RICE UNIVERSITY

Enhanced sampling method for free energy calculation and large scale conformational change

by

Cheng Zhang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Science

APPROVED, THESIS COMMITTEE:

Jianpeng Ma, Professor
Bioengineering
Biochemistry & Molecular Biology, Baylor College of Medicine

Bruce Johnson, Distinguished Faculty Fellow
Chemistry Executive Director, Rice Quantum Institute

Oleg Igoshin, Assistant Professor
Bioengineering

HOUSTON, TEXAS
APRIL 2009

ABSTRACT

Enhanced sampling method for free energy calculation and large scale conformational change

by

Cheng Zhang

A method of directly computing the partition function (or the corresponding free energy) and accelerating configurational sampling is developed. In an expanded ensemble, the method can quickly sample a broad distribution and yield accurate results for the partition function. The method is shown to be efficient and accurate in studying thermodynamic properties, searching low-energy configurations of difficult molecular systems and counting solutions of puzzles.

ACKNOWLEDGEMENTS

I am grateful to my advisor Professor Jianpeng Ma, who provided immeasurable support and guidance throughout my research. It is always a pleasure to have open scientific discussions with Professor Ma, because he always encourages new ideas (even those seemingly implausible ones), and is willing to share his unique and insightful understanding towards the essence of various problems. I personally believe that the encouragement and inspiration I received from these discussions is much more important than any knowledge and techniques I learned elsewhere.

It is fortunate to have Dr. Bruce Johnson and Dr. Oleg Igoshin on my thesis committee. I am indebted to them for their invaluable time and guidance.

I would also like to thank people in Professor Ma's lab, especially Mingyang Lu, Yuan Mei, Jialin Li, Fengyun Ni and Athanasios Dousis for their friendship and company. In addition, I want to express my gratitude to Dr. Qinghua Wang, and to other lab members, Dr. Yinghao Wu, Zhao Ge, Dr. Billy Poon, Dr. Mingzhi Chen, Jun Shen, Xiaorui Chen, Yufeng Gou, Xia Tian, Brian Kirk and Yushao Cheng, who provided assistance in numerous ways.

Finally, I wish to thank my parents. This thesis would not have been possible without their support and understanding.

Table of Contents

List of Tables

List of Figures

I. Introduction

In this thesis, a novel computational method is presented for directly calculating the partition function. Using a generalized ensemble, the method delivers an enhanced sampling on a large phase space as well as a fast convergence of the partition function. Due to its generality and efficiency, the method is successfully applied to a wide range of problems, e.g., calculating free energy of physical or chemical systems, searching for low-energy configurations of coarse-grained model proteins and counting solutions of puzzles.

The thesis is organized as the following. The rest of the chapter provides a brief overview of the background information. In Chapter II, we focus on a description of the method in its basic form. Some applications in thermodynamic systems follow in the Chapter III. An especially intriguing application to counting solutions for N-queens problems and Latin squares are given in Chapter IV. In Chapter V, a theory is proposed to study the convergence of the method. Finally, we compare the method with the replica exchange method in their performance of enhancing configurational sampling.

A. Partition function

In statistical mechanics, the partition function plays the role of defining a statistical ensemble. For example, in the canonical ensemble, the partition function $Z(N,V,T)$ is defined as a function of the number of particles N, the volume V, and the temperature T

$$Z(N,V,T)=\sum_X \exp(-\beta E_X), \tag{1}$$

where E_X is the energy of microscopic state X, and $\beta=1/k_BT$ is the reciprocal temperature (with k_B is the Boltzmann constant).

Since the partition function sums over every microscopic configuration X with a weight $\exp(-\beta E_X)$, it can be used to derive all macroscopic properties of the system. For example, the average energy can be expressed as the derivative of the natural logarithm of the partition function $\ln Z$

$$\langle E\rangle=\frac{\sum_X E_X \exp(-\beta E_X)}{\sum_X \exp(-\beta E_X)}=-\frac{\partial \ln Z}{\partial \beta}. \tag{2a}$$

Similarly, the heat capacity is related to second order derivative of $\ln Z$

$$C_V=\frac{\partial\langle E\rangle}{\partial T}=-k_B\beta^2\frac{\partial^2 \ln Z}{\partial \beta^2}. \tag{2b}$$

More generally, if the exponential factor βE has an implicit or explicit dependence on a variable λ (which can be N, V, T or other variables), the derivative $\partial \ln Z/\partial\lambda$ gives the corresponding thermal conjugate.

The partition function itself corresponds to the free energy in thermodynamics. For example, in the case of the canonical ensemble, the corresponding free energy is the Helmholtz free energy $F=-\ln Z/\beta$.

B. Monte Carlo and molecular dynamics simulations

Computer simulations can be used to study statistical properties of a system. Generally, two classes of methods are commonly used.

The first one is the Monte Carlo (MC) method. Here the objective is to populate a desired distribution in the given ensemble though a Markov chain of states. In the most basic form, i.e., the Metropolis algorithm [1], one starts from an arbitrary configuration of the system, and keeps proposing random modifications of the current configuration, and using the following probability

$$\mathrm{Acc} = \min\{1, \exp(-\beta\Delta E)\},$$

to accept the changed configuration (if the proposal is rejected, the old configuration is retained). Here ΔE is the change of energy involved in changing the configuration, and β is the reciprocal temperature. Such a process leads to a chain of states that ultimately recover the desired distribution [in the above case, it is the canonical distribution $p(X) = \exp(-\beta E_X)/Z(\beta)$]. An excellent introduction to the Monte Carlo method can be found in Ref. [2].

The other approach is the molecular dynamics (MD) method, which is basically a direct integration of the Newton's second equation

$$d\mathbf{x}/dt = \mathbf{v}$$
$$d\mathbf{v}/dt = \mathbf{F}/m$$

Unlike the Monte Carlo method, the molecular dynamics method is deterministic in nature and can be also used to calculate dynamics properties of a system. Although the Newton's equation aims at a microcanonical ensemble instead of a canonical ensemble, there are efforts in modifying the equation of motion to yield the desired canonical distribution [3-5].

From the trajectory generated from the above methods, one can calculate average properties, such as the average energy $\langle E \rangle$ or the heat capacity $C = \langle \Delta E^2 \rangle / T^2$. However, the partition function or the free energy itself cannot be directly calculated by performing a regular average at a canonical ensemble. Therefore the usual computer simulation methods cannot be used to calculate the free energy.

C. Calculating the partition function from the density of states

An indirect approach of conducting sampling is to calculate the density of state $g(E)$ first, then compute the partition function through a Laplace transform as

$$Z(\beta) = \int_0^\infty g(E) \exp(-\beta E)\, dE\,, \tag{3}$$

where we assumed that the energy of the ground state(s) is 0. The density of states $g(E)\, dE$ gives the number of microscopic configurations with its energy being within the interval (E, $E + dE$). In such an approach, one first construct an expanded ensemble with each microscopic state weighted by a factor $1/g(E)$, instead of the usual Boltzmann factor $\exp(-\beta E)$. In this ensemble, all energy intervals receive (statistically) equal number of visits and a flat energy histogram is reached. These methods include the multicanonical ensemble method [6-9], the entropic sampling method [10], the density of states method [11, 12], and the statistical temperature method [13, 14]. Such an approach usually works well on a smooth energy landscape. However for a complex system, the energy landscape becomes extremely rugged at the low energy end and the density of states is hard to calculate accurately (in such a case, even the ground state energy is difficult to estimate until the ground state is reached), this approach becomes inefficient.

D. Directly calculating the partition function

In the method presented here, we follow a different approach, that is, to directly calculate the partition function as a function of a macroscopic variable, e.g., N, V, or T. In this case, the flat energy histogram technique can be applied to any of the macroscopic variables. As one does not know the partition function in advance, the partition function is continuously updated throughout simulation until convergence. In this way, the method can handle a broader range of applications than other methods based on the density of states [11-14]. Techniques such as cluster algorithms or molecular dynamics can be incorporated. Further, as the partition function is smoother than the density of states, less sampling points are needed. In addition, it is easier both to reach convergence and to explore a rugged energy landscape of a complex system.

II. Method

A. Case study in the temperature space

To avoid unnecessary abstraction, we first demonstrate the method in a special case, where the objective is to compute the partition function as a function of the temperature T (or equivalently $\beta = 1/k_B T$). For numerical convenience, the temperature here is treated a discrete variable, i.e., the partition function is calculated at a few predefined values β_i's. Before simulation, the sampling temperatures β_i are distributed over the temperature range of interest. Each sampling temperature is associated with an estimated partition function $\tilde{Z}(\beta_i)$. Initially, the values $\tilde{Z}(\beta_i)$'s are uniformly set to 1.0. Since we usually have no information about the actual partition function, the objective is to converge $\tilde{Z}(\beta_i)$'s to the actual values $Z(\beta_i)$'s throughout the simulation.

Since we wish to compute the partition function at multiple temperatures, the simulation cannot be limited to a single temperature. Therefore, the temperature is now treated as a random variable and random transitions between different β_i's are incorporated into the simulation.

To couple sampling in the configurational space with that in the temperature space, we introduce two kinds of MC moves: energy moves under a fixed temperature and temperature moves under a fixed energy. The two moves can be thought as an extended MC move. Before such an extended MC step, we use a fixed probability to determine which kind of move the system takes. If the energy move is chosen, the Metropolis algorithm is used to generate a configurational change at the current temperature β_i. If the temperature move is

chosen, another temperature β_j is randomly chosen from the predefined sampling temperatures, and the following acceptance probability is used to accept the move,

$$\mathrm{Acc}(\beta_i \to \beta_j) = \min\left\{1, \frac{\exp(-\beta_j E)/\tilde{Z}(\beta_j)}{\exp(-\beta_i E)/\tilde{Z}(\beta_i)}\right\}. \tag{4}$$

Here E is the current energy; $\tilde{Z}(\beta_i)$ and $\tilde{Z}(\beta_j)$ are the estimated partition function at temperature β_i and β_j, respectively.

Such an acceptance probability is designed to follow the detailed balance condition

$$p(\beta_i, E)\,\mathrm{Acc}(\beta_i \to \beta_j) = p(\beta_j, E)\,\mathrm{Acc}(\beta_j \to \beta_i),$$

where $p(\beta_i, E) \sim \exp(-\beta_i E)/\tilde{Z}(\beta_i)$ is the probability distribution for a state with energy E to be in the temperature β_i. The ratio of overall frequencies of visiting β_i and β_j (denoted h_i and h_j by respectively) is given by summing over all configurations

$$\frac{h_i}{h_j} = \frac{\sum_X p(\beta_i, E_X)}{\sum_X p(\beta_j, E_X)} = \frac{Z(\beta_i)/\tilde{Z}(\beta_i)}{Z(\beta_j)/\tilde{Z}(\beta_i)}. \tag{5}$$

The overall frequencies correspond to the number of visits to β_i and β_j. If the estimated the partition function is exact, the right hand side gives unity, and thus one can expect that $h_i = h_j$, or the overall visits to β_i and β_j are the same. Since this argument applies to any pair of temperatures, we expect a uniform distribution of visits to different temperatures at the end of simulation.

Since we do not know the exact partition function *a priori*, the estimated values should be systematically modified throughout the simulation. After each MC step, the estimated partition at current temperature is added by a modification factor $\alpha > 0$ as,

$$\ln \tilde{Z}(\beta_i) \to \ln \tilde{Z}(\beta_i) + \alpha . \qquad (6)$$

As we shall see in Sec. V, by repeating the simply updating procedure for a fixed α, a convergence of the estimated partition function can be reached with an error roughly proportional to $\sqrt{\alpha}$. The error is due to that the acceptance probability is constantly modified, and the detailed balance is no longer preserved. Since the error is larger in a stage with a larger α, we should gradually decrease the magnitude of α in order to improve the accuracy of the estimated partition function. In practice, the whole simulation is separated into several stages, each marked by a different value of α. On passing from one stage to the next one, α is reduced to α / n. For convenience, we use $n = \sqrt{10}$ so that α is decreased by an order of magnitude every two stages.

Note, due the use of α, the acceptance probability Eq. (4) changes after every Monte Carlo step. And the detailed balance is no longer satisfied. However, at the end, α is reduced to a tiny number such that the violation of detailed balance is negligible.

For each α stage, if the simulation runs sufficiently long, each temperature on average receives equal number of visits, i.e., a flat temperature histogram can be achieved (the proof will be given in Sec V). Here the term "temperature histogram" refers to the number of visits to each discrete temperature instead of an interval. The simulation is allowed to enter the next α stage when the histogram fluctuation goes below a cutoff percentage. An alternative approach is to fix the number of simulation steps by $C/\sqrt{\alpha}$ for stage α, where the constant C can be estimated from a few initial α stages. The second approach ensures a better convergence.

B. Using the recursion with a non-uniform weight

In some circumstances a flat histogram is not the best choice. For example, for a system with a phase transition, the correlation time is much longer at the transition temperature than at any other temperatures. Therefore to enhance the random walk in the whole temperature range, temperatures around the transition temperature deserve higher statistical weights.

We now consider a variation of the updating scheme that allows the system to visit temperatures with non-uniform weights $w(\beta)$. To achieve this, the estimated partition function $\tilde{Z}(\beta)$ in the acceptance probability Eq. (4) is replaced by $\tilde{Z}(\beta)/w(\beta)$

$$\mathrm{Acc}(\beta_i \to \beta_j) = \min\left\{1, \frac{\exp(-\beta_j E)\, w(\beta_j)/\tilde{Z}(\beta_j)}{\exp(-\beta_i E)\, w(\beta_i)/\tilde{Z}(\beta_i)}\right\}, \tag{7}$$

whereas the updating scheme Eq. (6) is changed to

$$\ln \tilde{Z}(\beta_i) \to \ln \tilde{Z}(\beta_i) + \alpha / w(\beta_i).$$

In the modified scheme, the number of visits to different temperatures are expected to converge to a non-uniform distribution $w(\beta)$. We thus need a different condition to test the convergence in each α stage. A simple way is to construct a modulated temperature histogram $h^*(\beta)$, which is defined as the number of visits to temperature β divided by $w(\beta)$. The flatness of the modulated temperature histogram can be used to determine when to stop an α stage. Besides, one can also fix the number of simulation steps by $C/\sqrt{\alpha}$.

C. Sampling in the volume space

One can use a macroscopic variable other than the temperature as the sampling variable. In the following case, we shall assume that the volume is the sampling variable, while the temperature and particle number are held constant during the simulation.

A volume move is equivalent to a scale change of the system. Therefore it is convenient to work in a unit box and adopt the reduced coordinates, $\mathbf{s} = \mathbf{r}/\sqrt[3]{V}$. The partition function is factorized to the ideal gas part and a potential part, i.e.,

$$Z = Z_{\text{ig}} Z_{\text{pot}},$$

where

$$Z_{\text{ig}} = \frac{V^N}{N!}\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3N/2}$$

$$Z_{\text{pot}}(V) \equiv \frac{1}{V^N}\int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N)] = \int d\mathbf{s}^N \exp[-\beta U(\mathbf{s}^N;V)],$$

with $\hbar$ being the Planck's constant. Since the ideal gas part is analytically solved, we only need to compute the potential part of the partition function Z_{pot} (instead of Z itself). The acceptance probability for a volume change is

$$\text{Acc}(V_i \to V_j) = \min\left\{1, \frac{\exp[-\beta U(\mathbf{s}^N;V_j)]/\tilde{Z}_{\text{pot}}(V_j)}{\exp[-\beta U(\mathbf{s}^N;V_i)]/\tilde{Z}_{\text{pot}}(V_i)}\right\}.$$

After the simulation, the Helmholtz free energy is obtained through $F = F_{\text{ig}} - \ln Z_{\text{pot}}/\beta$.

III. Applications in thermodynamics

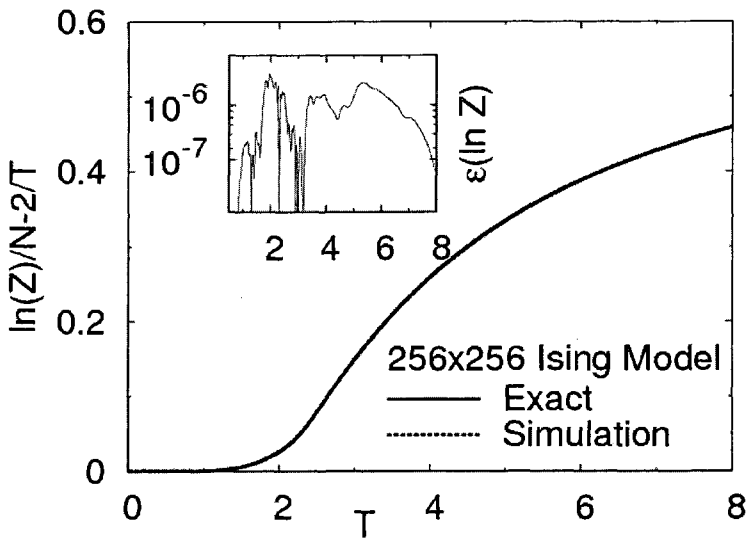
A. Ising model

The method was first tested on 256×256 Ising model. A wide temperature range, $T \in [0,8]$, was used in a single simulation. The temperature spacing for an efficient simulation should be inversely related to the heat capacity. This is because at the transition temperature (where the heat capacity diverges) the correlation time is much longer. Therefore more simulation effort should be spent there to equilibrate the whole system. For this large system, sampling temperatures were distributed based on the roughly estimated heat capacity (e.g., that from simulation of a smaller system). Accordingly, the entire temperature range was partitioned into 13 subranges. Sampling temperatures were evenly distributed within each subrange, but the temperature increments are different in different subranges. The temperature subranges and their increments were (0.1, 1.0 | 0.1), (1.0, 1.8 | 0.04), (1.8, 2.0 | 0.02), (2.0, 2.2 | 0.005), (2.2, 2.25 | 0.0025), (2.25, 2.3 | 0.002), (2.3, 2.35 | 0.005), (2.35, 2.5 | 0.01), (2.5, 2.7 | 0.02), (2.7, 3.6 | 0.05), (3.6, 5.0 | 0.07), (5.0, 6.0 | 0.1), (6.0, 8.0 | 0.2), where the notation for a subrange reads as (beginning temperature, ending temperature | increment). Totally, we used 218 sampling temperatures. Each time the probability of choosing temperature move over energy move was 0.1%. (This number should be larger for smaller systems.) The modification factor α was decreased from 1.0 to 10^{-9}, the number of MC steps for stage α was $100/\sqrt{\alpha}$ sweeps, so the whole simulation took 7.2×10^{6} sweeps. Thermodynamic quantities at temperatures other than the sampled ones can be calculated using multiple histogram method [15, 16]. Histograms from the last α stage were used. The exact results of the Ising model were also calculated using the method in reference [17]. The relative errors of the partition function, the energy, the entropy, and

the heat capacity were no larger than 0.00064%, 0.071%, 1.1%, and 3.9%, respectively. Fig. 1 shows the results for the partition function and heat capacity. For comparison, Wang-Landau (WL) density of state algorithm [11, 12] was applied to the same system using 15 independent simulations, and the maximum relative errors of free energy, the energy, the entropy, and the heat capacity were 0.0008%, 0.09%, 1.2%, and 4.5%, respectively. The simulation cost of WL method was 6.1×10^6 sweeps. However, in our method, the acceptance probabilities for metropolis energy moves can be pre-calculated to avoid expensive exponential computation. The above simulation was finished in 10 hours on an Intel Xeon processor (2.8GHz).

We now demonstrate the use a nonuniform weight. To focus sampling around the transition temperature, the frequency $w(\beta)$ can be associated with the heat capacity. Since the values of heat capacity are unknown in advance, they are updated at the end of each α stage (to be used in the next stage). The modified method was tested on the same 256×256 Ising system. The frequency $w(\beta)$ at temperature β was set as the square of the heat capacity per spin. Sampling temperatures were uniformly distributed over the whole range with a fixed increment $\Delta T = 0.002$. The probability of choosing temperature move over energy move was raised to 10%. The value of α was descended from 1 to $\sqrt{10}\times10^{-9}$. The simulation was kept running at each α stage until the fluctuation of temperature histogram was lowered below 50%. The last stage was purposely extended to 5×10^6 MC sweeps to accumulate more statistical data. Totally, 9.8×10^6 sweeps were used. The relative errors of the free energy, the energy and the heat capacity were no larger than 0.00045%, 0.055%, and 4.0%, respectively.

(a)



(b)

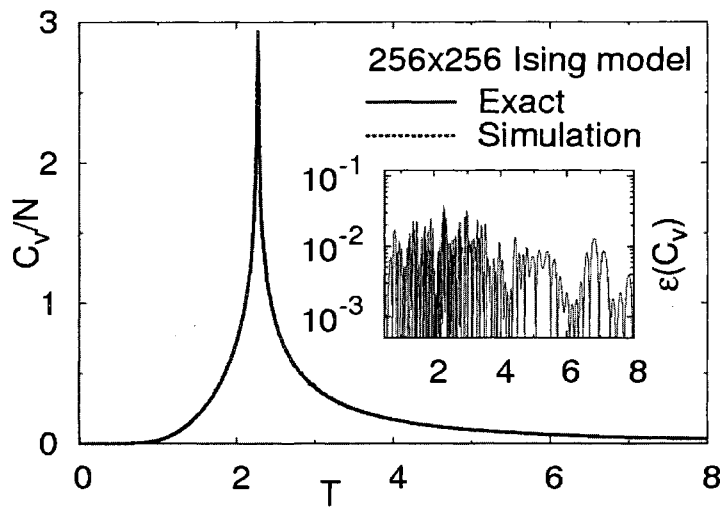


Figure 1. Results for a 256×256 Ising model. (a) The partition function as a function of temperature. The curve is shown for ln Z per spin with the contribution of the ground state subtracted. (b) The heat capacity per spin as a function of temperature. The relative errors are shown in the insets for both panels.

For a large temperature range, a randomly proposed temperature is easily rejected. In this case the following temperature transition scheme is more efficient. First, the relative probability $P_i = \exp(-\beta_i E)/\tilde{Z}(\beta_i)$ at each temperature β_i is calculated for the current energy E. Next the accumulated probability for each temperature $Q_i = \sum_{j\le i} P_j / \sum_j P_j$ is also calculated to form a series of brackets, $[Q_{i-1}, Q_i)$, $i = 1,2,\cdots$, with $Q_0 = 0$. If a uniform random number $r \in [0,1]$ falls in the ith bracket, β_i will be chosen as the next temperature.

As an example, the Swendsen-Wang cluster algorithm [18] was used as energy move on large $L \times L$ Ising models. To improve the efficiency, energy move and temperature move are merged in such a way that each energy move is immediately followed by a rejection-free temperature move. Simulations were performed on critical temperature windows estimated by $|T - T_c| \sim L^{-\nu}$. where the critical exponent $\nu = 1$ and T_c is the critical temperature. About 10 ~ 20 sampling temperatures were distributed in each window. Parameters and results were listed in Table I. The efficiency of the method is clear in terms of the number of simulation steps in reaching the desired accuracy.

Table I. Results for $L \times L$ Ising model using cluster algorithm. T_- and T_+ define a temperature window. Maximum relative error of $\ln Z$ is calculated by assuming that the partition function at T_- is correct.

L	T_-	T_+	MC steps (10^6)	$\varepsilon(\ln Z)$	$\varepsilon(C_V)$
64	2.0	2.9	0.7	4.0×10^{-6}	1.6%
128	2.15	2.6	1.5	2.2×10^{-6}	1.2%
256	2.2	2.4	3.0	6.0×10^{-7}	0.9%
512	2.24	2.34	2.9	1.6×10^{-7}	0.9%
1024	2.25	2.30	2.6	4.0×10^{-8}	2.0%

B. Lennard-Jones system

To illustrate the use of the volume as the sampling variable, we studied a 108-particle Lennard-Jones system with half box truncation. The choice of volume as the sampling variable is particularly useful in studying a system with a liquid-gas phase transition. This is because in such a case the volume (or the density) serves as the order parameter of the phase transition. For a fixed temperature, the liquid phase and gas phase can exist simultaneously under a particular pressure. The phase coexistence properties can be obtained in the following fashion. After the Helmholtz free energy $F(V)$ is obtained, the Gibbs free energy profile under pressure p can be derived as

$$G_p(V) = F(V) + pV .$$

The profile is related to the probability distribution of the system assuming a volume V under the given pressure p as,

$$p(V)\,dV \sim \exp[-\beta G_p(V)]\,dV$$

Usually, the probability distribution only has one dominant peak, either at the liquid phase with a high density or at the gas phase with a low density. At the point of phase coexistence, however, the probability distribution develops two peaks (one at the liquid phase, the other at the gas phase) with equal height. The two peaks correspond to two minima on the Gibbs free energy profile, see Fig. 2. The location of the two minima gives the liquid density ρ_+ and the gas density ρ_-. In this way, one can obtain the phase coexistence pressure as well as the liquid and gas densities under a given temperature.

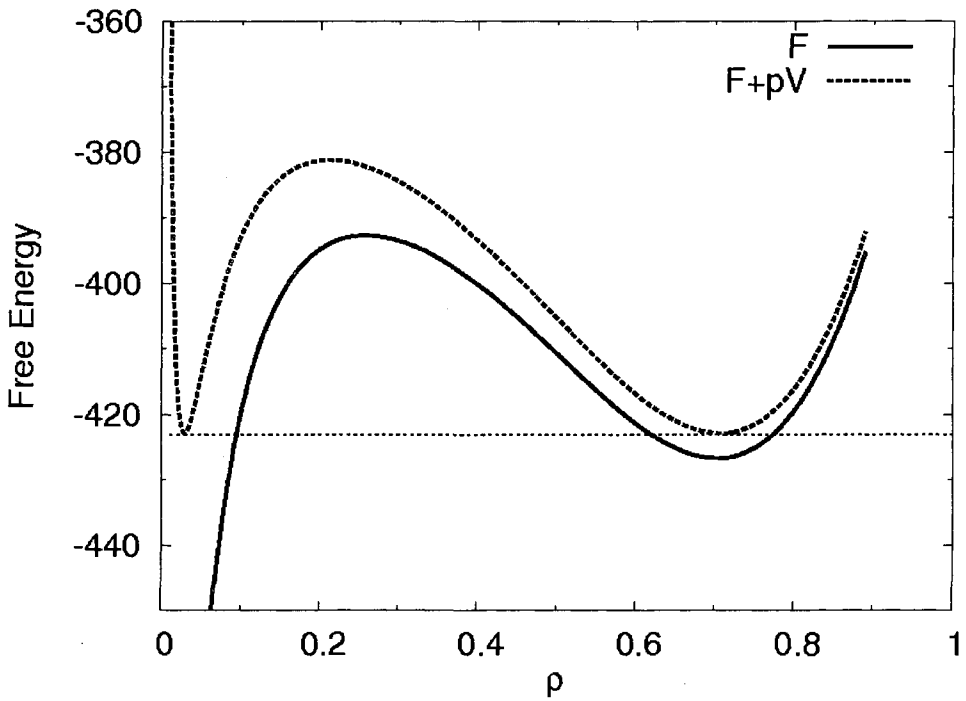


Figure 2. Free energy profile under the liquid-gas coexistence pressure. The solid line gives the Helmholtz free energy; the dashed line gives the Gibbs free energy profile under the coexistence pressure p.

Using the above method, we performed simulations under different temperatures, ranging $T = 0.85$ to $T = 1.20$ with an increment 0.01. To accurately determine the position of coexistence densities, increments for sampling densities were 0.002 and 0.0005 around the liquid and gas coexistence densities, i.e., ρ_+ and ρ_- respectively, whereas the transition region was filled by a larger increment 0.005. Typically about 300 volume sampling points were used in a single simulation. The phase diagram is shown in Fig. 3.

Once simulation data from different temperatures are collect, we can use the following relation

$$\rho_{\pm} - \rho_c \sim a|T_c - T| \pm b|T_c - T|^{\beta},$$

(where β is a critical exponent $\beta = 0.3258$ [19], not to be confused with the reciprocal temperature) to extrapolate the critical temperature T_c and the critical density ρ_c under the corresponding power-law regions. The estimated critical temperature T_c and critical density ρ_c were 1.304 and 0.315, respectively, see Fig. 4. The results from this small system are consistent with those from simulations on larger systems ($T_c = 1.3123 \pm 0.0006$ and $\rho_c = 0.3174 \pm 0.0006$) [20].

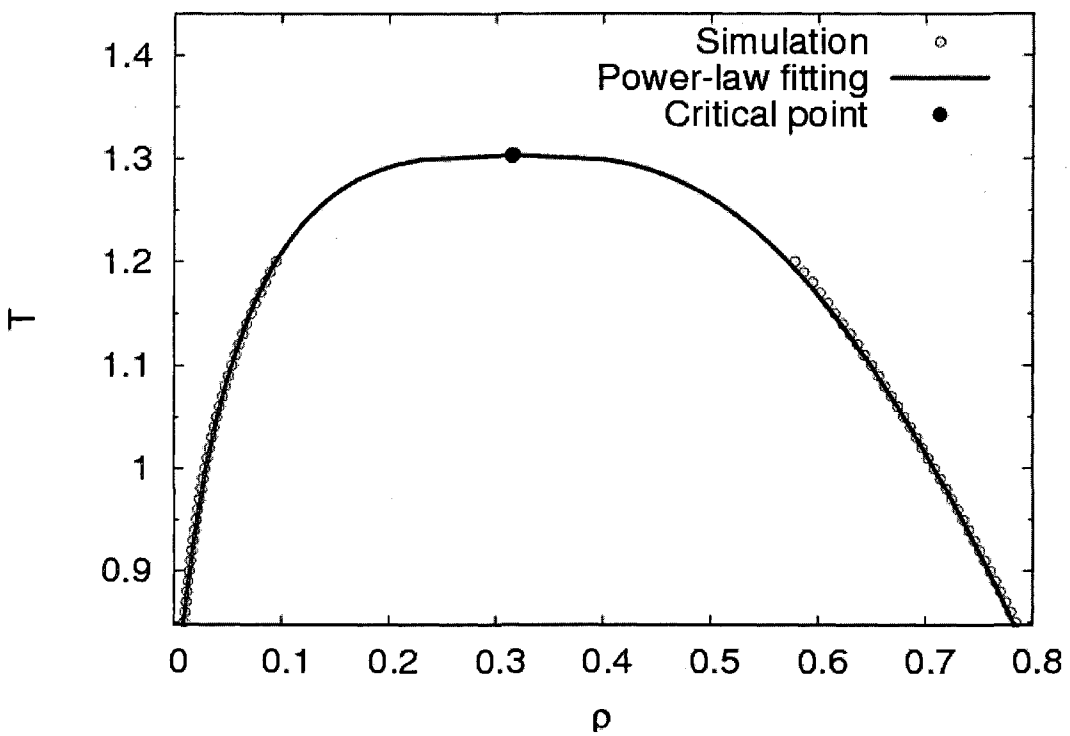


Figure 3. Phase diagram for the 108 LJ system. The empty circles are from simulations; the solid line is from power-law fitting, and the estimated critical point is marked by a solid circle.

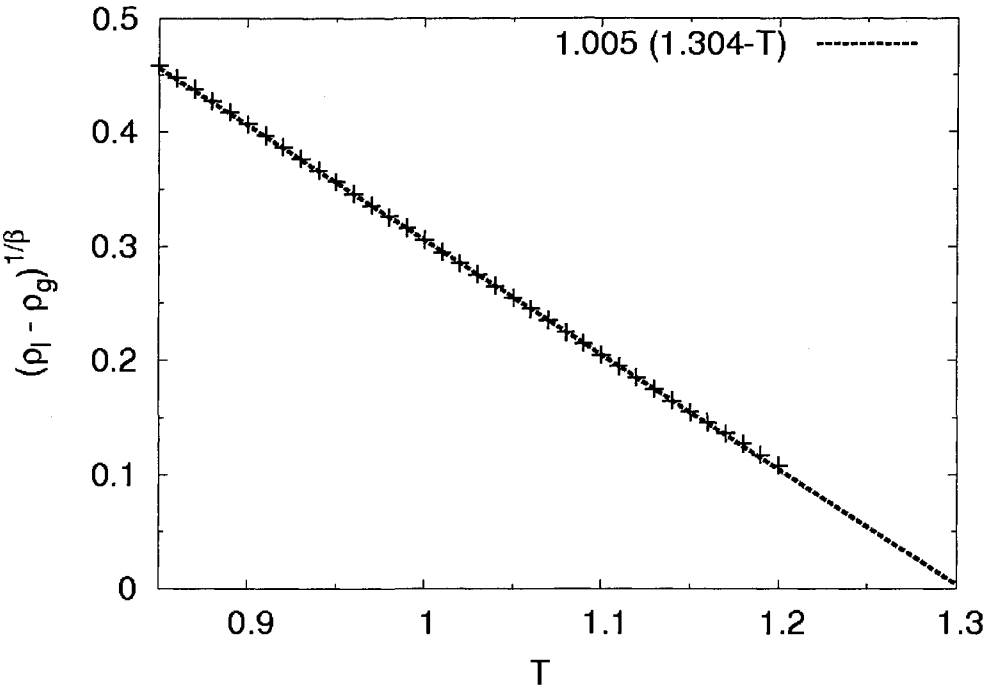


Figure 4. Critical temperature of the Lennard Jones system.

C. Off-lattice model proteins

Since the method is based on an expanded ensemble with multiple temperatures, it can be used to enhance sampling by overcoming energy barriers (a molecular system is often trapped in a local configurational space under a low temperature in the presence of energy barriers). Here we present an application in searching the ground state of a model protein, i.e., the configuration with the lowest potential energy. The model protein studied here is the *AB* off-lattice model [21-23]. The model specify polymers of only two kinds of residues *A* (hydrophobic) and *B* (hydrophilic) with Fibonacci sequences constructed as

$$S_0 = A,$$
$$S_1 = B,$$
$$S_{k+2} = S_k + S_{k+1}.$$

Here, the '+' represent a string concatenation operation. For example, $S_2 = S_0 + S_1 = AB$, $S_3 = S_1 + S_2 = BAB$, $S_4 = S_2 + S_3 = ABBAB$,.... The common used sequences are listed in Table II.

Table II. Sequence of *AB* model proteins

	# of res.	Sequence
S_6	13	*ABBABBABABBAB*
S_7	21	*BABABBABABBABBABABBAB*
S_8	34	*ABBABBABABBABBABABBABABBABBABABBAB*
S_9	55	*BABABBABABBABBABABBABABBABBABABBABBABABBABABBABABBABABBABABBAB*

Two types of energy functions were proposed for this model [21-23]. In both cases the bond length between two successive residues are fixed to unity. For the first type (model I), the energy function is composed of two kinds of interactions [21]

$$E = \sum_{i=2}^{N-1} 0.25(1+\cos\theta_i) + 4 \sum_{i,\, j,\, i\le j+2}^{N} \left(\frac{1}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right). \tag{8}$$

The first term on the right hand side of Eq. (8) describes a bond-bending energy, where θ_i is the bond angle formed by the three nearby residues $i - 1$, i, and $i + 1$. The second term on the right hand side of Eq. (8) describes the van de Waals long range interaction. The coefficient C_{ij} depends on the residue types (i.e., A or B) of the two involving residues, and is equal to +1, +0.5 and −0.5 for AA, BB and AB pairs respectively. This model can be applied to two dimensional systems as well as three dimensional systems.

The second type of energy function (model II) uses an additional energy term to describe the contribution from torsional angles [22],

$$E = \sum_{i=2}^{N-1} \cos\theta_i - \frac{1}{2} \sum_{i=2}^{N-2} \cos\psi_i + 4 \sum_{i,\, j,\, i\le j+2}^{N} C'_{ij} \left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}} \right). \tag{9}$$

Here the second term describes the torsional energy, where ψ_i is the angle formed by the bond between the $i - 1$th and ith residues and that between the $i + 1$th and $i + 2$th residues. The C'_{ij} is +1 for an AA pair and +0.5 for AB and BB pairs. Generally the energy function has no long range repulsion and therefore yields more globular structures.

Due to the complexity caused by bond constraints, molecular dynamics (MD) is a more suitable option than Monte Carlo (MC) in this case. Here we used a constant-temperature MD to generate energy moves, with SHAKE [24] for fixing the bond lengths. The thermostat we used here is a length-5 Nosé-Hoover chain [3-5]. Since the temperature of the system is not fixed, we used force-scaling [25] to dynamically adjust the instantaneous temperature [13]. In force scaling, the temperature of the thermostat is fixed at T_0, but the magnitude of the force is the scaled according to the actual temperature T as

$$\mathbf{f} = -\left(\frac{T_0}{T}\right)\nabla U .$$

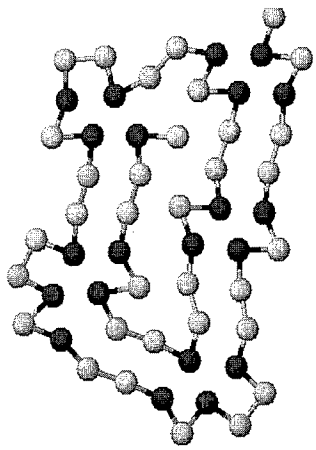
Here the thermostat temperature T_0 was set to 0.5. The scheme essentially scales the potential energy and puts the effective temperature for the potential energy to the targeting value T. Note, however, the scheme only guarantees a canonical distribution for the potential part of the energy, because the kinetic energy is not scaled.

To maximize the tempering frequency, we make a trial of temperature move after every MD step. The simulation was used to find ground states of AB protein models [21, 26]. We were able to find all known ground states [13, 22, 27, 28] as well as a few new configurations with lower energy. Table III lists the new lowest energy values and Fig. 5 shows the corresponding configurations. For example, comparing with the recent results from the statistical temperature method [13], in Fig. 5(a), the new ground state of the 2D 55mer has a different topology in the two inner strands; in Fig. 5(c), the new ground state of 3D 55mer (Model 1) has a more compact configuration.

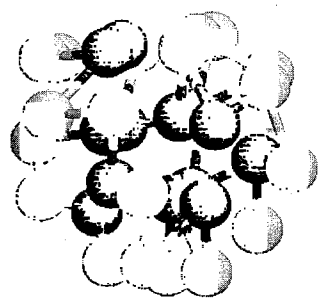
Table III. Lowest energies of AB proteins with Fibonacci sequences. Results were compared with those from the annealing contour Monte Carlo (ACMC), the energy landscape paving (ELP), the conformational space annealing (CSA), and the statistical temperature molecular dynamics (STMD).

Protein Models	ACMC	ELP	CSA	STMD	This work
2D, 55mer, Model I	-18.7407		-18.9110	-18.9202	-19.2570
3D, 55mer, Model I		-42.438	-42.3418	-42.5789	-44.8765
3D, 34mer, Model II	-94.0431	-92.746	-97.7321		-98.3571
3D, 55mer, Model II	-154.505	-172.696	-173.9803		-178.1339

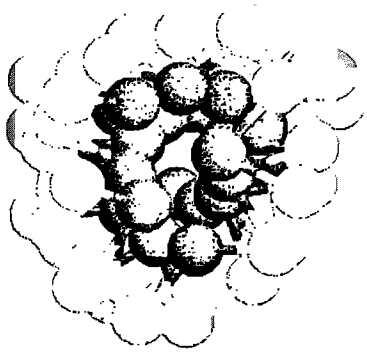
(a)



(b)



(c)



(d)

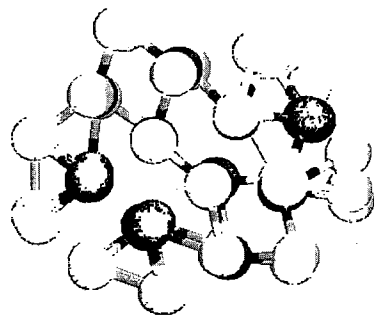


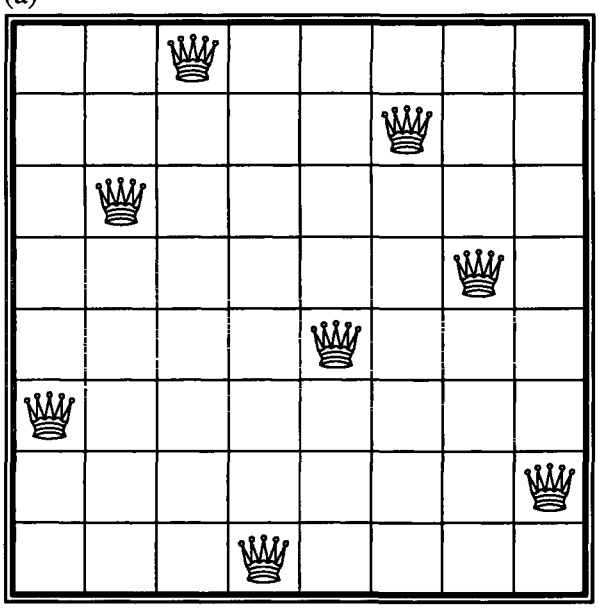
Figure 5. The lowest-energy configurations of AB proteins. (a) 2D, 55mer, model I; (b) 3D, 34mer, model II; (c) 3D, 55mer, model I; (d) 3D, 55mer, model II.

IV. Application in counting solutions

The method can be used in a different context: to counting solutions of constraint-satisfaction problems. In each constraint-satisfaction problem, one has a set of constraints. The objective is to calculate the number of ways of satisfying all the constraints.

In the following, we consider two typical problems. The first problem is the N-queens problem, in which the constraints are to avoid the N queens on an $N \times N$ chessboard to attack each other, see Fig. 6(a). The second problem is the Latin square problem, in which the objective is to fill an $L \times L$ table using L different symbols such that in every row or column, each number only occurs once, see Fig. 6(b).

(a)



(b)

8	7	6	5	4	3	2	1
7	6	5	4	3	2	1	8
6	5	4	3	2	1	8	7
5	4	3	2	1	8	7	6
4	3	2	1	8	7	6	5
3	2	1	8	7	6	5	4
2	1	8	7	6	5	4	3
1	8	7	6	5	4	3	2

Figure 6. (a) The N-queens problem (N = 8). (b) The L×L Latin square problem (L=8).

Due to their simplicity, the two problems are often used as standard benchmark tests. Many heuristics and combinatorial methods are developed to search for one or a few of their solutions, e.g., the minimizing conflicts algorithm [29], dynamic programming [30] and the

iterated map method [31]. However, to count all solutions is a more challenging task. A complete enumeration in general can only handle systems of a relatively small size (even with the help of a computer) because the number of solutions grows exponentially with the system size. For the N-queens problem, the largest system size solved to date is $N = 25$, where there are about 2.21×10^{15} solutions according to the result of a recent enumeration [32]. For Latin square problem, the largest system size solved to date is $L = 11$, where there are about 7.77×10^{47} solutions [33].

An alternative approach is to calculate the ratio between the number of solutions of the original problem and that of a simplified problem. If we know the exact number of solutions of the simplified problem, then the number of solutions of the original problem can be deduced.

Following this thought, we can apply our method to counting solutions if a special thermodynamic system can be constructed in such a way that its partition function corresponds to the numbers of solutions of the two problems. We describe such a construction in the following.

To connect the original problem (denoted as O) with the simpler problem (denoted as S), we carefully choose the problem S to be a generalized version of the problem O such that every solution of the problem O is a solution of the problem S. Here, the simpler problem has typically fewer constraints, and hence more (but easier-to-find) solutions. We then perform a Monte Carlo simulation in the configurational space spanned by all solutions of the problem S to compute the ratio of solutions of O and S.

To recognize a solution of the problem O, we use an energy function E that is nonnegative everywhere and is zero if and only if the configuration is a solution of the problem O, see Fig. 7. Thus we only need to count zero-energy configurations.

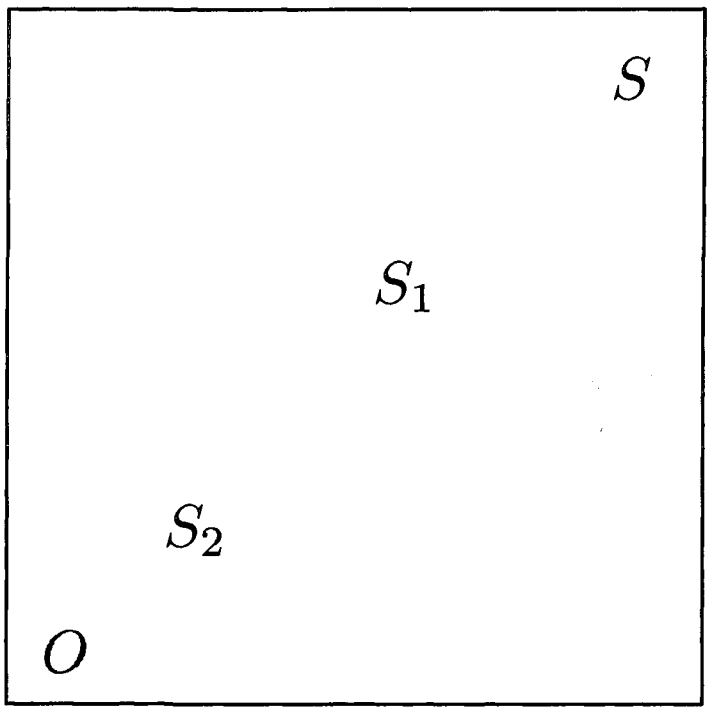


Figure 7. A schematic illustration of the method. To calculate the ratio between the number of solutions of the original problem O (the darkest area) and that of the simplified problem S (the whole square), a few intermediate problems S_i are introduced.

Since the numbers of solutions of O and S usually differ by many orders of magnitude for a large system, the ratio of the two becomes too small to be computed accurately. Therefore we need a set of intermediate problems $\{S_i\}$, each of which is associated with a reciprocal temperature β_i, which weights a configuration according to its energy E as $\exp(-\beta_i E)$. The weighted sum of solutions is reduced to the partition function $Z(\beta_i) = \sum \exp(-\beta_i E)$. Note the partition function has an interpretation of the number of

solutions in two extreme cases: the number of solutions for the problem S corresponds to the partition function at $\beta = 0$, and that for the problem O is the partition function at $\beta \to \infty$, where only zero-energy terms survive in the sum. Although several Monte Carlo methods were previously used to infer the partition function [34-36], these methods failed to be applied to large systems. The current method is naturally perfect for this purpose since it can directly and efficiently calculate the partition function.

A. *N*-queens problem

As the first application of the method, let us consider the *N*-queens problem. The *N*-rooks problem can serve as the problem S, where queens function as rooks such that they can attack each other only horizontally and vertically, but not diagonally. The problem S is a trivial one: each of its solutions corresponds to a permutation of the N column indices, because the row constraints are satisfied by placing only one rook on each row while the column constraints are satisfied by placing rooks from different rows at different columns.

We now specify the energy function that connects the simple problem and the original one. For each diagonal *d*, we count the number of resident queens as C_d. The energy of that diagonal is $E_d = C_d - 1$ for $C_d > 0$, or 0 if the diagonal is empty. The energy of the whole system is a sum of the energy of all diagonals. A zero-energy configuration clearly corresponds to a solution, where no diagonal contains more than one queen.

We used the energy move introduced by Sosic and Gu [37] to sample the configurational space. In each Monte Carlo step, we randomly choose two rows and try to swap their column indices of the queens there. Note, after a swap the horizontal and vertical constraints are still satisfied. Thus these swaps can be used to perform sampling on the

configurational space of the problem S. In Table IV, we listed the number of solutions for systems of several typical sizes. The simulation time of the final run is measured by sweeps (the number of Monte Carlo steps per queen). To date, the largest system whose exact number of solutions is known is $N = 25$ [32]; in which case our relative error is only 5×10^{-5}. The results on small systems serve as a check of our method. There is a dispute about the number of solution for $N = 24$. An alternative calculation [38] gives 226,732,487,925,864 solutions instead of the value 227,514,171,973,736 used in Table IV. Our long-time simulation result 2.2751×10^{14} clearly supports the latter result. More importantly, our method can be applied to larger systems, to which traditional counting algorithms fail to apply due to astronomically large numbers of solutions. In the largest system, there are about 1.32×10^{31560} solutions for N=10000 (in which case we used 82 temperatures from $\beta = 9.2$ to $\beta = 0$). The results on large systems are shown in Fig. 8. Our linear fitting result shows that for large systems $N > 100$, the number of solutions Q_N satisfies an empirical formula

$$N!/Q_N \approx 0.391\times2.57025^N,$$

where the maximal fitting error is less than 0.02 in this range.

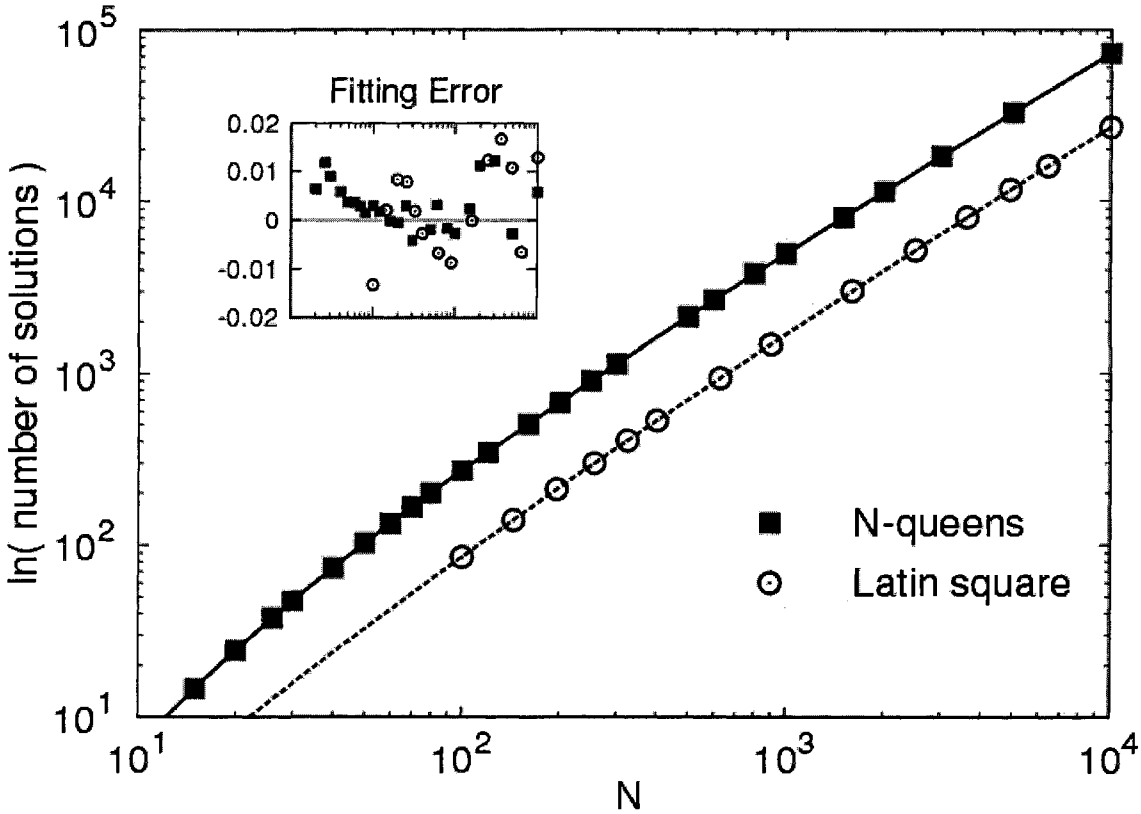


Figure 8. The numbers of solutions of the N-queens problem Q_N and that of the Latin square problem S_L versus the system size N (for a Latin square $N = L \times L$). The inset shows the error of fitting the formulas to the numerical results.

Table IV. The numbers of solutions Q_N of the N-queens problems, where N is the number of queens. The first six significant digits of the exact results are displayed in the last column for comparison.

N	Simulation cost (sweeps)	Q_N	Exact value
21	4×10^{10}	3.1468×10^{11}	3.14666×10^{11}
22	5×10^{10}	2.6910×10^{12}	2.69101×10^{12}
23	4×10^{10}	2.4234×10^{13}	2.42339×10^{13}
24	1×10^{11}	2.2751×10^{14}	2.27514×10^{14}
25	1×10^{11}	2.2080×10^{15}	2.20789×10^{15}
26	1×10^{11}	2.2319×10^{16}	
27	5×10^{10}	2.3489×10^{17}	
28	5×10^{10}	2.5645×10^{18}	
29	5×10^{10}	2.8899×10^{19}	
30	5×10^{10}	3.3731×10^{20}	
40	2×10^{10}	8.273×10^{31}	
50	2×10^{10}	2.456×10^{44}	
100	1×10^{10}	2.392×10^{117}	
200	1×10^{10}	2.041×10^{293}	
500	1×10^{10}	3.219×10^{929}	
1000	5×10^{9}	1.094×10^{2158}	
2000	2×10^{9}	9.45×10^{4915}	
5000	1×10^{9}	1.46×10^{14276}	
10000	1×10^{9}	1.32×10^{31560}	

B. Latin square problem

We now turn to the second problem: the Latin square problem. For convenience we choose 1,2,...,L as the L different symbols to fill the $L \times L$ table. To construct a problem S, we remove the constraints for columns, i.e., we no longer require each symbol to occur once in a column, while retaining the constraints for rows. Thus different rows act independently. The constraints for symbols within a row being mutually different require that each row configuration is a permutation of the L symbols. Thus there are $L!$ different arrangements for each individual row, and the total number of solutions of the problem S is $(L!)^L$ due to the mutual independence of different rows.

The energy function is the following. For any two rows i and j, if they share the same symbol on the kth column, they contribute +1 to the total energy, i.e.,

$$E = \sum_{i<j,k} \delta(s_{ik}, s_{jk}) .$$

Here s_{ik} is the symbol at the ith row and jth column; $\delta(a,b)$ is +1 if the two symbols a and b are the same, zero otherwise; the two indices i and j enumerate over every pair of rows, k every column. A Metropolis way of sampling the system is to randomly choose two positions on a row, and to try to swap their symbols. Similar to the previous case, these swaps preserves the constraints for columns within a row, thus is qualified as a sampler of the configurational space.

However, at a low temperature, the swap becomes inefficient due to frequent rejections. For example, at the lowest temperature $\beta = 8.4$ we used for the 100×100 system, the acceptance ratio of the Metropolis algorithm is less than 0.01%. To overcome the problem, we develop a reject-free cluster algorithm for this system and used it to generate

configurational changes. The cluster algorithm is of the same spirit of its counterpart on the Ising model [18]. It exploits the symmetry between any two symbols a and b, e.g., the system energy is unchanged if we exchange the two symbols at a suitable collection of rows (or a cluster).

A cluster is generated as the following. We first randomly choose two symbols a and b as well as a row index i, and add this row index into the cluster as a "seed". For each row in the cluster, we pick up the column j where the symbol s_{ij} is a, and search in other rows i' for the symbol b at the same column j, i.e., $s_{i'j} = b$. For each row i' found, we use a probability $1-\exp(-\beta)$ to add it into the cluster. Similarly, we pick up the column k where $s_{ik} = b$, and add every other row i'' where $s_{i''k} = a$ to the cluster using the same probability. This process is repeated until every row in the cluster is considered. An example of the cluster is shown in Fig. 9.

3	1	5	4	8	2	6	7
5	3	2	7	6	8	4	1
1	8	3	2	7	4	5	6
4	6	8	3	2	7	1	5
7	2	6	5	3	1	8	4
6	7	4	1	5	3	2	8
8	4	7	6	1	5	3	2
2	5	1	8	4	6	7	3

Figure 9. An example of a cluster generated from the bottom row is shown as the six marked cells. After it is generated, the symbols '5' and '7' within the cluster are exchanged.

The number of solutions of the Latin square problem is listed in Table V. We used Metropolis moves for small systems, but cluster moves for large systems at low temperatures. In this way we can access large systems, as shown in Fig. 8. The size of the largest system is 100 by 100, in which there are over 10^{11710} solutions. In this system, we used 85 temperatures from $\beta = 8.4$ to $\beta = 0$. We attempted to fit the number of solutions S_L to the formula

$$\ln(L!^L / S_L) \approx L^2 (0.99642 + 42.3252/L - 35.6031/L^2)/(1 + 48.9874/L + 149.97/L^2),$$

Here, the maximal fitting error is 0.03.

Table V. The numbers of solutions S_L of the $L \times L$ Latin squares. The first six significant digits of the exact results are displayed in the last column for comparison. We used the cluster algorithm for the last two systems.

Size	Simulation cost (sweeps)	S_L	Exact value
10×10	1×10^{10}	9.988×10^{36}	9.9824×10^{36}
11×11	1×10^{10}	7.773×10^{47}	7.7697×10^{47}
12×12	1×10^{10}	3.102×10^{60}	
13×13	1×10^{10}	7.500×10^{74}	
14×14	1×10^{10}	1.266×10^{91}	
15×15	1×10^{10}	1.728×10^{109}	
16×16	5×10^{10}	2.161×10^{129}	
17×17	5×10^{10}	2.804×10^{151}	
18×18	5×10^{10}	4.256×10^{175}	
19×19	5×10^{10}	8.354×10^{201}	
20×20	2×10^{10}	2.365×10^{230}	
50×50	1×10^{10}	5.66×10^{2250}	
100×100	1×10^{10}	1.55×10^{11710}	

C. Anomaly in the heat capacity

Besides being able to count the number of solutions, we find an interest property in the Latin square systems. The heat capacity C of the system also shows an interesting anomaly at a low temperature. As the system size increases, the system gradually develops two separate maxima, see Fig. 10. The anomaly of the heat capacity is a manifestation of highly disordered low energy landscape. The valley between the two maxima appears to coincide with the location where the system is most "difficult". Here the difficulty can be measured by the fraction of clusters that include all rows. This is because the system state is essentially unchanged if we exchange two symbols everywhere. An all-row cluster also indicates a configuration with the strongest correlation between different rows. As shown in the inset of Fig. 10, for the 100×100 Latin square, the maximum fraction 0.06 occurs at T = 0.14, where the heat capacity hits its local minimum.

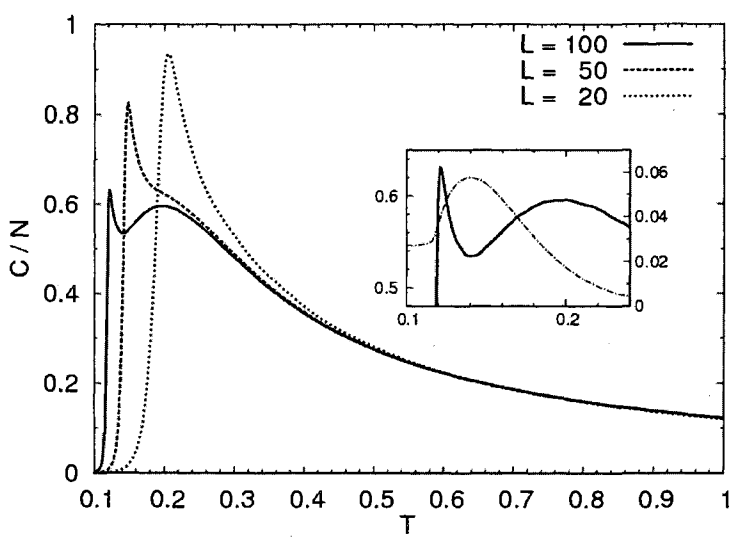


Figure 10. Heat capacity C_V per site of Latin squares versus temperature T. The heat capacity develops two peaks as one increases the system size. The inset shows that the valley between the two maxima of the heat capacity for the 100 × 100 system (the solid line, the left axis) corresponds to where the fraction of percolated clusters (the dash dot line, the right axis) reaches the maximum.

C. Discussion

We conclude the chapter with a few remarks. Essentially, in this type of application, the original problem O is generalized to a less-constrained problem S and its partition function is calculated. We achieved a high sampling efficiency by using collective Monte Carlo moves: in the N-queens problem, the column indices of the queens are swapped rather than altered individually; similarly, in the Latin square problem, symbols within a row are always exchanged. These collective moves not only help to conduct an efficient sampling at low temperatures, but also reduce the sampling space by making the problem S as close to the problem O as possible. In the N-queens problem, the use of the swap move reduces the sampling space of the problem S from N^N solutions to $N!$ solutions, while in the Latin square problem the sampling space is reduced from $L^{L\times L}$ to $(L!)^L$.

V. Theory on convergence

As indicated in previously, our method relies on a recursive updating scheme that constantly changes the acceptance probability as well as the distribution. Therefore it cannot be simply classified as a regular Monte Carlo method targeting a fixed distribution.

In this chapter, we propose a simple theory to understand the convergence of the method. Previously several theories were proposed for the Wang-Landau algorithm [39, 40]. However these theories limited themselves in the case of a small modification factor α, and left the convergence at a large modification factor unknown. We present our analyses based on an ensemble of identical systems. The approach not only simplifies the complicated mathematics involved in previous studies, but also gives a more accurate prediction on the accuracy of the updating scheme. The results also indicate the stability of the updating scheme by showing that the errors of the estimated partition function are well bounded even under a large modification factor.

A. Ensemble

We analyze the method using a special ensemble: a collection of identical systems experiencing the temperature transitions as well as the updating of the partition function. To study the convergence, we compute the evolution of the ensemble for systems sharing the same initial condition. After the ensemble enters a stationary state, it is also possible to calculate average quantities, which correspond to long time averages of a single system.

We first define the accuracy of the method in terms of calculating the partition function. Consider a system of two temperatures β and β'. Since one can uniformly shift

the partition function without affecting the partition function, we introduce a reduced parameter $X = [\ln \tilde{Z}(\beta) - \ln \tilde{Z}(\beta')] - [\ln Z(\beta) - \ln Z(\beta')]$. In terms of X, the acceptance probability can be written as

$$\mathrm{Acc}(\beta \to \beta') = \min\left\{1, \frac{\exp(-\beta' E)/Z(\beta')}{\exp(-\beta E)/Z(\beta)}\exp(X)\right\}$$

$$\mathrm{Acc}(\beta' \to \beta) = \min\left\{1, \frac{\exp(-\beta E)/Z(\beta)}{\exp(-\beta' E)/Z(\beta')}\exp(-X)\right\}.$$

Ideally, the factor X should be 0 after the convergence, i.e., if the estimated partition function is converged to the actual one. However during simulation, X changes each step due to the updating process, and therefore cannot be exactly 0. We shall use the ensemble averaged value of X to define the accuracy of the convergence scheme.

The value X determines the relative weight of visiting the two temperatures as well as the transition rates between them. Roughly, β' tends to receive more visits than β if $X>0$, less visits, if $X<0$. If we assume that by using an ever-changing acceptance probability, the disturbance to the Boltzmann distributions at the two temperatures is negligible, then the overall weight of visiting β' is $\exp(X)$ times that of visiting β. Similarly, the rates of mutual transitions r (from β to β') and r' (from β' to β) are expected to satisfy a detailed balance (on a coarse-grained level)

$$\frac{r(X)}{r'(X)} = \exp(X). \tag{10}$$

Particularly, $r(0) = r'(0) = r_0$. Although we shall not assume this condition in our general treatment, Eq. (10) usually yields simple examples with qualitatively correct results.

The current temperature is labeled by another variable Y, which takes +1 when the temperature is β, -1 when the temperature in β'. In this way, the pair (X,Y) furnishes a representation of the system state.

Instead of studying the evolution of X and Y of a single system, we consider the time evolution of the ensemble averages $\langle X \rangle$ and $\langle Y \rangle$. During each attempted temperature transition, the amount of updating X is αY, i.e., X is increased by α for a system in β, or decreased by α for a system in β'. Therefore,

$$\frac{d\langle X \rangle}{dt} = \alpha \langle Y \rangle, \tag{11}$$

where the time interval between two attempted temperature transitions is denoted by dt. The probability of transition can be summarized as $\frac{1}{2}[(r+r')+(r-r')Y]$, which becomes r if the temperature is β, or r' if it is β'. Since a successful temperature transition changes Y to $-Y$, the incremental change of $\langle Y \rangle$ during dt is

$$\frac{d\langle Y \rangle}{dt} = -\langle (r+r')Y \rangle - \langle (r-r') \rangle. \tag{12}$$

Eqs. (11) and (12) provide the basis to study the time evolution of $\langle X \rangle$ and $\langle Y \rangle$, e.g., properties of the equilibrium state of the ensemble, and the rate of approaching the equilibrium state. Although the exact solution is difficult to obtain, useful information can be extracted from the mean-field approximation, especially about how fast the updating scheme converges $\langle X \rangle$ from an arbitrary initial condition.

B. Mean-field approximation

As a mean-field theory, we replace all occurrence of X and Y in Eq. (12) by their averages $x = \langle X \rangle$ and $y = \langle Y \rangle$ respectively. This substitution greatly simplifies the problem because now the evolution of the ensemble is reduced to that of a representative system, whose motion can be described by two variables, or a point on the plane (x, y),

$$\begin{cases} \dfrac{dx}{dt} = \alpha\, y \\ \dfrac{dy}{dt} = -[r(x) + r'(x)]y - [r(x) - r'(x)] \end{cases} \tag{13}$$

The fixed point of the equations (where all ensemble averages stay as constants) is located at $x = y = 0$. In other words, the mean-field theory indicates that in the equilibrated ensemble, the populations of being in the two temperatures are equal, and $\langle X \rangle$ is zero.

The stability of the fixed point, i.e., whether the system will eventually be attracted to the fixed point if it starts from somewhere nearby, can be analyzed by linearly expanding Eqs. (13) at its surroundings

$$\begin{cases} \dfrac{dx}{dt} = \alpha\, y \\ \dfrac{dy}{dt} = -r_0[2y + (b + b')x] \end{cases},$$

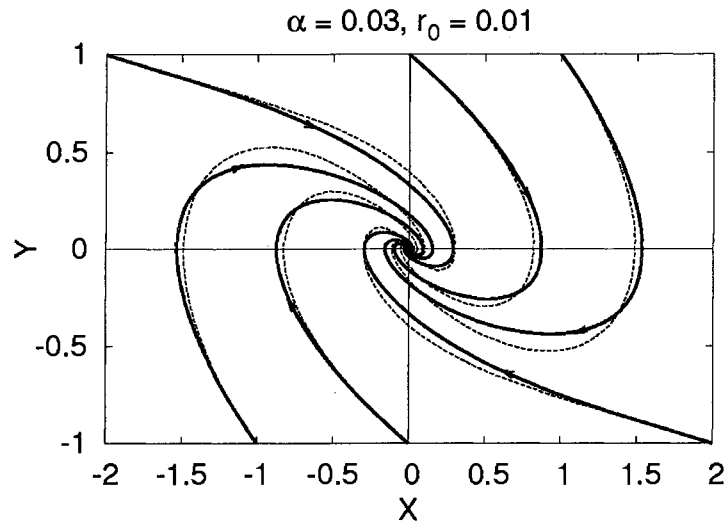
where we used the first-order expansion of the rates $r(x) \approx r_0(1 + bx)$, $r'(x) \approx r_0(1 - b'x)$. The eigenvalues of the system are $-r_0 \pm \sqrt{r_0^2 - \alpha(b + b')r_0}$. Since both eigenvalues are negative or have a negative real part, the fixed point is stable at its surroundings.

What happens at point far from the fixed point? The answer is that the fixed point is still attractive and the system is drawn to the fixed point with a much higher driving force. In other words, the updating scheme is globally stable. To see this, consider the following

Liapunov function $V(x,y)=R(x)+\frac{1}{2}\alpha y^2$ (where $R(x)=\int_0^x [r(z)-r'(z)]\,dz$), which satisfies $dV/dt=-\alpha(r+r')y^2<0$ everywhere $y\neq 0$. If at some point the system becomes unstable, there will be a stable closed orbit, and V must not change after the system travels along the orbit one period and returns to the original place. On the other hand, $\Delta V=\int_0^T (dV/dt)\,dt$ is negative after the integration along the orbit. The contradiction rules out the existence of any closed orbits. It is also clear that the Liapunov function generally shrinks faster where the absolute value of x or y is large.

We can visually study the time evolution of the ensemble by using the phase portrait, see Fig. 11. We used $r(x)=r_0\exp(x/2)$ and $r'(x)=r_0\exp(-x/2)$, which strictly preserves the detailed balance condition Eq. (10) with the coefficients of the linear expansion $b=b'=1/2$. Each solid line represents a trajectory of the representative system that evolves according to the mean-field equations Eqs. (13). The arrow indicates the direction of time evolution. By comparison, the corresponding trajectories of an ensemble consists of 10^5 independent systems from the same initial condition are shown as dashed lines. It can be seen that the mean-field equation captures major features of the ensemble evolution in this case.

(a)



(b)

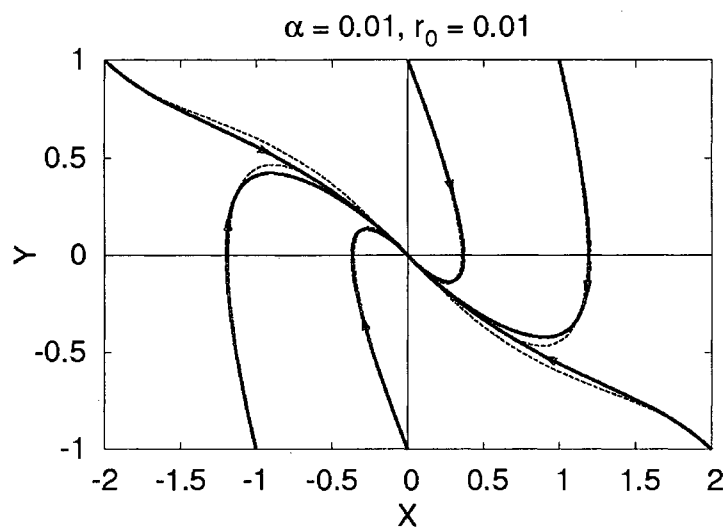


Figure 11. Phase portraits of an ensemble-average system. The solid lines are obtained by integrating the mean-field equations, Eqs. (13). The dashed lines represent the actual evolution of an ensemble of 10^5 systems. In both cases, the transition rate are $r(x) = r_0 \exp(x/2)$ and $r'(x) = r_0 \exp(-x/2)$. (a) $\alpha = 0.03$, $r_0 = 0.01$, the trajectory is a spiral. (b) $\alpha = r_0 = 0.01$, the critical case.

In Fig. 11(a), the conditions are $\alpha = 0.03$ and $r_0 = 0.01$, which represent a typical case for $(b + b')\alpha > r_0$. The trajectory is a damped oscillation between the two temperatures. Note, a typical system tends to wander to a large X region even if it start at $X = 0$. Mathematically, this oscillation is due to that $\sqrt{r_0^2 - \alpha(b + b')r_0}$ becomes an imaginary number. Qualitatively, it can be understood as the following. Before a typical system transits to the other temperature, it on average stays at a temperature $1/r_0$ steps, during which the value X is updated by α / r_0. If α / r_0 is sufficiently larger, the amount of updating leads to a significant driving force to repel the system from the current temperature. In this way, the system is forced to oscillate between the two temperatures.

In Fig. 11(b), the conditions are $\alpha = 0.01$ and $r_0 = 0.01$, which represent a typical case for $(b + b')\alpha = r_0$ [the behavior of the $(b + b')\alpha < r_0$ case is similar]. Quite different from the previous case, the ensemble tends to collapse onto the same curve, then monotonically approaches the fixed point, with an approaching rate $r_0 - \sqrt{r_0^2 - \alpha(b + b')r_0} \approx (b + b')\alpha / r_0$.

C. Series expansion

We can study the averages such as $\langle X \rangle$ (which represents the systematic error) and $\langle X^2 \rangle$ (random fluctuation), after the ensemble enters the stationary state, or the fixed point. The location of the fixed point obtained from the mean-field theory is inaccurate because we replaced all averages of products, such as $\langle X^2 \rangle$ and $\langle XY \rangle$, by products of averages, such as

$\langle X\rangle^2$ and $\langle X\rangle\langle Y\rangle$. Therefore, we shall make a correction by treating $\langle X^2\rangle$ and $\langle XY\rangle$ as independent variables just as $\langle X\rangle$ and $\langle Y\rangle$ in the following calculation.

We proceed under the condition $\alpha << 1$ and $r_0 << 1$ (however, the ratio of the two can still be comparable with 1.0). First, we generalize Eq. (11) to an arbitrary function $f(X)$,

$$\frac{d\langle f(X)\rangle}{dt} \approx \alpha\langle f'(X)Y\rangle .$$

At the fixed point, $f(X)$ cannot change, hence

$$\langle f'(X)Y\rangle = 0 . \tag{14}$$

Since $f(X)$ is an arbitrary function, Eq. (14) asserts that the product of Y and any function, e.g., $\langle XY\rangle$ or $\langle X^2Y\rangle$ averages to zero in the first order approximation. Another convenient fact is that $Y^2 \equiv 1$, since Y is either 1 or -1. Hence, $Y^5 \equiv Y^3 \equiv Y$, and $Y^6 \equiv Y^4 \equiv 1$. Consequently, any product that involves Y can be ignored, and we only need to calculate averages of X.

Similarly, let us consider another condition $d\langle Y\exp(\lambda X)\rangle / dt = 0$, which leads to

$$\langle (r - r')\exp(\lambda X)\rangle = \lambda\alpha\langle \exp(\lambda X)\rangle .$$

By expanding both sides of the equation in Taylor series of λ, and compare the coefficient of each λ^n, we obtain

$$\langle (r - r')X^n\rangle = n\alpha\langle X^{n-1}\rangle . \tag{15}$$

These equations completely define the different moments of X. We now consider the function $S(X) = r_0 X / [r(X) - r'(X)]$, where $r_0 = r(0) = r'(0)$. This function allows us to

calculate moments of X explicitly as $r_0\langle X^n\rangle=\langle(r-r')S(X)X^{n-1}\rangle$. By expanding $S(X)$ as a Taylor series

$$S(X)=\frac{r_0X}{r(X)-r'(X)}=s_0+s_1X+s_2X^2+s_3X^3+\cdots, \tag{16}$$

and applying (15), a set of linear equations between $\langle X^n\rangle$'s can be obtained,

$$r_0\langle X^n\rangle=\alpha\{(n-1)s_0\langle X^{n-2}\rangle+ns_1\langle X^{n-1}\rangle+(n+1)s_2\langle X^n\rangle+\cdots\}. \tag{17}$$

If we truncating the series at s_2, the leading two moments can be expressed in closed form [this is why we choose to solve Eqs. (17) instead of Eqs. (15)]

$$\langle X\rangle=\frac{a^*s_1}{1+2a^*(-s_2)}$$

$$\langle X^2\rangle=\frac{a^*}{1+3a^*(-s_2)}\left(s_0+\frac{2a^*s_1^2}{1+2a^*(-s_2)}\right), \tag{18}$$

where the ratio $a^*=\alpha/r_0$. If we further include s_3 and assume the odd-order coefficients s_1 and s_3 are small, the result for $\langle X\rangle$ is modified as

$$\langle X\rangle\approx\frac{a^*}{1+2a^*(-s_2)}\left(\frac{s_1+3a^*(s_0s_3-s_2s_1)}{1+3a^*(-s_2)}\right). \tag{19}$$

For a small a^*, the random fluctuation $\sqrt{\langle X^2\rangle}$ is proportional to $\sqrt{\alpha}$, while the average $\langle X\rangle$ is proportional α. This result agrees with previous analyses on the Wang-Landau algorithm.

For a large α, the theory gives qualitatively different behaviors. For example, we expect that $\langle X\rangle$ and $\langle X^2\rangle$ would manifest a saturation behavior at a large α, due to the fact that s_2 is usually negative (see Sec. D). It shows that the terminal value X (the error of the

estimated $\ln Z_\beta$) is properly bounded for a large α. Another example is that the $\langle X \rangle$ does not increase monotonically with α; it may even can change its sign at a particular a^*, if $s_0(s_3/s_1) - s_2 < 0$. This behavior can be seen from an example in Sec. E.

D. Properties of the transition rates

We now briefly discuss the meaning of the transition rates $r(X)$ and $r'(X)$. The rate r_0 represents the average rate of commuting between the two temperatures β and β'. Note r_0 is not equal to the acceptance ratio of temperature transitions (or the averaged acceptance probability for temperature transitions). This is because the acceptance ratio only counts the fraction of states that are capable of making a temperature transition, but it does not take into account the time of travelling between different parts of the energy distribution at a fixed temperature. Accordingly, r_0 is usually smaller than the acceptance ratio.

The series coefficients s_0, s_1, s_2, on the other hand, determines the dependence on X. Generally we expect that a positive X significantly boosts the magnitude of $r(X)$, while a negative X increases that of $r'(X)$. The purpose of the updating scheme is to exploit a faster transition rate at a large $|X|$. If the detailed balance condition Eq. (10) is strictly satisfied, s_0 must be unity. The coefficients s_1 and s_3 represents an asymmetrical difference between the rates $r(X)$ and $r'(-X)$. This can be easily seen: if $r(X) = r'(-X)$, the function $r(X) - r'(X)$ is an odd function, and $r_0 X / [r(X) - r'(X)]$ is an even function; thus the odd order terms in the Taylor-expansion vanish.

By contrast, the coefficient s_2 is usually negative for a sufficiently large temperature difference between β and β'. This is because transition probabilities from β to β' are expected to increase dramatically at a positive X, while transitions in the reverse direction are largely enhanced at a negative X. In mathematical terms, we have a) $r - r'$ should share the same sign with X; b) the magnitude $|r - r'|$ generally increases with $|X|$. Thus we expect that the third order coefficient of the Taylor expansion of $r(X) - r'(X)$ is strongly positive, which in turn leads to a negative coefficient s_2. For example if we assume the detailed balance condition Eq. (10) is strictly preserved, and the transition rates are perfectly symmetrical (i.e., $s_1 = s_3 = 0$), then $S(x) = (x/2)/\sinh(x/2) \approx 1 - x^2/24$, where $s_2 = -1/24$.

E. Numerical verification

We now test the above theory on the 32×32 Ising model. We used two pairs of temperatures that manifest different behaviors. The first pair of temperatures are $T = 2.5$ and $T = 3.0$; the second $T = 1.7$ and $T = 2.0$. The frequency of attempting temperature transitions is once every ten configurational sampling steps. In Fig. 12(a) and Fig. 12(b), we show $\langle X \rangle$ and $\langle X^2 \rangle$ versus α, respectively. For each temperature pair, we applied Eq. (19) and Eq. (18) to simultaneously fit the results of $\langle X \rangle$ and $\langle X^2 \rangle$ (solid lines) using four independent parameters: $s_0 / r_0 = 7.6 \times 10^3$, $s_1 / r_0 = 430$, $s_2 / r_0 = -106$, $s_3 / r_0 = -9.6$ for the first pair, and $s_0 / r_0 = 3 \times 10^3$, $s_1 / r_0 = -290$, $s_2 / r_0 = -62$, $s_3 / r_0 = 4.8$ for the second. Note, according to Eq. (16), one can only (and one only needs to) determine the ratios between s_i's and r_0, rather than their absolute values. If the detailed balance condition Eq. (10) is assumed to be correct, then $s_0 = 1$. A good agreement can be seen in Fig. 12. It is also obvious that

previous theories failed to apply for a moderately large α (dashed line). Particularly, as one can see from Fig. 12(a), the average $\langle X \rangle$ changes the sign from positive to negative in the first case, while it does not in the second case. This is because, in the first case, the quantity $s_0(s_3/s_1) - s_2$ is negative, while it becomes positive in the second case. In Fig. 12(b) we also observe a saturation of $\langle X^2 \rangle$ at a large α. A slight difference between our predictions and simulation results at a very large α is due to that Eq. (19) and Eq. (18) are simplified formulas under the assumption that s_1 and s_3 are small compared with s_2. This is however not entirely true in our testing cases. Despite this discrepancy, our theory successfully captures large-α behaviors that are unreachable from previous theories.

F. Optimal arrangement of α stages

According to Eqs. (18) and (19), the terminal errors of estimated partition function generally decrease with the modification factor α. Therefore it is necessary to eventually use a small α to obtain a good final accuracy. However, the updating is inefficient with a small α, due to that the rate of approaching the fixed point is proportional to α (as predicted by the mean-field theory). Thus the total number of steps needed to overcome an initial error ε is proportional ε/α. If the initial error is large, it is inefficient to start with a small α.

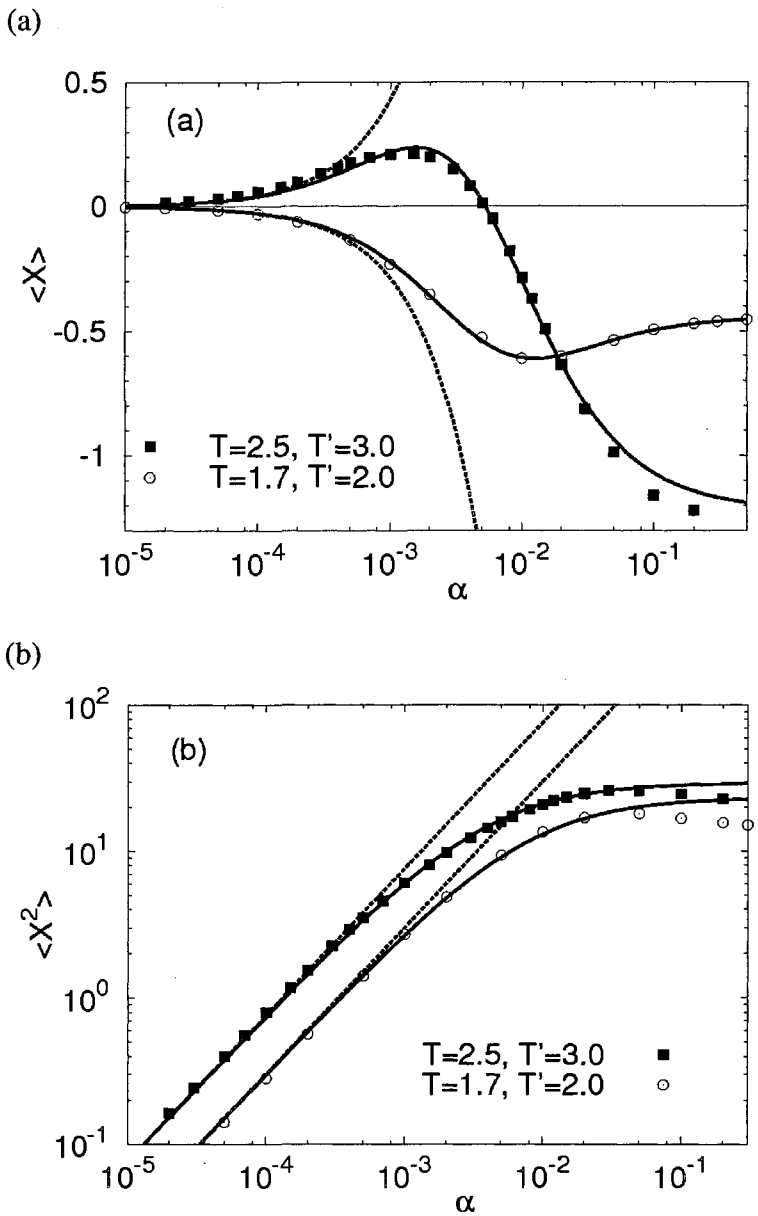


Figure 12. (a): $< X >$ (the systematic error) versus the modification factor α; the logarithmic scale is used in the α-axis. (b): $< X^2 >$ (the random fluctuation) versus α, the logarithmic scale is used in the both axes. Predictions from our theory (solid lines) and previous theories (dashed lines) are also shown.

A better way is to start with a large updating factor and to decrease it gradually to improve the estimate of the partition function. More specifically, we split the total amount of simulation time to several stages, each of which has a fixed updating factor α_n. The error is reduced to a relatively small level in larger α_n stages, and less simulation time is needed for later stages, where α_n is smaller.

We now discuss the optimal arrangement of these stages that gives the minimal time of reaching the final stage α_{fin}, which corresponds to a desired final accuracy. In the following, we assume that the random fluctuation is much larger than the systematic error, i.e., $\sqrt{\langle X^2 \rangle} >> \langle X \rangle$, and thus the contribution from systematic error is ignored.

Let us consider the characteristic time for X to reduce a typical error at the end of the previous stage $n-1$. The typical error ε_{n-1} from the previous stage, measured from $\sqrt{\langle X^2 \rangle}$, is roughly proportional to $\sqrt{\alpha_{n-1}}$ according to Eqs. (18). The minimal simulation time for stage n, should be proportional to $\sqrt{\alpha_{n-1}} / \alpha_n$. The total time for all stages is given by the sum $\sum_n \sqrt{\alpha_{n-1}} / \alpha_n$. By minimizing this quantity with respect to each α_n, we obtain a recursion relation,

$$\sqrt{\alpha_{n-1}} / \alpha_n = \tfrac{1}{2} \sqrt{\alpha_n} / \alpha_{n+1} .$$

The general solution to this recursion relation is $\alpha_n = c_0 (1/4)^n \exp(-c_1 / 2^n)$, where c_0 and c_1 are two constants. The second factor decays exponentially and is unimportant for the last few stages. The stable solution thus is a geometric series,

$$\alpha_n = \alpha_{n-1} / 4 . \qquad (20)$$

Note, the optimal solution gives a larger decrement factor 4 than the previously suggested values.

According to Eq. (20), the simulation time spent on stage n is proportional to $1/\sqrt{\alpha_n}$. We used this criterion previously by the fixing the simulation time of each α_n stage as

$$\tau_n = C/\sqrt{\alpha_n}. \tag{21}$$

An alternative criterion is to stop an α_n stage when the fluctuation of the temperature histogram (the number of visits to sampling temperatures) goes below certain cutoff percentage. This approach however tends to prematurely stops a stage and thus produces a larger error. The reason is that in the last few stages, the histogram is flattened faster than the partition function converges, i.e., the histogram can be sufficiently flat even if the partition function still contains a much larger error than its stationary value.

The difference of the two criteria can be seen in Fig. 13. In the numerical example on the 32×32 Ising model, we used two temperatures $T = 1.7$ and $T = 2.0$. We started with $\alpha_1 = 1.0$ and descend down to $\alpha_{14} = (1/4)^{13} \approx 1.5\times10^{-8}$. In Fig. 13(a), we show the error of the estimated partition function at the end of stage n versus α_n. In Fig. 13(b), we show the number of steps of spend on stage n (measured by Monte Carlo steps per spin) versus α_n. Results of fixing the simulation time of each α_n stage are represented by solid symbols, while those of using histogram flatness as the stopping criterion are represented by empty symbols. The results are averaged over a thousand independent simulations. In both cases, the average numbers of steps spent on different α_n stages satisfy Eq. (21). It is clearly seen that using fix simulation time for α_n each stage produces less terminal error than stopping a stage according to the histogram flatness, even by using less simulation steps. For example,

using 5% histogram flatness as stopping criterion, the terminal error of the last stage is 0.20, while the simulation steps used on the last stage is 7×10^4 sweeps. On the other hand, by fixing simulation time by $1/\sqrt{\alpha_n}$ sweeps, the terminal error is only 0.12, and the simulation steps spent on the last stage are 8×10^3 sweeps, which is less than that of using histogram flatness by an order of magnitude. Even if we increase the simulation time to $5/\sqrt{\alpha_n}$ sweeps (which increases the simulation time on the last stage to 4×10^4 sweeps), the terminal error is reduced to 0.037, which is less than one fifth of that of using 5% histogram flatness.

However, when the modification factor α is small, a longer time scale, the one for configurational sampling, should be considered. Thus the number of steps to overcome an error ε would be larger than the predicted value. At a very small α, the relation between the error and the simulation steps should scale as $\varepsilon \sim 1/\sqrt{t}$, as the scale of configurational sampling dominates. According to Eqs. (18), $\varepsilon \sim \sqrt{\alpha}$, we thus conclude that $\alpha \sim 1/t$. This result agrees with a recent suggestion for the terminal correction for the recursion, where the modification factor is modified in each step according to $\alpha = C/t$ (C is a constant) disregarding the flatness of the histogram [41-44].

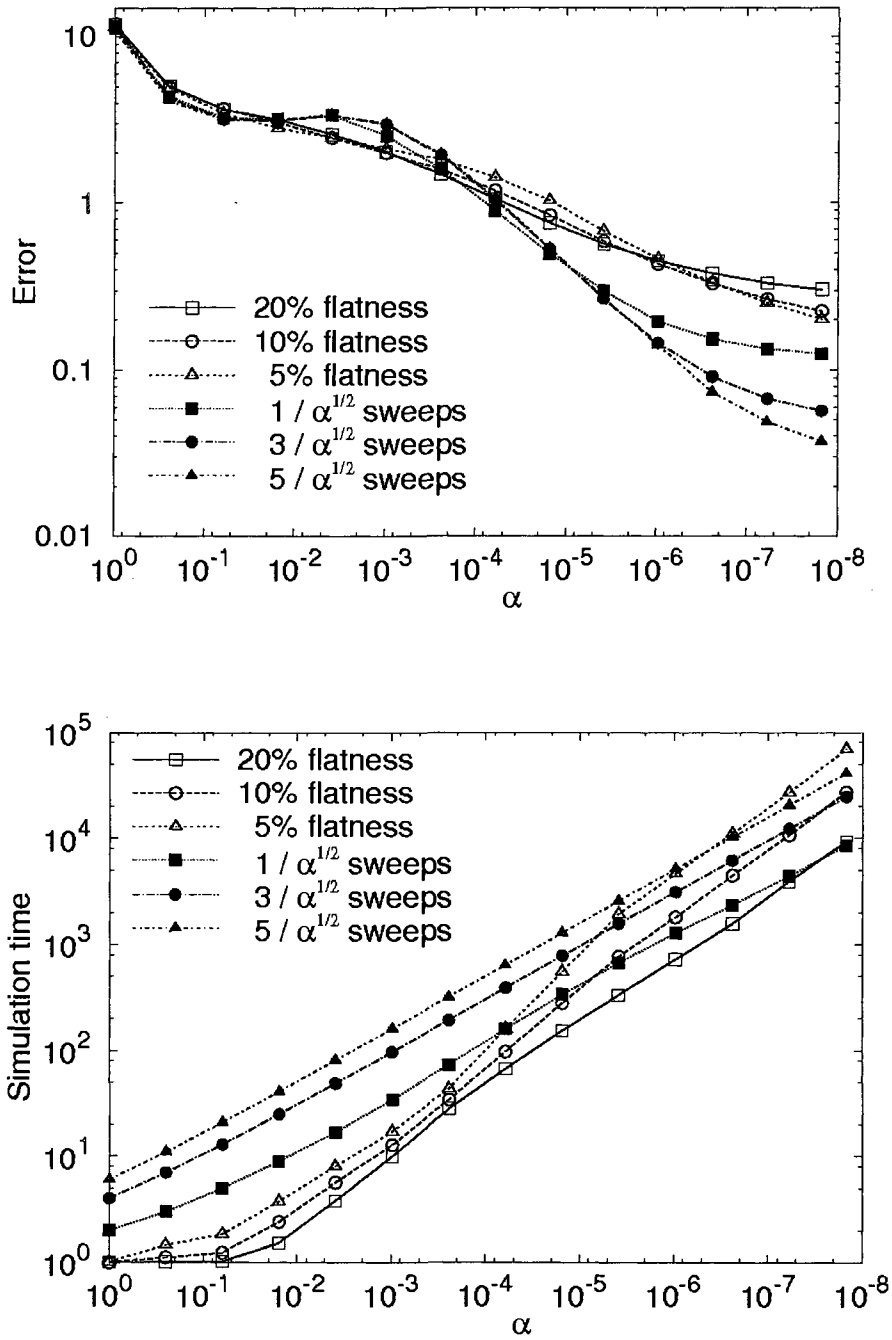


Figure 13. Rate of convergence using two different schemes.

VI. Comparison with the replica exchange method

One advantage of the method is its ability of enhancing sampling in the configurational space. In this aspect, the method bears some resemblance to a popular method, the replica exchange [45-48] method. In both methods, we simultaneously sample a system at different temperatures (for simplicity, we mainly discuss the case where the variable of simulation is the temperature). The advantage of employing multiple temperatures in a single simulation is that the system can overcome the problem of broken ergodicity. In a regular simulation, the system is likely trapped in local configurational space if the simulation is performed only at a low temperature. In the tempering methods, the system can frequently visit higher temperatures where the sampling is ergodic and go back to lower temperatures with a very different configuration. Thus both methods are useful in accelerating samplings at low temperatures and improving the statistics there.

For our method, we restrict ourselves to the case where the simulation is almost converged. Thus the enhancement in configurational space sampling due to a finite modification factor α is not included here. In such a case, our method is reduced to the simulated tempering [49, 50] with a set of optimal weighting factors. The numbers of visits to different temperatures are equal for a sufficiently long trajectory.

In replica exchange, also known as parallel tempering, independent simulations (replicas) at different temperatures are run simultaneously. Two replicas can randomly exchange their temperatures according to certain acceptance rule, which preserves the Boltzmann distribution at each temperature. In this way, the frequencies of visiting different temperatures are always the same.

The performance of the two methods was studied by many authors [51-65] focusing on different aspects. Particularly, Park found that the simulated tempering offers a larger acceptance ratio for temperature transitions than replica exchange [62]. Sindhikara *et al.* found that the sampling efficiency increases as one increases the frequency of attempting exchanges in replica exchange [64].

Here we focus on a measure of tempering efficiency based on the rate of traversing the expanded energy space, which allows us to systematically study the effects of temperature spacing and tempering frequency. We shall demonstrate that a temperature transition is generally more efficient than a temperature exchange under the same conditions.

We first review the acceptance probabilities of simulated tempering and replica exchange. Next, through the autocorrelation function, we define the tempering efficiency, according to which the two tempering methods are compared.

A. Acceptance probabilities in simulated tempering and replica exchange

Here we give details on the calculation of the acceptance ratios of simulated tempering and replica exchange. Without loss of generality, we assume that the pair of reciprocal temperatures denoted by β and β' satisfy $\beta > \beta'$.

In simulated tempering, several sampling temperatures are defined before simulation. Temperature transitions are randomly proposed to change the system temperature from the current value to another one. The probability of accepting a proposed temperature transition from β to β' is

$$\mathrm{Acc}_{\mathrm{ST}}(\beta \to \beta') = \min\left\{1, \frac{w(\beta')\exp(-\beta' E)/Z(\beta')}{w(\beta)\exp(-\beta E)/Z(\beta)}\right\}, \tag{22}$$

where E is the current potential energy (we shall drop the term "potential" in the following discussion for convenience); $Z(\beta)$ and $Z(\beta')$ are the values of the partition function at temperatures β and β', respectively; $w(\beta)$ and $w(\beta')$ are the weights of visiting temperatures β and β', respectively. Accordingly, the equilibrium weight of an individual configuration being at β is $w(\beta)\exp(-\beta E)/Z(\beta)$; after summing over all configurations, the total weight of visiting β is $w(\beta)$. The use of the partition function ensures that each sampling temperature is visited by the appropriate weight. In the following, we assume that all $w(\beta)$'s are the same unless the otherwise is specified.

In replica exchange, the acceptance probability for an attempted exchange between two replicas with temperatures β and β' ($\beta > \beta'$) is

$$\mathrm{Acc}_{\mathrm{RE}}(\beta, E; \beta', E') = \min\{1, \exp[(\beta - \beta')(E - E')]\}, \tag{23}$$

where E and E' are the energy of replicas at temperatures β and β', respectively.

An important measure is the averaged acceptance probability, or the acceptance ratio. It can be calculated in both cases under the Gaussian approximation, where the constant temperature energy distribution $p_\beta(E)$ is approximated as a Gaussian distribution. Such an approximation is a result of the second order expansion of the entropy, or the logarithm of the density of states: $S(E)/k_B = \ln g(E) \approx -\frac{1}{2}aE^2 + bE + c$; the density of states $g(E)$ is related to the energy distribution as $p_\beta(E) \propto g(E)\exp(-\beta E)$. The calculated acceptance ratios are

$$\mathrm{AR}_{\mathrm{ST}}(\beta \to \beta') = \mathrm{erfc}\left[\frac{1}{2\sqrt{2\,a}}|\beta - \beta'|\right], \tag{24}$$

$$\mathrm{AR}_{\mathrm{RE}}(\beta \to \beta') = \mathrm{erfc}\left[\frac{1}{2\sqrt{a}}|\beta - \beta'|\right], \tag{25}$$

for simulated tempering [53, 58] and replica exchange [51], respectively. Here, $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-x^2)\ dx$ is the complementary error function. The relation between the two can be expressed as:

$$\mathrm{erfc}^{-1}(\mathrm{AR}_{\mathrm{RE}}) = \sqrt{2}\ \mathrm{erfc}^{-1}(\mathrm{AR}_{\mathrm{ST}}), \tag{26}$$

which indicates a larger acceptance ratio provided by simulated tempering than by replica exchange. Note the Gaussian approximation is only correct for large systems. For a small system, however, we have to include higher order terms, such as E^3 and E^4, in the expansion of the entropy in the energy range of interest, and thus the energy distribution is no longer a Gaussian [51]. In such a case, the relation between the two acceptance ratios can deviate from the prediction of Eq. (5).

The above results can be generalized to tempering along several "temperatures" $\mathbf{b} = \{\beta_i\}$ instead of one. The partition function of such a generalized canonical ensemble is $Z = \sum \exp(-\mathbf{b} \cdot \mathbf{E})$, where the vector $\mathbf{E} = \{E_i\}$ contains energy terms E_i's correspond to different β_i's. The acceptance ratios of simulated tempering and replica exchange become:

$$\mathrm{AR}_{\mathrm{ST}} = \mathrm{erfc}\left(\frac{\sqrt{\Delta\mathbf{b} \cdot \mathbf{A}^{-1} \Delta\mathbf{b}}}{2\sqrt{2}}\right).$$

$$\mathrm{AR}_{\mathrm{RE}} = \mathrm{erfc}\left(\frac{\sqrt{\Delta\mathbf{b} \cdot \mathbf{A}^{-1} \Delta\mathbf{b}}}{2}\right),$$

where $\Delta\mathbf{b} = \mathbf{b} - \mathbf{b}'$, and $\mathbf{A}$ is the curvature matrix of the entropy, i.e., $S(\mathbf{E})/k_B \approx -\frac{1}{2}\mathbf{E} \cdot \mathbf{A}\mathbf{E} + \ldots$. A detailed derivation can be found from Appendix A.

An example of the generalized tempering is that along the pressure instead of the temperature, where vector $\mathbf{E}$ contains the energy and volume, and $\mathbf{b}$ the temperature and

pressure. Another example is the tempering along the volume (or the density) of a Lennard-Jones system, whose partition function written in the reduced coordinates $\mathbf{s} = \mathbf{r}/\sqrt[3]{V}$ is:

$$Z_{\text{pot}}(V) \equiv (1/V^N) \int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N)] = \int d\mathbf{s}^N \exp[-\beta U(\mathbf{s}^N; V)],$$

where N is the number of particles, V the volume; the contribution from the ideal gas part is removed. The potential between a particle pair, i and j, with a separation s_{ij} is $U(s_{ij};V) = (1/V^4)(1/s_{ij}^{12}) - (1/V^2)(1/s_{ij}^6)$. The repulsive and the attractive energy parts can be collected independently over all particle pairs to form two energy terms $\mathbf{E} = \{\sum_{i<j} 1/s_{ij}^{12}, \sum_{i<j} 1/s_{ij}^6\}$; the corresponding "temperatures" are given by $\mathbf{b} = \{1/V^4,\ 1/V^2\}$.

We tested the validity of Eq. (5) on three different systems. The first system is a 32×32 Ising model with two temperatures. The coupling constant is unity. The first temperature is $T = 3.0$. Different choices of the second temperature from 3.05 to 3.70 were used to cover a wide range of acceptance ratios. The second testing system is a 108-particle Lennard-Jones system. The density of the system is 0.3. The first temperature is $T = 2.0$, and the second temperature ranges from $T = 1.0$ to 1.9. The third case is the multiple-volume tempering on a 108-particle Lennard-Jones system, where the first density is $\rho = 0.35$, while the other varies from $\rho = 0.3$ to $\rho = 0.345$. The temperature is fixed at $T = 3.0$. In all cases, we used the single particle Metropolis algorithm to generate configurational changes. The results of the three simulations, and the prediction from the Gaussian model Eq. (26), are plotted in Fig. 14. A good agreement between the two confirms the validity of the Gaussian model in calculating acceptance ratios. Note there is generally a significant difference between the two acceptance ratios. For example, at the point where the acceptance ratio of

simulated tempering is 15%, that of replica exchange is about 4%. Generally the difference is more prominent as one increases the spacing between the two temperatures (or densities).

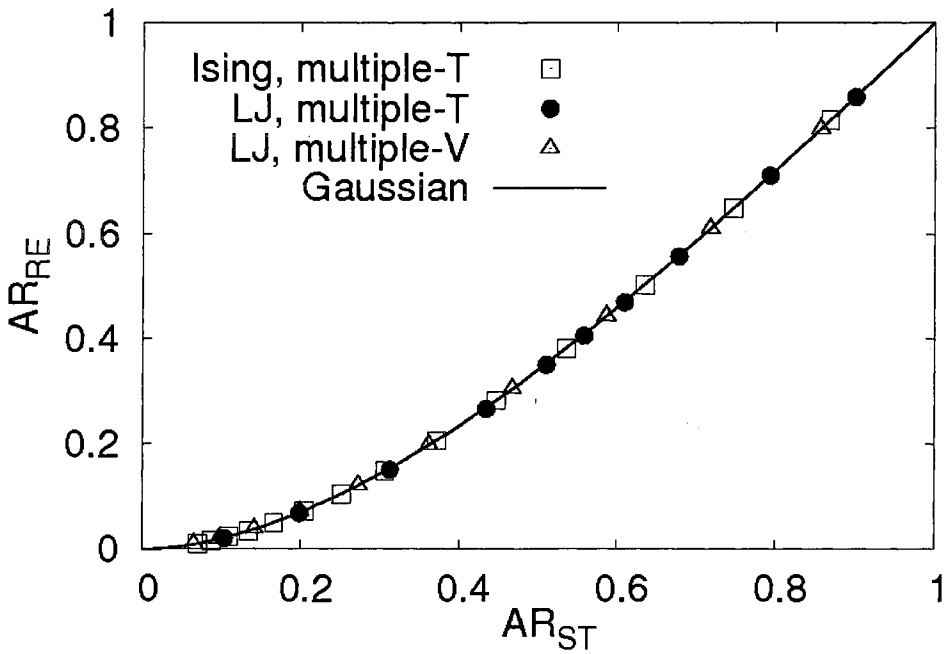


Figure 14. The acceptance ratio of simulated tempering (ST) versus that of replica exchange (RE). Results from the multiple-temperature simulations on the Ising model (squares), multiple-temperature and multiple-volume simulations on a Lennard-Jones system (solid circles and triangles respectively), as well as prediction from the Gaussian approximation (solid line) are shown.

B. Efficiency of crossing energy barriers

In this section, we compare the efficiency in terms of configurational space sampling. This aspect is not covered by the acceptance ratio and is most easily formulated using a transition matrix.

Transition matrix

An accurate measure of tempering efficiency can be established from a transition matrix and its correlation functions. Each step of a Monte Carlo simulation can be viewed as an implementation of a discrete-time master equation. If, at step t, the probability of the system to be in state m ($m = 1,2,\ldots,M$) is given by the mth component of an M-dimensional vector $\mathbf{p}(t) = \{p_m(t)\}$, then the probability vector at the next step $t+1$ is given by

$$\mathbf{p}(t+1) = \mathbf{A}\,\mathbf{p}(t),$$

where $\mathbf{A}$ is the transition matrix. The largest eigenvector of the transition matrix represents the equilibrium distribution and its eigenvalue is $\lambda_1 = 1.0$ [2]. The other eigenvectors represent directions of fluctuation modes. Starting from an initial distribution $\mathbf{p}(0)$, the distribution at time t is $\mathbf{p}(t) = \mathbf{A}^t\,\mathbf{p}(0)$. Then the autocorrelation function for $\mathbf{p}(0)$ is defined as

$$C(t) = \mathbf{p}(0)\cdot[\mathbf{p}(t) - \mathbf{p}(\infty)], \tag{27}$$

where the contribution from the final distribution $\mathbf{p}(\infty)$ is removed to make $C(\infty) = 0$. Since the vector $\mathbf{p}(0)$ can be decomposed to a linear combination of eigenvectors $\mathbf{v}_m$'s of the transition matrix as $\mathbf{p}(0) = \sum_{k=1}^{M} b_k \mathbf{v}_k$, the autocorrelation function must adopt the form

$C(t) = \sum_{k=2}^{M} c_k \lambda_k^t = \sum_{k=2}^{M} c_k \exp(-t/\tau_k)$, where $1/\tau_k = -\log(\lambda_k) \approx 1 - \lambda_k$ represents the decay rate of the kth fluctuation mode.

Among all fluctuation modes, the second or the slowest mode usually dominates the correlation function of interest, and hence is most important. As long as the mode is retained, it is convenient to define states on a coarse-grained level to eliminate irrelevant fluctuation modes and to simplify calculations.

Random walk in the configurational space

The main purpose of tempering is to enhance low-temperature sampling by helping the system to overcome high energy barriers. With the help of high temperatures, the system at a low temperature can readily switch between different energy wells. The tempering process therefore can be modeled as a combination of a well-switching process at high temperatures and a tempering process that delivers the system between the highest and the lowest temperatures.

In simulated tempering, a system state can be labeled by the current configuration and the temperature. In replica exchange, such a labeling applies to each individual replica, whose temperature is changed through exchanges with other replicas (however, the configuration of a replica or its energy remains unchanged during an exchange). In this way, each replica acts like a system in simulated tempering; the only difference is that now the acceptance probability of changing the temperature depends on the energy of another replica (not the partition function), as indicated in Eq. (23).

Consider tempering on a double well system with a low temperature β and a high temperature β' ($\beta > \beta'$). The system state can be categorized into four cases for the system

being at: a) temperature β and well 1, b) temperature β' and well 1, c) temperature β' and well 2, d) temperature β and well 2. In each Monte Carlo step at the high temperature β', the system can switch between wells with a probability p (the probability of well-switching at the low temperature is ignored). The transition probability between the two temperatures is P. The transition matrix is:

$$\begin{pmatrix} 1-P & P & 0 & 0 \\ P & 1-P-p & p & 0 \\ 0 & p & 1-P-p & P \\ 0 & 0 & P & 1-P \end{pmatrix}.$$

The decay rate of the second largest eigenvalue is:

$$1/\tau_2 \approx 1-\lambda_2 = p+P-\sqrt{P^2+p^2}\,.$$

This rate represents how fast the system moves in the temperature-energy space, and serves as the measure of tempering efficiency. Although derived from the two-temperature case, the model can be approximately used where more than two temperatures are involved. In such a case, β and β' are interpreted as the lowest and the highest temperature, respectively, and the well-switching process at a temperature other than β' is ignored.

Since the well switching rate p at β' is completely determined by details of the configurational sampling algorithm (e.g., MC or MD), the dependence on temperature transitions can only rest on P, the rate of delivering states between the highest and the lowest temperatures. A tempering method is more efficient if it more frequently shuttles the system state between the energy region of the highest temperature and that of the lowest.

Efficiency of traversing the energy space

In the case of simulated tempering, consider tempering between two temperatures β and β'. We divide the energy distribution at β into two exclusive parts: an overlapping region with the energy distribution at β' and the rest, see Fig. 15. A temperature transition from β to β' is possible if and only if the system resides in the overlapping region. In other words, the overlapping region separates states that are capable of making a temperature transition from those are not. Accordingly, the acceptance ratio $\mathrm{AR}_{\mathrm{ST}}$ equals to the probability of being in the overlapping region (or the fraction of states that are ready to make a transition). We use r' to denote the transition rate from the overlapping region to the rest part, and r that in the reverse direction during constant-temperature sampling at β. The detailed balance condition between the two parts imposes a relation between r and r':

$\mathrm{AR}_{\mathrm{ST}}\, r' = (1-\mathrm{AR}_{\mathrm{ST}})\, r$, or

$$\frac{r}{r+r'} = \mathrm{AR}_{\mathrm{ST}}.$$

The corresponding division can be made on the energy distribution at β'. For the sake of simplicity, we assume that r' and r in β' take the same values as those in the β case.

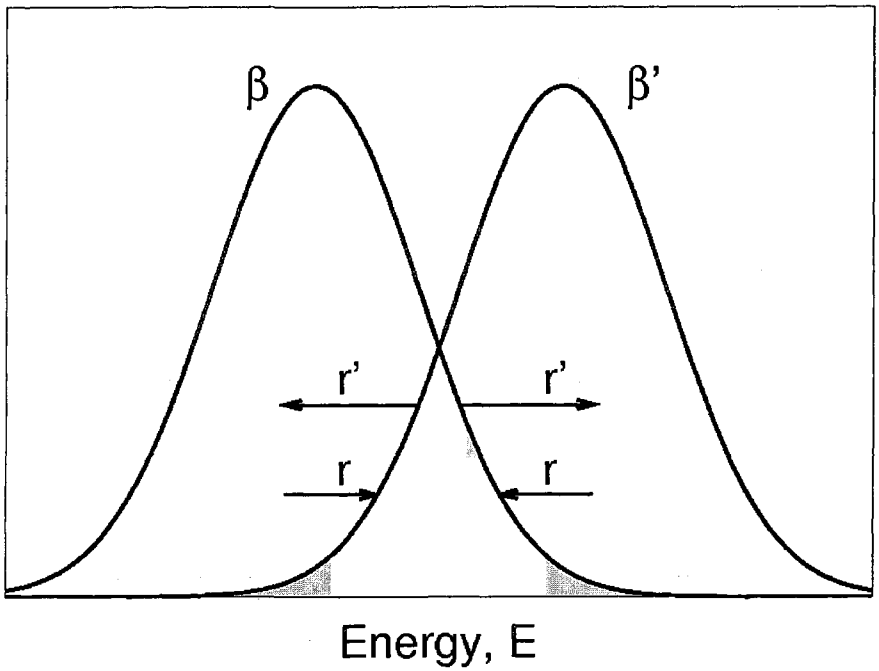


Figure 15. Model of tempering. A temperature transition is possible only when the system enters the overlapping region (shaded area). Configurational sampling randomly delivers the system into (r) or out of (r') the overlapping region.

The above divisions yield four coarse-grained states: a) being at β but not in the overlapping region, b) being at β and in the overlapping region, c) being at β' and in the overlapping region, d) being at β' but not in the overlapping region. While constant temperature sampling transfers the system between a) and b), or c) and d), temperature transitions transfer the system between b) and c). The transition matrix is:

$$\begin{pmatrix} 1-r & r' & & \\ r & 1-r'-f & f & \\ & f & 1-r'-f & r \\ & & r' & 1-r \end{pmatrix},$$

where f is the frequency of attempting temperature transitions. The decay rate of the second largest eigenvalue is

$$1/\tau_2 = \tfrac{1}{2}\left[2f + r + r' - \sqrt{(r+r')^2 + 4f(r'-r) + 4f^2}\right] \tag{28}$$

For frequent tempering limit, $f >> r, r'$, it becomes r; for infrequent tempering limit, $f << r, r'$, $1/\tau_2$ is reduced to $2rf/(r+r') = 2\mathrm{AR}_{\mathrm{ST}}\, f$.

In the case of replica exchange, we first assume that a temperature exchange is possible if and only if both replicas, I and II, reside in an overlapping region of the two energy distributions. It is because, if either replica leaves the overlapping region, a large energy difference between the two replicas will effectively inhibit the exchange through the acceptance probability Eq. (23). Due to the difference of the acceptance probabilities, the overlapping region here can take a different shape from that in simulated tempering. We now classify states of the system into two categories. In the first category, replica I is at temperature β and II at β'; in the second, I at β' and II at β. Each category contains three states: neither replica, only one replica, or both replicas being in the overlapping region. Similar to the case of simulated tempering, the rates of moving into or out of the overlapping region are denoted by r and r' respectively, but they assume slight different values from their counterparts in simulated tempering. The two rates relate to the acceptance ratio as:

$$\left(\frac{r}{r+r'}\right)^2 = \mathrm{AR}_{\mathrm{RE}}.$$

Assuming both r and r' are small, the transition matrix is:

$$\begin{pmatrix} 1-2r & r' & & & & & \\ 2r & 1-r-r' & 2r' & & & & \\ & r & 1-2r'-f & f & & & \\ & & f & 1-2r'-f & r & & \\ & & & 2r' & 1-r-r' & 2r & \\ & & & & r' & 1-2r & \end{pmatrix}.$$

The decay rate of the second largest eigenvalue is approximately,

$$1/\tau_2 \approx B/A - \sqrt{(B/A)^2 - C/A}, \tag{29}$$

where $A = 1 + \frac{2f}{3(r+r')}$, $B = \frac{1}{3}(r + r' + \frac{3r+r'}{r+r'}f)$, and $C = \frac{4fr^2}{3(r+r')}$. It is reduced to $\frac{1}{2}[3r + r' - \sqrt{r^2 + r'^2 + 6rr'}]$ in the frequent tempering limit, or $2\left(\frac{r}{r+r'}\right)^2 f = 2\mathrm{AR}_{\mathrm{RE}} f$ in the infrequent tempering limit.

In both simulated tempering and replica exchange, the decay rate is proportional to the acceptance ratio in the limit of infrequent tempering; hence simulated tempering is superior over replica exchange because it offers a larger acceptance ratio. This conclusion remains true even in the frequent tempering limit. This can be seen by assuming that the rate r takes a similar value in the two tempering methods and then making an algebraic comparison between the corresponding decay rates. A significant difference can be seen as we separate the two energy distributions far apart. This leads to a small overlapping region, and $r \ll r'$. Consequently, the ratio of the decay rate in replica exchange and that in simulated tempering becomes $2r/(r' + 3r)$, which is much less than 1.

To test this model, a numerical comparison was done on the 32×32 Ising model with two temperatures. The first was fixed at $T = 1.7$ while the second was variable. As before, the setup allowed us to compare the two methods at different temperature separations. Since

the eigenvalue λ_2 corresponds to the slowest mode in the energy space, it can be measured from the slowest mode of the energy autocorrelation function, $C_E(\tau) = \langle (E(t) - \overline{E}) \cdot (E(t+\tau) - \overline{E}) \rangle$, where $\overline{E} = [\overline{E}(\beta) + \overline{E}(\beta')]/2$ is the average energy at the two temperatures. Particularly we focused on the frequent tempering limit where the tempering frequency f is fixed at 1.0 (in the infrequent tempering limit we only need to compare the acceptance ratio directly). In Fig. 16, we show the decay rate $1/\tau_2 = 1 - \lambda_2$ as a function of the temperature separation. Simulated tempering invariably gives a larger decay rate (or a shorter correlation time) than replica exchange. The difference between the two methods is small when the two temperatures are close. With the increase of the separation, the advantage of simulated tempering over replica exchange also becomes apparent, e.g., at the rightmost point, the ratio of the two decay rates is larger than 10.

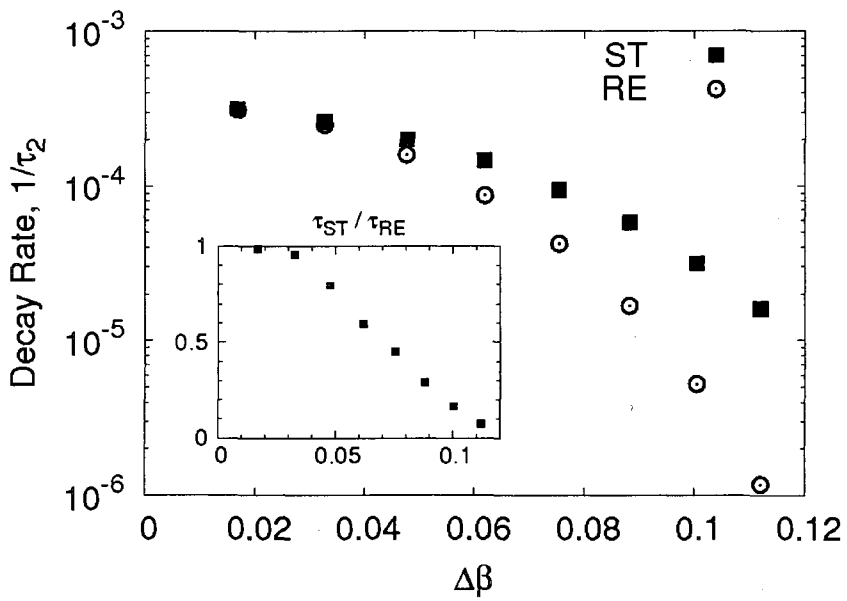


Figure 16. The rate of energy space traversing $1/\tau_2$ (measured from the slowest mode energy correlation function) versus the temperature separation $\Delta\beta$ for simulated tempering (ST) and replica exchange (RE). The ratio of the two rates is shown in the inset.

Next, we fixed the second temperature $T = 2.0$, but varied the tempering frequency f from 2×10^{-5} to 1.0. The acceptance ratios are 6.65% for simulated tempering and 1.03% for replica exchange, respectively. In the frequent tempering limit, the decay rate saturates to 5.8×10^{-5} for simulated tempering, and to 1.7×10^{-5} for replica exchange. The rates r and r' can be deduced from the acceptance ratio and the saturated decay rate, and then used in calculating model predictions from Eq. (28) and Eq. (29). A good agreement between the simulation results and the model prediction from can be seen in Fig. 17(a). It is also evident that simulated tempering gives a larger decay rate than replica exchange for any given tempering frequency f.

A numerical comparison with multiple temperatures was performed on the same Ising model. Seven temperatures were used, T = 1.5, 1.8, 2.0, 2.2, 2.3, 2.4, 2.6. The transition temperature is about 2.27. The results are shown in Fig. 17(b), where we also used

$$R(f) = R_0 f/(f + f_0), \tag{30}$$

to fit the curves. Here, the parameter R_0 is the saturated (maximal) decay rate, and f_0 is the tempering frequency where the rate drops to half of its maximal value. A good fitting can be seen in both simulated tempering and replica exchange. This is because Eq. (30) is the simplest form that captures the essential features of frequency dependence (for both simulated tempering and replica exchange), that is, a linear dependence in the low frequency limit, and a saturation behavior in the high frequency limit. Note, Eq. (30) can also be treated as a good approximation for the results of the two-temperature case: Eq. (28) is approximately $rf/[f+(r+r')/2]$, while Eq. (29) is roughly $C/(2B)$ with C proportional to f and B linear to f. We thus expect that Eq. (30) can be used to approximate the multiple temperature case as well. From Fig. 17(b), it can be seen that simulated tempering once

again outperforms replica exchange at any given frequency. Note we used a small separation between neighboring temperature pairs to reduce the difference of the decay rates in the high frequency limit. However, as the frequency is lowered, the difference between the two methods still grows significantly. At the saturated frequency $f = 1$, the decay rate of simulated tempering is about 1.5 times that of replica exchange. At $f = 10^{-4}$, the ratio of the two rises to about 10.

System with a transition temperature

Many systems possess a transition temperature that separates the system between disordered states at high temperatures and ordered states at low temperatures. The energy distribution at the transition temperature becomes a bottleneck region for traversing the energy space. The tempering efficiency can be improved by focusing sampling around the transition temperature to enhance the bridge between the high energy (disordered) states and the low energy (ordered) states.

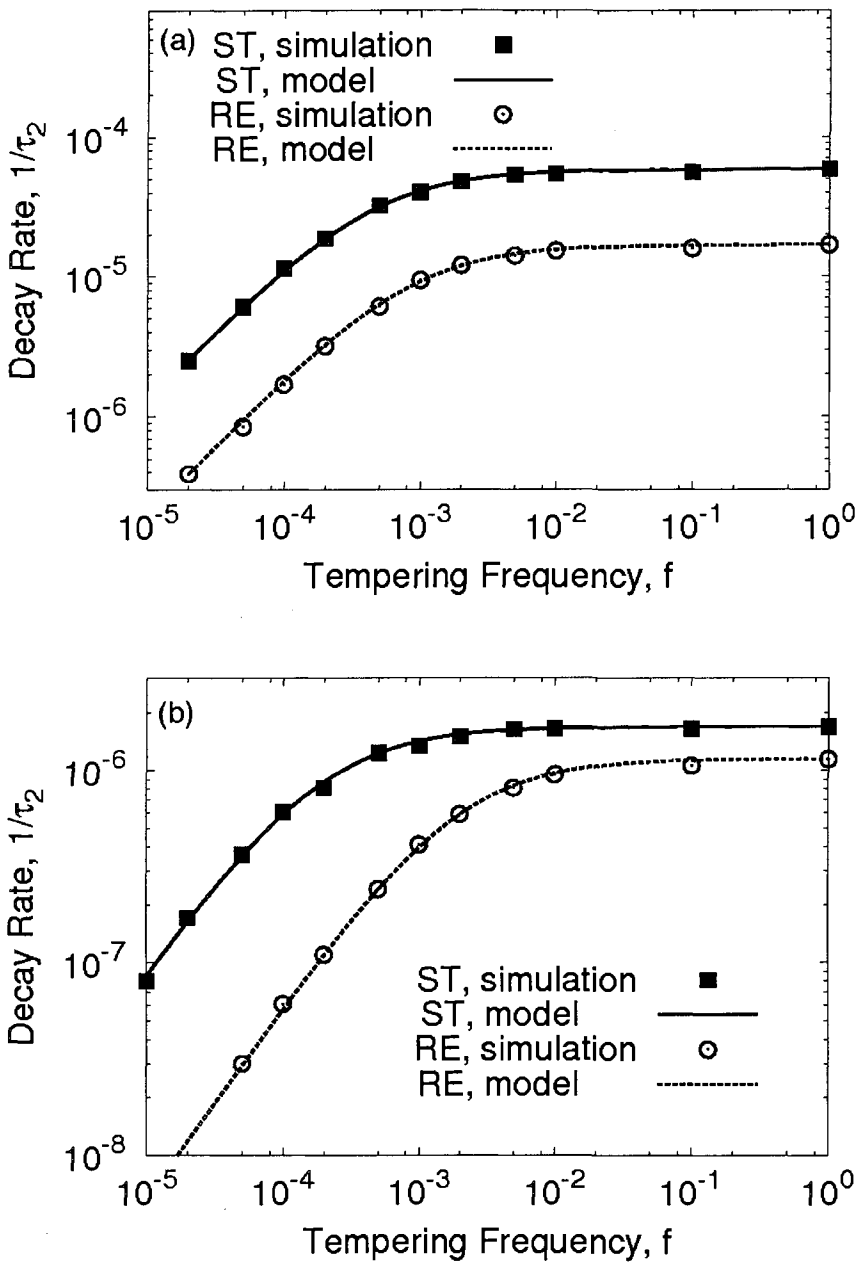


Figure 17. The rate of energy space traversing $1/\tau_2$ versus the frequency f of attempting temperature transitions in simulated tempering (ST) and replica exchange (RE). (a) The two-temperature case. The predictions from Eq. (28) and Eq. (29) are also shown in solid line and dashed line, respectively. (b) The multiple-temperature case. The lines are from the prediction of Eq. (30).

In simulated tempering, we can simply adjust the acceptance probability Eq. (22) by raising the weight $w(\beta)$ of the transition temperature. The effect of adjusting the weight at the transition temperature can be shown on the 32×32 Ising model. We used the seven-temperature set, T = 1.5, 1.8, 2.0, 2.2, 2.3, 2.4, 2.6. The integrated correlation time (obtained from integrating the normalized autocorrelation function from zero to infinity) at the sampling temperatures are shown in the inset of Fig. 18, where the peak is reached around the transition temperature $T_c \approx 2.3$. In Fig. 18, we compared the energy correlation functions in cases where the relative weight of visiting temperature $T = 2.3$ was adjusted to 1.0, 5.0, 10.0 or 20.0, while the weights of other temperatures were kept at 1.0 (the tempering frequency was $f = 1$ in all the cases). The correlation time was significantly reduced with the increase of the weight, e.g., its value at $w = 20.0$ was less than one third of that at $w = 1.0$. As a control, we also increased the weight of another sampling temperature T = 1.8 to 20.0 while keeping the weights to other temperatures (including $T = 2.3$) unity. From Fig. 18, it is evident that it failed to shorten the correlation time. This indicates that the tempering efficiency can be significantly improved only if the weight at the transition temperature is increased.

In replica exchange, it is also possible to the increase the weight to the transition temperature by adding additional replicas there. However, it is usually inconvenient if replicas are distributed in different computer nodes. In the above case, one has to use 19 additional replicas to reach the weight $w = 20.0$. Further, the efficiency can be compromised if an implementation is unable to afford a frequent communication between computer nodes.

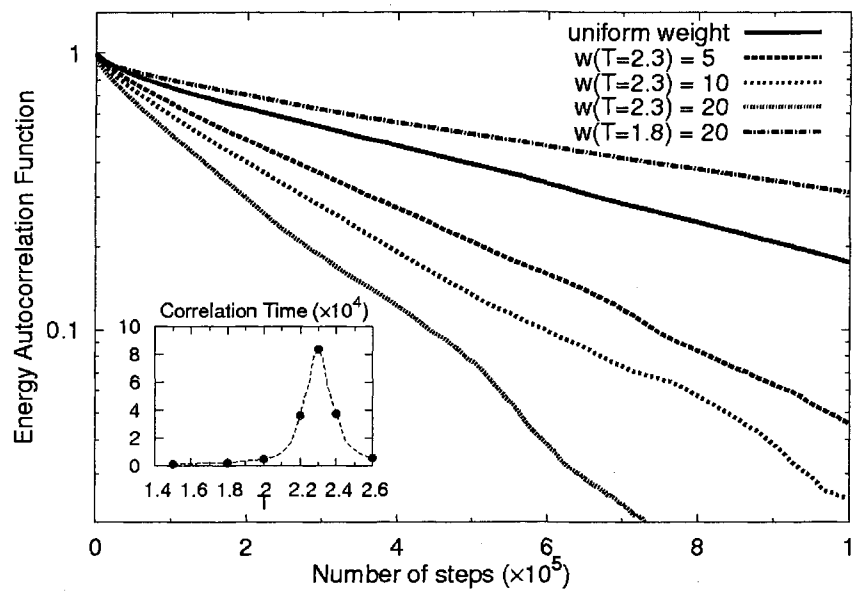


Figure 18. The energy autocorrelation functions under different weights to the transition temperature. The relative weight to the transition temperature $T = 2.3$ is changed to different values. By comparison, the result of increasing the relative weight to another sampling temperature $T = 1.8$ is also shown. The energy autocorrelation time under single-temperature (Metropolis) sampling is shown in the inset; the peak is reached around the phase-transition temperature $T \approx 2.3$.

A possible side-effect of raising the weight of the transition temperature is that temperatures other than the transition temperature may receive fewer visits, and thus reduce the total amount of statistics there. However, in a system that manifests a phase transition, it usually takes a long time for the system to switch between different local wells even in the presence of tempering. Thus if we expect that the system can accumulate sufficient statistics in a local well before it transits to another well, such a side-effect is negligible.

C. Discussions

We compared the tempering efficiency between simulated tempering and replica exchange. Simulated tempering consistently gives a higher rate of delivering the system between high temperature states and low temperature states as well as a higher rate of traversing the energy space. The difference is especially eminent if the energy distributions of neighboring temperatures are well separated, or/and if the tempering frequency is low. The fundamental feature that makes replica exchange less efficient than simulated tempering is that in replica exchange a successful temperature exchange requires two replicas to be simultaneously in their common energy space, while in simulated tempering a temperature transition can happen whenever the system falls in the region. Besides, it is usually easier for simulated tempering to adopt higher tempering frequency. This is because, in replica exchange, a high tempering frequency requires heavy cost of computer node communication. In addition, simulated tempering is able to concentrate simulation effort on the "bottleneck" temperature to reach the maximal efficiency without adding additional replicas.

In the above discussion, we assumed that the partition function used in simulated tempering is exact. But in reality we usually use an approximated partition function for sampling parameters. In this case, what changes is that each temperature may not receive exactly equal number of visits. To see this, let us assume that the sampling parameter we used is $Z(\beta)\exp[\varepsilon(\beta)]$ instead of $Z(\beta)$, where $\varepsilon(\beta)$ represents the error for the sampling temperature β. From Eq. (22), it is clear that using this set of parameters is equivalent to using the exact partition function but with a non-uniform weighting factor $w(\beta) = \exp[-\varepsilon(\beta)]$. However, the Boltzmann distribution of each individual sampling

temperature is still preserved and thus quantities calculated from each temperature contain no systematic error after equilibration.

However, a sufficiently accurate partition function (e.g., $|\varepsilon(\beta)|<0.5$) is still needed for an efficient simulation. Otherwise some temperatures can absorb most of the sampling weights from the rest temperatures, which in turn leads to a slower rate of traversing the energy space (with the exception mentioned above where the focused temperature is the transition temperature). For a large and realistic system such as a protein, to obtain an accurate partition function is nontrivial, especially for low temperatures where the partition function is dominated by the ground state. In such systems, the technique of using an updating factor to gradually converge the partition function is particularly helpful, because the partition function obtained in this way can properly reflect the contribution from low energy states which are usually found in a late stage of simulation. Accordingly we also believe that the updating of the partition function should be used until the system is properly equilibrated.

There are concerns in literature that a high frequency may actually compromise the tempering efficiency in replica exchange [65]. This effect is not observed in our model and numerical results. Our results shows that the tempering efficiency (measured by the decay rate of traversing energy space) increases monotonically with the frequency of attempting temperature transitions/exchanges, although it encounters a saturation and the pace of increase slows down significantly when the frequency grows to be comparable with the reciprocal of the autocorrelation time of the energy. The efficiency drop may due to cost from communications between different replicas or other reasons that escaped our calculation.

Appendix A. Computing the acceptance ratios

Consider a generalized canonical ensemble. Suppose there are N extensive quantities of a multi-particle system, $\mathbf{E} = \{E_1, E_2, \cdots, E_N\}$, that can be expressed as functions of a microscopic configuration. In a local $\mathbf{E}$ region under concern, the logarithm of the density of states, i.e., the entropy, can be expanded as a hyper-parabola.

$$\ln g(\mathbf{E}) \approx -\tfrac{1}{2}\mathbf{E}\cdot\mathbf{A}\,\mathbf{E} + \mathbf{B}\cdot\mathbf{E} + C\,.$$

Here the symmetrical matrix $\mathbf{A}$ is positive-definite due to the extensive nature of $\mathbf{E}$. A general canonical ensemble is characterized by a couple of thermodynamic quantities $\mathbf{b} = \{\beta_1, \beta_2, \cdots, \beta_N\}$ conjugated to $\mathbf{E}$. Under the Gaussian approximation, the partition function for the generalized canonical ensemble is readily calculated,

$$Z(\boldsymbol{\beta}) = \int g(\mathbf{E})\exp(-\mathbf{b}\cdot\mathbf{E})d\mathbf{E} = \sqrt{\frac{(2\pi)^N}{|\mathbf{A}|}}\exp[C + \frac{1}{2}(\mathbf{B}-\mathbf{b})\cdot\mathbf{A}^{-1}(\mathbf{B}-\mathbf{b})]\,.$$

To compute the acceptance ratio of the replica exchange method, one needs to perform a double integral over $\mathbf{E}$ and $\mathbf{E}'$

$$\mathrm{AR}_{\mathrm{RE}}(\mathbf{b},\mathbf{b}') = \int_{-\infty}^{\infty} d\mathbf{E}\, p_{\mathbf{b}}(\mathbf{E}) \int_{-\infty}^{\infty} d\mathbf{E}'\, p_{\mathbf{b}'}(\mathbf{E}')\mathrm{Acc}_{\mathrm{RE}}(\mathbf{b},\mathbf{E};\mathbf{b}',\mathbf{E}')\,.$$

It is convenient to introduce a change of variables: $\overline{\mathbf{E}} = (\mathbf{E}+\mathbf{E}')/2$ and $\Delta\mathbf{E} = \mathbf{E}-\mathbf{E}'$. The integral over $\overline{\mathbf{E}}$ can be immediately separated and it yields a constant independent of $\mathbf{b}$ and $\mathbf{b}'$. The rest part of the integral is proportional to

$$\int \exp(-\tfrac{1}{4}\Delta\mathbf{E}\cdot\mathbf{A}\,\Delta\mathbf{E} - \tfrac{1}{2}\Delta\mathbf{b}\cdot\Delta\mathbf{E} - \tfrac{1}{4}\Delta\mathbf{b}\cdot\mathbf{A}^{-1}\,\Delta\mathbf{b})\,\min\{1, \exp(\Delta\mathbf{b}\cdot\Delta\mathbf{E})\}\,d(\Delta\mathbf{E})\,,$$

where $\Delta\mathbf{b} = \mathbf{b} - \mathbf{b}'$. Since $\mathbf{A}$ is symmetrical and positive-definite, there exists a matrix $\mathbf{T}$, that decomposes $\mathbf{A}$ as $\mathbf{A} = \mathbf{T}\,\mathbf{T}^T$, where $\mathbf{T}^T$ is the transpose matrix of $\mathbf{T}$. By introducing $\delta\mathbf{E} = \frac{1}{2}\mathbf{T}^T\,\Delta\mathbf{E}$ and $\delta\mathbf{b} = \mathbf{T}^{-1}\Delta\mathbf{b}$, the above formula is simplified as

$$\int \exp(-|\delta\mathbf{E} + \tfrac{1}{2}\delta\mathbf{b}|^2)\,\min\{1, \exp(2\delta\mathbf{b}\cdot\delta\mathbf{E})\}\,d(\delta\mathbf{E}).$$

Evaluated for $\delta\mathbf{b}\cdot\delta\mathbf{E} > 0$ and for $\delta\mathbf{b}\cdot\delta\mathbf{E} \le 0$ separately, the integral becomes constant $\times$ erfc $(|\delta\mathbf{b}|/2)$. The constant before the complementary error function must be one in order to make the acceptance ratio unity under the special case $\delta\mathbf{b} = 0$. The final result for the acceptance ratio of replica exchange is

$$\mathrm{AR}_{\mathrm{RE}} = \mathrm{erfc}\left(\frac{|\mathbf{T}^{-1}\Delta\mathbf{b}|}{2}\right) = \mathrm{erfc}\left(\frac{\sqrt{\Delta\mathbf{b}\cdot\mathbf{A}^{-1}\Delta\mathbf{b}}}{2}\right).$$

The acceptance ratio of simulated tempering can be calculated in a similar fashion

$$\mathrm{AR}_{\mathrm{ST}}(\mathbf{b}\to\mathbf{b}') = \int_{-\infty}^{\infty} d\mathbf{E}\, p_{\mathbf{b}}(\mathbf{E})\,\mathrm{Acc}_{\mathrm{ST}}(\mathbf{b}\to\mathbf{b}';\mathbf{E}).$$

Under a shift of origin $\mathbf{E}\to\mathbf{E} + \mathbf{A}^{-1}[\mathbf{B} - (\mathbf{b}+\mathbf{b}')/2]$, the acceptance probability is simplified as $\min\{1, \exp(\Delta\mathbf{b}\cdot\mathbf{E})\}$. By specializing in the two cases $\Delta\mathbf{b}\cdot\mathbf{E} > 0$ and $\Delta\mathbf{b}\cdot\mathbf{E} \le 0$, the integral is readily evaluated and the acceptance ratio of simulated tempering is

$$\mathrm{AR}_{\mathrm{ST}} = \mathrm{erfc}\left(\frac{|\mathbf{T}^{-1}\Delta\mathbf{b}|}{2\sqrt{2}}\right) = \mathrm{erfc}\left(\frac{\sqrt{\Delta\mathbf{b}\cdot\mathbf{A}^{-1}\Delta\mathbf{b}}}{2\sqrt{2}}\right).$$

References

1. Metropolis, N., et al., *Equation of state calculations by fast computing machines.* J. Chem. Phys., 1953. **21**: p. 1087-1092.
2. Newman, M.E.J. and G.T. Barkema, *Monte Carlo methods in statistical physics.* 1999, Oxford, New York Clarendon Press ; Oxford University Press. xiv, 475 p.
3. Nose, S., *A Molecular-Dynamics Method for Simulations in the Canonical Ensemble.* Molecular Physics, 1984. **52**(2): p. 255-268.
4. Hoover, W.G., *Canonical Dynamics - Equilibrium Phase-Space Distributions.* Physical Review A, 1985. **31**(3): p. 1695-1697.
5. Martyna, G.J., M.L. Klein, and M. Tuckerman, *Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics.* Journal of Chemical Physics, 1992. **97**(4): p. 2635-2643.
6. Baumann, B., *Noncanonical path and surface simulation.* Nucl. Phys., 1987. **B285**: p. 391-409.
7. Berg, B.A. and T. Neuhaus, *Multicanonical Ensemble - a New Approach to Simulate 1st-Order Phase-Transitions.* Physical Review Letters, 1992. **68**(1): p. 9-12.
8. Berg, B.A. and T. Celik, *New Approach to Spin-Glass Simulations.* Physical Review Letters, 1992. **69**(15): p. 2292-2295.
9. Berg, B.A. and W. Janke, *Multioverlap simulations of the 3D Edwards-Anderson Ising spin glass.* Physical Review Letters, 1998. **80**(21): p. 4771-4774.
10. Lee, J., *New Monte-Carlo Algorithm - Entropic Sampling.* Physical Review Letters, 1993. **71**(2): p. 211-214.

11. Wang, F.G. and D.P. Landau, *Efficient, multiple-range random walk algorithm to calculate the density of states.* Physical Review Letters, 2001. **86**(10): p. 2050-2053.

12. Wang, F.G. and D.P. Landau, *Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram.* Physical Review E, 2001. **64**(5): p. -.

13. Kim, J., J.E. Straub, and T. Keyes, *Statistical-temperature monte carlo and molecular dynamics algorithms.* Physical Review Letters, 2006. **97**(5): p. -.

14. Kim, J., J.E. Straub, and T. Keyes, *Statistical temperature molecular dynamics: Application to coarse-grained beta-barrel-forming protein models.* J Chem Phys, 2007. **126**(13): p. 135101.

15. Ferrenberg, A.M. and R.H. Swendsen, *New Monte-Carlo Technique for Studying Phase-Transitions.* Physical Review Letters, 1988. **61**(23): p. 2635-2638.

16. Ferrenberg, A.M. and R.H. Swendsen, *Optimized Monte-Carlo Data-Analysis.* Physical Review Letters, 1989. **63**(12): p. 1195-1198.

17. Ferdinand, A.E. and M.E. Fisher, *Bounded and Inhomogeneous Ising Models .I. Specific-heat Anomaly of a Finite Lattice.* Physical Review, 1969. **185**: p. 832-846.

18. Swendsen, R.H. and J.S. Wang, *Nonuniversal Critical-Dynamics in Monte-Carlo Simulations.* Physical Review Letters, 1987. **58**(2): p. 86-88.

19. Ferrenberg, A.M. and D.P. Landau, *Critical-Behavior of the 3-Dimensional Ising-Model - a High-Resolution Monte-Carlo Study.* Physical Review B, 1991. **44**(10): p. 5081-5091.

20. Perez-Pellitero, J., et al., *Critical point estimation of the Lennard-Jones pure fluid and binary mixtures.* Journal of Chemical Physics, 2006. **125**(5): p. -.

21. Irback, A., et al., *Local interactions and protein folding: A three-dimensional off-lattice approach.* Journal of Chemical Physics, 1997. **107**(1): p. 273-282.

22. Liang, F.M., *Annealing contour Monte Carlo algorithm for structure optimization in an off-lattice protein model.* Journal of Chemical Physics, 2004. **120**(14): p. 6756-6763.

23. Hsu, H.P., V. Mehra, and P. Grassberger, *Structure optimization in an off-lattice protein model.* Physical Review E, 2003. **68**(3): p. -.

24. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.

25. Nakajima, N., H. Nakamura, and A. Kidera, *Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides.* Journal of Physical Chemistry B, 1997. **101**(5): p. 817-824.

26. Stillinger, F.H., T. Headgordon, and C.L. Hirshfeld, *Toy Model for Protein-Folding.* Physical Review E, 1993. **48**(2): p. 1469-1477.

27. Bachmann, M., H. Arkin, and W. Janke, *Multicanonical study of coarse-grained off-lattice models for folding heteropolymers.* Physical Review E, 2005. **71**(3): p. -.

28. Kim, S.Y., S.B. Lee, and J. Lee, *Structure optimization by conformational space annealing in an off-lattice protein model.* Physical Review E, 2005. **72**(1): p. -.

29. Minton, S., et al., *Minimizing Conflicts - a Heuristic Repair Method for Constraint Satisfaction and Scheduling Problems.* Artificial Intelligence, 1992. **58**(1-3): p. 161-205.

30. Rivin, I. and R. Zabih, *A Dynamic-Programming Solution to the N-Queens Problem.* Information Processing Letters, 1992. **41**(5): p. 253-256.

31. Elser, V., I. Rankenburg, and P. Thibault, *Searching with iterated maps.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(2): p. 418-423.

32. Sloane, N.J.A., *A000170: Number of ways of placing n nonattacking queens on n X n board.*, in *The On-Line Encyclopedia of Integer Sequences.* p. A000170.

33. Sloane, N.J.A., *A002860: Number of Latin squares of order n; or labeled quasigroups.*, in *The On-Line Encyclopedia of Integer Sequences.* p. A002860.

34. Pinn, K. and C. Wieczerkowski, *Number of magic squares from parallel tempering Monte Carlo.* International Journal of Modern Physics C, 1998. **9**(4): p. 541-546.

35. Lima, A.R. and M.A. de Menezes, *Entropy-based analysis of the number partitioning problem.* Physical Review E, 2001. **6302**(2): p. -.

36. Hukushima, K., *Extended ensemble Monte Carlo approach to hardly relaxing problems.* Computer Physics Communications, 2002. **147**(1-2): p. 77-82.

37. Sosic, R. and J. Gu, *Fast Search Algorithms for the N-Queens Problem.* Ieee Transactions on Systems Man and Cybernetics, 1991. **21**(6): p. 1572-1576.

38. Sloane, N.J.A., *A140393: Number of ways of placing n nonattacking queens on n X n board (version computed by NQueens@home).* in *The On-Line Encyclopedia of Integer Sequences.* p. A140393.

39. Morozov, A.N. and S.H. Lin, *Accuracy and convergence of the Wang-Landau sampling algorithm.* Physical Review E, 2007. **76**(2): p. -.

40. Zhou, C.G. and R.N. Bhatt, *Understanding and improving the Wang-Landau algorithm.* Physical Review E, 2005. **72**(2): p. -.

41. Zhou, C.G. and J. Su, *Optimal modification factor and convergence of the Wang-Landau algorithm.* Physical Review E, 2008. **78**(4): p. -.

42. Belardinelli, R.E., S. Manzi, and V.D. Pereyra, *Analysis of the convergence of the 1/t and Wang-Landau algorithms in the calculation of multidimensional integrals.* Physical Review E, 2008. **78**(6): p. -.

43. Belardinelli, R.E. and V.D. Pereyra, *Wang-Landau algorithm: A theoretical analysis of the saturation of the error.* Journal of Chemical Physics, 2007. **127**(18): p. -.

44. Belardinelli, R.E. and V.D. Pereyra, *Fast algorithm to calculate density of states.* Physical Review E, 2007. **75**(4): p. -.

45. Swendsen, R.H. and J.S. Wang, *Replica Monte-Carlo Simulation of Spin-Glasses.* Physical Review Letters, 1986. **57**: p. 2607-2609.

46. Geyer, C.J., *Proceedings of the 23rd symposium on the interface* Markov chain Monte Carlo maximum likelihood, ed. E.M. Keramidas. 1991, New York: American Statistical Association.

47. Hukushima, K. and K. Nemoto, *Exchange Monte Carlo method and application to spin glass simulations.* Journal of the Physical Society of Japan, 1996. **65**: p. 1604-1608.

48. Hansmann, U.H.E., *Parallel tempering algorithm for conformational studies of biological molecules.* Chem. Phys. Lett., 1997. **281**: p. 140-150.

49. Marinari, E. and G. Parisi, *Simulated Tempering - a New Monte-Carlo Scheme.* Europhysics Letters, 1992. **19**: p. 451-458.

50. Lyubartsev, A.P., et al., *New Approach to Monte-Carlo Calculation of the Free-Energy - Method of Expanded Ensembles.* Journal of Chemical Physics, 1991. **96**: p. 1776-1783.

51. Kofke, D.A., *On the acceptance probability of replica-exchange Monte Carlo trials.* Journal of Chemical Physics, 2002. **117**: p. 6911-6914.

52. Kone, A. and D.A. Kofke, *Selection of temperature intervals for parallel-tempering simulations.* Journal of Chemical Physics, 2005. **122**: p. -.

53. Rathore, N., M. Chopra, and J.J. de Pablo, *Optimal allocation of replicas in parallel tempering simulations.* Journal of Chemical Physics, 2005. **122**: p. 024111.

54. Earl, D.J. and M.W. Deem, *Optimal allocation of replicas to processors in parallel tempering simulations.* Journal of Physical Chemistry B, 2004. **108**: p. 6844-6849.

55. Earl, D.J. and M.W. Deem, *Parallel tempering: Theory, applications, and new perspectives.* Physical Chemistry Chemical Physics, 2005. **7**: p. 3910-3916.

56. Zuckerman, D.M. and E. Lyman, *A second look at canonical sampling of biomolecules using replica exchange simulation. (vol 2, pg 1200, 2006).* Journal of Chemical Theory and Computation, 2006. **2**: p. 1693-1693.

57. Nadler, W. and U.H.E. Hansmann, *Dynamics and optimal number of replicas in parallel tempering simulations.* Physical Review E, 2007. **76**: p. 065701 (R).

58. Nadler, W. and U.H.E. Hansmann, *Generalized ensemble and tempering simulations: A unified view.* Physical Review E, 2007. **75**: p. 026109.

59. Trebst, S., D.A. Huse, and M. Troyer, *Optimizing the ensemble for equilibration in broad-histogram Monte Carlo simulations.* Physical Review E, 2004. **70**: p. 046701.

60. Trebst, S., M. Troyer, and U.H.E. Hansmann, *Optimized parallel tempering simulations of proteins.* Journal of Chemical Physics, 2006. **124**: p. 174903.

61. Katzgraber, H.G., et al., *Feedback-optimized parallel tempering Monte Carlo.* Journal of Statistical Mechanics-Theory and Experiment, 2006: p. P03018.

62. Park, S., *Comparison of the serial and parallel algorithms of generalized ensemble simulations: an analytical approach.* Physical Review E, 2008. **77**: p. 016709.

63. Nymeyer, H., *How efficient is replica exchange molecular dynamics? An analytic approach.* Journal of Chemical Theory and Computation, 2008. **4**(4): p. 626-636.

64. Sindhikara, D., Y.L. Meng, and A.E. Roitberg, *Exchange frequency in replica exchange molecular dynamics.* Journal of Chemical Physics, 2008. **128**: p. 024103.

65. Abraham, M.J. and J.E. Gready, *Ensuring mixing efficiency of replica-exchange molecular dynamics simulations.* Journal of Chemical Theory and Computation, 2008. **4**(7): p. 1119-1128.