Extended state-space Monte Carlo methods

Sheldon B. Opps and Jeremy Schofield

Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 3H6

In this paper various extensions of the parallel-tempering algorithm are developed and their properties are analyzed. The algorithms are designed to alleviate quasiergodic sampling in systems which have rough energy landscapes by coupling individual Monte Carlo chains to form a composite chain. As with parallel tempering, the procedures are based upon extending the state space to include parameters to encourage sampling mobility. One of the drawbacks of the parallel-tempering method is the stochastic nature of the Monte Carlo dynamics in the auxiliary variables which extend the state space. In this work, the possibility of improving the sampling rate by designing deterministic methods of moving through the parameter space is investigated. The methods developed in this article, which are based upon a statistical quenching and heating procedure similar in spirit to simulated annealing, are tested on a simple two-dimensional spin system (*xy* model) and on a model *in vacuo* polypeptide system. In the coupled Monte Carlo chain algorithms, we find that the net mobility of the composite chain is determined by the competition between the characteristic time of coupling between adjacent chains and the degree of overlap of their distributions. Extensive studies of all methods are carried out to obtain optimal sampling conditions. In particular, the most efficient parallel-tempering procedure is to attempt to swap configurations after very few Monte Carlo updates of the composite chains. Furthermore, it is demonstrated that, contrary to expectations, the deterministic procedure does *not* improve the sampling rate over that of parallel tempering.

PACS number(s): 02.70.Rr, 05.10.Ln

I. INTRODUCTION

Over the years, Markov chain Monte Carlo (MC) methods [1] have evolved into an important and commonly used tool for evaluating expectations of observables with respect to some distribution. In particular, MC methods have been frequently utilized in the field of statistical physics to evaluate equilibrium and nonequilibrium ensemble averages in systems where other methods, such as molecular dynamics, yield relatively poor estimates [2]. MC methods are based on the construction of a Markov chain of configurations of the system, in which the probability of each configuration in the chain is determined by a targeted distribution. If all points in the state space are accessible in the Markov chain, then averages over the chain sequence converge to expectations with respect to the target distribution as the length of the Markov chain goes to infinity. However, there is no guarantee that the estimates of the expectations converge quickly and one sometimes is faced with physical systems in which accurate estimates require Monte Carlo chains of intractable length. Such a situation typically occurs when movement of the Markov chain through state space is inhibited by regions of low probability, leading to configurational trapping in isolated modes of the system [3]. Such ''quasiergodic'' sampling is often observed in simulation studies of systems displaying first-order phase transitions [4] or in systems exhibiting frustration and rough energy landscapes (spin glasses [5], conformational studies of biological molecules [6]).

Over the last few years a number of importance sampling techniques have been developed to improve the rate of convergence of Monte Carlo calculations for systems exhibiting quasiergodic behavior. Most approaches have been based on widening the sampling distribution using umbrella-sampling techniques [7] or other generalized ensemble methods, such as multicanonical methods [8]. Recently, a sampling scheme known as parallel tempering [9] has been proposed which is designed to increase the mobility of the Markov chain by sampling on an extended state space. These algorithms have one major drawback in common: The exploration of auxiliary parameter space can be mapped onto a stochastic process, resulting in a nondeterministic Monte Carlo dynamics over the width of the generalized distribution. For systems with very rough energy landscapes and deep metastable minima, the sampling distributions must have sufficient width to allow for migration out of these minima. Since the exploration is stochastic in nature, the migration from one tail region of the distribution to another tail region can take a prohibitively long time.

In this article, we examine the feasability of improving the sampling rate of parallel tempering Monte Carlo simulations by incorporating a method of guiding the extended system through the auxiliary parameter space in a deterministic fashion to promote the swapping of configurations and thereby the exploration of configurational space. The method, termed ''annealed swapping,'' is based upon the combination of annealed-importance sampling [10] and multiple Markov chain methods [11]. The annealed-swapping algorithm is tested on two model systems which are characterized by rough energy landscapes. It is shown that the method does not lead to greater sampling mobility than that exhibited by the simple parallel-tempering algorithm in spite of the increased directionality of the dynamics in the extended state-space. In the following section, various extended state space Monte Carlo algorithms are presented and discussed in detail. In Sec. III, the methods are tested and compared on a simple two-dimensional spin system (*xy* model) and a model system composed of a single, ten-

residue polyglycine chain *in vacuo*. The results of the numerical study of the algorithms are summarized in Sec. IV.

II. ALGORITHMS

A. Quenched importance sampling

The simulated quenching procedure [10] is generally used to investigate structure on glassy energy hypersurfaces which are characterized by many deep metastable attractive basins. In biological structure applications, the algorithm allows one to identify important low-energy structures and is sometimes capable of finding the minimum-energy structure. However, one of the drawbacks of the simulated quenching procedure [10] is that statistical information about the relative statistical importance of each structure is not accessible. A Monte Carlo calculation, on the other hand, provides a form of importance sampling which, in principle, enables ensemble averages to be calculated. However, typical realizations of simple Monte Carlo schemes, such as the Metropolis Monte Carlo scheme, spend long periods of time sampling relatively small regions of configuration space and only rarely move to other regions. Because of the poor mobility of Markov chains, the calculations converge very slowly to the correct ensemble averages.

Recently, a Monte Carlo algorithm [12] incorporating Markov chain transitions has been developed which uses a procedure similar to simulated quenching. Unlike the original implementation of simulated quenching, the ''quenched importance-sampling'' algorithm calculates statistical weights for the quenching process which permits an importance sampler to be defined. The algorithm therefore offers an approach which retains the benefits of a quenching process while permitting thermodynamic functions and statistical averages to be calculated.

The basic principles of the quenched importance-sampling algorithm are a combination of previous methods: namely, a quenching schedule in which one moves from a tractable distribution (such as a high-temperature Boltzmann distribution) to a target distribution via a sequence of intermediate distributions and the use of an extended state space [9,11]. Consider the expectation of an observable $a(\mathbf{x})$ with respect to a distribution $P_0(\mathbf{x})$:

$$\bar{a} = \int d\mathbf{x} a(\mathbf{x}) P_0(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is a possibly multidimensional point in state space. Expectations such as Eq. (1) are often evaluated using importance-sampling methods in which one samples a finite number N of points $\mathbf{x}^i$ from some distribution P_s and approximates the expectation as

$$\bar{a} \approx \frac{\sum_{i=1}^{N} w^{(i)} a(\mathbf{x}^i)}{\sum_{i=1}^{N} w^{(i)}}, \tag{2}$$

where $w^{(i)} = P_0(\mathbf{x}^i)/P_s(\mathbf{x}^i)$ is the statistical weight of sample point $\mathbf{x}^i$. Quenched importance sampling uses a Markov chain of configurations (path) generated along a quenching schedule to produce the independent sampling points $\mathbf{x}^i$ with weights which depend on the path. The quenching component of the sampling procedure consists of the use of a series of intermediate distributions P_j with $j = 1, \ldots, n$ to construct the sampling point. The Markov chain transitions for each distribution are represented by functions $T_j(\mathbf{x}_1 \rightarrow \mathbf{x}_2)$ giving the probability (density) of moving to state $\mathbf{x}_2$ from the state $\mathbf{x}_1$. We will assume that the T_j Markov transition probabilities obey a detailed-balance condition and have a corresponding limiting distribution P_j so that each T_j generates a sequence of states according to this distribution. The quenched importance sampling is carried out as follows.

(i) Generate a state point $\mathbf{x}_n$ according to the distribution P_n.

(ii) Carry out a Markov chain Monte Carlo simulation according to the distribution P_{n-1} starting from $\mathbf{x}_n$ and ending in state $\mathbf{x}_{n-1}$ utilizing some updating scheme (such as Metropolis).

(iii) Repeat the previous step for all intermediate distributions $j = n-2, \ldots, 1$.

(iv) Set the initial sampling point $\mathbf{x}^i = \mathbf{x}_1$ and calculate the weight factor for this point, given by

$$w^{(i)} = \frac{P_{n-1}(\mathbf{x}_n)}{P_n(\mathbf{x}_n)} \frac{P_{n-2}(\mathbf{x}_{n-1})}{P_{n-1}(\mathbf{x}_{n-1})} \cdots \frac{P_1(\mathbf{x}_2)}{P_2(\mathbf{x}_2)} \frac{P_0(\mathbf{x}_1)}{P_1(\mathbf{x}_1)}. \tag{3}$$

(v) Repeat steps (i)–(iv) N times to generate the set of independent sampling points used to calculate the average in Eq. (2).

The validity of this method is demonstrated in Appendix A. To obtain better estimates for the observables, the N sampled points $\mathbf{x}^i$ can also be used as initial states for a Markov chain Monte Carlo simulation which has the target limiting distribution P_0. The expectation value $\bar{a}$ can be calculated as weighted averages, using the $w^{(i)}$, of the simple average of $a(\mathbf{x})$ along the Markov chains which start from the sampled initial points.

B. Parallel tempering

Another method for increasing the rate of exploration of configuration space is based on sampling along a set of Markov chains run in parallel. This approach, and minor variants, has been called ''multiple Markov chains'' [13,14] or ''parallel tempering'' [15,16] in the literature. The basic idea of the parallel-tempering algorithm is that the mobility of isolated Markov chains with different limit distributions can be enhanced by coupling them together to form a composite Markov chain whose limit distribution is the product distribution of the separate chains. The method works particularly well for chains in which the convergence is slow, since the other chains in the composite chain which have greater mobility act as a randomizing heat bath which destroys correlation.

This algorithm, which we shall refer to as ''simple swapping,'' is implemented as follows: N chains, with limit dis-

tributions $P_1, \ldots, P_N$ spanning a range of parameters (such as temperature), are evolved using a standard Monte Carlo updating scheme simultaneously and independently for some number of updates N_{up}. The number of updates can be fixed or chosen from a distribution. After the N_{up} updates, two of the chains i and j with corresponding limit distribution P_i and P_j are selected at random and one tests to see if the current configurations on the chains can be swapped. Suppose that when the swap is attempted the states on the ith and jth chain are $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively. If one accepts the attempted swaps with probability $P_A^S(i \rightarrow j)$ given by

$$P_A^S(i \rightarrow j) = \min\left(1, \frac{P_i(\mathbf{x}_j) P_j(\mathbf{x}_i)}{P_i(\mathbf{x}_i) P_j(\mathbf{x}_j)}\right), \quad (4)$$

the composite Markov chain for the entire process has the limit distribution $P_1 \cdots P_N$ and each of the elementary or principal chains i has limit distribution P_i, as if it were entirely isolated. However, the swapping induces coupling between these principal chains and they are not independently Markov. In practice, it is often best to select only adjacent chains i and $i \pm 1$ to test for swapping, as the probability of acceptance vanishes as the overlap of the limiting of the chains i and j decreases.

One might think that a large number of principal chains is optimal because the swap acceptance ratios will be large. This, however, is not necessarily true since in the limit of an infinite number of principal chains, where the swap acceptance ratios are essentially 1, a given configuration evolves in parameter space in a random walk. For systems such as protein-folding studies or the study of collapse transitions in polymer models, which require a large range of temperatures due to the depth of the attractive basins in the potential energy surface, the diffusive nature of the algorithm may lead to slow exploration of the entire parameter space. It is therefore worthwhile to consider variants of the simple-swapping algorithm in which the exploration of the parameter space is not diffusive.

C. Annealed swapping

One means of circumventing diffusive exploration is to develop a Monte Carlo updating scheme in the generalized state space (state space plus the parameter space) which enables configurations on adjacent chains to be swapped with non-negligible probability. This may be achieved by constructing an updating scheme utilizing an annealing schedule which includes both a quenching procedure, such as the one described in Sec. II A, and an analogous heating method. Such an annealing process, which we shall refer to as "annealed swapping," can be incorporated into an updating scheme as follows: Suppose at a given iteration in the Markov chains the state of the system on chain 1 with limiting distribution Π_0 is $\mathbf{x}_0$ and the state of the system on chain N with limiting distribution Π_N is $\mathbf{x}_N$. First one decides upon a set of n intermediate distributions $P_1, \ldots, P_n$, where n is some positive integer greater than or equal to 2, which interpolate between Π_0 and Π_N, with $P_1 \equiv \Pi_0$ and $P_n \equiv \Pi_N$. These distributions scan a range of the parameter space, such
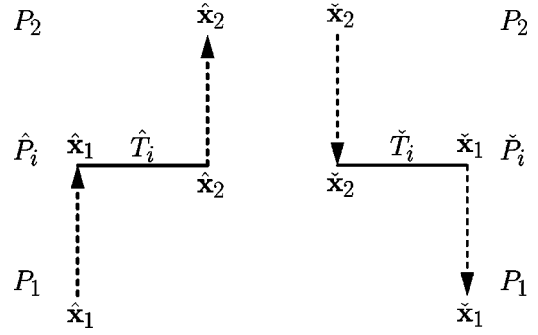

FIG. 1. A schematic of the annealing path $\{\hat{x}_1, \hat{x}_2, \check{x}_2, \check{x}_1\}$. The solid lines indicate Monte Carlo updates according to $\hat{T}_i$ and $\check{T}_i$.

as temperatures. The updating scheme consists of quenching (cooling down) from the configuration $\check{\mathbf{x}}_n = \mathbf{x}_N$ along all the intermediate distributions with a standard updating scheme for a specified number of steps, ending with distribution $P_1 = \Pi_0$. Simultaneously, the configuration $\mathbf{x}_0 = \hat{\mathbf{x}}_1$ is heated *up* along the set of intermediate distributions, ending with distribution $P_n = \Pi_N$. These procedures generate a set of intermediate configurations. For example, the quenching process may generate the set $\{\check{\mathbf{x}}_n, \check{\mathbf{x}}_{n-1}, \ldots, \check{\mathbf{x}}_1\}$ with statistical weight $\check{w}$, where $\check{w}$ is defined as in Eq. (3),

$$\check{w}^{(i)} = \frac{P_{n-1}(\check{\mathbf{x}}_n)}{P_n(\check{\mathbf{x}}_n)} \cdots \frac{P_1(\check{\mathbf{x}}_1)}{P_2(\check{\mathbf{x}}_1)}. \quad (5)$$

The heating procedure may generate the set $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_n\}$ with statistical weight $\hat{w}$ defined analogously:

$$\hat{w}^{(i)} = \frac{P_2(\hat{\mathbf{x}}_1)}{P_1(\hat{\mathbf{x}}_1)} \cdots \frac{P_n(\hat{\mathbf{x}}_n)}{P_{n-1}(\hat{\mathbf{x}}_n)}. \quad (6)$$

In the heating procedure, the state $\hat{\mathbf{x}}_2$ is generated by carrying out a specified number of simple MC updates, starting from the state $\hat{\mathbf{x}}_1$, according to the limiting distribution P_2. At the end of these processes, one accepts the trial state $\check{\mathbf{x}}_1$ as the new state for chain 1 and the state $\hat{\mathbf{x}}_n$ as the new state for chain N with probability

$$P_A^A = \min(1, \hat{w} \times \check{w}). \quad (7)$$

The annealed-swapping method is schematically depicted in Fig. 1 for two principal chains. It is demonstrated in Appendix B that this updating scheme obeys detailed balance for the composite Markov process and that each principal chain has the correct limiting distribution. Note that for the special case of no intermediate distributions ($n = 2$), the procedure described above is precisely the simple-swapping algorithm. It is also interesting to note that the process described above may be modified slightly if one is interested in only one target distribution. In this case, the algorithm describes an MC updating scheme for the target chain constructed by linking heating and quenching processes (connecting the

states $\check{\mathbf{x}}_2$ with $\hat{\mathbf{x}}_2$ in Fig. 1) into a continuous *annealing cycle*, known as tempered transitions [17].

The annealed-swapping process over intermediate distributions is incorporated to ensure that $P_A^A(i,j) \gg P_A^S(i,j)$ for adjacent principal chains i and j. The increased mobility in parameter space comes with a computational cost due to the intermediate updating steps along the heating and quenching paths. Increasing the number of intermediate distributions (or intermediate updating steps) in the annealed-swapping algorithm typically increases the parameter space mobility while increasing the computational time per updating step. These factors must be balanced against one another on a case-by-case basis in order to optimize the performance of the algorithm. It is important to recognize that there is a qualitative difference between the annealed-swapping algorithm and the simple-swapping procedure. The incorporation of heating and quenching processes into the updating scheme provides a *directed* way of increasing the rate of parameter space (and thereby configurational space) exploration, unlike the simple-swapping scheme in which the exploration of parameter space is a random-walk process at best.

There are a number of adjustable parameters which may effect the efficiency of the algorithm on a particular model. One may vary the number of principal chains comprising the composite system as well as how the chains are distributed in parameter space. It is also possible to vary the form of the distributions used for each principal Markov chain to incorporate umbrella-sampling [7] distributions in the annealed-swapping algorithm. In addition, the number of intermediate heating and quenching chains, the number of equilibration steps at each intermediate distribution, and the selection process for choosing the pair of initial principal chains to heat and quench may be optimized for each model. In the next section, we investigate these issues on both a simple two-dimensional spin system and a ten-residue, *in vacuo* polyglycine molecule.

III. RESULTS

A. *xy* model

In the *xy* model, spins (or planar magnets) of magnitude S are fixed in the *xy* plane with orientations specified by an angle ϕ_i ($0 \leq \phi_i < 2\pi$) with respect to the x axis. The spins interact via a coupling constant J according to the Hamiltonian,

$$H = -JS^2 \sum_{\langle ij \rangle} \cos(\phi_i - \phi_j), \qquad (8)$$

which is invariant to global rotations. For the present research, the system consisted of $N_s = 100$ spins arranged on a square lattice. Each spin was allowed to interact with its four nearest neighbors.

Simulations were performed utilizing the annealed-swapping algorithm on a set of chains spanning a temperature range of 200–800 K. Although the annealed-swapping method is designed to be implemented in parallel on a cluster of computer nodes, the *xy* model was sufficiently fast to run

in serial mode on alpha workstations. For more complex systems, such as for the peptide chains discussed below, there is significant gain in utilizing the parallel design. The total number of sweeps was set to 1×10^6, where each sweep roughly consists of updating the orientations of the 100 spins in the system ($= 1 \times 10^8$ updates in total). It is convenient to reexpress the total number of updates for the simulation in terms of the number of steps per swap attempt which, for n intermediate chains with m updates along a chain, consists of $2mn$ steps.

To determine the optimal combination of the above parameters and as a means of monitoring the efficiency of the algorithm, the ''trajectory'' of a composite Markov chain was followed as it cycled from lowest to highest temperatures. A single cycle was measured as the completed path, beginning at the lowest temperature, traversing to the high-temperature chain, and finally returning to the lowest temperature. It was determined that the number of cycles per CPU second, from here on referred to as the ''mixing rate,'' was optimized for a system of $N = 3$ principal chains (corresponding to $T = 200$, 500, and 800 K), $n = 80$ intermediate distributions (threads), and $m = 400$ updating steps along each thread. For 1×10^8 total updates, this corresponds to ~ 781 total swaps with 128 000 updates per swap. Note that although increasing the number of intermediate threads improves the overlap between adjacent principal chains, one must balance this improvement against the increased cost in CPU time.

In order to compare with the annealed-swapping technique, simulations were also executed for the simple-swap method with the total number of updates held constant. For simple swapping, the number of principal chains, N, and the number of updates, m', along a chain were also adjustable parameters. It was observed that the greatest mixing rates were obtained by performing many swap attempts with short updating (along a particular chain). We anticipate that this is a general characteristic of the method and is not system dependent. The number of cycles per second was optimized for 6 chains (corresponding to $T = 200$, 280, 390, 500, 630, and 800 K) with $m' = 20$ updates per swap along a principal chain, giving a total of 833 333 swaps with 120 total steps per swap. Although the swap acceptance ratios were quite small ($\sim 8\%$), the composite chain thoroughly explored the full temperature range and only occasionally became trapped within a given temperature interval.

There are a number of observables that can be used to compare the efficiency of the different simulation methods. We have chosen to monitor the potential energy U, heat capacity C_v, and the orientational order parameter S_{xy}, defined as

$$S_{xy} = \langle \overline{S_x}^2 + \overline{S_y}^2 \rangle^{1/2}, \qquad (9)$$

where

$$\overline{S_x} = \frac{1}{N_s} \sum_{i=1}^{N_s} S_{xi}, \qquad (10)$$
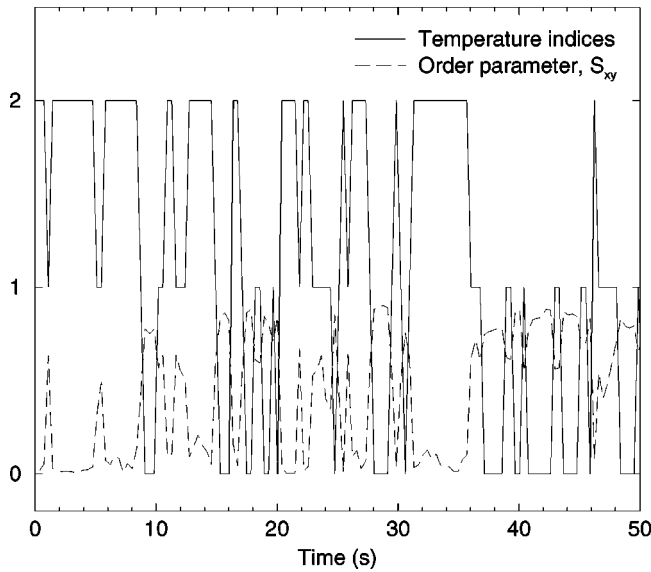
FIG. 2. The time-series evolution (first 50 sec) of the order parameter S_{xy} for a single composite chain using the annealed-sampling algorithm with the xy model.

with $\mathbf{S}_i$ representing the orientation of a given spin i in the xy plane. Note that for the ordered state, corresponding to uniformly aligned spins, the system has continuous rotational symmetry. This order parameter has the property that it is independent of the direction of alignment and is of $\mathcal{O}(1)$. For the disordered state, S_{xy} is $\mathcal{O}(1/\sqrt{N_s})$. In Fig. 2, we display the time-series evolution (first 50 sec) of the order parameter for a given composite chain using the annealed-swapping algorithm. The actual pathway of the chain through the different temperature indices (0=low temperature, 2=high temperature) is also included for reference. It is clear that the order parameter ''follows'' this pathway: as the chain reaches high temperature, the spins become disordered and at low temperature the system is ordered. The corresponding history of the potential energy can be seen in Fig. 3.

By tracking the states associated with a given temperature index, one can also monitor the potential energy for an individual chain. Additionally, the probabilities or weights for swapping between two adjacent chains may be recorded. In Fig. 4, the histograms of such weights are presented for both simple and annealed-swapping routines. Note that for annealed swapping there is an appreciable fraction of the distribution which is greater than zero, indicating significant overlap between the chains and higher swap acceptance ratios. In contrast, the simple-swap distribution is predominantly below zero and results in very small acceptance probabilities. The improved overlap between chains for the annealed-swapping method allows for greater mobility of the composite chain and is one of the attractive features of this method. However, it has been determined that there is only a marginal difference between the mixing rates for the two algorithms. This is demonstrated through the autocorrelation function for S_{xy}, which is plotted in Fig. 5 for both algorithms. Observe that the correlations decay more rapidly for the simple-swapping method compared to the annealed swapping.
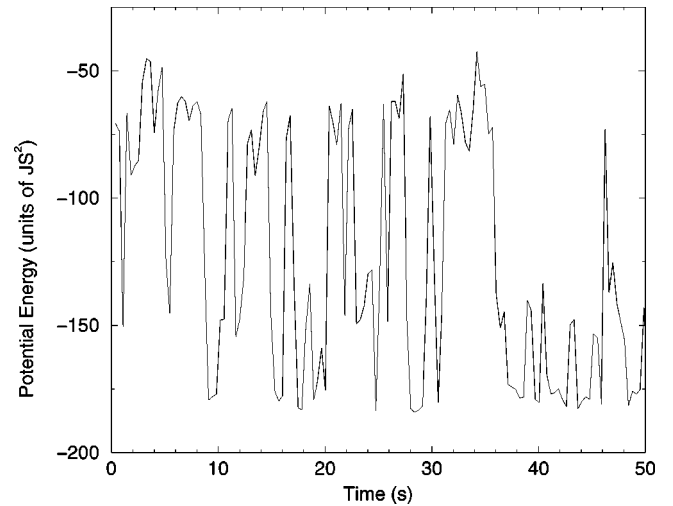

FIG. 3. The history (first 50 sec) of the potential energy for a single composite chain with the annealed-sampling algorithm using the xy model.

Thus, although the annealed-swapping algorithm affords greater overlap between neighboring chains, the simple-swapping method results in a faster mixing of the chains, despite smaller acceptance probabilities. A possible explanation for these discrepancies is that, although the annealed swapping results in greater overlap between chains, there is a greater CPU cost for swap attempts. For annealed swapping, much time is expended performing standard updating on the intermediate chains. If the move is rejected, then much work is wasted since the weights generated on the intermediate threads are not utilized. Thus, in this scheme, failed attempts are quite costly. In contrast, for simple swapping, failed swap attempts are computationally cheaper than accepted attempts (which require a shuffling of temperatures and chain labels), since no updating is required. One can take advantage of this mismatch by attempting more frequent swap
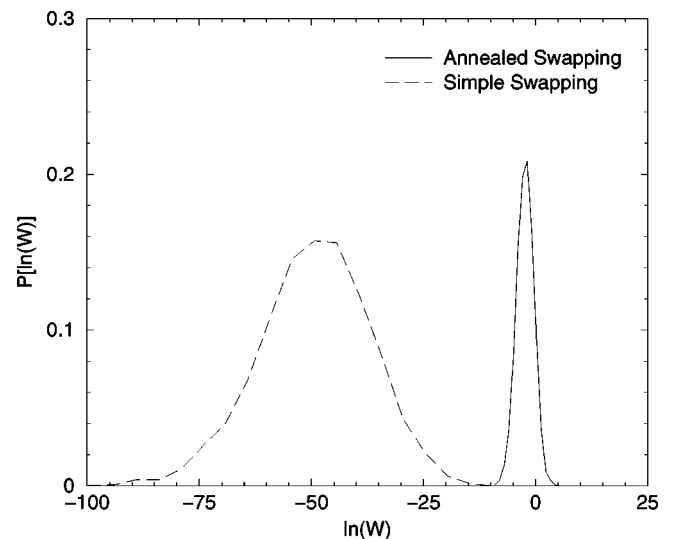

FIG. 4. The distributions of the logarithm of the weights obtained from simulations of the xy model for both annealed and simple swapping.
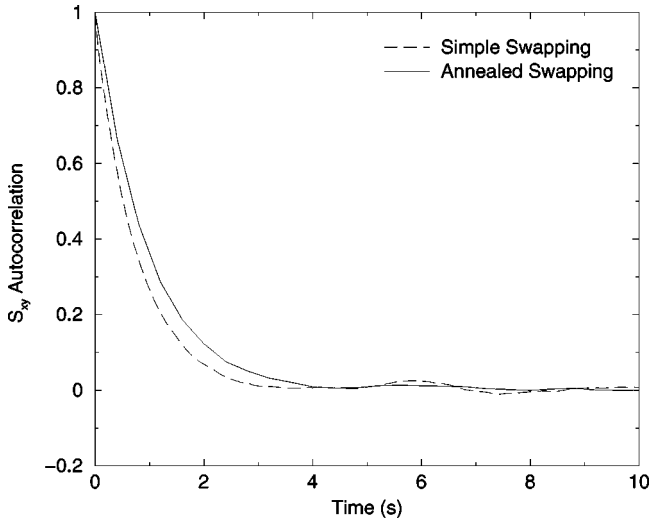
FIG. 5. The spin-spin autocorrelation function calculated from simulations of the xy model for both annealed- and simple-swapping algorithms.

FIG. 6. A scatter plot of weights vs energy for a simulation utilizing quenched importance sampling with the xy model.

moves, which, although resulting in lower acceptance ratios, allows for greater mixing of the chains per CPU second.

To reduce the number of failed swap attempts and improve the acceptance probabilities for annealed swapping, one can increase the number n of intermediate threads and adjust the number m of updates. However, this also comes at a cost due to the increased time for regular Metropolis updating. As mentioned above, we have performed detailed analysis and ascertained that there is an optimal number of intermediate chains, beyond which one finds diminishing returns. Note that these findings are not altered by modifications of the standard updating scheme along an individual chain. For instance, we have incorporated hybrid MC updating as a means of improving the mobility along an individual chain, but this did not lead to an improvement in the cycles per second, since the time required for updating along the intermediate chains remained roughly the same as for standard Metropolis MC updates. Variants of the annealed-swapping method have also been tested. A method which involves a hybrid mix of both simple and annealed swapping led to an increase in the number of cycles per second, but did not surpass the mixing rates for the simple-swapping method alone.

If one is solely interested in a single target distribution P_0 which has a number of isolated modes, then the quenching method outlined in Sec. II A is a perfectly valid procedure. Recall that this method allows one to generate a sequence of states $x^1, \ldots, x^N$ distributed according to P_0 with associated weights $w^1, \ldots, w^N$. In order to accurately calculate averages, as given in Eq. (2), it is required that the state space be well sampled, implying a fairly uniform distribution of weights. For comparative purposes, the procedure was tested by quenching from the high-temperature to the low-temperature chain. One of the key findings was that only a small fraction of the weights gave an appreciable contribution to the average in Eq. (2). A scatter plot of the weights versus the final energies, as seen in Fig. 6, illustrates another shortcoming of the method. Observe that there is consider-
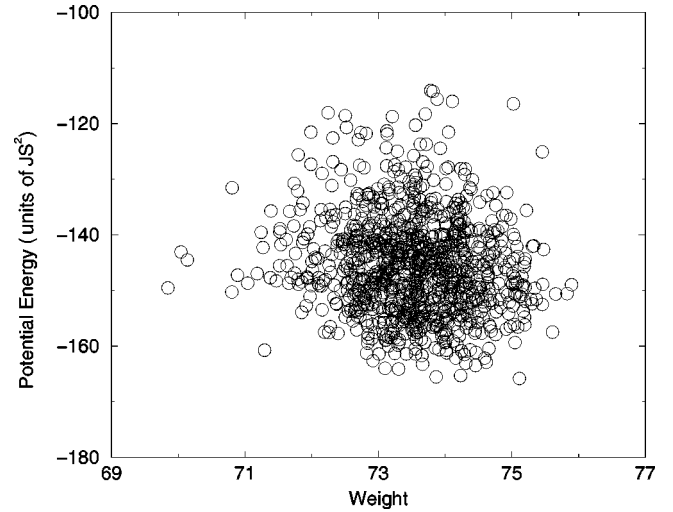
able dispersion in the energies for fixed weight, which implies significant memory effects and would lead to large variances and poor estimates of the average.

B. Peptide model

The xy model is a useful diagnostic for comparing the various algorithms, since it is relatively fast and exhibits many of the features of interest. However, it is useful to consider more complex models in order to gauge the effectiveness of these methods.

We have performed simulations on gly10, a peptide composed of ten glycine residues, in which the CH_2 methylene groups were treated as single monomer units under the united atom scheme. The potential energy function employed consists of both nonbonded V_{nb} and bonded V_b interactions which are based on the classical ''Charmm'' potentials [18]. The nonbonded potentials include van der Waals and electrostatic interactions,

$$V_{nb} = \sum_{i,j}{}' \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{K_d q_i q_j}{r_{ij}} \right), \qquad (11)$$

where the prime indicates a reduced sum over all pairs of atoms spaced by at least four bonds and A_{ij} and B_{ij} are the Lennard-Jones parameters. Note that the hydrogen-bond interactions are incorporated into the electrostatic interactions with dielectric constant $K_d = 332.0638$ $(K_{cal}$ Å $)/mol$. The bonded potentials include bond length (bl), bond angle (ba), proper dihedral (pd), and improper dihedral (id) interactions,

$$V_b = \sum_{i=1}^{N-1} K_i^{bl}(b_i - b_i^0)^2 + \sum_i^{N_{ba}} K_i^{ba}(\theta_i - \theta_i^0)^2$$

$$+ \sum_i^{N_{id}} K_i^{id}(\gamma_i - \gamma_i^0)^2 + \sum_i^{N_{pd}} \sum_{n=0}^{3} K_{i,n}^{pd}\cos(n\phi_i), \quad (12)$$

where the maximum values for the various force constants are given as $K_{max}^{bl} = 640$ Kcal/(mol Å^2), $K_{max}^{id} = 147$ Kcal/mol, $K_{max}^{ba} = 123.5$ Kcal/mol, and $K_{max}^{pd} = 6$ Kcal/mol.

Based upon the relative strengths of the bonded terms in Eq. (11), the magnitudes of the various types of trial moves in a simulation can be ordered according to the following scaling:

$$\frac{\Delta b_{max}}{b_0} < \frac{\Delta \cos(\theta_{max})}{\pi} \sim \frac{\Delta \omega_{max}}{2\pi} = \frac{\Delta \gamma_{max}}{2\pi} \ll \frac{\Delta \phi_{max}}{2\pi}, \quad (13)$$

where Δb_{max}, $\Delta \cos(\theta_{max})$, $\Delta \omega_{max}$, $\Delta \gamma_{max}$, and $\Delta \phi_{max}$ correspond to the maximum displacements for the bond length, cosine of the bond angle, peptide-bond dihedral, improper dihedral, and ''soft'' backbone dihedral degrees of freedom, respectively. For the present study, the bond lengths were constrained to their ground-state values and the peptide-bond dihedrals were set to π.

Simulations were performed with the annealed-swapping method and, for comparison, the simple-swapping algorithm over a temperature range of 400–1000 K. The total number of MC steps for the simulations was 5×10^7 with a burn-in length of $\sim 1 \times 10^3$ steps. For the annealed-swapping simulations, $N = 3$ principal chains (with $T = 400$, 600, and 1000 K) were used with $n = 80$ intermediate threads and $m = 156$ updating steps along each thread, corresponding to 1000 swaps with 5×10^4 total updates per swap. The simple-swapping runs utilized 6 principal chains (with $T = 400, 470, 550, 650, 800, 1000$ K) with 50 updates per swap along a chain, giving 166 666 swaps with 300 steps per swap. Although the temperature range for these studies was comparable to those for the xy model, the swap acceptance ratios were considerably *higher* despite the relatively short updating. These issues will be elaborated upon below.

Of the various thermodynamic quantities which were calculated during the simulations, we focus on the potential energy U and the radius of gyration R_g, defined as

$$R_g^2 = \frac{1}{M_t} \left\langle \sum_{i=1}^{N} M_i (\mathbf{R}_i - \mathbf{R}_{c.m.})^2 \right\rangle, \quad (14)$$

where M_i is the mass of atom i, $M_t = \Sigma_i M_i$, $\mathbf{R}_{c.m.}$ is the center of mass of the system, and the summation is over all the N atoms in the peptide. This order parameter gives a measure of the ''size'' of the peptide, which for a random coil scales as $\sim \sqrt{N}$. In Fig. 7, data are presented for R_g for a particular composite chain utilizing the annealed-swapping procedure. Observe that at low temperatures the peptide adopts a compact structure and at higher temperatures the chain becomes more extended. Although the radius of gyration is insensitive to changes in confirmations at low temperatures, it is evident that the composite chain samples the full temperature range and becomes thoroughly mixed.

To compare the annealed and simple-swapping methods, we consider the weights associated with swapping between two adjacent chains. The histograms of these weights are plotted in Fig. 8. There is a significant proportion of the annealed-swapping distribution which lies greater than zero, resulting in good overlap between the chains. Although the
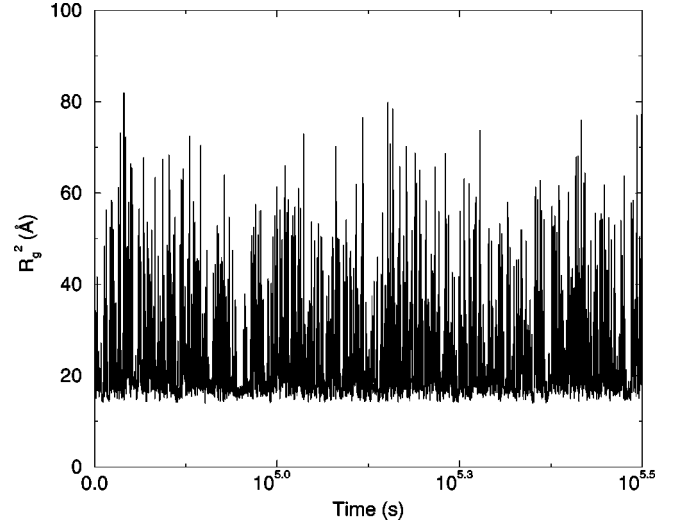

FIG. 7. The time evolution of the radius of gyration, R_g, for a single composite chain for a simulation of gly10 employing the annealed-sampling method.

mean of the simple-swap distribution is located well below zero, the distribution is skewed and has a relatively long tail extending greater than zero. This long tail results in somewhat larger acceptance ratios than those obtained in the xy model, as noted above. Despite better overlap between adjacent chains for the annealed-swapping procedure, the mixing rates are an order of magnitude smaller than for the simple-swapping algorithm, in sharp contrast to the xy model. These differences can be seen through the autocorrelation function for R_g, presented in Fig. 9. Note that the correlation time for the annealed-swapping method is significantly greater than that for the simple-swapping procedure.

Hence, despite better overlap between adjacent chains, the annealed-swapping technique did not perform as well as the simple-swapping algorithm and the discrepancy between the two methods was much greater for the peptide model com-
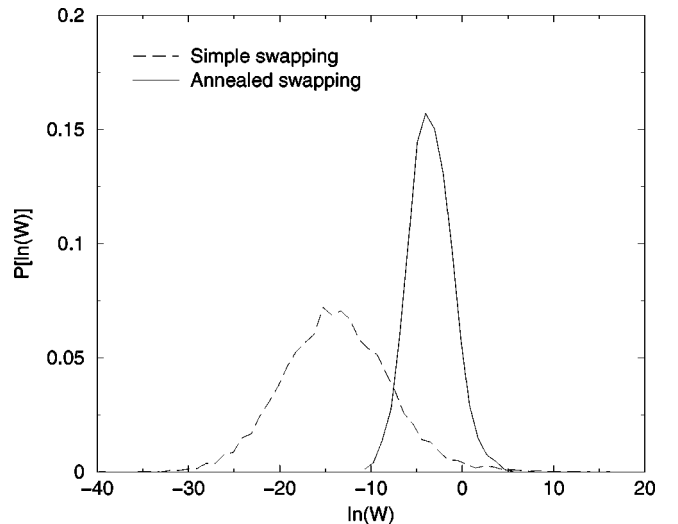

FIG. 8. Histograms of the logarithm of weights obtained from simulations of gly10 for both the simple- and annealed-swapping methods.
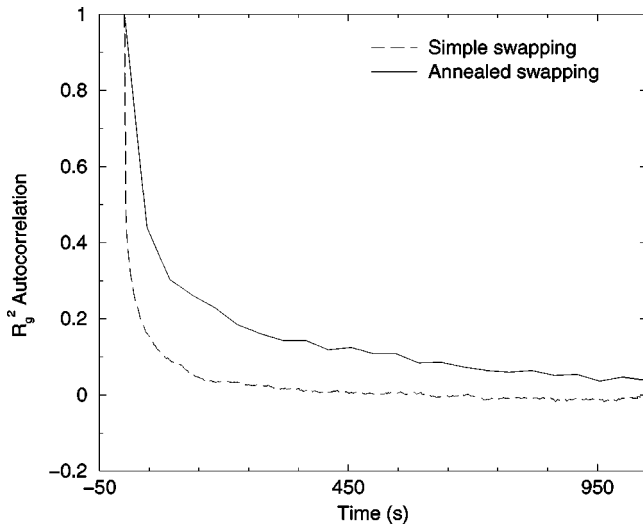
FIG. 9. The autocorrelation for the radius of gyration calculated for simulations of gly10 for both annealed- and simple-swapping algorithms.

pared to the xy model. It was anticipated that the increased complexity of the peptide model would result in a decrease in the mixing rates for both sampling methods, but that the *ratio* of these rates would remain constant. To test this hypothesis, additional simulations were executed with the xy model where a larger neighbor list was employed (~ 18 neighbors per spin). The increased computational effort resulted in a decrease in the number of cycles per second for both sampling methods, but the ratio of the mixing rates remained approximately the same.

The lack of such scaling invariance between the models might be explained by differences in the sampling methodology. In the simulation of the xy model the procedure for updating in the system was adjusted to ensure acceptance ratios of roughly 30%, which should yield optimal mobility (for single-variable updates) at each of these temperatures. Such tuning is difficult to implement in the simulations of the peptide system since different *types* of updates are extremely important at different temperatures to obtain optimal mobility. The lack of adjustment of updating schemes in the peptide simulation leads to significantly different acceptance ratios for each principal chain, resulting in a mismatch in the swapping acceptance ratios and poorer overall mobility of the composite Markov chain. For the annealed-swapping routine, decreased exploration along a given chain tends to exacerbate hysteresis effects with the composite chain becoming trapped within a certain temperature interval. It is believed that such bottlenecking contributes to the dramatically slower mixing rates and is responsible for the major change in the ratio of mixing rates between the two models.

To support or refute these ideas, additional simulations were run with the xy model, where the maximal displacement in the angle ϕ was reduced in magnitude and set to the same value for all temperatures. It was observed that the ratio of mixing rates became comparable to that for the peptide model where, again, the swapping mismatch had a greater effect on the performance of the annealed-swapping

technique with a significant decrease in the mixing rate. These results indicate that differences in the mixing rates between the models are most likely due to details of the sampling procedure. By adjusting the peptide simulation to incorporate different updating schemes for each temperature, it is anticipated that the ratios of mixing rates will be similar for the two models. However, it is also possible that fundamental differences in the models govern the behavior of the swapping and lead to deviations in the mixing rates. It is conceivable that the energy landscape for the peptide model is much ''rougher'' than the xy model and there is a greater propensity to become trapped in metastable states. Unfortunately, without more knowledge about the details of the energy landscapes, it is difficult to test this hypothesis.

IV. CONCLUSIONS

In this article, a number of extended state-space Monte Carlo algorithms which generalize the parallel-tempering method have been examined. These methods have been designed to incorporate a deterministic procedure to sample over the auxiliary parameter space in order to test whether directed motion can improve the sampling rate over that observed in parallel tempering (here called simple swapping). In the present study, the deterministic procedure was based on heating and quenching procedures, which resemble the simulated annealing method. However, unlike the simulated-annealing algorithm, the directed heating and quenching procedures allow statistical information to be recovered and thereby permit an importance sampler to be formulated. It should be mentioned that canonical jump walking (CJW) [19] is another sampling method which shares the deterministic character of the annealed swapping algorithm. However, it can be rigorously shown that CJW does not satisfy detailed balance and is only approximately correct in the asymptotic limit of a very long auxiliary chain. As such, there is no guarantee that, for finite-length simulations, one has obtained the correct limiting distribution. In contrast, as demonstrated in Appendix B, the annealed-sampling method does satisfy detailed balance.

Although the annealed-swapping procedure increases the swap probability by effectively improving the overlap of the energy distributions of adjacent chains, it was shown that for both the xy and peptide models the mixing rates were greater for simple-swapping. We interpret this remarkable result as a reflection of the robustness of the simple-swapping method. The deviations in the rates also increased for the peptide model, where the annealed swapping had significantly slower mixing, but such discrepancies may be attributed to differences in the sampling methodology. Modifications of the peptide code, although not trivial, should allow for ratios of mixing rates which are comparable to those obtained for the xy model. Nonetheless, even for the xy model, the lack of any order-of-magnitude improvement in the performance of annealed-swapping compared to simple swapping must be explained. With annealed swapping, a large percentage of the CPU time is dedicated to performing simple Metropolis updates on the intermediate threads. Hence, although the number of failed swap attempts is reduced by the improved

overlap between adjacent chains, much time is consumed on average to obtain an accepted swap. Furthermore, as the swap acceptance ratio increases, there is a tendency for the composite Markov chain to cycle locally between two adjacent chains and become trapped. The overall efficiency of the annealed-swapping method is determined by the interplay between the increased computational cost of performing the heating and cooling processes and the improved net overlap of the distribution of proposed configurations. For the systems studied in this work, these factors approximately balance one another and little differences in efficiency can be observed between the annealed-swapping and simple-swapping procedures.

In order to provide a fair comparison of all methods, extensive optimization studies were carried out. An interesting result of these investigations was that swap attempts should be executed extremely frequently in the simple-swapping approach. This result seems surprising in light of the fact that rapid swaps require short intervals of local Metropolis updates which induces strong correlations in the swap dynamics of the chains. It appears that, although increasing the length of the local updates tends to destroy correlations, the rate at which correlations are lost is not sufficient to compensate for the reduced swap time. The optimal procedure of frequent swap attempts was observed in all systems studied and appears to be fairly general. This result is encouraging as it suggests that little optimization of the parallel tempering algorithm is required for each new system.

In the present study, the auxiliary parameter space was taken to be the inverse temperature. In this context, the deterministic procedure used to promote the coupling between adjacent chains corresponds to heating and quenching processes. Although this implementation of the directed procedure was not successful in improving the rate of exploration of configurational space over the sampling rate of parallel tempering, it is possible that a more judicious choice of control parameter would help equilibrate the system along paths linking high-energy to low-energy states, leading to greater mobility of composite chain. Further work along these lines is under way. In addition, the use of Markov chain transition matrices operating on an extended state space provides a useful tool in the construction of sampling approaches and has been exploited to improve the efficiency of *ab initio*-based Monte Carlo simulations [20].

APPENDIX A

In this appendix, we demonstrate that the quenched importance sampling algorithm outlined in Sec. II A obeys detailed balance [12]. The validity of the algorithm can be established by viewing the expectation of $a(\mathbf{x})$ in Eq. (1) as an average over a probability f in an extended space $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$,

$$\bar{a} = \int d\mathbf{x}_1 \cdots d\mathbf{x}_n a(\mathbf{x}_1) f(\mathbf{x}_1, \ldots, \mathbf{x}_n), \qquad \text{(A1)}$$

where we define the probability of the extended state-space point $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ to be the joint probability

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = P_0(\mathbf{x}_1) T_1(\mathbf{x}_1 \rightarrow \mathbf{x}_2) \cdots T_{n-1}(\mathbf{x}_{n-1} \rightarrow \mathbf{x}_n). \qquad \text{(A2)}$$

Since the T_j are valid transition probabilities, the marginal distribution of f is P_0 so that Eq. (A1) holds. Since the detailed balance condition

$$T_j(\mathbf{x}_2 \rightarrow \mathbf{x}_1) = T_j(\mathbf{x}_1 \rightarrow \mathbf{x}_2) \frac{P_j(\mathbf{x}_1)}{P_j(\mathbf{x}_2)} \qquad \text{(A3)}$$

is assumed to hold for the T_j, Eq. (A2) can be expressed as

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{P_0(\mathbf{x}_1)}{P_1(\mathbf{x}_1)} T_1(\mathbf{x}_2 \rightarrow \mathbf{x}_1) \cdots \frac{P_{n-2}(\mathbf{x}_{n-1})}{P_{n-1}(\mathbf{x}_{n-1})}$$
$$\times T_{n-1}(\mathbf{x}_n \rightarrow \mathbf{x}_{n-1}) P_{n-1}(\mathbf{x}_n). \qquad \text{(A4)}$$

The quenching procedure generates the extended state point $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ with probability

$$g(\mathbf{x}_1, \ldots, \mathbf{x}_n) = P_n(\mathbf{x}_n) T_{n-1}(\mathbf{x}_n \rightarrow \mathbf{x}_{n-1}) \cdots T_1(\mathbf{x}_2 \rightarrow \mathbf{x}_1), \qquad \text{(A5)}$$

which implies that the correct statistical weight $w^{(i)}$ for the point $\mathbf{x}_i = \mathbf{x}_1$ is

$$w^{(i)} = \frac{f(\mathbf{x}_1, \ldots, \mathbf{x}_n)}{g(\mathbf{x}_1, \ldots, \mathbf{x}_n)}, \qquad \text{(A6)}$$

which yields the weight factor in Eq. (3).

APPENDIX B

In this appendix, we demonstrate that the annealed-swapping algorithm outlined in Sec. II C obeys detailed balance for a composite Markov chain consisting of two principal chains, chain 1 and chain 2. Suppose the principal chains are initially in states $\hat{\mathbf{x}}_1$ and $\check{\mathbf{x}}_2$, respectively (see Fig. 1). The composite chain is defined to be a Markovian sequence of extended phase points $\tilde{\mathbf{x}}_1 = (\hat{\mathbf{x}}_1, \check{\mathbf{x}}_2)$ in which the extended phase points are distributed in the chain according to the limiting distribution $P(\tilde{\mathbf{x}}_1) = P_1(\hat{\mathbf{x}}_1) P_2(\check{\mathbf{x}}_2)$. The annealed-swapping method consists of taking the extended state $\tilde{\mathbf{x}}_1$ and generating a new trial state $\tilde{\mathbf{x}}_2 = (\check{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ by a heating and quenching process. The trial state is then accepted or rejected according to a Metropolis-Hastings procedure. Without loss of generality, we will assume that there is only one intermediate distribution $\hat{P}_i$ in the heating process and one intermediate distribution $\check{P}_i$ for the quenching process. We assume that transition probabilities $\hat{T}_i(\mathbf{x}_1 \rightarrow \mathbf{x}_2)$ and $\check{T}_i(\mathbf{x}_1 \rightarrow \mathbf{x}_2)$ obey microscopic reversibility (or detailed balance):

$$\hat{T}_i(\mathbf{x}_2 \to \mathbf{x}_1) = \hat{T}_i(\mathbf{x}_1 \to \mathbf{x}_2)\frac{\hat{P}_i(\mathbf{x}_1)}{\hat{P}_i(\mathbf{x}_2)}, \tag{B1}$$

where $\hat{T}_i(\mathbf{x}_2 \to \mathbf{x}_1)$ is the probability of moving to state $\mathbf{x}_1$ from state $\mathbf{x}_2$ in the intermediate-heating Markov chain.

We must show that the following detailed-balance condition is satisfied:

$$P(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) = P(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1), \tag{B2}$$

where $P(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) = P(\tilde{\mathbf{x}}_1)T(\tilde{\mathbf{x}}_1 \to \tilde{\mathbf{x}}_2)$ is the probability of observing the sequence $\{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2\}$ in the composite chain, and $T(\tilde{\mathbf{x}}_1 \to \tilde{\mathbf{x}}_2)$ is the transition probability of moving from state $\tilde{\mathbf{x}}_1$ to state $\tilde{\mathbf{x}}_2$. The transition probability can be expressed as

$$T(\tilde{\mathbf{x}}_1 \to \tilde{\mathbf{x}}_2) = P_g(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2)P_a(\tilde{\mathbf{x}}_2 | \tilde{\mathbf{x}}_1), \tag{B3}$$

where $P_g(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) = P_g(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)P_g(\check{\mathbf{x}}_2, \check{\mathbf{x}}_1)$ is the joint probability of generating $\tilde{\mathbf{x}}_2$ from $\tilde{\mathbf{x}}_1$ by the annealing procedure and $P_a(\tilde{\mathbf{x}}_2 | \tilde{\mathbf{x}}_1)$ is the path-dependent probability of accepting the proposed state $\tilde{\mathbf{x}}_2$. One can define $P_a(\tilde{\mathbf{x}}_2 | \tilde{\mathbf{x}}_1) = \min[1, A(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1)]$, where the acceptance matrix A is defined to be

$$A(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1) = \frac{P(\tilde{\mathbf{x}}_2)P_g(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1)}{P(\tilde{\mathbf{x}}_1)P_g(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2)}, \tag{B4}$$

and it then follows that

$$P(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2) = \min[P(\tilde{\mathbf{x}}_1)P_g(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2), P(\tilde{\mathbf{x}}_2)P_g(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1)]$$
$$= P(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1), \tag{B5}$$

showing that detailed balance is satisfied for the composite Markov chain.

Noting that the conditional probability $P(\tilde{\mathbf{x}}_2)P_g(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1)$ of generating the state $\tilde{\mathbf{x}}_1$ by the annealing procedure is

$$P(\tilde{\mathbf{x}}_2)P_g(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1) = P_1(\check{\mathbf{x}}_1)\check{T}_i(\check{\mathbf{x}}_1 \to \check{\mathbf{x}}_2)P_2(\hat{\mathbf{x}}_2)\hat{T}_i(\hat{\mathbf{x}}_2 \to \hat{\mathbf{x}}_1)$$

$$= P_1(\check{\mathbf{x}}_1)\frac{\check{P}_i(\check{\mathbf{x}}_2)}{\check{P}_i(\check{\mathbf{x}}_1)}\check{T}_i(\check{\mathbf{x}}_2 \to \check{\mathbf{x}}_1)$$

$$\times P_2(\hat{\mathbf{x}}_2)\frac{\hat{P}_i(\hat{\mathbf{x}}_1)}{\hat{P}_i(\hat{\mathbf{x}}_2)}\hat{T}_i(\hat{\mathbf{x}}_1 \to \hat{\mathbf{x}}_2)$$

$$= \frac{P_1(\check{\mathbf{x}}_1)}{P_1(\hat{\mathbf{x}}_1)}\frac{\check{P}_i(\check{\mathbf{x}}_2)}{\check{P}_i(\check{\mathbf{x}}_1)}\frac{P_2(\hat{\mathbf{x}}_2)}{P_2(\check{\mathbf{x}}_2)}\frac{\hat{P}_i(\hat{\mathbf{x}}_1)}{\hat{P}_i(\hat{\mathbf{x}}_2)}$$

$$\times P(\tilde{\mathbf{x}}_1)P_g(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2), \tag{B6}$$

the acceptance matrix can be written as $A(\hat{\mathbf{x}}_2, \check{\mathbf{x}}_1) = \hat{w} \times \check{w}$, where the weights $\hat{w}$ and $\check{w}$ are the statistical weights for the heating and cooling processes, respectively, and are given by

$$\hat{w} = \frac{\hat{P}_i(\hat{\mathbf{x}}_1)}{P_1(\hat{\mathbf{x}}_1)}\frac{P_2(\hat{\mathbf{x}}_2)}{\hat{P}_i(\hat{\mathbf{x}}_2)}, \tag{B7}$$

$$\check{w} = \frac{\check{P}_i(\check{\mathbf{x}}_2)}{P_2(\check{\mathbf{x}}_2)}\frac{P_1(\check{\mathbf{x}}_1)}{\check{P}_i(\check{\mathbf{x}}_1)}. \tag{B8}$$

Note that all distributions in $A(\hat{\mathbf{x}}_2, \check{\mathbf{x}}_1)$ appear as ratios, so that no normalization factors need to be calculated.

The arguments above establish that the limiting distribution of the composite Markov chain is $P = P_1 P_2$. This, in turn, implies that the states composing principal chain j formed by taking the jth component of the extended phase points in the composite chain are distributed according to P_j for long chains.

Now consider adding another Monte Carlo update according to $\check{T}_i$ in the annealing procedure so that the path connecting $\check{\mathbf{x}}_2$ to $\check{\mathbf{x}}_1$ is now $\{\check{\mathbf{x}}_2, \check{\mathbf{x}}_1', \check{\mathbf{x}}_1\}$. Since

$$\check{T}_i(\check{\mathbf{x}}_2 \to \check{\mathbf{x}}_1')\check{T}_i(\check{\mathbf{x}}_1' \to \check{\mathbf{x}}_1)$$

$$= \frac{\check{P}_i(\check{\mathbf{x}}_1')}{\check{P}_i(\check{\mathbf{x}}_2)}\frac{\check{P}_i(\check{\mathbf{x}}_1)}{\check{P}_i(\check{\mathbf{x}}_1')}\check{T}_i(\check{\mathbf{x}}_1 \to \check{\mathbf{x}}_1')\check{T}_i(\check{\mathbf{x}}_1' \to \check{\mathbf{x}}_2)$$

$$= \frac{\check{P}_i(\check{\mathbf{x}}_1)}{\check{P}_i(\check{\mathbf{x}}_2)}\check{T}_i(\check{\mathbf{x}}_1 \to \check{\mathbf{x}}_1')\check{T}_i(\check{\mathbf{x}}_1' \to \check{\mathbf{x}}_2), \tag{B9}$$

it is straightforward to show that any number of Monte Carlo updates may be done at each step in the annealing process without modifying the acceptance probability matrix A.

The derivation outlined here is easily generalized to show that the proper acceptance matrix A for arbitrary numbers $\hat{N} - 2$, $\check{N} - 2$ of intermediate heating and quenching distributions between initial and final distributions P_i and P_f is $A = \hat{w} \times \check{w}$, where

$$\hat{w} = \prod_{j=1}^{\hat{N}-1} \frac{\hat{P}_{j+1}(\hat{\mathbf{x}}_j)}{\hat{P}_j(\hat{\mathbf{x}}_j)}, \tag{B10}$$

$$\check{w} = \prod_{j=1}^{\check{N}-1} \frac{\check{P}_j(\check{\mathbf{x}}_j)}{\check{P}_{j+1}(\check{\mathbf{x}}_j)}, \tag{B11}$$

with $\hat{P}_{\hat{N}} = P_f = \check{P}_{\check{N}}$ and $\hat{P}_1 = P_i = \check{P}_1$.

[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[2] See, for example, K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics* (Spring-Verlag, Berlin, 1992).

[3] J. P. Valleau and S. G. Whittington, in *Statistical Mechanics*, edited by B. J. Berne (Plenum, New York, 1977), Chap. 4, pp. 114–119.

[4] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991); Phys. Rev. Lett. **68**, 9 (1992); J. P. Valleau, J. Chem. Phys. **99**,

4718 (1993).

[5] W. Kerler and P. Rehberg, Phys. Rev. E **50**, 4220 (1994); B. A. Berg and T. Celik, Phys. Rev. Lett. **69**, 2292 (1992).

[6] U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem. **14**, 1333 (1993); M.-H. Hao and H. A. Scheraga, J. Phys. Chem. **98**, 4940 (1994).

[7] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977); J. P. Valleau, J. Chem. Phys. **99**, 4718 (1993).

[8] B. A. Berg, Int. J. Mod. Phys. C **3**, 1083 (1992).

[9] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992); A. P. Lyubartsev, A. A. Martinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyminov, J. Chem. Phys. **96**, 1776 (1992).

[10] S. Kirkpatrick, C. D. Gelatt, and P. M. Vecchi, Science **220**, 671 (1983).

[11] C. J. Geyer, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (Interface, Fairfax Station, VA, 1991), p. 156.

[12] R. M. Neal (unpublished).

[13] C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).

[14] M. C. Tesi, E. J. Janse van Rensberg, E. Orlandini, and S. G. Whittington, J. Stat. Phys. **82**, 155 (1996).

[15] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).

[16] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[17] R. M. Neal, Stat. Comput. **6**, 353 (1996).

[18] *QUANTA Parameter Handbook* (Polygen Corp., Waltham, MA, 1990).

[19] D. D. Frantz, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **93**, 2769 (1990).

[20] R. Iftimie, D. Salahub, D. Wei, and J. Schofield, J. Chem. Phys. **113**, 4852 (2000).