Parallel continuous simulated tempering and its applications in large-scale molecular simulations

Tianwu Zang,¹ Linglin Yu,¹ Chong Zhang,¹ and Jianpeng Ma^{1,2,a)}

¹*Applied Physics Program and Department of Bioengineering, Rice University, Houston, Texas 77005, USA*

²*Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, BCM-125, Houston, Texas 77030, USA*

(Received 25 April 2014; accepted 30 June 2014; published online 25 July 2014)

In this paper, we introduce a parallel continuous simulated tempering (PCST) method for enhanced sampling in studying large complex systems. It mainly inherits the continuous simulated tempering (CST) method in our previous studies [C. Zhang and J. Ma, *J. Chem. Phys.* **130**, 194112 (2009); **132**, 244101 (2010)], while adopts the spirit of parallel tempering (PT), or replica exchange method, by employing multiple copies with different temperature distributions. Differing from conventional PT methods, despite the large stride of total temperature range, the PCST method requires very few copies of simulations, typically 2–3 copies, yet it is still capable of maintaining a high rate of exchange between neighboring copies. Furthermore, in PCST method, the size of the system does not dramatically affect the number of copy needed because the exchange rate is independent of total potential energy, thus providing an enormous advantage over conventional PT methods in studying very large systems. The sampling efficiency of PCST was tested in two-dimensional Ising model, Lennard-Jones liquid and all-atom folding simulation of a small globular protein trp-cage in explicit solvent. The results demonstrate that the PCST method significantly improves sampling efficiency compared with other methods and it is particularly effective in simulating systems with long relaxation time or correlation time. We expect the PCST method to be a good alternative to parallel tempering methods in simulating large systems such as phase transition and dynamics of macromolecules in explicit solvent. © 2014 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4890038>]

I. INTRODUCTION

The main objective of computer simulation in equilibrium statistical physics is to calculate the ensemble average of interested physical properties $\langle A \rangle$ (for example, average potential energy $\langle E \rangle$ and mean square fluctuation $\langle (\Delta E)^2 \rangle$, which is related to heat capacity C_V). The simulation in canonical ensemble can sample the phase space at constant temperature, but the system is easy to get trapped in local minima for a long time.¹ This phenomenon is called broken ergodicity and it limits the accuracy of computed physical properties with trajectories of finite length.

During the past decades, generalized ensemble² was introduced to overcome the broken ergodicity issue. As its name implies, a generalized ensemble uses a more general form of probability distribution $W(X, \beta)$ instead of the Boltzmann distribution $P_{\text{Boltz}}(X; \beta) = Z(\beta)^{-1} e^{-\beta E(X)}$, where X is the configuration, β is the reciprocal temperature, $E(X)$ is the potential energy, and $Z(\beta)$ is the canonical partition function. Among all the generalized ensembles, the idea of multicanonical ensemble^{3,4} has been widely used because of its high barrier-crossing efficiency in the configurational space (CS). The multicanonical ensemble generates a flat histogram of potential energy and its mathematical form of $W(X, \beta)$ can be written as $W(X) \propto \Omega^{-1}(E(X)) = e^{-\beta T S(E(X))}$, where

$\Omega(E(X))$ represents the density of states, T is the temperature and $S(E(X))$ represents the microcanonical entropy. Multicanonical ensemble has already been shown to be very effective in the simulation of many complex systems such as the spin glass,^{3,5} simple liquid systems,⁶ small peptides,⁷ and structure-based model of proteins.⁸

Another typical approach of generalized ensemble is called simulated tempering (ST),⁹ which generates a flat histogram of temperature, i.e., $W(T) = \text{const.}$ or a flat histogram of reciprocal temperature, i.e., $W(\beta) = \text{const.}$ In small peptide simulation, the performance of flat energy histogram method and flat temperature histogram method was shown to be similar.¹⁰ However, a flat histogram in temperature space does not need to estimate $\Omega(E)$ or $S(E)$ of the system,^{11–13} which makes it easier to be applied to larger systems, such as the protein folding simulation in explicit solvent.¹³ ST is particularly useful in the situation that the energy change upon folding of polypeptide is smaller than the total energy fluctuation of the entire system when a large water box is present. In conventional ST, the system jumps between various discrete temperatures.⁹ Since the acceptance ratio for every attempt of jump is related to the overlap between the energy histograms of two canonical ensembles,¹⁴ it decreases quickly when the system size increases. In our previous studies,^{12,13} a single-copy continuous simulated tempering (CST) method was proposed to overcome this limitation and allow the application of ST in larger systems.

^{a)} Author to whom the correspondence should be addressed. Electronic mail: jpma@bcm.tmc.edu. Phone: 713-798-8187. Fax: 713-796-9438.

The CST method has already been shown to be very successful in folding simulation of small proteins in explicit solvent.^{13,15} However, for systems with long relaxation time such as folding a long polypeptide¹⁶ or correlation time such as phase transition,¹⁷ a single copy method is often far from enough because the residing time of the simulation in the important temperature range (e.g., near room temperature in folding) is shorter than the relaxation time of the system (the time required for the polypeptide chain to structurally rearrange).

Parallel tempering (PT),^{18,19} or replica exchange, provides another way of searching the CS. In PT, multiple copies are used and each copy samples in canonical ensemble at different temperatures. At certain time intervals, an attempt of temperature exchange between neighboring copies is made with the acceptance ratio of $\min\{1, e^{\Delta\beta\Delta E}\}$, where $\Delta\beta$ and ΔE represent the temperature and potential energy difference between involved copies, respectively. In the conventional PT, one of the major limitations in simulating large systems is that the acceptance ratio falls sharply when the system size increases as ΔE becomes quite large.²⁰ To avoid this, an increase of copy $\propto \sqrt{N}$ (N is the degree of freedom)²⁰ is required to keep a reasonable acceptance ratio, which consumes the computing resources significantly. Although progresses²¹ have also been made to increase the exchange probability and extend the scope of PT, this issue remains to be a major obstacle in the application of conventional PT in large systems.

In this paper, we introduce a new method for importance sampling in molecular simulation of large complex systems. This method inherits features from CST method^{12,13} in our previous studies and adopts the spirit of PT in employing multiple copies with different temperature distributions and a specific exchange protocol to satisfy the detailed balance. The temperature distribution in each copy is specified in a large temperature range, and the energy histograms are much wider than the ones in canonical ensemble, which brings a sufficient overlap between them and thus guarantees a higher exchange rate. Furthermore, an important feature of our method is that the exchange rate is independent of total potential energy E , therefore it does not suffer from the increasing of copy number with the growth of the system size compared to conventional PT (in practice, typical number of copies is 2–3). With the combination of CST and PT, the PCST method can effectively simulate the systems with long relaxation times or correlation times.

We present the paper as follows: in Sec. II, we first make a review of the CST method^{12,13} along with the estimator based on integral identities¹² developed in our previous studies, and then introduce the parallel probability distribution and the exchange protocol in the PCST method. In Sec. III, the performance of the PCST method is demonstrated in three typical systems, the two-dimensional (2D) Ising model, the Lennard-Jones fluid, and small protein folding (trp-cage) simulation in explicit solvent. The significant performance improvement of PCST over CST is clearly demonstrated in folding simulations. In Sec. IV, we make a summary of the results and discuss the perspectives.

II. METHOD

A. Generalized ensemble

To avoid the direct calculation of density of states $\Omega(E)$ or microcanonical entropy $S(E)$ for large, complex systems, a generalized ensemble that is a collection of canonical ensembles at different temperatures was proposed in previous study,¹² whose probability distribution function, $P(X, \beta)$, follows:

$$P(X, \beta) = CZ^{-1}(\beta)e^{-\beta E(X)}\omega(\beta), \quad (1)$$

where X is the configuration, β is the reciprocal temperature defined in a large range ($\beta_{\min}, \beta_{\max}$), $\omega(\beta)$ is the weight of canonical ensemble at the temperature β in the generalized ensemble, C is the normalization factor of $\omega(\beta)$, $E(X)$ is the potential energy, and $Z(\beta)$ is the canonical partition function. It was also shown that¹³ a choice of $\omega(\beta) \propto \sqrt{C_V(\beta)}/\beta$ would optimize the energy histogram overlap between canonical ensembles at neighboring temperatures. For example, in a system with a roughly constant heat capacity (such as ideal gas or Lennard-Jones system), the optimized $\omega(\beta)$ would be β^{-1} and for other systems like proteins in which heat capacity was observed as $\propto \beta$, a suggested $\omega(\beta)$ would be $\beta^{-1/2}$. Generally speaking, $\omega(\beta)$ can be represented as $\beta^{-\gamma}$ if $\sqrt{C_V(\beta)} \propto \beta^{1-\gamma}$.

B. Parallel generalized ensemble for importance sampling

There are many cases in which one needs to pay special attention to sampling around an important but narrow temperature range, such as the temperature around the critical temperature T_c in simulation of phase transition. Another example is the protein folding simulation, in which the long simulation around room temperature is needed for more thorough sampling near the native state as the relaxation timescales of polypeptides are quite long for large systems.¹⁶ On the other hand, simply biasing the single copy simulation in generalized ensemble toward the lower temperature is not enough, as this will increase the possibility for the system to be trapped in an energy basin at lower temperature and also decrease the ability of barrier crossing at high temperature. To solve the dilemma between barrier-crossing efficiency and demand of importance sampling around key temperature range, we propose a multiple-copy method with the probability distribution of the revised generalized ensemble as follows:

$$P(X, \beta) = CZ^{-1}(\beta)e^{-\beta E(X)}\beta^{-\gamma}f_i(\beta) \quad (2)$$

for the i th copy. The function $f_i(\beta)$ can be designed to either increase or decrease the weight of canonical ensemble at temperature β and the sampling time on certain temperature ranges. For example, a Gaussian function $e^{-(\beta-\beta_i^0)^2/2\sigma_i^2}$ can focus the importance sampling around the interested temperature β_i^0 .

In order to facilitate barrier crossing at high temperature and thorough sampling at lower temperature, we introduce a parameter exchange protocol between high temperature copies and low temperature copies. In this protocol, the

acceptance ratio for copy exchange between the i th copy and the j th copy is defined as

$$\begin{aligned} & \text{Acc}(X_i\beta_i\beta_i^0\sigma_i, X_j\beta_j\beta_j^0\sigma_j \rightarrow X_j\beta_j\beta_i^0\sigma_i, X_i\beta_i\beta_j^0\sigma_j) \\ &= \min \left\{ 1, \frac{P(X_j, \beta_j; \beta_i^0, \sigma_i)P(X_i, \beta_i; \beta_j^0, \sigma_j)}{P(X_i, \beta_i; \beta_i^0, \sigma_i)P(X_j, \beta_j; \beta_j^0, \sigma_j)} \right\} \\ &= \min \{1, e^{-(\Delta\beta\Delta q_0 + \Delta\beta^2\Delta q_1)}\}, \end{aligned} \quad (3)$$

where $\Delta\beta \equiv \beta_i - \beta_j$, $\Delta\beta^2 \equiv \beta_i^2 - \beta_j^2$, $\Delta q_0 \equiv \beta_i^0/\sigma_i^2 - \beta_j^0/\sigma_j^2$, and $\Delta q_1 \equiv -1/(2\sigma_i^2) + 1/(2\sigma_j^2)$.

Specially, if the variances of different distributions are the same, i.e., $\sigma_i = \sigma_j$, then the acceptance ratio becomes

$$\begin{aligned} & \text{Acc}(X_i\beta_i\beta_i^0, X_j\beta_j\beta_j^0 \rightarrow X_j\beta_j\beta_i^0, X_i\beta_i\beta_j^0) \\ &= \min \left\{ 1, \frac{P(X_j, \beta_j; \beta_i^0)P(X_i, \beta_i; \beta_j^0)}{P(X_i, \beta_i; \beta_i^0)P(X_j, \beta_j; \beta_j^0)} \right\} \\ &= \min\{1, e^{-\Delta\beta\Delta\beta^0/\sigma^2}\}, \end{aligned} \quad (4)$$

where $\Delta\beta^0 \equiv \beta_i^0 - \beta_j^0$.

It is easy to show that this protocol satisfies the detailed balance. An important feature of the parameter exchange protocol is that the acceptance ratio is only related to the parameters in generalized ensemble and the current temperature, whereas it does not have an explicit dependence on the current potential energy E . Therefore, compared with the conventional replica-exchange method,¹⁸ this method does not suffer from a decrease of successful exchange rate with the increase of system size. Moreover, the acceptance ratio is user-adjustable via changing the parameters in generalized ensemble. It is noted that at the limit $\sigma \rightarrow 0$, the generalized ensemble is converted into isolated canonical ensembles at temperature β_i^0 for the i th copy. In this case, if the temperature gap $\Delta\beta^0$ is relatively large, then the exchange ratio will decrease to zero.

C. Temperature random walk to guild a continuous tempering

In continuous tempering, a temperature random walk scheme is used instead of the acceptance ratio scheme in conventional ST to generate the desired $P(X, \beta)$. We first write down a general form of one-dimensional Fokker-Planck equation, which follows:

$$\frac{\partial \rho(T)}{\partial t} = -\frac{\partial}{\partial T} [\mu(T)\rho(T)] + \frac{\partial^2}{\partial T^2} [D(T)\rho(T)], \quad (5)$$

where $\rho(T)$ is the instant temperature probability distribution, $D(T)$ is a diffusion term, and $\mu(T)$ represents a drift force in temperature space with the mathematical form of $\partial_T[D(T)P(X, T)]/P(X, T) = D(T)\partial_T \ln P(X, T) + \partial_T D(T)$. Based on the equivalence of Brownian motion and one-dimensional F-P equation in this case, a temperature space random walk can be written as

$$dT = \left[D(T) \frac{\partial \ln P(X, T)}{\partial T} + \frac{\partial D(T)}{\partial T} \right] dt + \sqrt{2D(T)} dW, \quad (6)$$

where dW is the Weiner process. It can be easily shown that the stationary distribution of this F-P equation is $P(X, T)$. This Brownian equation is valid for any form of $D(T)$, while we choose $D(T) = T^2$ to encourage the diffusion at high temperatures.

Considering that $P(X, T) = P(X, \beta)|\partial_T \beta| = P(X, \beta)/k_B T^2$, the same Brownian equation but in the β representation can be derived as

$$\begin{aligned} d\left(\frac{1}{\beta}\right) &= -\frac{\partial \ln P(X, \beta)}{\partial \beta} dt + \frac{\sqrt{2}}{\beta} dW \\ &= \left(E - \bar{E}(\beta) - \frac{\partial \ln \omega(\beta)}{\partial \beta} - \frac{\partial \ln f_i(\beta)}{\partial \beta} \right) dt \\ &\quad + \frac{\sqrt{2}}{\beta} dW, \end{aligned} \quad (7)$$

where $\bar{E}(\beta) \equiv -\partial_\beta \ln Z(\beta)$ denotes the average potential energy at current temperature in canonical ensemble and E is the potential energy of current configuration. In reality, $\bar{E}(\beta)$ can be evaluated during the simulation, and advanced techniques such as integral identity estimator¹³ and adaptive averaging¹² can be used to extract more information from the data and thus enhance the convergence speed and accuracy of $\bar{E}(\beta)$. For $f_i = e^{-(\beta - \beta_i^0)^2/2\sigma_i^2}$, the Langevin equation becomes

$$\frac{d(1/\beta)}{dt} = E - \bar{E}(\beta) + \frac{\gamma}{\beta} + \frac{\beta - \beta_i^0}{\sigma_i^2} + \frac{\sqrt{2}}{\beta} \xi, \quad (8)$$

where γ is either 0.5 or 1 according to the specific system in Eq. (2) and ξ is the standard Gaussian noise. Therefore, this equation updates the temperature adaptively and generates the desired energy and temperature distributions asymptotically.

During the tempering, many small bins (β_i, β_{i+1}) are created in a large temperature range ($\beta_{\min}, \beta_{\max}$). If the current temperature falls into the i th bin, then the current potential energy is collected to calculate the average energy in this bin $\bar{E}_i$, which is defined as $\frac{1}{\beta_{i+1} - \beta_i} \int_{\beta_i}^{\beta_{i+1}} \bar{E}(\beta) d\beta$. Since the bin size is very small, we assume that the average potential energy in this bin is constant, which is already shown to be valid even for finite-size phase transition problems.¹² In this case, the Langevin equation can be approximately written as

$$\frac{d(1/\beta)}{dt} = E - \bar{E}_i + \frac{\gamma}{\beta} + \frac{\beta - \beta_i^0}{\sigma_i^2} + \frac{\sqrt{2}}{\beta} \xi. \quad (9)$$

Since the energy data are not reliable in the early stage of simulation when the system has not equilibrated properly, adaptive averaging scheme¹³ is used to improve the accuracy as larger weights are assigned toward newer collected statistics. Suppose the k th sample of potential energy in the i th small bin is $E_i^{(k)}$ and its weight is $w_i^{(k)}$. If we define $S_d^{(n)} = \sum_{k=1}^n w_i^{(k)} E_i^{(k)}$, then the weighted mean and variance of potential energy in this bin can be written as

$$\begin{aligned} \langle E \rangle_i^{(n)} &= S_1^{(n)} / S_0^{(n)} \\ \langle (\Delta E)^2 \rangle_i^{(n)} &= \langle E^2 \rangle_i^{(n)} - \langle E \rangle_i^{(n)2} = (S_2^{(n)} - S_1^{(n)2} / S_0^{(n)}) / S_0^{(n)} \\ &\equiv \sigma^{(n)} / S_0^{(n)}. \end{aligned} \quad (10)$$

The adaptive averaging scheme uses $w_i^{(k)} = 1/(1 - \frac{C}{k})$, where C is a constant that is smaller than 1 and can gradually reduces $w_i^{(k)}$ from $1/(1 - C)$ to 1 as $k \rightarrow \infty$. In practice, $S_0^{(n)}$, $S_1^{(n)}$, and $\sigma^{(n)}$ are calculated incrementally as

$$\begin{aligned} S_0^{(n)} &= S_0^{(n-1)} + w_i^{(n)} \\ S_1^{(n)} &= S_1^{(n-1)} + w_i^{(n)} E_i^{(k)} \\ \sigma^{(n)} &= \sigma^{(n-1)} + w_i^{(n)} (E_i^{(k)} - S_1^{(n-1)}/S_0^{(n-1)}) (E_i^{(k)} - S_1^{(n)}/S_0^{(n)}) \end{aligned} \quad (11)$$

However, the convergence speed and accuracy of $\langle E \rangle_i^{(n)}$ and $\langle \Delta E^2 \rangle_i^{(n)}$ are still very poor since only limited number of sample is available in every small bin for a short simulation. To overcome this limitation, a multiple-bin estimator based on integral identity¹³ (MEII) was proposed. The effectiveness of MEII in enhancing the accuracy of estimating $\bar{E}_i$ was clearly shown in previous work.¹³ We briefly summarize the main points of MEII in Sec. II D.

D. Multiple-bin estimators based on integral identity

The main idea of MEII is to borrow the statistics in neighboring bins to provide a much more accurate estimation of $\bar{E}_i$. We will call the estimation of $\bar{E}_i$ as $\tilde{E}_i$. In canonical ensemble, the heat capacity has a direct relationship with the fluctuation of potential energy at the same temperature,

$$C_V(T) \equiv \left(\frac{\partial \bar{E}}{\partial T} \right)_{N,V} = \frac{1}{k_B T^2} \overline{(\Delta E)^2}. \quad (12)$$

We can also write the equation above as

$$\frac{\partial \bar{E}}{\partial \beta} = -\overline{(\Delta E)^2}, \quad (13)$$

which brings the integral identity for any continuous function $\phi(\beta)$ with zero values at the boundary of a larger temperature window between β_- and β_+ as

$$\begin{aligned} \int_{\beta_-}^{\beta_+} \phi'(\beta) \bar{E}(\beta) d\beta &= \phi(\beta) \bar{E}(\beta) \Big|_{\beta_-}^{\beta_+} - \int_{\beta_-}^{\beta_+} \phi(\beta) \frac{\partial \bar{E}(\beta)}{\partial \beta} d\beta \\ &= \int_{\beta_-}^{\beta_+} \phi(\beta) \overline{(\Delta E)^2} d\beta. \end{aligned} \quad (14)$$

Given the approximation that $\bar{E}(\beta) = \bar{E}_i$ and $\overline{(\Delta E)^2} = \overline{(\Delta E)^2}_i$ in each small bin, the integrals in both sides of the above equation become the sums over all the small bins as

$$\sum_j \bar{E}_j(\beta) \int_{\beta_j}^{\beta_{j+1}} \phi'(\beta) d\beta = \sum_j \overline{(\Delta E)^2}_j \int_{\beta_j}^{\beta_{j+1}} \phi(\beta) d\beta.$$

Therefore, the integral identity can be written in the vector form as

$$\Phi^T \mathbf{E} = \Phi^T \mathbf{F}, \quad (15)$$

where $\Phi_j \equiv \int_{\beta_j}^{\beta_{j+1}} \phi'(\beta) d\beta$; $\phi_j \equiv \int_{\beta_j}^{\beta_{j+1}} \phi(\beta) d\beta$; $\mathbf{E}_j \equiv \bar{E}_j$; $\mathbf{F}_j \equiv \overline{(\Delta E)^2}_j$.

Thus, the energy average in the i th bin $\bar{E}_i$ is exactly

$$\bar{E}_i = -\Phi_j \mathbf{E}_j + \phi_j \mathbf{F}_j + \bar{E}_i = \psi_j \mathbf{E}_j + \phi_j \mathbf{F}_j, \quad (16)$$

where $\psi_j \equiv -\Phi_j + \delta_{ij}$.

In real simulation, while $\mathbf{E}_j$ and $\mathbf{F}_j$ are substituted by trajectory average $\langle E_j \rangle$ and $\langle (\Delta E)^2 \rangle_j$, the form of $\phi(\beta)$ can be selected to keep $\Phi^T \mathbf{F}$ being zero to reduce the sampling error. In this case, an unbiased estimation of any interested $\bar{E}_i$ becomes

$$\tilde{E}_i = \psi_j \mathbf{E}_j. \quad (17)$$

In previous work,²² the form of $\phi(\beta)$ was given as

$$\phi(\beta) = \begin{cases} a_- \frac{\beta - \beta_-}{\beta_{i+1} - \beta_-}, & \beta \in (\beta_-, \beta_i) \\ a_- \frac{\beta - \beta_-}{\beta_{i+1} - \beta_-} - \frac{\beta - \beta_i}{\beta_{i+1} - \beta_i}, & \beta \in (\beta_i, \beta_{i+1}) \\ a_+ \frac{\beta - \beta_+}{\beta_+ - \beta_{i+1}}, & \beta \in (\beta_{i+1}, \beta_+) \end{cases} \quad (18)$$

so

$$\begin{aligned} \psi_j &= \begin{cases} \frac{a_-}{a_- \Delta \beta_j / (\beta_{i+1} - \beta_-)}, & j \leq i \\ \frac{a_+}{a_+ \Delta \beta_j / (\beta_+ - \beta_{i+1})}, & j > i \end{cases} \quad \text{and} \\ \phi_i &= \begin{cases} a_- \frac{\bar{\beta}_j - \beta_-}{\beta_{i+1} - \beta_-} \Delta \beta_j, & j < i \\ a_- \frac{\bar{\beta}_i - \beta_-}{\beta_{i+1} - \beta_-} \Delta \beta_i - \frac{1}{2} \Delta \beta_i, & j = i \\ -a_+ \frac{\bar{\beta}_j - \beta_+}{\beta_+ - \beta_{i+1}} \Delta \beta_j, & j > i \end{cases} \end{aligned} \quad (19)$$

where $\bar{\beta}_j \equiv \frac{\beta_j + \beta_{j+1}}{2}$, $\Delta \beta_j \equiv \beta_{j+1} - \beta_j$, and $a_- + a_+ = 1$ ($a_- > 0$, $a_+ > 0$). Usually $\Delta \beta_j$, which is the bin size of the j th bin, is a constant and does not change with the index j . The parameters a_- and a_+ also represent the “weight” of the sum of $\psi_j \mathbf{E}_j$ on the left and right side of the i th bin, and they are determined by the equation $\phi_j \mathbf{F}_j = 0$. In practice, the value of ϕ_j and ψ_j do not change with time and only need to be calculated once. Moreover, the window (β_-, β_+) is chosen so that the interested bin is approximately in the middle of the window. In summary, the implementation of MEII can be divided into two major steps:

- (i) Given the current $\mathbf{F}_j = \langle (\Delta E)^2 \rangle_j$ of every small bins in the window, get a_- and a_+ using the equation $\phi_j \mathbf{F}_j = 0$ with the form of ϕ_j defined in Eq. (19);
- (ii) Given the current $\mathbf{E}_j = \langle E \rangle_j$ of every small bins in the window and parameters $\{a_-, a_+\}$, get $\tilde{E}_i$ using the equation $\tilde{E}_i = \psi_j \mathbf{E}_j$ with the form of ψ_j defined in Eq. (19).

In this way, $\tilde{E}_i$ can serve as an unbiased estimation of the exact average potential energy $\bar{E}_i$ in the i th bin, and it will converge to $\bar{E}_i$ asymptotically with long time simulation.

E. Simulation protocols

There are three time intervals in parallel generalized ensemble simulation: dt_{CE} , the timestep in canonical ensemble at a fixed temperature, i.e., the simulation timestep; dt_{walk} , the timestep for integrating the Langevin equation guiding the temperature random walk; and dt_{ex} , the time interval to

attempt an exchange of parameters between different copies. The detailed simulation protocols are outlined below.

- (1) For each copy, run a short trajectory in canonical ensemble with the timestep dt_{CE} , and record the potential energy (if available).
- (2) For every time interval of dt_{walk} , (a) if current temperature β falls into the i th bin, then estimate the average energy $\bar{E}_i$ using MEII; (b) update temperature based on the Langevin equation, Eq. (9).
- (3) For every time interval of dt_{ex} , (a) collect the statistics from every copy and combine them; (b) based on the current temperature β and parameters $\{\beta^0, \sigma\}$ in selected pairs of copy, calculate the acceptance ratio using Eq. (3) or Eq. (4); (c) if accepted, exchange the states (coordinates x , velocities v , temperature β , and thermostat-related parameters, etc.) of two copies while keeping generalized ensemble parameters unchanged.

III. RESULTS

A. Two-dimensional Ising model

We first tested our method on a 32×32 Ising model to show how our generalized ensemble works in phase transition problems. We have set the temperature range to be (0.25, 0.65) which covers the critical temperature of the phase transition. The bin size $\delta\beta$ equals to 0.0002. The parameter γ in Eq. (2) was set to zero, which corresponds to flat- β histogram. For each temperature bin, MEII was implemented with the window size of 201 bins. The Langevin equation was integrated after every 100 Monte Carlo moves, with an integrating step $\Delta t = 2 \times 10^{-5}$.

We performed the simulation with length of 1.024×10^9 steps which corresponds to 10^6 flips per site. Two copies are applied and the parameter $\{\beta_0, \sigma\}$ in each copy was set to be $\{0.5, 0.1\}$ and $\{0.4, 0.1\}$. The heat capacity compared with analytical results is shown in Figure 1(a). It can be seen that the errors in heat capacity are relatively small compared to the analytical results. The temperature histograms of each copy are shown in Figure 1(b). It is shown that peak positions and widths of temperature histograms follow the desired Gaussian distribution very well. This indicates that the method can get an accurate estimation of physical properties for this phase transition problem.

B. Lennard-Jones fluid

To demonstrate the convergence of physical properties with time, we tested different methods on the Lennard-Jones fluid with 1000 particles. Three methods were used in total: the previous CST method, PCST method with and without parameter exchange. The PCST method without parameter exchange is equivalent to two independent copies using CST with high temperature and low temperature biases. The molecular simulation package GROMACS 4.6.3²³ was used for molecular dynamics simulation. In the reduced units, the thermostat temperature and density were set as

FIG. 1. (a) The estimated heat capacity per site in two-dimensional Ising model. The embedded figure shows the absolute error from the reference value generated by analytical results. (b) The temperature histogram low temperature (red line) and high temperature copy (green line).

0.9 and 0.78, respectively. Following the reduced unit set in Argon,²⁴ the thermostat temperature and the length of cube in GROMCAS configuration were 107.82 K and 3.699 nm, respectively. The force parameters $\{C_6, C_{12}\}$ in Lennard-Jones potential were $\{6.209 \times 10^{-3}, 9.677 \times 10^{-6}\}$ kJmol⁻¹nm⁶. A force-rescaling scheme¹³ was applied to sample at different temperature using the fixed-temperature thermostat. The temperature range for tempering was set as $\beta = (0.5, 1.21)$. Two copies were used and the parameter $\{\beta_0, \sigma\}$ in each copy was set to be $\{0.65, 0.3\}$ and $\{0.95, 0.3\}$, respectively. Molecular dynamics simulation was implemented with an integration time step $\Delta t = 0.0002$. We first ran the simulation with 1×10^9 steps and set the estimated average energy as reference. For each method, we calculated the absolute error ΔE at different time from the reference one by averaging all the absolute errors in every involved bins in the targeted temperature range.

We show the speed of convergence of estimated average energy at higher temperature range ($\beta = 0.5-0.7$) and lower temperature range ($\beta = 1.0-1.21$) in Figure 2. At high temperature range, the CST method with high temperature

FIG. 2. The convergence of potential energy per atom in low temperature range and high temperature range (embedded). The results of four methods: single-copy CST without any temperature bias, CST with high and low temperature bias, and our PCST, are shown in the figure, respectively. The energy difference from the reference one is calculated using the average difference in each single bin in $\beta \in (0.5, 0.7)$ and $(1.0, 1.21)$, respectively. For single-copy methods, the values of each point are taken from the average of 20 simulations, while for multi-copy method (with two copies), the values are taken from 10 simulations.

bias (“high T” in Figure 2) shows the highest convergence speed, while the one with low temperature bias (“low T” in Figure 2) shows the highest convergence speed at low temperature range. For the original CST method (without any Gaussian temperature bias), the convergence at the high temperature range is similar to the one with high temperature bias, while the convergence at the low temperature range is not as good as the one with low temperature bias. In PCST method, the convergence at low temperature range is enhanced, while the convergence at high temperature can still be kept similar to the CST method with high temperature bias. The result demonstrates that the exchange protocol in PCST method can accelerate the convergence of physical properties at the whole temperature range, especially at the low temperature range compared with the other methods.

C. Small protein folding

We tested our method in folding a 20 amino acid helical protein, tryptophan cage²⁵ (pdb code 1L2Y, sequence NLYIQ WLKDG GPSSG RPPPS), which has already been studied by the CST method in our previous work.¹³ It has an N-terminal helix, a short 3_{10} -helix, and a C-terminal polypyrroline region. The polypyrroline region is relatively flexible and does not form any secondary structure. For comparison, we also used the CST method¹³ for the folding study. We used AMBER99SB²⁶ force field in all simulations.

We have implemented our method into a modified GROMACS 4.6.3 package.²³ In all the simulations, the trp-cage protein was put in a cubic $46 \times 46 \times 46$ Å box with 3143 TIP3P model²⁷ of water molecules and one Cl^- ion to keep the total charge zero. The cut-off distances of Lennard-Jones interaction, electrostatic interaction, and neighbor list were 15 Å. Particle mesh Ewald (PME) method²⁸ was used to calculate the electrostatic force, with the 1.19 Å grid spacing of Fourier transform. The parallel LINCS algorithm²⁹ was used to constrain the hydrogen bonds in protein and the SETTLE algorithm³⁰ was used for constraints in water molecules. The simulation was performed in an (N, V, T) ensemble with thermostat temperature 300 K. Velocity rescaling³¹ scheme was applied on the thermostat. A force-rescaling scheme¹³ was applied to sample at different temperature using the fixed-temperature thermostat. The timestep for molecular dynamics integration was 0.002 ps. The timestep dt_{walk} for integrating the Langevin equation was 0.04 ps, which was the same as the neighbor list refreshing interval. For our PCST method, the time interval dt_{ex} of exchange attempt was 10 ns (5×10^6 steps). Note we have only two copies in this simulation, so the argument “-replex” in GROMACS should be set as 2.5×10^6 because an exchange attempt exists only when t/dt_{ex} is odd.

For both CST and PCST methods, we performed five independent trajectories with length of $2 \mu\text{s}$ for each method. The temperature range ($\beta_{\text{min}}, \beta_{\text{max}}$) was set as $(0.24, 0.41)$ kJ^{-1}mol and the length of small bin for data collection was $\delta\beta = 0.0001$. The window width for MEII was 200 times of the bin size, which was approximately 10% of the total temperature range. The parameter C in adaptive averaging¹² was 0.1. Following the convention in the CST method,¹³ the factor γ in generalized ensemble was set as 1, which would make the system bias toward the high temperature¹³ and accelerate the barrier crossing. In our PCST method, this bias was removed and it was kept as 0.5. The parameters $\{\beta_0, \sigma\}$ in Eq. (8) were set as $\{0.38, 0.05\}$ for the low temperature copy and $\{0.27, 0.13\}$ for the high temperature copy (with the peaks of temperature distribution in 316 K and 445 K), respectively.

The folding results are indicated by alpha-carbon root mean square deviations (C_α -RMSD). The native state is marked by the blue line which represents 3 Å RMSD from the NMR structure.²⁵ The results of five trajectories of CST method are shown in Figure 3(a). All five trajectories reach the native state for a few times. However, the length of time for the system to stay near the native state in each folding event is not very long. The average fraction of time to stay

FIG. 3. The time series of C_{α} -RMSD of five independent trajectories using (a) CST method, (b) PCST method, with the red line and black line representing high temperature copy and low temperature copy, respectively. The blue line in each trajectory indicates the 3 Å RMSD. The total simulation time was 2000 ns for each trajectory.

in the near native state of five trajectories was around 3%, with the smallest value 0.8% (traj 2) and the largest value 6% (traj 5). This may be partly due to high temperature bias was applied to satisfy the barrier-crossing efficiency, and it significantly reduces the time of sampling in the low temperature range. As we mentioned previously, to increase the low temperature sampling, it is far from enough to simply convert the high temperature bias to low temperature bias, e.g., to change the generalized ensemble parameter γ from 1.0 to 0.5 or even smaller values, as the system may be trapped in energy basins.

In Figure 3(b), we show the C_{α} -RMSD of five trajectories using the PCST method. It can be seen that the occurrence of folding events was significantly increased comparing with that of the CST method. Two copies are used in our simulation, one for high temperature and the other for low temperature. For the low temperature copy, the average fraction of time near the native state was 23.7%, with the smallest value 12% and the largest value 42%. For the high temperature copy, the average fraction was 12.8%, with the smallest value 7% and the largest value 20%. The results were achieved by the collaboration of two copies: the low temperature copy is mainly responsible for searching in the folding temperature range, while the high temperature copy is mainly responsible for barrier crossing.

The relationship between RMSD and radius of gyration (R_g) of C_{α} atoms in the PCST method for both copies is shown in Figure 4. As the peaks of temperature distribution in two copies are 445 K and 316 K, respectively, while the experimental folding temperature is 315 K,^{25,32} we can approximately use the RMSD- R_g relationship of two copies to observe the compaction of structure above or around the folding temperature. It is clear that both simulation copies have substantial populations at near native state. In Figure 4(a), the copy samples at high temperature have a significant population in a region around 5 Å RMSD from the native state. In Figure 4(b), for the copy samples around the folding tem-

perature, the near native state region is the only favored region. It is worth mentioning, in high temperature copy simulation, that the R_g of the population peak around 5 Å RMSD is similar to the native state, indicating that this peak

FIG. 4. Normalized joint distribution of the radius of gyration (R_g) versus RMSD of C_{α} atoms in (a) high temperature copy and (b) low temperature copy in the PCST method.

FIG. 5. The C_{α} -RMSD distributions. (a) is for PCST method. The dashed and solid lines denote the distribution in the high temperature copy and low temperature copy, respectively. (b) is for the CST method.

corresponds to a state that the hydrophobic collapse has already happened (molten globular state) and that an enormous entropic barrier is present moving from molten globular state to the native state. The barrier between this state and the native state is very eminent. Such results indicate that the generalized ensemble sampling methods are useful in exploring free energy landscape and characterizing the folding transition states, even though the timescale in this kind of simulations is not realistic.

The normalized RMSD distributions of two methods are shown in Figures 5(a) and 5(b). Compared to the CST method, the PCST method allows the system to sample near the native state more thoroughly, so the population in the low RMSD range is enhanced. Furthermore, in the distribution of the PCST method, two distinct peaks were observed around 1 Å and 2 Å RMSD, respectively. This seems to suggest that there are two substates near the native state. This phenomenon can also be easily seen in Figure 3(b). We show the amplified parts of trajectory 3 and 5 (Figure 6(a)) along with the molecular structures of the two substates (Figure 6(b)). Between the two substates, the N-terminal helices of the two structures are almost the same, and differences between the two are mainly caused by the C-terminal polypyrroline region. Our observation is consistent with the previous computational^{33,34} and experimental^{34,35} studies about the existence of these two substates.

To further characterize the nature of the two substates, we analyzed the C_{α} - $C_{\alpha}(\text{helix})$ RMSD distribution for the PCST

FIG. 6. (a) Enlargement of parts of traj 3 and traj 5 from Fig. 3(b). The selected time windows are 300–500 ns for traj 3 and 1500–1800 ns for traj 5. The blue line in each trajectory indicates 3 Å RMSD. The RMSD below 3 Å clearly shows two different substates. (b) Molecular structure of two substates. The two structures are aligned with respect to the N-terminal helix (residues 1–10). Two important residues, Pro17 and Pro19, whose sidechains show big difference between two substates, are explicitly drawn. Another residue Ser14, from which large difference of C-alpha trace between the two substates can be seen, is also indicated. The difference between the two substates is mainly in the C-terminal proline-rich loop.

method (Figure 7), where $C_{\alpha}(\text{helix})$ represents the C_{α} atoms in the N-terminal helix formed by the first 10 residues. Results of both copies show that the two substates have almost the same $C_{\alpha}(\text{helix})$ RMSD, and the difference in the C_{α} RMSD is caused by the long polypyrroline in C-terminal region. We have found that there seems to be two distinct types of pathway from the collapsed state (the 5 Å RMSD region mentioned before) to the native state, one type is direct folding to the either substate of the native state (indicated by arrow I and II in Figure 7) and the other type is first folding the N-terminal helix to the native state followed by the folding of flexible polypyrroline region to the native state (indicated by arrow III in Figure 7). As the two substates were not observed in the CST method, the results of PCST method demonstrate that the existence of the low temperature copy provides more thorough sampling of the free energy landscape near the native state.

FIG. 7. Normalized joint distribution of the RMSD of C_α with respect to the N-terminal helix versus that with respect to the whole chain, in (a) high temperature copy and (b) low temperature copy using PCST methods. In both cases, three pathways to the native state are indicated as I, II, and III.

IV. CONCLUDING DISCUSSION

In this paper, we present a PCST method which enables parallel copies of simulation to explore the CS collaboratively. A Gaussian distribution in temperature space with different parameters is added on each copy so that it can focus on searching the temperature range around the peak of Gaussian distribution. An exchange protocol of parameters is introduced to eliminate the broken ergodicity issue that might exist in the low temperature copies. PCST inherits the ideas in CST method^{12,13} in using continuous tempering, while borrows the spirits in parallel tempering (PT) methods^{16,17} in using multiple copies of simulation. The employment of various distributions in different copies overcomes the dilemma between barrier-crossing efficiency and importance sampling in particular temperature range, especially in the low temperature range. As the overlap between temperature distributions is sufficient, the exchange rate between copies can be kept at a relatively high value. At the same time, the exchange rate does not dramatically depend on the system size, which allows one to use much smaller number of copies compared to the conventional PT method. Therefore, it can serve as a good alternative for conventional PT in simulating large systems such as phase transition and dynamics of macromolecules in explicit solvent.

In the simulation of two-dimensional Ising model and Lennard-Jones fluid, we have shown that the desired temperature distribution in each copy can be generated correctly in the long simulation. Not only physical properties can be estimated accurately, the convergence speed of them in the whole temperature range is enhanced as well compared to the single-copy method. Especially for the low temperature copy, the introduction of exchange protocol ensures that it can sample various regions and minima in the CS, which accelerates the convergence of physical properties.

PCST also shows its special effectiveness in protein folding simulation, in which the long residing time in low temperature range is required. In the application of folding trp-cage protein, we observed a significant enhancement of occurrence of the native state compared to the CST method in all the trajectories. In addition, the abundant availability of the native states brings the observation of the existence of two substates in the native region. The molecular structures of two substates are in agreement with the previous computational^{33,34} and experimental^{34,35} studies. Moreover, three possible folding pathways are observed in the analyzing of $C_\alpha - C_\alpha(\text{helix})$ RMSD relationship, which indicates that various mechanisms might exist in the folding of trp-cage protein. The detection of two substates and multiple possible pathways demonstrates that the PCST method can sample the free energy landscape of protein folding thoroughly.

In PCST method, the number of simulation copy is much smaller than conventional PT, or replica exchange. In protein folding simulations in explicit solvent, for instance, the PCST method typically needs 2–3 copies, while the PT methods commonly use 30–40 copies or more. The centers and widths of the Gaussian distribution in temperature space are empirical parameters to be determined based on the nature of system. For example, in the folding of trp-cage, the low temperature copy distribution was positioned at 316 K which was around the melting temperature of this polypeptide^{22,29} so it is set to facilitate enhanced sampling in the compact state of the system. The center of Gaussian for high temperature copy was at 445 K, which was set to facilitate effective barrier crossing. The widths of two Gaussians were adjusted to ensure 30%–40% exchange rate. In folding of even larger systems with higher energy barriers, one may need even wider total temperature range, then one copy could be set at low temperature, one copy at high temperature, and some intermediate copies would also be desirable to bridge the entire temperature range.

ACKNOWLEDGMENTS

The authors thank the helpful discussions with Zhenwei Luo. J.M. acknowledges support from a NIH Grant No. (GM067801) and a Welch Grant No. (Q-1512). This work was supported in part by Blue BioU at Rice University under NIH Award No. NCRR S10RR02950 and an IBM Shared University Research (SUR) Award in partnership with CISCO, Qlogic and Adaptive Computing. The authors also acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC and grid resources. Use of matplotlib³⁶ and PyMOL³⁷ is also gratefully acknowledged.

- ¹D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, 2001).
- ²A. Mitsutake, Y. Sugita, and Y. Okamoto, *Biopolymers* **60**, 96–123 (2001).
- ³B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9–12 (1992).
- ⁴B. Baumann, *Nucl. Phys. B* **285**, 391–409 (1987).
- ⁵F. G. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050–2053 (2001); P. Dayal *et al.*, *Phys. Rev. Lett.* **92**, 097201 (2004).
- ⁶Q. L. Yan, R. Faller, and J. J. de Pablo, *J. Chem. Phys.* **116**, 8745–8749 (2002); Q. L. Yan and J. J. de Pablo, *Phys. Rev. Lett.* **90**, 035701 (2003); M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, *Phys. Rev. E* **66**, 056703 (2002).
- ⁷U. H. E. Hansmann and Y. Okamoto, *J. Comput. Chem.* **14**, 1333–1338 (1993); M. H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 4940–4948 (1994); A. Irback and F. Potthast, *J. Chem. Phys.* **103**, 10298–10305 (1995).
- ⁸P. Jiang, F. Yasar, and U. H. E. Hansmann, *J. Chem. Theory Comput.* **9**, 3816–3825 (2013); C. Zhang and M. W. Deem, *J. Chem. Phys.* **138**, 034103 (2013).
- ⁹E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451–458 (1992).
- ¹⁰U. H. E. Hansmann and Y. Okamoto, *Phys. Rev. E* **54**, 5863–5865 (1996).
- ¹¹C. Zhang and J. Ma, *Phys. Rev. E* **76**, 036708 (2007).
- ¹²C. Zhang and J. Ma, *J. Chem. Phys.* **130**, 194112 (2009).
- ¹³C. Zhang and J. Ma, *J. Chem. Phys.* **132**, 244101 (2010).
- ¹⁴C. Zhang and J. Ma, *J. Chem. Phys.* **129**, 134112 (2008).
- ¹⁵Y. X. Liu *et al.*, *J. Phys. Chem. Lett.* **3**, 1117–1123 (2012); C. Zhang and J. Ma, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8139–8144 (2012).
- ¹⁶R. Gilmanishin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3709–3713 (1997); W. A. Eaton *et al.*, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359 (2000).
- ¹⁷R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* **58**, 86–88 (1987); U. Wolff, *ibid.* **62**, 361–364 (1989).
- ¹⁸R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* **57**, 2607–2609 (1986).
- ¹⁹M. Falcioni and M. W. Deem, *J. Chem. Phys.* **110**, 1754–1766 (1999); Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141–151 (1999); D. J. Earl and M. W. Deem, *Phys. Chem. Chem. Phys.* **7**, 3910–3916 (2005).
- ²⁰R. M. Neal, *Stat. Comput.* **6**, 353–366 (1996).
- ²¹Y. M. Rhee and V. S. Pande, *Biophys. J.* **84**, 775–786 (2003); J. Kim, T. Keyes, and J. E. Straub, *J. Chem. Phys.* **132**, 224107 (2010); J. Kim, J. E. Straub, and T. Keyes, *J. Phys. Chem. B* **116**, 8646–8653 (2012).
- ²²R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford at the Clarendon Press, 1938).
- ²³H. J. C. Berendsen, D. Vanderspoel, and R. Vandrunen, *Comput. Phys. Commun.* **91**, 43–56 (1995); E. Lindahl, B. Hess, and D. van der Spoel, *J. Mol. Model.* **7**, 306–317 (2001); D. Van der Spoel *et al.*, *J. Comput. Chem.* **26**, 1701–1718 (2005); B. Hess *et al.*, *J. Chem. Theory Comput.* **4**, 435–447 (2008).
- ²⁴L. A. Rowley, D. Nicholson, and N. G. Parsonage, *J. Comput. Phys.* **17**, 401–414 (1975).
- ²⁵J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen, *Nat. Struct. Biol.* **9**, 425–430 (2002).
- ²⁶J. M. Wang, P. Cieplak, and P. A. Kollman, *J. Comput. Chem.* **21**, 1049–1074 (2000); V. Hornak *et al.*, *Proteins* **65**, 712–725 (2006).
- ²⁷W. L. Jorgensen *et al.*, *J. Chem. Phys.* **79**, 926–935 (1983).
- ²⁸U. Essmann *et al.*, *J. Chem. Phys.* **103**, 8577–8593 (1995).
- ²⁹B. Hess, *J. Chem. Theory Comput.* **4**, 116–122 (2008).
- ³⁰S. Miyamoto and P. A. Kollman, *J. Comput. Chem.* **13**, 952–962 (1992).
- ³¹G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- ³²L. L. Qiu *et al.*, *J. Am. Chem. Soc.* **124**, 12952–12953 (2002).
- ³³K. A. Marino and P. G. Bolhuis, *J. Phys. Chem. B* **116**, 11872–11880 (2012); Q. Shao, J. Y. Shi, and W. L. Zhu, *J. Chem. Phys.* **137**, 125103 (2012).
- ³⁴H. Meuzelaar *et al.*, *J. Phys. Chem. B* **117**, 11490–11501 (2013).
- ³⁵P. Rovo *et al.*, *J. Pept. Sci.* **17**, 610–619 (2011).
- ³⁶J. D. Hunter, *Comput. Sci. Eng.* **9**, 90–95 (2007).
- ³⁷Schrodinger, LLC, 2010.