On the Efficiency of Exchange in Parallel Tempering Monte Carlo Simulations

Cristian Predescu*

Department of Chemistry and Kenneth S. Pitzer Center for Theoretical Chemistry, University of California, Berkeley, California 94720

Mihaela Predescu

Department of Mathematics, Bentley College, Waltham, Massachusetts 02452

Cristian V. Ciobanu

Division of Engineering, Colorado School of Mines, Golden, Colorado 80401

Received: October 28, 2004; In Final Form: December 19, 2004

We introduce the concept of effective fraction, defined as the expected probability that a configuration from the lowest index replica successfully reaches the highest index replica during a replica exchange Monte Carlo simulation. We then argue that the effective fraction represents an adequate measure of the quality of the sampling technique, as far as swapping is concerned. Under the hypothesis that the correlation between successive exchanges is negligible, we propose a technique for the computation of the effective fraction, a technique that relies solely on the values of the acceptance probabilities obtained at the end of the simulation. The effective fraction is then utilized for the study of the efficiency of a popular swapping scheme in the context of parallel tempering in the canonical ensemble. For large dimensional oscillators, we show that the swapping probability that minimizes the computational effort is 38.74%. By studying the parallel tempering swapping efficiency for a 13-atom Lennard−Jones cluster, we argue that the value of 38.74% remains roughly the optimal probability for most systems with continuous distributions that are likely to be encountered in practice.

I. Introduction

Since its introduction in the context of Monte Carlo simulations,[1−3] the technique of exchanging configurations between replicas characterized by slightly different parameters has been widely employed as a means to accelerate the convergence of Monte Carlo simulations. Perhaps the most widely utilized technique is parallel tempering in the canonical ensemble,[3−5] where the index parameter is the temperature. However, the index parameter may also be the energy, as in the case of parallel tempering in the microcanonical ensemble;[6] the chemical potential, as in the hyperparallel tempering method;[7] a delocalization parameter, as in the q-jumping Monte Carlo method;[8] or suitable modifications of the potential, as in the Hamiltonian replica exchange method.[9,10]

Irrespective of the index parameter that is utilized, the main features of the replica exchange technique are as follows.[2−5] Instead of one, a number of N parallel replicas of the system are simulated using Monte Carlo sampling or molecular dynamics, for example. This is called the parallel step. The N different replicas have distributions that are characterized by different values of an index parameter, for example, different temperatures. The replicas are usually ordered with the index parameter in a monotonic fashion. The Markov chains utilized for sampling in the parallel step have different equilibration times. For definiteness and perhaps after reordering, we may assume that the lowest index Markov chain equilibrates the slowest, whereas the highest index Markov chain is the fastest. In the swapping step, exchanges between replicas of different indexes are attempted periodically or randomly in order to induce mixing between the Markov chains utilized for the parallel step. The replica exchanges are accepted with the Metropolis et al acceptance probability[11,12] and rejected with the remaining probability. This way, detailed balance is enforced and the sampled distributions remain stationary with respect to the swapping step. The induced mixing helps the lowest index Markov chain to equilibrate faster, by "borrowing" the property of fast equilibration from the highest index Markov chain.

The rate at which configurations leave the lowest index Markov chain and successfully climb the ladder of index parameter values up to the highest index chain plays a crucial role in ensuring an efficient mixing. The fraction of configurations that do so during the Monte Carlo simulation will be called the "effective fraction". We point out that the exchange step is not capable of destroying or creating new configurations. In the worst case, an unsuitable configuration originating from the lowest index replica has to reach the highest index replica in order to be destroyed and replaced by a more favorable configuration. It is then apparent that the higher the effective fraction is, the better the mixing. The purpose of the present paper is to provide a means to compute the effective fraction. The work is facilitated by the observation that in the limit that the states of the parallel Markov chains become uncorrelated between two successive swapping events, the study of the mixing efficiency turns out to be independent of the particular sampling technique utilized for the parallel step.

In section II, we perform a complete mathematical treatment of the problem of computing the effective fraction for the most general cyclic swapping strategy. At the end of the section, we

10.1021/jp045073+ CCC: $30.25
Published on Web 02/10/2005

make the simplifying assumption that the Monte Carlo states between two successive swapping events are independent. Using this, we demonstrate that the computation of the effective fraction is reduced to the case of a cyclic composition of discrete state temporally homogeneous Markov chains with transition probabilities given by the average swapping probabilities. Although section II appears as an unnecessary complication given the end result, short of an exact mathematical treatment, there seems to be no other way to demonstrate how the problem of computing the effective fraction simplifies in the limit of uncorrelated exchanges. Familiarity with the general theory of discrete state Markov chains, for instance as provided in ref 13, is assumed. We have detailed the main steps involved in the computation of the effective fraction in section III, even if at the risk of reiterating some of the information already provided in section II.

The effective fraction concept can be utilized to study the efficiency of various swapping strategies. It can also be utilized for the determination of optimal acceptance probabilities. We illustrate this use by studying the optimal efficiency for parallel tempering simulations in the canonical ensemble. We work with the popular two-step swapping strategy, in which exchanges to the nearest neighbors of lower and higher index are attempted periodically, in a successive fashion. We demonstrate that the optimal acceptance probability for high-dimensional oscillators is about 38.74%, an interesting result, which, by means of a numerical example, is shown to hold for more complex distributions. The result is interesting not only because of its apparent applicability to many of the systems with continuous distributions but also because the optimal acceptance probability is actually not very high. This observation appears to somewhat contradict the general belief that low acceptance probabilities automatically imply a low mixing efficiency. In fact, we demonstrate that the acceptance probabilities may lie anywhere in the interval (7%, 82%), with a penalty of at most 100% in the computational cost.

We end our journey with a set of recommendations regarding the use of replica exchange methods. Perhaps the most important one is that the values of the acceptance probabilities alone should not be used as a means of judging the quality of the mixing efficiency. Rather, the effective number of swapping events, which is the product between the effective fraction and the total number of swapping attempts, must be calculated and employed for such purposes.

II. Effective Fraction of Swaps as a Measure of Exchange Efficiency

In standard implementations of parallel Monte Carlo techniques, (statistically) independent Markov chains $\{X_n^i;\ n \geq 0\}$ defined on the same state space Ω and having the same or different stationary normalized distributions $\{\rho_i(x);\ x \in \Omega,\ 1 \leq i \leq N\}$ are run in parallel on different machines. In the replica exchange Monte Carlo, swaps (also called exchanges) are attempted between states X_n^i and X_n^j of two "neighboring" Markov chains of indexes i and j. The swaps are then accepted with the conditional probability

$$a(X_n^i, X_n^j) = \min\left\{1, \frac{\rho_i(X_n^j)\rho_j(X_n^i)}{\rho_i(X_n^i)\rho_j(X_n^j)}\right\} \tag{1}$$

or rejected with the remaining probability

$$r(X_n^i, X_n^j) = 1 - a(X_n^i, X_n^j) \tag{2}$$

This acceptance/rejection rule ensures that the detailed balance condition is satisfied and that the product probability distribution $\rho(x_1)\rho(x_2)\cdots\rho(x_N)$ is stationary with respect to the swapping step.[2,3] The expected (average) values of the acceptance and rejection probabilities are defined by

$$a_{ij} = \mathbb{E}[a(X_n^i, X_n^j)] = \int_\Omega \int_\Omega \rho_i(x)\rho_j(y)a(x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{3}$$

and

$$r_{ij} = \mathbb{E}[r(X_n^i, X_n^j)] = 1 - a_{ij} \tag{4}$$

respectively. These averages are to be evaluated during the Monte Carlo simulation itself.

The term "neighboring" refers to the fact that the Markov chain distributions are normally chosen to depend continuously upon an index parameter T, so that $\rho_i(x) = \rho(x; T_i)$. The chains are indexed in such a way that the parameter increases (or decreases) with i, i.e.,

$$T_{\min} = T_1 < T_2 < \cdots < T_N = T_{\max}$$

For instance, in the case of parallel tempering Monte Carlo simulations in the canonical ensemble, the index parameter is the temperature. By decreasing the difference between the index parameter values of two successive Markov chains, one increases the exchange probability. Indeed, by continuity, it is clear from eq 1 that

$$a(X_n^i, X_n^j) \to 1 \text{ as } T_i \to T_j$$

In replica exchange, the normal course of the simulation is interrupted by swaps between neighboring configurations. The simulation occurs as follows

$$\ldots X_{n-1}^{\ i} \xrightarrow{\text{Met}} X_n^i \xrightarrow{\text{swap}} X_n'^i \xrightarrow{\text{Met}} X_{n+1}^i \ldots$$

where Met refers to a normal Metropolis update, although other sampling techniques may be utilized for the parallel step. If X_n^i is involved in swapping at time n, then X_n^i always denotes the state of the chain i prior to the replica exchange. If needed, the state after exchange will be denoted by a prime sign.

A cyclic swapping strategy consists of a sequence of direct transition matrices $\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_P$ that describe parallel exchanges between some pairs of replicas. The parallel exchanges are attempted at time intervals $k_1, k_2, \ldots, k_P$. We shall call a succession of such P swapping events a cycle. The entire sequence is then repeated periodically every $\mathscr{K} = k_1 + k_2 + \cdots + k_P$ Monte Carlo steps. There are two main requirements that we need to enforce in choosing the transition matrices, with the second one, which ensures ergodicity of the swapping event, being stated later in the present section. The first requirement is that any given matrix $\mathbf{T}_k$ must contain only direct transitions between disjoint pairs of particles, so that any given particle is involved in exchanges with at most one other particle. An example of such a matrix is given by eq 5. On the main diagonal, a matrix $\mathbf{T}_k$ may contain either 1 or some rejection probabilities for direct transitions. The off-diagonal terms are either 0 or the corresponding acceptance probabilities for direct transitions. If two particles i and j are involved in direct transitions during an exchange described by $\mathbf{T}_k$ at some Monte Carlo step n, then the entries $t_{ii}^{(k)}$ and $t_{jj}^{(k)}$ of the matrix $\mathbf{T}_k$ must equal $r(X_n^i, X_n^j)$, whereas the entries $t_{ij}^{(k)}$ and $t_{ji}^{(k)}$ must equal $a(X_n^i, X_n^j)$. The particles i and j are not allowed to participate at any other

exchanges during this particular swapping step.

$$\mathbf{T}_k(X_n^1, X_n^2, ..., X_n^N) = \begin{pmatrix} r(X_n^1, X_n^3) & 0 & a(X_n^1, X_n^3) & 0 & 0 & ... \\ 0 & r(X_n^2, X_n^4) & 0 & a(X_n^2, X_n^4) & 0 & ... \\ a(X_n^1, X_n^3) & 0 & r(X_n^1, X_n^3) & 0 & 0 & ... \\ 0 & a(X_n^2, X_n^4) & 0 & r(X_n^2, X_n^4) & 0 & ... \\ 0 & 0 & 0 & 0 & 1 & ... \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \quad (5)$$

The entries of the direct transition matrices are random variables depending upon the state of the Monte Carlo replicas prior to an exchange. It is convenient to restrict our attention only to the states that are involved in swaps and arrange them as the sequence $Y_1, Y_2, ...$ Accordingly, if $[n]$ is the largest integer strictly smaller than n/P and $\{n\} = n - [n]P$ is the remainder, then

$$Y_n = (X^1_{[n]\mathscr{K}+k_{\{n\}}}, X^2_{[n]\mathscr{K}+k_{\{n\}}}, ..., X^N_{[n]\mathscr{K}+k_{\{n\}}}) \quad (6)$$

for $n = 1, 2, 3, ...$ The collection of future states, from the perspective of somebody that starts observing the simulation at time n, is denoted by

$$F_n = (Y_n, Y_{n+1}, Y_{n+2}, ...) \quad (7)$$

Of course, as far as swapping is concerned, the entire simulation is represented by F_1.

The Monte Carlo simulation is assumed to have equilibrated. Moreover, the transition or the group of Metropolis transitions responsible for the parallel step are assumed to be time-independent. In these conditions, the sequences $Y_n, Y_{n+1}, Y_{n+2}, ...$ are stationary with period P for all $n \geq 1$, that is, they have the same distribution as the shifted sequences $Y_{n+P}, Y_{n+P+1}, Y_{n+P+2}, ...$ The reason that the sequence $Y_1, Y_2, Y_3, ...$ is not stationary with respect to shifting by 1 is that the transition matrices $\mathbf{T}_1, \mathbf{T}_2, ..., \mathbf{T}_P$ are generally different. However, the marginal distributions of $Y_1, Y_1, Y_3, ...$ are the same for an equilibrated Monte Carlo simulation and are given by the stationary distribution $\rho(x_1)\rho(x_2)...\rho(x_N)$. As a consequence, the expected values of the entries of the direct transition matrices $\mathbf{T}_k(Y_n)$ are independent of n, a property already utilized in eqs 3 and 4. We shall denote these common expected values by $\tau_{ij}^{(k)}$, whereas the entire matrix of expected values is denoted by $\mathscr{T}_k$.

Because they are stochastic, in fact doubly stochastic by symmetry, the matrices $\mathbf{T}_k(Y_n)$ may be interpreted as representing transition probabilities of a discrete state temporally inhomogeneous Markov chain $\{\Theta_n(F_1); n \geq 1\}$ with random transition probabilities, acting on the space of indexes $\mathscr{I} = \{1, 2, ..., N\}$. More precisely, considering again the unique decomposition $n = [n]P + \{n\}$, with $[n]$ the largest integer strictly smaller than n/P, the transition probabilities of the Markov chain $\Theta_n(F_1)$ are defined by the equation

$$P[\Theta_{n+1}(F_1) = j | \Theta_n(F_1) = i] = t_{ij}^{(\{n\})}(Y_n) \quad (8)$$

Markov chains with random transition probabilities are collections of temporally inhomogeneous Markov chains, each characterized by transition probabilities depending on the particular course of the simulation, as specified by the values of the random vectors $Y_1, Y_2, Y_3, ...$ that make up F_1. Conditioned on the starting position $\Theta_1(F_1) = i$, the random variable $\Theta_{n+1}(F_1)$ represents the state (the replica index) reached by a configuration that originally started from replica i, given a particular course of the simulation F_1 and given that n swapping events have occurred. The significance of the transition probabilities appearing in eq 8 is as follows. Given a particular course of the simulation F_1, the transition probability $t_{ij}^{(\{n\})}(Y_n)$ represents the probability that a configuration that has reached replica i, after n swapping attempts, ends up in replica j, after yet another attempt. As we shall see, our development builds heavily on this interpretation.

The second requirement that must be enforced in choosing the transition matrices $\mathbf{T}_k$ is that the product

$$\mathscr{T}_P \mathscr{T}_{P-1} \cdots \mathscr{T}_1 \quad (9)$$

of the matrices made up of the expected values of the entries of $\mathbf{T}_k$ must induce an ergodic discrete state Markov chain on the index space $\mathscr{I} = \{1, 2, 3, ...\}$. Ergodicity (or irreducibility) means that starting from any arbitrary index i, we may reach any other state j in a finite number of steps. Clearly, ergodicity of the product matrix given by eq 9 is a natural requirement for any replica exchange strategy, because accessibility from a state to any other state is required. We shall accept without proof that the ergodicity of this product matrix is a sufficient condition to ensure ergodicity of the temporally inhomogeneous Markov chain $\Theta_n(F_1)$. For temporally inhomogeneous Markov chains with random transition probabilities, the condition of ergodicity can be stated as the requirement that starting from any position i and at any time n, the probability to reach any other state j in a finite number of steps is nonzero, for almost all histories of the simulation.

As previously mentioned, the purpose of replica exchange is to reduce the equilibration time of the Monte Carlo chain characterized, say, by the index parameter $T_{\min}$, by using the property of fast equilibration of the chain characterized by the index parameter $T_{\max}$. Of course, many of the intermediate chains do participate and help the lowest temperature chain to equilibrate faster. However, we shall address the worst scenario setting, in which the configurations in the lowest temperature replica have to climb the entire ladder of temperatures in order to get destroyed and be replaced with more favorable configurations. In other words, the implicit assumption is that a configuration receives a status of "energetically unfavorable" just because it originates from the lowest temperature replica. The configuration cannot change its status as the result of the modifications undergone while climbing through the ladder of temperatures. It may only change its status upon arrival at the highest temperature replica.

Consistent with our worst scenario model, a direct measure of the efficiency of the replica exchange method is represented by the fraction f_N of configurations that start in the replica of lowest index 1 and successfully reach the replica of highest index N, during the simulation. For the remainder of this section, our goal is the computation of this quantity, which shall be called the effective fraction. Although the "unfavorable" tag that a configuration from the lowest index replica receives may be justifiable only at the beginning of the simulation, we shall assume that the Markov chain is equilibrated. To parallel similar situations in the chemical literature, the reader is reminded of the classical transition state theory for reaction rates, where the standard model is a steady supply of reactant molecules that are, nevertheless, in thermodynamic equilibrium during the entire reaction. A more stringent assumption, which requires special care, is that the time intervals $k_1, k_2, ..., k_P$ between consecutive replica exchanges are sufficiently large that the

variables Y_1, Y_2, Y_3, ... may be assumed independent. We shall make this assumption at the very end of the section, to obtain formulas that are feasible to evaluate.

For definiteness, let us assume that we are at the swapping time n. Again, $[n]$ represents the largest integer strictly smaller than n/P, whereas $\{n\}$ is the remainder$\{n\} = n - [n]P$. The remainder $\{n\}$ always takes values in the state space $\mathcal{S} = \{1, 2, ..., N\}$. For any $n \geq 1$, the matrix $\mathbf{T}_{\{n\}}$ contains the transition probabilities of the Markov chain $\Theta_n(F_1)$ at time n. The probability that the configuration in the lowest temperature replica jumps directly into a state $j \in \{2, 3, ... N\}$ is $t_{1j}^{(\{n\})}(Y_n)$. This is so because the acceptance/rejection step for exchanges between the replicas 1 and j is the result of a random variable that only depends on the current state of the two replicas (according to eq 1) and is independent of the other random variables controlling the course of the simulation. As already stated, given a particular course of the simulation F_1, the time evolution of the index of any tagged configuration is described by the Markov chain $\Theta_n(F_1)$. The configuration may also remain in the state $j = 1$ with probability $t_{11}^{(\{n\})}$. If so, then it will get a second chance to thermalize at the next swapping time.

A configuration starting in some position $j \geq 2$ at the next swapping time $n + 1$ may return to the configuration 1 without first passing through the configuration N. Let $Q_j(F_{n+1})$ be the probability of this event. Notice that consistent with the Markov property of the Monte Carlo simulation, $Q_j(F_{n+1})$ depends only on the future swapping states. The configuration may also go to the replica $j = N$ without first passing through the $j = 1$. We let $P_j(F_{n+1})$ be the probability of the second event. Clearly, $P_N(F_{n+1}) = 1$ and $Q_N(F_{n+1}) = 0$. More generally, ergodicity requires that

$$Q_j(F_{n+1}) + P_j(F_{n+1}) = 1 \tag{10}$$

that is, starting from j at any time $n + 1$, the configuration eventually hits either the lowest index or the highest index configurations, with probability one. Indeed, for instance, a nonvanishing probability for the event that starting at j we may never reach 1 contradicts the ergodicity of the Markov chain $\Theta_n(F_1)$.

The event that a configuration starting at a position $j \geq 2$ returns to replica 1 without first passing through replica N is not an effective event. In this case, the configuration will get a second chance to jump out of replica 1 and reach the highest index replica in future exchanges. The reader should notice that the configuration will eventually hit replica 1 at some time $m + n$ in the future, by ergodicity. Even if it reaches replica N after that, it should not be counted as effective at the current time n, because we risk to double count the single event of reaching N: once for the present time n and again for the future time $m + n$. The complementary event that the configuration reaches replica N without first passing through replica 1 is effective. Summing over all intermediate states $j = 2$, it follows that the probability that the configuration exits replica 1 at time n and then reaches replica N, without passing through replica 1 first, is given by the formula

$$\sum_{j=2}^{N} t_{1j}^{(\{n\})}(Y_n) P_j(F_{n+1}) \tag{11}$$

If the entire Monte Carlo simulation consists of N_S swapping events, then the fraction of configurations that successfully change their status from unfavorable to favorable is

$$\frac{1}{N_S} \sum_{n=1}^{N_S} \sum_{j=2}^{N} t_{1j}^{(\{n\})}(Y_n) P_j(F_{n+1})$$

Remembering that the collection of variables Y_1, Y_2, Y_3, ... is stationary with period P and invoking the ergodic theorem, we learn that the above expression converges to

$$f_N = \frac{1}{P} \sum_{n=1}^{P} \sum_{j=2}^{N} \mathbb{E}[t_{1j}^{(n)}(Y_n) P_j(F_{\{n+1\}})] \tag{12}$$

almost surely.

Although the effective fraction f_N can be computed during the Monte Carlo simulation itself, the procedure is rather tedious. To obtain an expression for f_N that is easier to evaluate, we shall make the simplifying assumption that the time intervals $k_1, k_2, ..., k_P$ between consecutive replica exchanges are so large that the variables Y_1, Y_2, Y_3, ... are independent. Replacing the expected values in eq 12 byproducts of expected values, we obtain

$$f_N = \frac{1}{P} \sum_{n=1}^{P} \sum_{j=2}^{N} \mathbb{E}[t_{1j}^{(n)}(Y_n)]\, \mathbb{E}[P_j(F_{\{n+1\}})] = \frac{1}{P} \sum_{n=1}^{P} \sum_{j=2}^{N} \tau_{1j}^{(n)}\, p_j^{(\{n+1\})} \tag{13}$$

where we have introduced the notation $p_j^{(n)} = \mathbb{E}[P_j(F_n)]$ for the average probabilities that a configuration starting in the position $j \geq 2$ at time $n \in \{1, 2, ..., P\}$ first hits replica N.

We are now ready to establish a recurrence relation between the average hitting probabilities $p_j^{(n)}$. If we start from state $j = 2, 3, ..., N - 1$ at time $n \leq P$, we may jump to 1 with probability $t_{j1}^{(n)}(Y_n)$. In this case, the conditional probability to reach N before reaching 1 is zero. We may also jump to N with probability $t_{jN}^{(n)}(Y_n)$. In this case, the conditional probability to first reach N is 1. Finally, we may jump to $i = 2, 3, ..., N - 1$ with probability $t_{ji}^{(n)}(Y_n)$. The conditional probability to first reach N from the new position, while starting at time $n + 1$, is $P_i(F_{n+1})$. Summing over all intermediate states i, we must obtain the conditional probability to first reach N while starting from j at time n. Therefore,

$$P_j(F_n) = t_{jN}^{(n)} + \sum_{i=2}^{N-1} t_{ji}^{(n)}(Y_n) P_i(F_{n+1}) \tag{14}$$

for all $2 \leq j \leq N - 1$ and $1 \leq n \leq P$. Taking the expected values of the left- and right-hand sides and again using independence and stationarity with period P, we obtain the system of equations

$$p_j^{(n)} = \tau_{jN}^{(n)} + \sum_{i=2}^{N-1} \tau_{ji}^{(n)}\, p_i^{(\{n+1\})} \tag{15}$$

for all $2 \leq j \leq N - 1$ and $1 \leq n \leq P$. Of course, $p_N^{(n)} = 1$ for all $1 \leq n \leq P$.

The system of equations obtained in the preceding paragraph may or may not uniquely determine the average hitting probabilities. For example, let us imagine the swapping strategy in which direct exchanges are only attempted between replica 1 and the rest of the replicas. In this case, $\tau_{ji}^{(n)} = \delta_{ji}$ for all i, j

≥ 2. Any arbitrary hitting probabilities $p_j^{(n)}$ satisfy eq 15. However, in this situation, it is clear that $p_j^{(n)} = 0$ for all $2 \leq j \leq N-1$. More generally, the system of equations given by eq 15 is to be solved iteratively starting with the initial guess $p_j^{(1)} = 0$, for all $2 \leq j \leq N-1$. At the next step of iteration, with the help of eq 15, we compute the values $p_j^{(n)}$ for $n = P$ and $2 \leq j \leq N-1$. The procedure is repeated down to $n = 1$, and new cycles of computation are performed until self-consistency is attained. The resulting solution is used together with eq 13 to compute the effective fraction f_N. To conclude, if we assume that the Monte Carlo states involved in exchanges are independent, then all that is needed for the computation of the effective fraction f_N are the expected values $\tau_{ij}^{(n)}$ of the direct transition matrices.

III. More on the Computation of the Effective Fraction

In the preceding section, we have argued that the effective fraction f_N constitutes a measure of the swapping efficiency that is more adequate than, for example, the values of the acceptance and rejection probabilities alone. The effective fraction represents the fraction of configurations originating from the lowest index replica that successfully climb the ladder of index parameter values and reach the highest index replica, during the Monte Carlo simulation. If the replica exchanges are performed sufficiently infrequently that the exchanged configurations may be considered statistically independent, the effective fraction is solely determined by the values of the average acceptance/rejection exchange probabilities, as explained below.

Using the values of the average acceptance/rejection probabilities determined during the Monte Carlo simulation, one constructs the transition matrices $\{\mathscr{T}_k;\ 1 \leq k \leq P\}$ that make up a swapping cycle. For example, a popular swapping strategy involves pairs of neighboring replicas only. In the first step, parallel swaps are attempted between the pairs (1, 2), (3, 4), etc. At the second step, parallel swaps are attempted between the pairs (2, 3), (4, 5), etc. The whole exchange cycle is then repeated periodically. The entries $\tau_{ij}^{(1)}$ and $\tau_{ij}^{(2)}$ of the two transition matrices are those from the arrays

$$\mathscr{T}_1 = \begin{pmatrix} r_{12} & a_{12} & 0 & 0 & \dots \\ a_{12} & r_{12} & 0 & 0 & \dots \\ 0 & 0 & r_{34} & a_{34} & \dots \\ 0 & 0 & a_{34} & r_{34} & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \tag{16}$$

and

$$\mathscr{T}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & r_{23} & a_{23} & 0 & \dots \\ 0 & a_{23} & r_{23} & 0 & \dots \\ 0 & 0 & 0 & r_{45} & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \tag{17}$$

respectively.

The selection of the swapping strategy must be done in such a way that (i) any given replica must not be involved in more than one exchange at any particular swapping time and (ii) the Markov chain induced by the product matrix

$$\mathscr{T}_P \mathscr{T}_{P-1} \cdots \mathscr{T}_1$$

must be ergodic. These two properties are obviously respected by the two transition matrices given by eqs 16 and 17, provided that none of the acceptance probabilities $a_{12}, a_{23}, \dots, a_{N-1,N}$ is zero.

Next, we compute the average probabilities $p_j^{(n)}$ that a configuration starting at the position $j = 2, 3, \dots, N-1$ and at the simulation time $n = 1, 2, \dots, P$ reaches the replica N without hitting replica 1 first. To do so, we solve iteratively the system of equations

$$p_j^{(n)} = \tau_{jN}^{(n)} + \sum_{i=2}^{N-1} \tau_{ji}^{(n)}\, p_i^{(\{n+1\})} \tag{18}$$

for $j = 2, 3, \dots, N-1$ and $n = 1, 2, \dots, P$. The reader is reminded that

$$\{n+1\} = \begin{cases} 1, & \text{if } n = P \\ n+1, & \text{otherwise} \end{cases} \tag{19}$$

The iteration is started with $n = P$, using the initial guess $p_j^{(1)} = 0$ for all $j = 2, 3, \dots, N-1$. After the hitting probabilities $p_j^{(n)}$ for $n = P$ are computed, the procedure is repeated for $n = P-1$, down to $n = 1$. New cycles of computation are then initiated, until self-consistency is attained.

Once the solution of eq 18 is established, the effective fraction f_N is computed with the help of the formula

$$f_N = \frac{1}{P} \sum_{n=1}^{P} \sum_{j=2}^{N} \tau_{1j}^{(n)}\, p_j^{(\{n+1\})} \tag{20}$$

The values $p_N{}^{(n)}$ needed in eq 20 are equal to 1, by definition.

If the acceptance probabilities for exchanges between neighboring replicas are tuned to be equal to the common value $a_{i,i+1} = a$, then an explicit formula for f_N can be derived in the case of the two-step swapping strategy presented earlier in the section. The formula reads

$$f_N(a) = \frac{a}{2} \frac{1}{(1-a)N + 2a - 1} \tag{21}$$

and can be proven by induction as follows. Jumps out of the replica 1 cannot be realized during the action of the transition matrix $\mathscr{T}_2$. When the transition matrix $\mathscr{T}_1$ acts, there is a direct jump only in replica $j = 2$, a jump that happens with probability a. Because $P = 2$, eq 20 reduces to

$$f_N = \frac{a}{2} p^{(2)}{}_2(N)$$

where we have explicitly stated that the hitting probability $p_2^{(2)}$ depends on N.

It remains to establish that a particle starting at the position 2 during the second swapping event reaches replica N before hitting replica 1 with the probability

$$p_2^{(2)}(N) = \frac{1}{(1-a)N + 2a - 1} \tag{22}$$

Trivially, $p_2^{(2)}(2) = 1$, so eq 22 is verified for $N = 2$. Let $N \geq 3$. All events that eventually reach N without first hitting 1 also hit $N-1$ without first hitting 1. The probability that the second, larger class of events happens is $p_2^{(2)}(N-1)$. The events always arrive at replica $N-1$ through an exchange between replicas $N-2$ and $N-1$. At the next exchange, the configurations in $N-1$ may jump into N directly, with a probability a. These events are effective. They may also get rejected with a probability of $1-a$. In this second case, they may hit 1 before hitting N, an event that happens with probability $p_2^{(2)}(N)$ by the symmetry of the problem, or may hit N before

hitting 1 with the remaining probability, $1 - p_2^{(2)}(N)$, a case in which they are again effective. Combining everything, we end up with the recursive formula

$$p_2^{(2)}(N) = p_2^{(2)}(N-1)\{a + (1-a)[1 - p_2^{(2)}(N)]\}$$

The reader may easily check that eq 22 does, indeed, verify the recursion. The proof is concluded by the principle of induction. Perhaps the most important step in the proof is the observation that the probability to hit 1 without first hitting N, while starting from the replica $N - 1$, is the same as the probability to hit N without first hitting 1, while starting from replica 2, by symmetry.

Equation 21 shows that the effective fraction decreases with the increase of the number of replicas, an expected result. However, for a given system, the number of replicas necessary to achieve a given acceptance probability is also a function of a. Therefore, the effective fraction becomes a function depending solely on the acceptance probability a. We shall use this observation in the context of parallel tempering in the canonical ensemble to study the optimal value for the acceptance probability.

IV. Optimal Acceptance Probabilities for Parallel Tempering in the Canonical Ensemble

If the length of a replica exchange Monte Carlo simulation consists of N_{MC} parallel steps and if complete cycles of swaps are performed every $\mathscr{K}$ steps, with P swapping events per cycle, then the total number of effective swapping events is

$$N_{ef} = f_N \frac{N_{MC} P}{\mathscr{K}} \tag{23}$$

where f_N is the effective fraction and N is the number of replicas. The total number of effective swapping events represents a direct measure of the quality of the exchange. We believe that it is the correct number to be reported for any replica exchange Monte Carlo simulation.

The computational effort to obtain N_{ef} effective exchanges is proportional to the number of replicas, more precisely to $N_{MC}N$. Thus, the efficiency of the swapping strategy is measured by the ratio

$$\frac{N_{MC} N}{N_{ef}} = \frac{N}{f_N} \frac{\mathscr{K}}{P} \tag{24}$$

which gives the computational time per effective swapping event. Among all swapping strategies that ensure the same number of effective swaps N_{ef}, the most efficient ones are those that maximize the fraction N/f_N.

In this section, we study the optimal acceptance probability for parallel tempering simulations in the canonical ensemble. We work with the two-step swapping strategy considered in the preceding section. We also assume that the acceptance probabilities for swaps between neighboring replicas have been tuned to be equal to the common value $a_{i,i+1} = a$. Our goal is to establish the acceptance probability that maximizes the efficiency of the replica exchange method, as measured by the ratio

$$N/f_N = 2N[(1-a)N + 2a - 1]/a \tag{25}$$

We first study the case of high-dimensional harmonic oscillators, for which an analytical formula for the dependence of the number of replicas N with the acceptance probability is available. Then, we perform a numerical study for a 13-particle Lennard−Jones cluster.

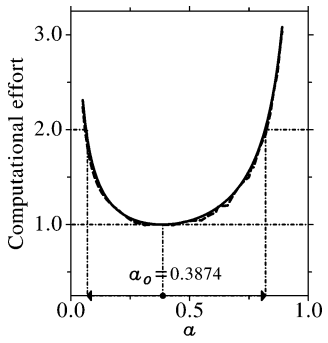

Figure 1. Computational costs per effective swapping event as functions of the acceptance probability. The costs are relative to their values at $a_o = 0.3874$. The theoretical curve for high-dimensional oscillators (solid line) and the efficiency curve obtained experimentally for the LJ_{13} cluster (dashed line) are purposefully superimposed to demonstrate their likeness. Also shown is the interval of acceptance probabilities for which the increase in the computational effort is at most 100%, relative to the optimal value.

In the case of high dimensional oscillators, the number of replicas necessary to achieve a given acceptance probability a is given by the formula[14]

$$N(a) \approx t/\mathrm{erf}^{-1}(1-a) \tag{26}$$

where

$$t = (2d)^{1/2} \ln(\beta_{\max}/\beta_{\min})/4$$

is a parameter independent of the acceptance probability a. The function $\mathrm{erf}^{-1}(x)$ represents the inverse of the error function.

Substituting the expression of $N(a)$ in eq 25 and using that $(1-a)N(a) \gg 2a - 1$, we obtain the following asymptotic formula

$$N/f_N \approx 2t^2(1/a - 1)[\mathrm{erf}^{-1}(1-a)]^{-2} \tag{27}$$

which is valid for large values of t. We readily see that the minimum of N/f_N is independent of the value of the parameter t. This minimum, which is unique, as apparent from Figure 1, can be determined numerically and has the approximate value of $a_o = 0.3874$. Thus, the optimal acceptance probability for harmonic oscillators is about 38.74%.

Let us analyze a little more closely the issue of efficiency by studying an example. Assume that for a certain value of t, the number of replicas necessary to achieve the optimal acceptance probability $a_o = 0.3874$ is N. If, to the contrary, we choose to run our simulation with an acceptance of $a = 0.666$, then eq 26 demands that the number of replicas be $2N$. Thus, the computational effort is twice as large. The number of effective swapping events is

$$N_{ef} = f_N \frac{N_{MC} P}{\mathscr{K}} = \frac{0.305}{t} \frac{N_{MC} P}{\mathscr{K}}$$

If we apply the optimal strategy, the apparent number of effective replicas is

$$N_{ef} = \frac{0.193}{t} \frac{N_{MC} P}{\mathscr{K}}$$

a smaller number. However, the optimal strategy utilizes only half of the number of processors. We can utilize the other half

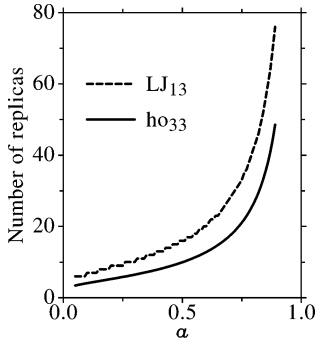

Figure 2. Number of replicas necessary to achieve a prescribed acceptance probability a. The number of replicas necessary for the LJ_{13} cluster (dashed lines) is roughly 70% larger than for a corresponding 33-dimensional harmonic oscillator (solid line). The six missing degrees of freedom, up to $3 \times 13 = 39$, correspond to translations and rotations of the cluster. Because they do not affect the exchange acceptance probabilities, these degrees of freedom are not considered.

to run a second, statistically independent, simulation. The number of effective swaps for the same computational effort becomes twice as large, that is,

$$N_{ef} = \frac{0.386}{t} \frac{N_{MC} P}{\mathscr{K}}$$

The ratio $0.386/0.305 = 1.267$ measures the loss of efficiency when the acceptance $a = 0.666$ is utilized. It says that, if the nonoptimal strategy is utilized, one must invest 26.7% more computational effort to obtain the same number of effective swapping events.

The reader may argue that 26.7% is not always a critical saving. It is common practice to attempt a modification of the algorithm only if the computational time can be reduced to half or less. From Figure 1, where we present the computational costs N/f_N relative to their optimal values, we see that the efficiency remains good for a large range of acceptance probabilities. In fact, the computational costs increase by more than 100% of the optimal value only if the acceptance ratio falls outside the interval (7%, 82%). Thus, except for certain special simulations for which a correct allocation of computational resources is critical, there seems to be little to be gained by tuning the acceptance probabilities to the optimal value, as long as the acceptance probabilities are maintained in the (7%, 82%) interval.

We have also computed an efficiency curve for the Ne realization of the Lennard−Jones cluster made up of 13 particles. The physical parameters of the system and the Monte Carlo simulation technique are the same as in ref 14. In the range of temperatures employed, 3−30 K, the system undergoes two phase transitions: a solid−liquid transition at about 10 K and a liquid−gas transition at about 18 K, respectively. Thus, the simulation can be considered representative of what a chemical physicist is likely to encounter in practice. In the interval of the temperature specified, we have run different parallel tempering simulations for different numbers of replicas, using the two-step swapping strategy. For each number of replicas, we have determined the schedule of temperatures that makes all acceptance probabilities constant, using the adaptive procedure of Hukushima and Nemoto.[3]

The resulting acceptance probabilities have been utilized to construct a function that expresses the dependence of the number of replicas with the acceptance probability. This function is plotted in Figure 2, where we also plot the number of replicas for a 33-dimensional harmonic oscillator, for comparison purposes. Because of the increase in the heat capacity at the two phase−transition points, the number of replicas necessary to attain a prescribed acceptance probability are roughly 70% larger, in the case of the LJ_{13} cluster. Nevertheless, the two curves are quite similar in shape, an early hint that the shapes of the efficiency curves are also similar, upon normalization. Because $N(a)$ is a discontinuous function (the number of replicas can only be an integer), we have decided to linearly interpolate the values of $N(a)$, so as to obtain a continuous dependence and avoid certain oscillations in the plots. The interpolated values have been utilized together with eq 25 to produce an efficiency curve. As readily seen from Figure 1, the computational costs, relative to the optimal value, are almost identical to those for high dimensional oscillators. Therefore, we expect that the conclusions drawn for harmonic oscillators remain generally true of many of the systems with continuous distributions that may be encountered in practical applications.

V. Conclusions

Contrary to the general belief, the values of the acceptance probabilities alone do not represent a measure of the quality of the simulation. Rather, they measure the efficiency of the simulation. The quality of the simulation is more adequately measured by the number of effective swapping events, which is the product between the total number of swapping events and the effective fraction. The effective fraction, the computation of which has been detailed in section III, represents the expected probability that a configuration originating from the lowest index replica eventually reaches the highest index replica. We believe the scientists in the field are better served if they utilize the total number of effective swapping events as a measure of the quality of the simulation. The invested computational effort is minimal, and the result is also useful for other purposes as, for instance, in studying the relative efficiency of alternative swapping strategies. We have provided an example of such a study, by determining the optimal acceptance probability for parallel tempering simulations in the canonical ensemble.

For parallel tempering simulations of systems with continuous distributions, the acceptance probability that ensures the highest efficiency (the highest number of effective swaps per unit of computational cost) is around 38.74%, the optimal acceptance probability for high dimensional oscillators. We hope that this has been proven convincingly enough by the study that we have performed on the swapping efficiency for the 13-atom Lennard−Jones cluster. Another interesting finding is that the acceptance probability may take any value between 7 and 82%, with an increase of at most a factor of 2 in the computational effort.

For those simulations where an efficient allocation of computational resources is required, as is the case when parallel tempering is utilized as a stochastic minimization technique, an acceptance probability close to 38.74% should be employed. The computational resources that are saved can be better utilized to improve the quality of the sampling for the parallel step or to run additional statistically independent simulations. In cases where the number of intermediate temperatures is larger than the optimal number because it is dictated by other considerations (for instance, to ensure that a temperature plot is sufficiently detailed), the efficiency can be increased by employing swapping strategies that propose more aggressive moves, to the next-to-nearest neighbor, for example.

An important assumption in the present development has been that the sampling responsible of the parallel step (Metropolis or otherwise) effectively decorrelates the states involved in successive swapping attempts. Of course, this is unlikely to be

true of most simulations that are encountered in practice. For instance, the build-up of correlation in the parallel step can be significant for simulations in the grand canonical ensemble, where the insertion of a new particle is frequently rejected at high densities. Because the state of the expanded ensemble after a swap may resemble the old state with a high probability, the fraction of effective moves is in reality smaller. The definition introduced in the present work is an accurate representation of the exchange efficiency, as this relates to the overall convergence rate of the simulation, only in the limit that the replica exchange is so infrequent that the states decorrelate to a good degree between two succesive swaps. Further research is necessary in order to understand how this correlation affects the effective fraction or the optimal acceptance probabilities. Nevertheless, the computational resources saved by employing an optimal swapping strategy may find a better utilization in reducing the correlation of the Markov chain responsible for the parallel step, for example, by increasing the number of intermediate Metropolis steps.

Acknowledgment. This work was supported in part by the National Science Foundation Grant CHE-0345280, the Director, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division, U.S. Department of Energy under Contract DE AC03-65SF00098, and the U.S.−Israel Binational Science Foundation Award 2002170.

References and Notes

(1) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607.
(2) Geyer, C. J. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*; Keramigas, E. M., Ed.; Interface Foundation: Fairfax, 1991; p 156.
(3) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604.
(4) Neirotti, J. P.; Calvo, F.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **2000**, *112*, 10340.
(5) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.
(6) Calvo, F.; Neirotti, J. P.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **2000**, *112*, 10350.
(7) Yan, Q.; de Pablo, J. J. *J. Chem. Phys.* **1999**, *111*, 9509.
(8) Andricioaei, I.; Straub, J. E. *J. Chem. Phys.* **1997**, *107*, 9117.
(9) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042.
(10) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058.
(11) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. M.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(12) Kalos, M.; Whitlock, P. *Monte Carlo Methods*; Wiley-Interscience: New York, 1986.
(13) Grinstead, C. M.; Snell, J. L. *Introduction to Probability: Second Revised Edition*; AMS: Providence, 1997.
(14) Predescu, C.; Predescu, M.; Ciobanu, C. V. *J. Chem. Phys.* **2004**, *120*, 4119.