.75 GPa and temperatures of 750-850 K. The arrow points in the direction of uniformly increasing contour values; the dotted line indicates the “isotherm” containing the maximum acceptance probability. Three such isotherms are plotted in the right panel, including the one containing the  $\bar{A}_{rw}$  maximum at  $T=790$  K.



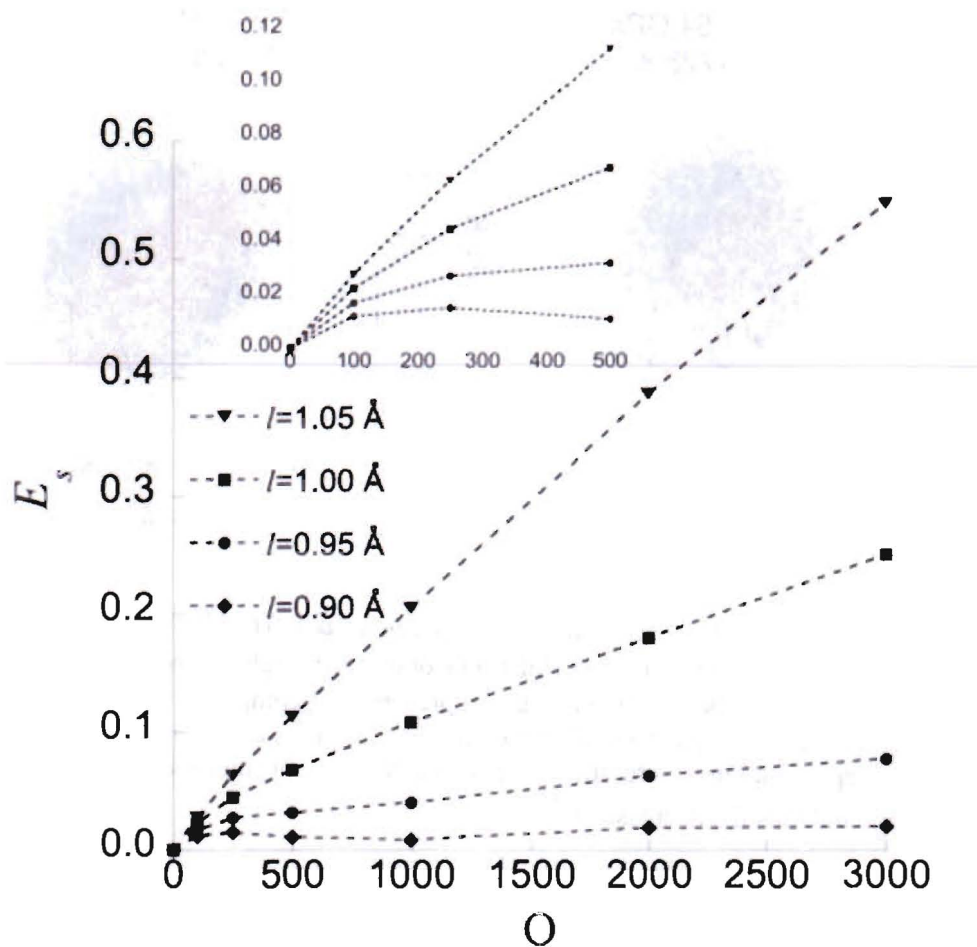
**Figure 7.** Convergence of optimized ( $P=P_{opt}, T=T_{opt}$ ) as a function of the number of reweighting points  $N_{rw}$ . For each value of  $N_{rw}$ , ( $P_{opt}, T_{opt}$ ) are found by solution of (30) using  $\bar{A}_{rw}$  built from (28). As the quality of the reference potential increases ( $l \rightarrow 1.10 \text{ \AA}$ ), so does the speed of convergence.



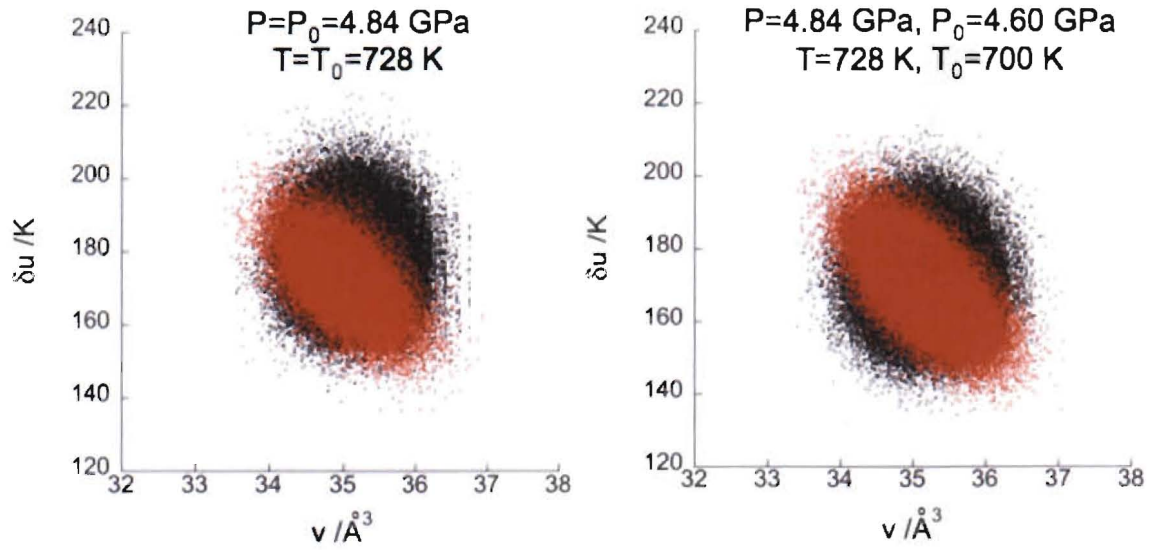
**Figure 8.** Convergence of the optimal acceptance probability  $\bar{A}_{opt}(P', T', P_{opt}, T_{opt})$  as a function of reweighted sampling points  $N_{rw}$ .  $\bar{A}_{rw}$  is built from (28) and optimized using (30). All four potentials roughly converge within 1000 steps, although convergence becomes slightly oscillatory as the potential deviates further from the full potential. For  $N_{rw} \geq 1000$ , the statistical errors are roughly the size of the symbols.



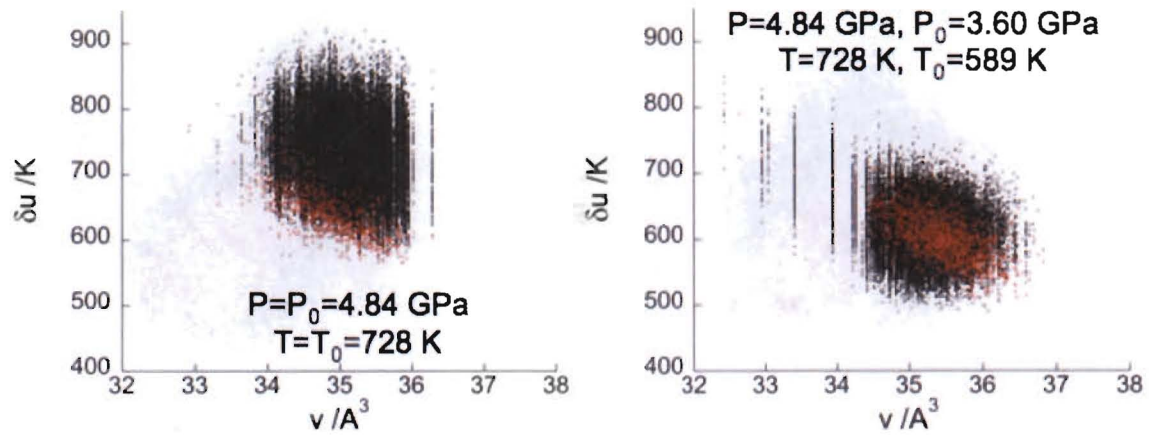
**Figure 9.** Average acceptance probabilities  $\bar{\alpha}$  for the optimized  $N(\text{MC})^2$  procedure as a function of the number of reference steps  $O$  taken between full energy evaluations. Results should be compared with the those of the unoptimized procedure, shown in Figure 4.



**Figure 10.** Sampling efficiency  $E_s$  for the optimized  $N(\text{MC})^2$  procedure as a function of the number of reference steps  $O$  take between full energy evaluations. Results should be compared those of the unoptimized procedure, shown in Figure 5.

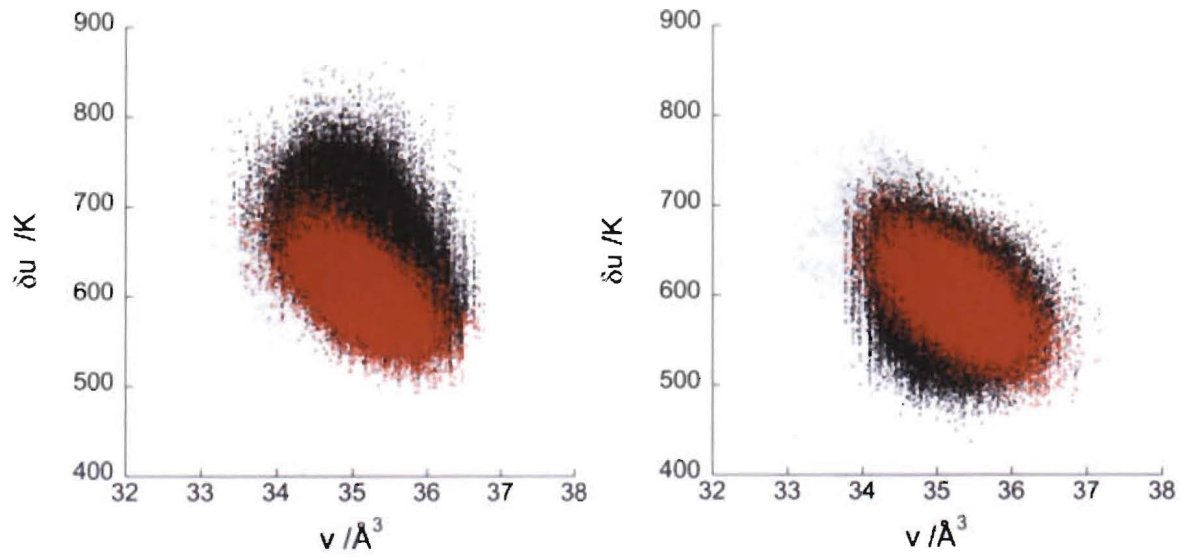


**Figure 11.** Distribution of energy difference per particle ( $\delta u \equiv ((U-U^0)/N)$ ) vs. volume per particle ( $v \equiv V/N$ ) for the unoptimized (left) and optimized (right)  $N(\text{MC})^2$  procedure at  $l=1.05 \text{ \AA}$  and offset  $O=1000$ . The black data represent trial composite steps, the red data are accepted composite steps. The  $\Delta W$  appearing in (23) are found from the difference in trial and accepted points. Note the increase in overlap of these distributions upon optimization, and thus the decrease in  $\langle\langle \Delta W \rangle\rangle$ .

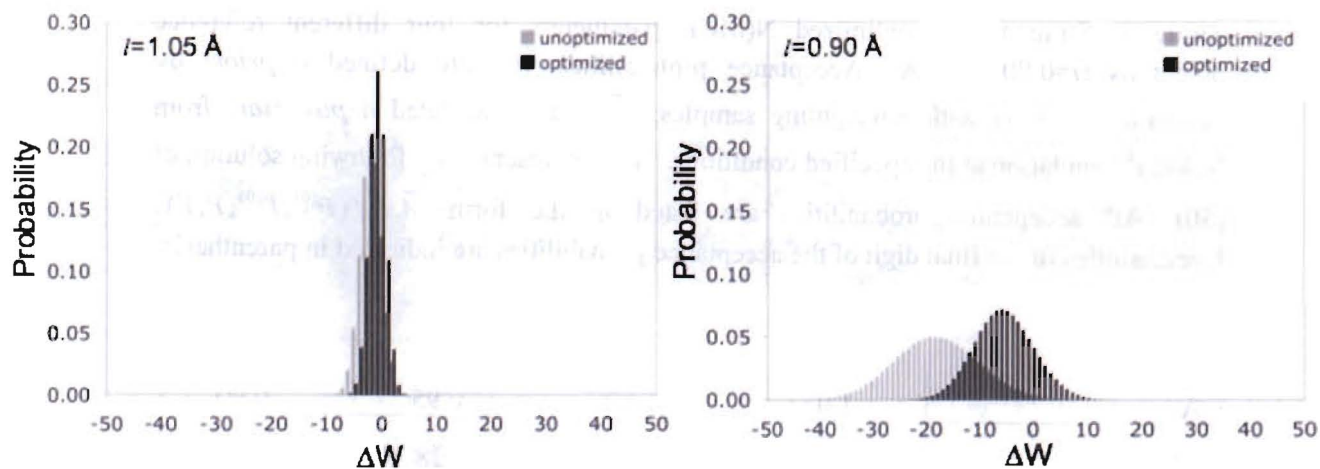


**Figure 12.** Distribution of energy difference per particle ( $\delta u \equiv (U - U^0)/N$ ) vs. volume per particle ( $v \equiv V/N$ ) for the unoptimized (left) and optimized (right)  $N(\text{MC})^2$  procedure at  $l=0.90 \text{ \AA}$  and offset  $O=1000$ . The black data represent trial composite steps, the red data are accepted composite steps. The  $\Delta W$  appearing in (23) are found from the difference in trial and accepted points. Note the change in overlap of these distributions upon optimization.





**Figure 13.** Distribution of energy difference per particle ( $\delta u$ ) vs. volume for the unoptimized (left) and optimized (right)  $N(\text{MC})^2$  procedure at  $l=0.90 \text{ \AA}$  and offset  $O=250$ . The black data represent trial composite steps, the red data are accepted composite steps. The  $\Delta W$  appearing in (23) are found from the difference in trial and accepted points. Even for a poor reference potential, significant acceptance probabilities can be achieved by lowering  $O$  (cf. Figure 12). Note that for lower values of  $O$ , the correlation of  $\delta u$  and  $v$  remains even in the trial distribution.



**Figure 14.** Direct visualization of the  $\Delta W$  appearing in (23), taken as the difference in  $\delta W$  for the trial and accepted distributions shown in Figure 11 (left panel) and Figure 12 (right panel). Upon optimization, the mean value of  $\Delta W$  shifts closer to zero and the width of the distribution shrinks in both cases. Both of these factors contribute to a higher acceptance probability for composite moves.

**Table 1.** Summary of optimized N(MC)<sup>2</sup> parameters for four different reference potentials ( $l=0.90-1.05$  Å). Acceptance probabilities  $\bar{A}_{rw}$  are defined *a priori* by evaluation of (28) with reweighting samples;  $\bar{A}_{MC}$  are calculated *a posteriori* from N(MC)<sup>2</sup> simulation at the specified conditions.  $\bar{A}_{max}$  represents  $\bar{A}_{rw}$  following solution of (30). All acceptance probabilities are listed in the form  $\bar{A} \equiv f(P^{(0)}, T^{(0)}, P, T)$ . Uncertainties in the final digit of the acceptance probabilities are indicated in parenthesis.

$l/\text{Å}$	1.05	1.00	0.95	0.90
$T'/\text{K}$	728	728	728	728
$P'/\text{GPa}$	4.84	4.84	4.84	4.84
$T_{opt}/\text{K}$	756	796	846	867
$P_{opt}/\text{GPa}$	5.08	5.35	5.74	6.08
$T_{opt}^{(0)}/\text{K}$	700	660	610	589
$P_{opt}^{(0)}/\text{GPa}$	4.60	4.33	3.94	3.60
$\bar{A}_{rw}(P', T', P', T')$	0.367(5)	0.069(3)	0.010(1)	0.001(0)
$\bar{A}_{MC}(P', T', P', T')$	0.355(6)	0.084(3)	0.018(3)	0.006(1)
$\bar{A}_{max}(P', T', P_{opt}, T_{opt})$	0.512(5)	0.195(5)	0.061(3)	0.024(2)
$\bar{A}_{MC}(P', T', P_{opt}, T_{opt})$	0.553(7)	0.279(9)	0.077(9)	0.030(5)
$\bar{A}_{MC}(P_{opt}^{(0)}, T_{opt}^{(0)}, P', T')$	0.548(6)	0.252(7)	0.078(6)	0.021(5)

**Table 2.** Improvements in efficiency for optimized relative to unoptimized  $N(\text{MC})^2$ , as a percentage of efficiency for the latter and as a function of the reference system Markov chain length  $O$ .

$O$	$l=1.05 \text{ \AA}$	$l=1.00 \text{ \AA}$	$l=0.95 \text{ \AA}$	$l=0.90 \text{ \AA}$
100	17	35	31	20
250	26	67	100	60
500	31	89	167	175
1000	51	179	471	350
2000	56	210	473	533
3000	54	182	359	250

**Table 3.** First two moments in the distribution of  $\Delta W$  appearing in (23), as a function of the reference system Markov chain length  $O$  and for all of the reference potentials surveyed. The double brackets indicate averaging over initial and final points ( $i$  and  $j$  in (27)), and values are listed before and after ( $opt$ ) optimization.

	$\langle\langle\Delta W\rangle\rangle$	$\langle\langle\Delta W\rangle\rangle_{opt}$	$\sigma(\Delta W)$	$\sigma_{opt}(\Delta W)$
$l=1.05 \text{ \AA}$				
O=100	-0.27	-0.08	1.88	1.32
O=250	-0.59	-0.19	1.95	1.34
O=500	-0.86	-0.35	2.08	1.37
O=1000	-1.38	-0.51	2.34	1.43
O=2000	-1.59	-0.63	2.46	1.47
O=3000	-1.68	-0.71	2.53	1.77
$l=1.00 \text{ \AA}$				
O=100	-1.08	-0.31	3.84	2.72
O=250	-2.27	-0.73	4.46	2.80
O=500	-3.27	-1.33	4.98	3.01
O=1000	-5.10	-1.89	6.32	3.23
O=2000	-5.89	-2.40	6.96	3.61
O=3000	-5.99	-2.60	6.99	3.67
$l=0.95 \text{ \AA}$				
O=100	-1.92	-0.67	5.70	4.47
O=250	-4.46	-1.55	7.34	4.83
O=500	-7.09	-2.85	9.18	5.41
O=1000	-10.72	-4.02	12.02	6.26
O=2000	-11.58	-4.98	12.68	6.88
O=3000	-11.36	-5.11	12.42	6.89
$l=0.90 \text{ \AA}$				
O=100	-2.70	-0.82	7.69	6.13
O=250	-6.17	-1.80	9.40	7.19
O=500	-12.46	-3.95	14.80	9.44
O=1000	-17.25	-5.11	18.96	11.02
O=2000	-16.66	-5.53	17.83	10.82
O=3000	-18.05	-7.39	19.33	11.68