Acceleration of Convergence to the Thermodynamic Equilibrium by Introducing Shuffling Operations to the Metropolis Algorithm of Monte Carlo Simulations

Hongzhi Sun[†] and István P. Sugár*

Departments of Biomathematical Sciences and Physiology & Biophysics, Mount Sinai School of Medicine, New York, New York 10029

Received: April 22, 1996; In Final Form: December 30, 1996[⊗]

This paper introduces a broad class of operations, called "shuffling trials", used to design nonphysical pathways. It is shown that in general the equilibrium distribution of a system can be attained when physically possible pathways are interrupted regularly by nonphysical shuffling trials. Including properly chosen shuffling trials in the commonly applied Metropolis algorithm often considerably accelerates the convergence to the equilibrium distribution. Shuffling trials are usually global changes in the system, sampling efficiently every set of the metastable configurations of the system. Shuffling trials are generated by symmetric stochastic matrices. Since ergodicity is not a required property, it is particularly easy to construct these matrices in accordance with the specificity of the system. The design, application, and efficiency of the shuffling trials in Monte Carlo simulations are demonstrated on an Ising model of two-dimensional spin lattices.

I. Introduction

Monte Carlo methods are designed to simulate stochastic processes of complex systems. In simulating a process, physically possible pathways are realized many times in order to obtain the time dependent frequency distribution function over the configurations of the system. According to the general theorems of Markov processes, the unique stationary distribution of the system is guaranteed to be attained as the time goes to infinity under conditions given in the Method section.[1,2] In practice, however, during the simulations the system is frequently trapped in a set of metastable configurations and the stationary distribution cannot be attained within a feasible computation time.

It is not necessary to stick to the physically possible pathways if only the stationary distribution (*e.g.* equilibrium distribution) of the system is of interest. In this specific case any pathway resulting in the stationary distribution of the system is permissible.

In this paper a broad class of stochastic operations, called "shuffling trials", is introduced. It is pointed out that the equilibrium distribution of a system can be attained not only by a Metropolis−Monte Carlo algorithm but also by a Metropolis−Monte Carlo algorithm that is regularly interrupted by shuffling trials. By using properly designed shuffling trials, the equilibrium distribution can be attained much faster.

Recently several specific methods were proposed for designing nonphysical pathways in order to obtain the stationary distribution of spin lattices within reasonable computation time.[3−5] Each method is a variation of the Swendsen−Wang[4] cluster algorithm, which changes the states of a cluster of spins simultaneously. The algorithm developed by Kandel *et al.*[3] for the Ising lattice can also be considered as an example of the application of shuffling, where a standard Metropolis−Monte Carlo algorithm was regularly interrupted by a Swendsen−Wang cluster algorithm.[4] In our terminology the cluster algorithm is a type of shuffling trial, resulting in global, nonphysical changes in the system.

* Corresponding author.

[†] Current address: Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853.

[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1997.

II. Method

A. General Formulation of Metropolis−Monte Carlo Simulation Using Markov Chain. This section is a reiteration of some concepts and results of the Markov chain theory in order to make the paper more readable and self-contained.

On the Conditions of Correct Simulation. Let $i = 1, 2, ..., W$ index the configurations of a system. The probability distribution over these configurations can be given by a row vector $\bar{\pi}$, where the ith element of the vector, $\bar{\pi}_i$, is the probability of finding the system in configuration i.

During a Monte Carlo simulation a stochastic operation P is repeatedly applied on the system, where P is a $W \times W$ stochastic matrix (*i.e.*, the sum of the elements in each row is 1) and the matrix element P_{ij} is the transition probability from configuration i to configuration j. The operation generates a *correct simulation* if for any initial distribution π^0 the distribution of the system approaches the equilibrium distribution π, *i.e.*

$$\lim_{n\to\infty} \pi^0 P^n = \pi \tag{1}$$

where $\pi = (\pi_1, \pi_2, ...\pi_W)$. In statistical mechanics, the equilibrium distribution π depends on the type of the contact between the system and its surrounding and also on the definiton of the system's configurations.[9] For example, if there is only thermal contact and each quantum mechanical state of the system corresponds to one configuration, then

$$\pi_i = \mathrm{e}^{-E_i/kt} \Big/ \sum_{j=1}^{W} \mathrm{e}^{-E_j/kt} \tag{2a}$$

where E_i is the energy of the ith quantum state, t is the equilibrium temperature of the system, k is the Boltzmann constant, and W is the number of quantum states. For statistical mechanical purposes the description of the system is frequently less detailed. In this case the quantum states of the system are grouped together in such a way that configuration i corresponds to g_i quantum states, all of which have approximately the same energy E_i. Then

S1089-5647(96)01161-3 CCC: $14.00

$$\pi_i = g_i e^{-E_i/kt} \Big/ \sum_{j=1}^{W} g_j e^{-E_j/kt} \tag{2b}$$

where g_i is the degeneracy of the ith configuration and W represents the total number of configurations. The following condition of correctness is equivalent to the condition given by eq 1 (see Appendix A.1):

$$\lim_{n\to\infty} P^n = B \tag{3}$$

where B is the equilibrium distribution matrix (each row of matrix B is equal to the row vector π).

Transition Probability Matrix of Metropolis−Monte Carlo Simulations. There are two basic steps in a Metropolis−Monte Carlo simulation.

Step I: Candidate Configuration Generation. If the system is in configuration i, the next candidate configuration j is generated by probabilistic rules. Commonly these rules of candidate configuration generation allow local changes in the system that are physically possible within a short time. Here we mention two frequently applied rules: (1) randomly pick a molecule of the system and change its intramolecular state (this rule is the basis of the so-called Glauber method[6]); (2) randomly pick a pair of nearest neighbor unlike molecules and exchange them (this rule is the basis of the Kawasaki method[7]).

Step II: Decision Making (Metropolis Criterion). The candidate configuration j is accepted with probability $\pi_j/(\pi_j + \pi_i)$; otherwise, configuration i is kept. It is important to note that the probability of acceptance can be defined other ways provided that the principle of detailed balance is satisfied.[8]

One repeat of step I and step II together defines one trial T. According to Markov chain theory,[10] step I corresponds to candidate configuration generation matrices $M^{(1)}$, $M^{(2)}$, ... Each of these matrices of size $W \times W$ represents a probabilistic rule for generating the next candidate configuration. For example, according to the kth rule, if the system is in configuration i, the probability of generating configuration j as a candidate configuration is $M_{ij}^{(k)}$, the (i,j)th element of the $M^{(k)}$ matrix. Two trials $T_{(k)}$ and $T_{(l)}$ are different if they are constructed by different candidate configuration generation matrices $M^{(k)}$ and $M^{(l)}$. Therefore, one trial corresponds exactly to one probabilistic rule in step I. The algorithm of the Monte Carlo simulation repeats the trials according to a certain pattern $T_{(\nu_1)}T_{(\nu_2)}...T_{(\nu_h)}$, called the *trial pattern*, which generates a Markov chain. The subsequent application of the rules of candidate configuration generation according to this pattern leads to the definition of a resultant generating matrix $\hat{M}$:

$$\hat{M} = M^{(\nu_1)}M^{(\nu_2)}...M^{(\nu_h)} \tag{4}$$

In the case of Metropolis−Monte Carlo simulation if the candidate configuration generation matrices satisfy the following three properties, then the algorithm is correct.

Normality

$$\sum_{j=1}^{W} M_{ij}^{(k)} = 1 \quad \text{for } i = 1, 2, ...W \text{ and } k = 1, 2, ... \tag{5}$$

Symmetry

$$M_{ij}^{(k)} = M_{ji}^{(k)} \quad \text{for } i, j = 1, 2, ... W \text{ and } k = 1, 2, ... \tag{6}$$

Ergodicity. The system can go from any initial configuration to any other configuration by repeatedly applying matrix $\hat{M}$ a finite number of times.

In the Markov chain theory[10] a trial $T_{(k)}$ has a mathematical equivalent, the transition probability matrix, $M^{E(k)}$. This matrix can also be called a configuration selection matrix, where the k superscript refers to the kth rule of candidate configuration generation in step I. The (i,j)th element of this $W \times W$ square matrix is

$$M_{ij}^{E(k)} = M_{ij}^{(k)} \frac{\pi_j}{\pi_j + \pi_i} \quad \text{for} \quad i \neq j$$
$$M_{ii}^{E(k)} = 1 - \sum_{j\neq 1}^{W} M_{ij}^{E(k)} \tag{7}$$

The subsequent applications of transition probability matrices according to the *trial pattern* leads to the definition of the resultant transition probability matrix $\hat{M}^E$:

$$\hat{M}^E = M^{E(\nu_1)}M^{E(\nu_2)}...M^{E(\nu_h)} \tag{8}$$

The Metropolis−Monte Carlo Algorithm Generates Correct Simulations. In this subsection we give a rigorous proof that the resultant transition probability matrices, $\hat{M}^E$, satisfy the condition of correct simulations given by eq 3. (Note that the correctness of our shuffling algorithm is based partially on this proof too.) The proof is based on the following propositions.

Proposition 1: Commutativity. It follows from the *normality* and *symmetry* of the candidate configuration generation matrices that the transition probability matrices $M^{E(k)}$ commute with the equilibrium distribution matrix B (Appendix A.2):

$$BM^{E(k)} = M^{E(k)}B = B \tag{9}$$

By using this relationship for every k, one can get a similar relationship for the resultant transition probability matrix, $\hat{M}^E$:

$$B(\hat{M}^E) = (\hat{M}^E)B = B \tag{10}$$

Note that eq 10 is equivalent to the *principle of detailed balance*.[10]

Proposition 2: Positivity. There is a positive integer l such that $(\hat{M}^E)^m$ is positive; that is, every matrix element is positive for $m \geq l$.

This proposition follows from the ergodicity of the matrix $\hat{M}$ and from the fact that at least one of the diagonal elements of $\hat{M}^E$ is positive[11] (Appendix A.3).

Proposition 3: The Limit Property. If A is a positive stochastic matrix, then the following limit exits:

$$\lim_{n\to\infty}(A)^n = C \tag{11}$$

In the limit matrix C the rows are distributions and are identical with each other.[12]

Now we can prove the correctness of the Metropolis−Monte Carlo simulations. According to proposition 2 $(\hat{M}^E)^m$ is a positive matrix for every $m \geq l$, and thus proposition 3 is applicable for $A = (\hat{M}^E)^m$:

$$\lim_{n\to\infty}(\hat{M}^E)^n = C \tag{12}$$

Let us multiply eq 12 with the equilibrium distribution matrix

and apply repeatedly the relationship of proposition 1 ($B\hat{M}^E = \hat{M}^E B = B$):

$$BC = B\lim_{n\to\infty}(\hat{M}^E)^n = \lim_{n\to\infty} B(\hat{M}^E)^n = \lim_{n\to\infty} B(\hat{M}^E)^{n-1}$$
$$= ... = \lim_{n\to\infty} B(\hat{M}^E) = \lim_{n\to\infty} B = B \quad (13)$$

This shows that $B = BC$. On the other hand, $BC = C$ is also true, because all elements in a column of C are identical, and B is a stochastic matrix where the sum of every row is 1. According to these two relationships, we get $C = B$ and from eq 12.

$$\lim_{n\to\infty}(\hat{M}^E)^n = B \quad (14)$$

Thus, the condition of correct simulations given by eq 3 is satisfied.

B. Introducing Shuffling Trials into the Metropolis−Monte Carlo Algorithm. *Shuffling Trials.* Shuffling trials are stochastic operations that help us to design Monte Carlo algorithms that converge faster to the equilibrium distribution of the system than the conventional Metropolis−Monte Carlo algorithms.

In the conventional algorithms the probabilistic rules of candidate configuration generation allow *local changes* in the system that are physically possible within a short time. However, by using the trials generated by these rules, the configurations are often strongly correlated within long segments of the generated Markov chain; that is, the system is frequently trapped in metastable states for long periods of the simulation. Because of this, the equilibrium distribution of the system is often not attainable within a feasible computation time. The probability of long, strongly correlated segments in the Markov chain can be reduced by performing trials generated by using probabilistic rules of candidate configuration generation which allow global, nonphysical, changes in the system. By doing this, we give up simulating properly the dynamics of the system in order to get the equilibrium distribution within a reasonable time. The rules for generating *global changes* should be specifically designed for each system in order to efficiently sample every set of the metastable configurations (see examples such as phase transition in a spin lattice, in this paper, and DNA strand separation[13]).

We generate global changes by means of the shuffling generating matrices $S^{(1)}$, $S^{(2)}$, ... These are $W \times W$ stochastic, symmetric matrices, like $M^{(1)}$, $M^{(2)}$, ..., but ergodicity is not required. With these simple properties it is particularly easy to design shuffling generating matrices in accordance with the specificity of the system.

A candidate configuration, generated by a shuffling $S^{(k)}$, is accepted or rejected according to the *Metropolis criterion.* In this way a shuffling trial $U_{(k)}$ is realized. The (i,j)th element of the respective transition probability matrix $S^{E(k)}$ is

$$S_{ij}^{E(k)} = S_{ij}^{(k)}\frac{\pi_j}{\pi_i + \pi_j} \quad \text{for} \quad i \neq j$$
$$S_{ii}^{E(k)} = 1 - \sum_{j\neq i}^{W} S_{ij}^{E(k)} \quad (15)$$

Subsequent application of several shuffling trials according to a certain pattern $U_{(u_1)}U_{(u_2)}...U_{(u_g)}$ leads to the definition of the resultant transition probability matrix $\hat{S}^E$,

$$\hat{S}^E = S^{E(u_1)}S^{E(u_2)}...S^{E(u_g)} \quad (16)$$

This $\hat{S}^E$ differs from $\hat{M}^E$ in that $\hat{S}$ need not satisfy the *ergodicity* requirement, where $\hat{S} = S^{(u_1)}S^{(u_2)}...S^{(u_g)}$.

Metropolis−Monte Carlo Algorithm Combined with Shuffling Generates Correct Simulations. Shuffling trials can be easily designed, and one can also prove that a Metropolis−Monte Carlo simulation interrupted regularly by shuffling trials is a correct simulation, *i.e.*, leading to the equilibrium distribution of the system. The proof is based on propositions 1 through 3 and on the following two propositions:

Proposition 4: Commutativity. Similarly to proposition 1, because of the *normality* and *symmetry* of the shuffling generating matrices $S^{(k)}$, the respective transition probability matrices, $S^{E(k)}$, commute with the equilibrium distribution matrix.

$$BS^{E(k)} = S^{E(k)}B = B \quad (17)$$

A similar relationship holds for the resultant transition probability matrix $\hat{S}^E$:

$$B\,\hat{S}^E = \hat{S}^E B = B \quad (18)$$

Proposition 5: Positivity. If A is a $W \times W$ positive matrix then $AS^{E(k)}$ is positive for every $k = 1, 2, ...$ (Appendix A.4).

Now we can prove the correctness of a simulation where the Metropolis−Monte Carlo algorithm is interrupted by one shuffling trial, $U_{(i)}$, after every mth repeat of the trial pattern. Or more precisely, we have to prove that transition probability matrix, $(\hat{M}^E)^m S^{E(i)}$, satisfies the condition of the correct simulations given by eq 3, where m is chosen such that $(\hat{M}^E)^m$ is a positive matrix.

According to propositions 2, $(\hat{M}^E)^m$ is a positive stochastic matrix if $m \geq l$. Thus the product of $(\hat{M}^E)^m$ and a shuffling matrix $S^{E(i)}$ is a positive matrix too (see proposition 5). Since $(\hat{M}^E)^m S^{E(i)}$ is a positive stochastic matrix, proposition 3 is applicable,

$$\lim_{n\to\infty}[(\hat{M}^E)^m S^{E(i)}]^n = D \quad (19)$$

where the rows of the limit matrix D are distributions and are identical to each other. Let us multiply eq 19 by the equilibrium distribution matrix and apply the commutativity properties ($B\hat{M}^E = B$ and $BS^{E(i)} = B$; see propositions 1, 4) repeatedly. Now similar to the derivation of eqs 13 and 14, we get

$$\lim_{n\to\infty}[(\hat{M}^E)^m S^{E(i)}]^n = B \quad (20)$$

In a similar way, the correctness of the simulation can be proved for the more general case when the Metropolis algorithm is interrupted by a series of shuffling trials,

$$\lim_{n\to\infty}[(\hat{M}^E)^m \hat{S}^E]^n = B \quad (21)$$

Random Permutation of the Trial Pattern. At this stage, the shuffling algorithm can be generalized further. In a shuffling algorithm the trial pattern can be written as $T_{(1)}T_{(2)}...T_{(n)}U_{(1)}U_{(2)}...U_{(g)}$, where $T_{(i)}$ is a common trial and $U_{(j)}$ is a shuffling trial. The resultant transition probability matrix corresponding to trial pattern $T_{(1)}T_{(2)}...T_{(n)}U_{(1)}U_{(2)}...U_{(g)}$ can be denoted by $\hat{H}^E$. Correctness of the algorithm implies that matrix $\hat{H}^E$ is positive. A random change of the trial order (for both T and U) in the trial pattern is called a random permutation of the pattern.

Proposition 6. A shuffling Monte Carlo algorithm in which random permutations of a trial pattern are performed (instead of repeating the trial pattern $T_{(1)}T_{(2)}...T_{(n)}U_{(1)}U_{(2)}...U_{(g)}$) is a correct algorithm if the probability of the occurrence of trial pattern $T_{(1)}T_{(2)}...T_{(n)}U_{(1)}U_{(2)}...U_{(g)}$ is positive.

To simplify the proof of this proposition, suppose $n=1$ and $g=1$ in the trial pattern. In this Monte Carlo algorithm trial pattern *TU* and *UT* will be performed with probability p and 1-p, respectively, where $0<p<1$. Let the resultant transition matrices of *TU* and *UT* be $\hat{H}^E$ and $\hat{Q}^E$, respectively. In this algorithm, transition matrix $\hat{H}^E$ occurs with probability p and transition matrix $\hat{Q}^E$ occurs with probability 1-p, i.e., the resultant transition matrix in the algorithm is $\hat{R}^E = p\hat{H}^E + (1 - p)\hat{Q}^E$. Since $\hat{H}^E$ is a positive matrix and every entry in $\hat{Q}^E$ is non-negative, $\hat{R}^E$ is a positive matrix too. The correctness of the algorithm follows from the positivity of matrix $\hat{R}^E$ and from the normality and symmetry of the respective candidate configuration generation matrices (see proposition 3).

III. Results

The design, application, and efficiency of the shuffling trials in Monte Carlo simulations are demonstrated in this section by using the Ising model of spin lattices.[9]

Ising Model of Spin Lattices. In the model a $65 \times 65(=N)$ triangular lattice of identical spins, with magnetic moment μ, is considered. An external, homogeneous magnetic field H is applied perpendicularly to the lattice plane. Each spin can have either parallel (up-spin state) or antiparallel (down-spin state) orientation to the magnetic field. The actual configuration of the spin lattice is given by an N element vector, **s**. The ith element of the vector is $s_i = 1$ or $s_i = -1$ if the state of the ith spin is up-spin or down-spin, respectively. The spin lattice is in thermal and magnetic contact with the surrounding, and thus the equilibrium probability of configuration **s** is[14]

$$\pi(\mathbf{s}) = \frac{\exp([J\sum_{\{ik\}} s_i s_k + \mu H \sum_{i=1}^{N} s_i]/kt)}{\sum_{\{\mathbf{s}'\}} \exp([J\sum_{\{ik\}} s'_i s'_k + \mu H \sum_{i=1}^{N} s'_i]/kt)} \qquad (22)$$

In eq 22 the summations $\sum_{\{s'\}}$ and $\sum_{\{ik\}}$ are taken over all of the configurations and over every nearest neighbor pair of spins, respectively, and J is the exchange energy. To reduce the edge effects of the finite spin lattice, periodic boundary conditions are utilized.

The Common Glauber Method. In the Glauber method a Monte Carlo cycle (MCC) is defined such that the number of opportunities for lattice points to change state is equal to the number of points in the lattice, N. The points of the lattice are picked randomly, with probability $1/N$. The algorithm, according to the Markov chain theory, can be represented by a $2^N \times 2^N$ candidate configuration generation matrix M, where 2^N is the total number of configurations in the lattice. During a candidate configuration generation there is direct access to N different configurations from the current configuration; thus every row of the matrix M contains N elements, each equals $1/N$, while the remaining $2^N - N$ elements equal zero. Every element in the main diagonal is zero, and the random selection of lattice points ensures the symmetry of this matrix. Any configuration of the spin lattice can be reached from any other configuration within N trials because any two configurations differ at no more than N lattice points. Since the matrix M satisfies the normality, symmetry, and ergodicity conditions, the respective transition probability matrix M^E (constructed according to eq 7) generates a formally correct simulation.

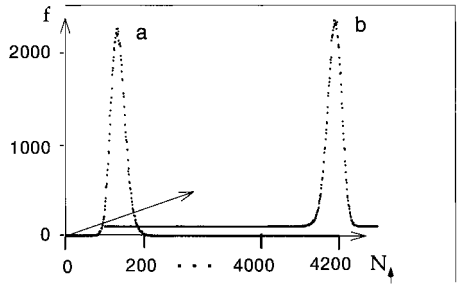

Figure 1. Frequency distributions of the number of up-spin lattice points, $N_\uparrow$, in the case of the common Glauber method. Parameters of the simulations: $J/kt = 1/3$, $\mu H/kt = 0$, lattice size $= 65 \times 65$, number of MCC $= 10^5$. Initial configuration is all down-spin and all up-spin for curve a and curve b, respectively.

During the simulations, however, one may get long, strongly correlated sequences of configurations when $J/kt > 0.274$. (Note that $J/kt_c = 0.274\,65$ for an infinite triangular spin lattice,[15] where t_c is the critical temperature.) In this case the convergence to the equilibrium distribution becomes so slow that the equilibrium distribution is not attainable within a feasible computation time.

As an example, we ran simulations by using the following parameters: $J/kt = 1/3$ and $\mu H/kt = 0$. The obtained frequency distributions of the number of up-spin lattice points $N_\uparrow$ show that the system remained in a phase of mostly down-spins when the simulation was started from an all down-spin configuration (Figure 1, curve a), and it remained in an up-spin phase when the simulation was started from an all up-spin configuration (Figure 1, curve b). Thus the *common Glauber method* resulted in different frequency distributions for different initial conditions; that is, the equilibrium distribution was not attained even after 10^5 MCC.

The Glauber Method with Shuffling. Including properly chosen shufling trials in the commonly applied Metropolis algorithm can considerably accelerate the convergence to the equilibrium distribution. To design optimal shuffling operations for a given physical system, one has to know *a priori* or explore computationally the regions of low-energy configurations in the configurational space. In addition to computational search, symmetry considerations and common sense frequently help to explore regions of low-energy configurations. Then shuffling operations taking the system from one region of low-energy configurations to another region of low-energy configurations accelerate convergence to the equilibrium distribution. Application of the shuffling method in simulating DNA strand separation[13] provides specific examples for the optimal design of shuffling operations.

In the case of spin lattices it is known from analytical studies[16] that below the critical temperature there are two regions of low-energy configurations. Onsager[16] pointed out that in a zero external field, below the critical temperature, the largest eigenvalue of the transition matrix of the spin lattice is doubly degenerated; that is, spins are mostly up or mostly down with similar probability. Thus introducing shuffling trials into the algorithm of the common Glauber method, each reversing the actual state of every spin in the lattice, one can accelerate the convergence to the equilibrium distribution. This global, nonphysical operation promotes changes from one to the other region of low-energy configurations without energy-intensive local initiation of the new phase. The operation is truly shuffling because the respective shuffling generating matrix S is a symmetric, stochastic $2^N \times 2^N$ matrix. In each row of S there

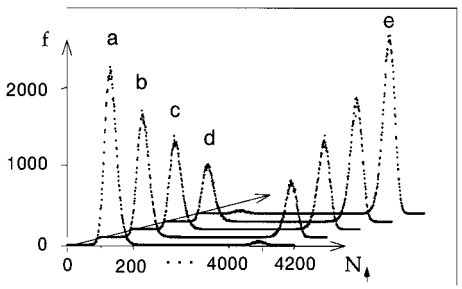

Figure 2. Frequency distributions of the number of up-spin lattice points, $N_\uparrow$, in the case of the Glauber method with shuffling. Parameters of the simulations: $\mu H/kt = -5 \times 10^{-4}$, -10^{-4}, 0, 10^{-4}, and 5×10^{-4} for curves a, b, c, d, and e, respectively, while the other parameters are given in the legend to Figure 1.

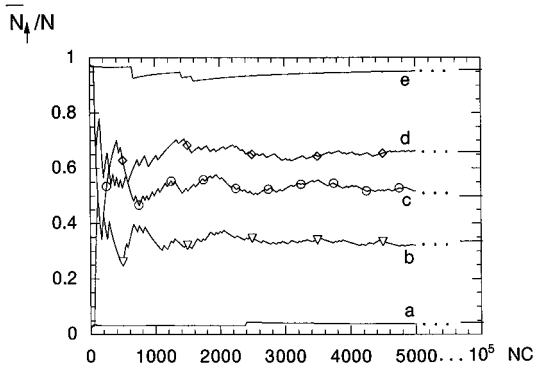

Figure 3. Convergence to the equilibrium in the case of the Glauber method with shuffling. The average number of up-spin lattice points, $\bar{N}_\uparrow$ (eq 23), is plotted against the number of MC cycles NC. Initial state: all up-spin. Parameters of the simulations: $\mu H/kt = -5 \times 10^{-4}$, -10^{-4}, 0, 10^{-4}, and 5×10^{-4} for curves a, b, c, d, and e, respectively, while the other parameters are given in the legend to Figure 1. The symbols on curves b, c, and d help to identify the curves.

is only one nonzero element, which is 1. The transition probability matrix S^E is constructed from S according to eq 15. As an example the M and S matrix is explicitely given for a 2 $\times$ 2 spin lattice in Appendix A.6.

Simulations were run by using the spin lattice model described before. In these simulations, however, the Glauber method was interrupted by shuffling after every 29th MCC, applying the first shuffling at the end of the first MCC. The respective transition probability matrix $(M^E)^N S^E (M^E)^{28N}$ generates a correct simulation because $(M^E)^{29N}$ is a positive matrix (see proposition 6 and Appendix A.5). Figure 2 shows the frequency distributions of the number of up-spin lattice points, $N_\uparrow$, obtained at five different external magnetic field strengths. In these cases the frequency distribution curves were independent from the initial configurations of the simulations; that is, the equilibrium distributions were attained. At each magnetic field strength the convergence to the equilibrium is shown in Figure 3, monitoring the convergence by the average of $N_\uparrow$,

$$\bar{N}_\uparrow(\mathrm{NC}) = \sum_{i=1}^{\mathrm{NC}} N_\uparrow(i)/\mathrm{NC} \qquad (23)$$

where NC is the number of MCCs. The curves in Figure 3 show singularity each time when a shuffling trial is accepted.

Systems with Attainable and Unattainable Thermodynamic Equilibrium. The symmetry of the equilibrium distribution in Figure 2c implies that below the critical temperature in a zero external field the average magnetization is zero. Is this in contradiction with the observed spontaneous magnetization of ferromagnets below the critical temperature? No, because below the critical temperature a ferromagnet is unable to attain the thermodynamic equilibrium. Depending on the initial configuration, the ferromagnet remains in either mostly up-spin or mostly down-spin phase; that is, the respective spin distribution depends on the initial configuration. However, at given external conditions, independently from the initial configuration, each system has one and only one equilibrium distribution, the Boltzmann distribution over every configuration of the system.

Calculating the spontaneous magnetization of a ferromagnet, one has to take the average of the spin lattice magnetization over a nonequilibrium distribution such as given in Figure 1a or b. To get an analytical expression for the average magnetization over a nonequilibrium distribution, Yang[17] introduced a vanishingly weak positive external magnetic field which cut out all configurations of the lattice for which the total spin was negative.

In general one can argue that on a sufficiently large system a sufficiently low temperature cannot be reached without destroying the thermodynamic equilibrium.[18] However there are exceptions. In the case of spin glasses individual systems can attain the thermodynamic equilibrium at low temperature by means of tunneling between the low-energy regions of the configurational space.[19]

In the case of small systems, such as biomacromolecules, the equilibrium distribution is usually attainable at physiological temperatures.

IV. Conclusions

The stationary distribution of a Markov process is not altered if the Markov chain is interrupted regularly by shuffling trials. This covers a broad set of nonphysical operations on the system, satisfying both the normalization and the detailed balance conditions. Wisely chosen shuffling trials may dramatically accelerate the convergence of the Monte Carlo simulations to the equilibrium distribution. The general method of shuffling is broadly applicable in simulating the thermodynamic equilibrium of complex systems. In this paper the shuffling method is applied for a two-state Ising model of spin lattices. Application of the shuffling method in simulating DNA strand separation has been published elsewhere.[13]

Acknowledgment. I.S. thanks Drs. R. L. Biltonen and T. E. Thompson for initiating this research and Dr. J. F. Nagle for drawing our attention to Houtappel's paper. The authors thank Mr. K. Kelliher for his contributions to the text. This project was supported by Grant No. GM 23573, awarded by The National Institute of General Medical Sciences.

Appendix

A.1. We first show that the condition given by eq 3 can be derived from eq 1.

Let e_i be the unit row vector whose ith component is 1 and the rest are zeros. The dimension of the vector is W. Let us choose e_i as the initial distribution π^0. Substituting the vector e_i into eq 1, we get

$$\lim_{n\to\infty} e_i P^n = \lim_{n\to\infty}(P_{i1}(n),P_{i2}(n),...,P_{iW}(n)) = \pi \qquad (24)$$

where $P_{ij}(n)$ is the (i,j)th element of the P^n matrix. According to this expression, the ith row vector of the limit matrix $\lim_{n\to\infty} P^n$ exists, and it is equal to π. Similarly, one can prove that every row of the limit matrix exists and is equal to π. Thus the limit matrix exists, and it is the equilibrium distribution matrix B.

Now we prove that eq 1 follows from eq 3. We have for all e_i's

$$\pi = e_i B = e_i \lim_{n\to\infty} P^n = \lim_{n\to\infty} e_i P^n \tag{25}$$

Any distribution π^0 can be expressed in the following way:

$$\pi^0 = \sum_{j=1}^{W} \pi_j^0 e_j \tag{26}$$

where $\pi_j^0 \geq 0$ and $\sum_{j=1}^{W} \pi_j^0 = 1$. Then the following is true:

$$\lim_{n\to\infty} \pi^0 P^n = \sum_{j=1}^{W} \pi_j^0 e_j \lim_{n\to\infty} P^n = \sum_{j=1}^{W} \pi_j^0 e_j B = \sum_{j=1}^{W} \pi_j^0 \pi = \pi \tag{27}$$

A.2. We show that if M is a stochastic symmetric matrix (ergodicity is not required) and if M^E is obtained according to eq 7, then $BM^E = M^E B = B$.

Let us consider the (i,j)th element of the product matrix BM^E.

$$(BM^E)_{ij} = \sum_{m=1}^{W} \pi_m M_{mj}^E = \pi_j M_{jj}^E + \sum_{m\neq j} \pi_m M_{mj}^E$$

$$= \pi_j (1 - \sum_{m\neq j} M_{jm}^E) + \sum_{m\neq j} \pi_m M_{mj}^E \tag{28}$$

$$= \pi_j + \sum_{m\neq j} (\pi_m M_{mj}^E - \pi_j M_{jm}^E) = \pi_j = B_{ij}$$

Notice that the bracketed two terms in the summation result in zero since $M_{jm}^E = M_{jm}\pi_m/(\pi_m + \pi_j)$ and $M_{mj}^E = M_{mj}\pi_j/(\pi_m + \pi_j)$ (see eq 7) and $M_{jm} = M_{mj}$ (see eq 6). Thus we proved that $BM^E = B$. On the other hand it is clear that $M^E B = B$ because the sum of the elements in each row of M^E is 1 and the column entries of the B matrix are identical. By combining the above two results we get $BM^E = M^E B = B$.

A.3. We prove that $\hat{M}^E$ satisfies the second condition in proposition 2. Let i be the configuration such that the associated π_i reaches the global maximum of the system. We show that if $M^{E(k)}$ is constructed according to eq 7 from $M^{(k)}$ which satisfies eqs 5 and 6, then the (i,i)th element of $M^{E(k)}$, denoted by $M_{ii}^{E(k)}$, is positive.

This is the case because according to eq 7, $M_{ij}^{E(k)} < M_{ij}^{(k)}$, and from eq 5 it follows that

$$1 = \sum_{j=1}^{W} M_{ij}^{(k)} \geq \sum_{j\neq i}^{W} M_{ij}^{(k)} > \sum_{j\neq i}^{W} M_{ij}^{E(k)} \tag{29}$$

After substituting this inequality into the second equation of eq 7, we get that $M_{ii}^{E(k)} > 0$. The (i,i)th element in $\hat{M}^E$, denoted by $\hat{M}_{ii}^E$, is also positive because

$$\hat{M}_{ii}^E \geq M_{ii}^{E(\nu_1)} M_{ii}^{E(\nu_2)} \ldots M_{ii}^{E(\nu_h)} > 0 \tag{30}$$

Thus, if $\hat{M}$ in eq 4 is ergodic, then $\hat{M}^E$ is also ergodic,[11] and so there exists an integer $l > 0$ such that $(\hat{M}^E)^l$ is positive. $(\hat{M}^E)^m$ is also positive for $m \geq l$ (see A.4).

A.4. If A is a positive matrix, S is a stochastic symmetric matrix (ergodicity is not required), and S^E is constructed according to eq 15, then AS^E is a positive matrix.

Let us assume that the (i,j)th element of the product matrix is zero, and thus the product matrix is not positive:

$$(AS^E)_{ij} = \sum_{m=1}^{W} A_{im} S_{mj}^E = 0 \tag{31}$$

Since every $A_{im} > 0$, the above equation can be satisfied only if the jth column of S^E is zero. From the constructing rules of matrix S^E, it follows that the jth row of S^E will be zero too. However, this contradicts the definition of stochastic matrix, *i.e.*, $\sum_{m=1}^{W} S_{jm}^E = 1$. Thus the initial assumption is incorrect; that is, the product matrix AS^E is positive.

A.5. $(M^E)^{2N}$ is a positive matrix, where M^E is the transition probability matrix of the common Glauber dynamics, and N is the total number of lattice points.

Lemma 1. If there is a configuration i such that $M_{ii}^E > 0$, then for any configuration j of the spin lattice the (i,j)th element, $M_{ij}^E(N)$, of the product matrix $(M^E)^N$ is positive.

Configuration i of the spin lattice can be changed into any chosen configuration j by changing the state of $t \in [1,N]$ lattice points; that is, the matrix M is ergodic $M_{i,l_1}^E M_{l_1,l_2}^E \ldots M_{l_{t-1},j}^E > 0$, where l_1, l_2, ..., l_{t-1} are the intermediate $t-1$ configurations between configurations i and j, and thus the following inequality holds for the (i,j)th element of the product matrix $(M^E)^t$:

$$M_{ij}^E(t) \geq M_{i,l_1}^E M_{l_1,l_2}^E \ldots M_{l_{t-1},j}^E > 0 \tag{32}$$

Therefore, $M_{ij}^E(N) \geq (M_{ii}^E)^{N-t} M^E{}_{i,l_1} M^E{}_{l_1,t_2} \ldots M^E{}_{l_{t-1},j} > 0$

Lemma 2. If there is a configuration i such that $M_{ii}^E > 0$, then $(M^E)^{2N}$ is a positive matrix.

From lemma 1 we have $M_{ij}^E(N) > 0$ for any j. This means one can select a chain of $N-1$ intermediate configurations (l_1, l_2, ..., l_{N-1}) leading from configuration i to j such that

$$M_{ij}^E(N) \geq M_{i,l_1}^E M_{l_1,l_2}^E \ldots M_{l_N-1,j}^E > 0 \tag{33}$$

According to the construction of M^E (see eqs 6 and 7), if $M_{lk}^E > 0$ then $M_{kl}^E > 0$. Thus one can return from configuration j to i on the same path with a positive probability, *i.e.*,

$$M_{ji}^E(N) \geq M_{j,l_{N-1}}^E M_{l_{N-1},l_{N-2}}^E \ldots M_{l_1,i}^E > 0 \tag{34}$$

This shows that $M_{ji}^E(N)$ is also positive. Finally, any element $M_{kl}^E(2N)$ in $(M^E)^{2N}$ satisfies the following inequality:

$$M_{kl}^E(2N) \geq M_{ki}^E(N) M_{il}^E(N) > 0 \tag{35}$$

Thus $(M^E)^{2N}$ is positive. The assumption of this proof that there is a configuration i such that $M_{ii}^E > 0$ follows from A.3.

A.6. Example of $N = 4$. In the simplest case the spin lattice is a 2×2 lattice ($N = 4$) and the number of configurations is 16. By labeling the configurations with four-digit binary numbers, one can give a natural order to the configurations. For example, 0000(=0) means that every lattice point is in the up-spin state, 0001(=1) means that the first three lattice points are in the up-spin state, while the last one is in the down-spin state, etc. The 16×16 configuration generating matrix M has the following form:

$$M = \begin{bmatrix} A & I/4 & I/4 & 0 \\ I/4 & A & 0 & I/4 \\ I/4 & 0 & A & I/4 \\ 0 & I/4 & I/4 & A \end{bmatrix}$$

where

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 0 & 1/4 & 1/4 & 0 \\ 1/4 & 0 & 0 & 1/4 \\ 1/4 & 0 & 0 & 1/4 \\ 0 & 1/4 & 1/4 & 0 \end{bmatrix} \tag{36}$$

Here the rows and columns of matrix M are indexed according

to the code numbers of the configurations. The 16 × 16 shuffling generating matrix *S* has a much simpler structure than the *M* matrix: every element along the antidiagonal of matrix *S* is 1, while the other elements are zeros.

References and Notes

(1) Chung, K. L. *Markov Process with Stationary Transition Probability;* Springer-Verlag, Heidelberg, 1976.
(2) Peskun, P. H. *J. Comput. Phys.* **1981**, *40*, 327.
(3) Kandel, D.; Ben-Av, R.; Domany, E. *Phys. Rev. Lett.* **1990**, *65*, 941.
(4) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *58*, 86.
(5) Wolff, U. *Phys. Rev. Lett.* **1989**, *62*, 361.
(6) Glauber, R. J. *J. Math. Phys.* **1963**, *4*, 294.
(7) Kawasaki, K. *Phase Transitions and Critical Phenomena*; Academic Press: London, 1972; Vol. 2, p 443.
(8) (a) Metropolis, M.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. N.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087. (b) Binder, K. *Monte Carlo Methods in Statistical Physics*; Springer-Verlag: Berlin, 1986; p 6. (c) Von Neumann, J.; Ulam, S. *Bull. Am. Math. Soc.* **1945**, *51*, 660. (d) Metropolis, M.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335.
(9) Davidson, N. *Statistical Mechanics*; McGraw-Hill Book Co. Inc.: New York, 1962.
(10) Hastings, W. K. *Biometrika* **1970**, *57*, 1.
(11) Feller, W. *An Introduction to Probability Theory and Its Applications*, 3rd ed.; John Wiley and Sons, Inc.: New York, London, Sydney, 1966; Vol. 1, Chapter XV. Especially, p 426, problems 21 and 22.
(12) Moran, A. P. *Introduction to Probability Theory*; Clarendon Press: Oxford, 1968; pp 110−112.
(13) Sun, H.; Mezei, M.; Fye, R.; Benham, C. J. *J. Chem Phys.* **1995**, *103*, 8653.
(14) Wannier, G. H. *Statistical Physics*; John Wiley & Sons: New York, 1966; p 336.
(15) Houtappel, R. M. F. *Physica* **1950**, *16*, 425.
(16) Onsager, L. *Phys. Rev.* **1944**, *65*, 117.
(17) Yang, C. N. *Phys. Rev.* **1952**, *85*, 808.
(18) Fischer, D. A.; Huse, D. A. *J. Phys. I* (*France*) **1991**, *1*, 621.
(19) Binder, K.; Young, A. P. *Rev. Mod. Phys.* **1986**, *58*, 801.