Comparison of the Serial and Parallel Algorithms of Generalized Ensemble Simulations: An Analytical Approach

Sanghyun Park*
Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA
(Dated: December 18, 2007)

This paper addresses issues related to weights and acceptance rates in generalized ensemble simulations, while comparing two algorithms: serial (e.g., simulated tempering or expanded ensemble method) and parallel (e.g., parallel tempering or replica exchange). We derive a cumulant approximation for weights and discuss its effectiveness in practical applications. We compare the acceptance rates of the serial and parallel algorithms and prove that the serial algorithm always has higher acceptance rates. The duality between forward and backward transitions plays a crucial role in the derivations throughout the paper.

I. INTRODUCTION

Computer simulation, such as molecular dynamics and Monte Carlo, is a powerful technique for studying complex systems. However, simulations of complex systems are often hindered by trapping in local energy minima and slow relaxations. One method to overcome this difficulty is simulated tempering [1, 2], which attempts to reduce relaxation times at low temperatures by repeatedly heating and cooling the system. The idea of simulated tempering can be readily extended to other parameters as in the expanded ensemble method [2] or the simulated scaling method [3]. We will collectively refer to these methods as the serial generalized ensemble method (GEM).

A successful application of serial GEM typically requires rapid and uniform exploration of the given ensemble space. In order to satisfy this criterion, acceptance rates must be not only high but also symmetric between forward and backward transitions. This symmetry can be achieved by assigning weights, that is, by performing weighted sampling of the ensemble space. The weights that yield symmetric acceptance rates are in fact determined by the relative free energies of the ensembles. Thus, we will refer to such weights as 'free energy weights'. Symmetric acceptance rates, however, do not necessarily lead to rapid sampling. To achieve rapid sampling of the ensemble space through high acceptance rates, we need to choose ensembles appropriately so that neighboring ensembles overlap significantly.

A few years after the emergence of simulated tempering, researchers developed parallel versions [4–7], known as parallel tempering or replica exchange. We will refer to these methods as parallel GEM. Because it does not require weights, parallel GEM has been quite popular, especially in molecular simulations. Parallel GEM, on the other hand, needs many computing nodes and frequent communication among them. Thus, parallel GEM is not quite suitable for certain computing environments such as distributed computing, which was the motivation for the recent effort to develop a serial variant of replica exchange [8].

Mitsutake and Okamoto [9] compared the serial and parallel algorithms in a tempering simulation of a peptide and found that simulated tempering has higher acceptance rates than does parallel tempering. In other words, their results suggest that, given the same set of ensembles, serial GEM tends to explore the ensemble space more rapidly, which can be a significant advantage if it turns out to be generally true. For the determination of weights in simulated tempering, they propose to use a short, initial parallel tempering run [9, 10]. More recently, Park and Pande [11] proposed an even simpler method of weight determination based on average energies and demonstrated its effectiveness in a simulation of a peptide in an explicit solvent.

In this paper, we address two issues. First, we derive a cumulant approximation for free energy weights (Section III). This leads to a systematic understanding of the weight determination method of Park and Pande [11], which was originally derived based on a heuristic argument, and to an explanation why such a simple method is so effective. Second, we derive general formulas for acceptance rates of serial and parallel GEM and prove that, with free energy weights, serial GEM always has higher acceptance rates (Section IV). As it turns out, these two issues are closely related, not only because of the link between weights and acceptance rates, but also because the duality between forward and backward transitions is at the heart of both issues. We start by reviewing the serial and parallel algorithms of GEM.

II. GENERALIZED ENSEMBLE METHOD

A generalized ensemble refers to a set of ensembles each associated with a different reduced Hamiltonian[1]

*Electronic address: sanghyun@mcs.anl.gov

[1] We use terms such as reduced Hamiltonian and reduced free energy to denote quantities that have been divided by $k_B T$.

$h_n(x)$ and the corresponding partition function

$$Z_n = \int dx \, \exp[-h_n(x)] \ , \tag{1}$$

where x denotes a microstate of the system and $n = 1, \ldots, K$, with K being the number of ensembles in the generalized ensemble. Let us list a few examples. Simulated tempering deals with a generalized ensemble with respect to temperature:

$$h_n(x) = \beta_n H(x) \ , \tag{2a}$$

where $\beta_n = 1/k_B T_n$ is the nth inverse temperature and H is the original Hamiltonian of the system. We can also construct a generalized ensemble for pressure:

$$h_n(x) = \beta[H(x) + P_n V(x)] \ , \tag{2b}$$

where P_n is the pressure of the nth ensemble and $V(x)$ is the volume. When we want to calculate free energy as a function of a parameter λ, a generalized ensemble with respect to λ can be useful:

$$h_n(x) = \beta H(x, \lambda_n) \ . \tag{2c}$$

It is also possible to construct a generalized ensemble for multiple parameters; for instance, we can combine the above three cases into

$$h_n(x) = \beta_n[H(x, \lambda_n) + P_n V(x)] \ . \tag{2d}$$

In this paper, we adhere to the most general context without specifying any form of $h_n(x)$, except when we discuss particulars of simulated tempering (Eq. 2a) and the free energy calculation (Eq. 2c).

The idea of GEM is to enhance the sampling of microstates by allowing the system to explore all the given ensembles. This is achieved by means of a random walk on the ensemble space (serial GEM) or exchanges of ensembles (parallel GEM). Below we describe these two algorithms.

A. Serial algorithm

Given K different ensembles, a generalized Hamiltonian for serial GEM is defined as

$$\mathcal{H}^S(x, n) = h_n(x) - g_n \ , \tag{3}$$

where $n = 1, \ldots, K$. The generalized partition function is then given as

$$\mathcal{Z}^S = \sum_{n=1}^{K} \int dx \, \exp[-\mathcal{H}^S(x, n)] = \sum_{n=1}^{K} Z_n e^{g_n} \ . \tag{4}$$

We use the superscript S to denote the serial algorithm. In this generalized ensemble, the nth ensemble is weighted by e^{g_n}; g_n is the log weight, but we call it the weight for simplicity.

With the generalized Hamiltonian of Eq. 3, a serial GEM simulation is performed as follows [1, 2]. A simulation is started in one of the K ensembles, and at regular intervals a transition is attempted to a randomly chosen ensemble.[2] Transitions are accepted according to the Metropolis criterion [12]; a transition from the mth to the nth ensemble, when the system is at microstate x, is accepted with probability

$$A^S_{m\to n}(x) = \min\left\{1, \exp[-\Delta\mathcal{H}^S_{m\to n}(x)]\right\} \ , \tag{5}$$

where

$$\Delta\mathcal{H}^S_{m\to n}(x) = h_n(x) - h_m(x) - (g_n - g_m) \ . \tag{6}$$

Notice that adding a constant to the weights has no effect; only the relative weights (differences of weights) matter.

In serial GEM, a random walk is performed on the ensemble space. The frequency that the nth ensemble is visited, as can be seen from Eq. 4, is proportional to $Z_n e^{g_n}$. Therefore, a uniform sampling of ensembles is obtained if and only if

$$g_n = f_n + \text{const.} \ , \tag{7}$$

where f_n is the reduced free energy of the nth ensemble

$$f_n = -\ln Z_n \ . \tag{8}$$

The presence of an arbitrary constant means that the weights and the reduced free energies are equal up to an additive constant. The weights that satisfy this property will be referred to as the free energy weights and will be denoted by $\hat{g}_n$.

In a free energy calculation with respect to a parameter λ (Eq. 2c), the reduced free energy f_n is related to the free energy $F_n = -k_B T \ln Z_n$ by $f_n = \beta F_n$. Thus, the relative reduced free energy $f_n - f_m$ is proportional to the relative free energy $F_n - F_m$, and finding the free energy weights is equivalent to calculating the free energy profile for λ. The free energy profile, therefore, naturally comes out of a serial GEM simulation with respect to λ.

B. Parallel algorithm

In parallel GEM, a generalized Hamiltonian is defined for a set of replicas:

$$\mathcal{H}^P(\mathbf{x}) = \sum_{n=1}^{K} h_n(x_n) \ , \tag{9}$$

[2] Typically, one allows nearest-neighbor transitions only.

where $\mathbf{x} := (x_1, \ldots, x_K)$ denotes microstates of the replicas. The generalized partition function is then given as

$$\mathcal{Z}^P = \int d\mathbf{x}\, \exp[-\mathcal{H}^P(\mathbf{x})] = \prod_{n=1}^{K} Z_n \ . \tag{10}$$

The superscript P denotes the parallel algorithm.

A parallel GEM simulation proceeds as follows [4, 5]. A set of replicas is simulated in parallel, one replica for each ensemble. At regular intervals, an exchange is attempted between a chosen pair of ensembles.[3] Exchanges are accepted according to the Metropolis criterion [12]; an exchange between the mth and the nth ensemble is accepted with probability

$$A^P_{m\leftrightarrow n}(\mathbf{x}) = \min\left\{1, \exp[-\Delta\mathcal{H}^P_{m\leftrightarrow n}(\mathbf{x})]\right\} \ , \tag{11}$$

where

$$\Delta\mathcal{H}^P_{m\leftrightarrow n}(\mathbf{x}) = h_m(x_n) + h_n(x_m) - h_m(x_m) - h_n(x_n) \ . \tag{12}$$

In this algorithm, no weighting is needed because the sampling of ensembles is already uniform; at any instant there is one replica for each ensemble.

III. CUMULANT APPROXIMATION FOR WEIGHTS

In this section, we derive an approximate formula for free energy weights from the cumulant expansion of reduced free energies and discuss the validity of the approximation in the context of simulated tempering and free energy calculations. This section, therefore, concerns only the serial GEM, in which the weight is a relevant concept. Since transitions take place pairwise, namely from one ensemble to another, what we need is the relative weights between pairs of ensembles for which transitions are allowed. Therefore, without loss of generality, we consider transitions between ensembles 1 and 2.

As shown in Eq. 7, uniform sampling of the two ensembles is obtained with

$$\Delta\hat{g} = \Delta f \ , \tag{13}$$

where $\Delta\hat{g} := \hat{g}_2 - \hat{g}_1$ and $\Delta f := f_2 - f_1$. The relative reduced free energy Δf can be written as

$$\Delta f = -\ln\langle e^{-\Delta h}\rangle_1 \ , \tag{14}$$

where $\Delta h := h_2 - h_1$ and $\langle\cdot\rangle_1$ denotes an average over ensemble 1. This is known as the free energy perturbation formula [13] and is a special case of Jarzynski's equality [14]. The right-hand side can be expanded in terms of cumulants:

$$\begin{aligned}\Delta f &= -\sum_{k=1}^{\infty} \frac{(-1)^k}{k!} Q_1^k(\Delta h) \\ &= \langle\Delta h\rangle_1 - \frac{1}{2}\mathrm{var}_1(\Delta h) + \cdots \ ,\end{aligned} \tag{15}$$

where $Q_1^k(\Delta h)$ is the kth-order cumulant of Δh over ensemble 1. This represents an expansion of Δf with respect to ensemble 1. Similarly, Δf can also be written in terms of ensemble 2:

$$\Delta f = \ln\langle e^{\Delta h}\rangle_2 \ , \tag{16}$$

which can be expanded as

$$\begin{aligned}\Delta f &= \sum_{k=1}^{\infty} \frac{1}{k!} Q_2^k(\Delta h) \\ &= \langle\Delta h\rangle_2 + \frac{1}{2}\mathrm{var}_2(\Delta h) + \cdots \ .\end{aligned} \tag{17}$$

Symmetrizing Eqs. 15 and 17 and using Eq. 13, we find

$$\Delta\hat{g} = \frac{1}{2}(\langle\Delta h\rangle_1 + \langle\Delta h\rangle_2) + \frac{1}{4}[\mathrm{var}_2(\Delta h) - \mathrm{var}_1(\Delta h)] + \cdots \ , \tag{18}$$

which, upon truncation, can be used for approximate estimation of free energy weights. This formula may be considered a cumulant expansion of Bennett's acceptance ratio method [15].

The effectiveness of this approximation scheme depends on whether we can truncate the cumulant expansions, Eqs. 15 and 17, at a low order without losing much accuracy. Cumulant expansion has been discussed previously in the context of Jarzynski's equality [14, 16, 17]. If the distribution in question is nearly Gaussian, cumulant expansion generally leads to a good approximation. One complication, however, is that the exponential average may be dominated by a distant tail region of the distribution. In such a case, cumulant expansion may yield a poor approximation if the distribution is far from Gaussian in the region that dominates the exponential average, no matter how close it is to Gaussian in the central region.

In the present case, there are two relevant distributions: ρ_1 and ρ_2, the distributions of Δh over ensemble 1 and 2, respectively. These two distributions are not independent. In fact, one completely determines the other because they are related by

$$\begin{aligned}\rho_1(\epsilon) &= \frac{1}{Z_1}\int dx\, e^{-h_1(x)}\, \delta(\Delta h(x) - \epsilon) \\ &= \frac{e^{-\Delta f}}{Z_2}\int dx\, e^{-h_2(x)+\epsilon}\, \delta(\Delta h(x) - \epsilon) \\ &= e^{\epsilon - \Delta f}\, \rho_2(\epsilon) \ ,\end{aligned} \tag{19}$$

which is a special case of Crooks' fluctuation theorem [18]. This duality implies that $\rho_1(\epsilon)$ and $\rho_2(\epsilon)$,

[3] Typically, one allows nearest-neighbor exchanges only.

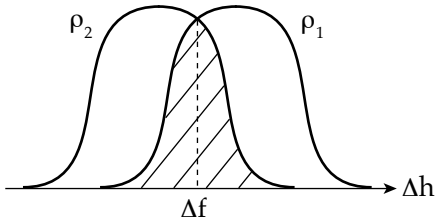

FIG. 1: Schematic diagram of ρ_1 and ρ_2, the distributions of Δh over ensemble 1 and 2. The two distributions intersect at a single point, Δf. As shown in Section IV, the acceptance rate of serial GEM is identical to the shaded area of overlap.

where they are nonzero, intersect at a single point $\epsilon = \Delta f$. And, by applying Jensen's inequality to Eqs. 14 and 16, we find that $\langle \Delta h \rangle_2 \leq \Delta f \leq \langle \Delta h \rangle_1$. Thus, the two distributions must be situated as shown schematically in Fig. 1. In Section IV, we show that the acceptance rate of serial GEM is identical to the area of overlap between ρ_1 and ρ_2.

These properties of ρ_1 and ρ_2 have important implications for the validity of the cumulant approximation for $\Delta\hat{g}$. Since $\rho_1(\epsilon)\, e^{-\epsilon}$ is proportional to $\rho_2(\epsilon)$, the region that dominates the exponential average $\langle e^{-\Delta h} \rangle_1$ coincides with the central region of ρ_2. Similarly, $\langle e^{\Delta h} \rangle_2$ is dominated by the central region of ρ_1. (This has been discussed by Jarzynski [19] in a more general context.) Now, since the overlap determines the acceptance rate, the central regions of ρ_1 and ρ_2 cannot be far apart unless the two ensembles have been chosen so poorly as to yield very low acceptance rates. In other words, the aforementioned situation in which an exponential average is dominated by a distant tail region will not occur as long as we ensure (e.g., by adding intermediate ensembles if necessary) that reasonable acceptance rates are obtained.

In simulated tempering, $\Delta h(x) = \Delta\beta\, H(x)$, where $H(x)$ is the original Hamiltonian of the system and $\Delta\beta := \beta_2 - \beta_1$. The cumulant approximation of Eq. 18 thus becomes

$$\Delta\hat{g} = \frac{\Delta\beta}{2}(\langle H \rangle_1 + \langle H \rangle_2) + \frac{\Delta\beta^2}{4}[\mathrm{var}_2(H) - \mathrm{var}_1(H)] + \cdots . \tag{20}$$

The first term is $O(\Delta\beta)$, but the second term is $O(\Delta\beta^3)$ because $\mathrm{var}_2(H) - \mathrm{var}_1(H)$ is $O(\Delta\beta)$; the absence of $O(\Delta\beta^2)$ terms is a consequence of the symmetrization of Eqs. 15 and 17. If we keep only the first term, we recover the method of Ref. [11], which was originally derived based on a heuristic argument of detailed balance. When the system contains many degrees of freedom, the distribution of H is likely to be nearly Gaussian at least in its central region. In such cases, assuming that there is significant overlap between the distributions of H at the two temperatures, the cumulant approximation is expected to be excellent, as was demonstrated in Ref. [11]. Inclusion of higher orders in Eq. 20 could improve the estimate of $\Delta\hat{g}$, but it is generally unnecessary if weights are to be adjusted through adaptive weighting.

The situation is not quite the same in free energy calculations where $\Delta h(x) = \beta[H(x, \lambda_2) - H(x, \lambda_1)]$. Depending on how the parameter λ is coupled to the system, Δh may or may not contain a significant portion of the system's degrees of freedom. Consequently, the preceding argument does not always apply in free energy calculations. When Δh contains only a small number of degrees of freedom, its distribution may not be close to Gaussian, and the higher orders of the cumulant expansion may be necessary for accurate estimation of free energy weights. We note, however, that when Δh contains a small number of degrees of freedom, Δf between neighboring ensembles tends to be fairly small, and adaptive weighting can readily find the free energy weights even starting with a naive initial guess, for example, $g_n = 0$ [20].

To summarize: When Δh contains a large number of degrees of freedom, Δf between neighboring ensembles can be large, and without good initial weights adaptive weighting will take very long to find the free energy weights. But, since the distribution of Δh is close to Gaussian, we can indeed obtain good initial weights using the cumulant approximation. When Δh contains a small number of degrees of freedom, the distribution of Δh may not be close to Gaussian. But, Δf between neighboring ensembles is small, and the convergence of weights through adaptive weighting is fast even without good initial weights.

IV. ACCEPTANCE RATS

A successful GEM simulation requires rapid exploration of the given ensemble space, which means high acceptance rates for transition attempts in serial GEM and exchange attempts in parallel GEM. In this section, we derive general formulas for the acceptance rates in serial and parallel GEM and address the question of which algorithm has higher acceptance rates. As in Section III, we focus on two ensembles, 1 and 2, without loss of generality.

A. Serial GEM

Transition attempts in serial GEM are accepted with the probability of Eq. 5, expressed as a function of microstate x. Thus, the average acceptance rate for the $1 \to 2$ transition is

$$\begin{aligned} \langle A^S \rangle_{1\to 2} &= \int dx\, \frac{e^{-h_1(x)}}{Z_1} \min\left\{1, e^{-\Delta h(x) + \Delta g}\right\} \\ &= \int_- dx\, \frac{e^{-h_1(x)}}{Z_1} + \int_+ dx\, \frac{e^{-h_2(x) + \Delta g}}{Z_1} , \end{aligned} \tag{21}$$

where $\int_-$ and $\int_+$ denote integrals restricted to the regions $\Delta h(x) < \Delta g$ and $\Delta h(x) > \Delta g$, respectively. This

expression can be rewritten in terms of ρ_1 and ρ_2:

$$\langle A^S\rangle_{1\to 2} = \int_{-\infty}^{\Delta g} d\epsilon\, \rho_1(\epsilon) + \frac{e^{\Delta g}}{e^{\Delta f}} \int_{\Delta g}^{\infty} d\epsilon\, \rho_2(\epsilon)\,. \tag{22}$$

The average acceptance rate for the backward transition can be written similarly:

$$\langle A^S\rangle_{2\to 1} = \frac{e^{\Delta f}}{e^{\Delta g}} \int_{-\infty}^{\Delta g} d\epsilon\, \rho_1(\epsilon) + \int_{\Delta g}^{\infty} d\epsilon\, \rho_2(\epsilon)\,. \tag{23}$$

For an arbitrary choice of Δg, $\langle A^S\rangle_{1\to 2}$ and $\langle A^S\rangle_{2\to 1}$ are different. Only with the choice $\Delta\hat{g} = \Delta f$ do they become identical:

$$\langle \hat{A}^S\rangle = \int_{-\infty}^{\Delta f} d\epsilon\, \rho_1(\epsilon) + \int_{\Delta f}^{\infty} d\epsilon\, \rho_2(\epsilon)\,, \tag{24}$$

where the hat indicates the use of free energy weights. The subscripts $1 \to 2$ and $2 \to 1$ have been dropped because the acceptance rate now is the same in both directions. Since $\rho_1(\epsilon)$ and $\rho_2(\epsilon)$ intersect at a single point $\epsilon = \Delta f$, $\langle \hat{A}^S\rangle$ is identical to the area of overlap between the two distributions (Fig. 1). For reasonable acceptance rates, significant overlap between ρ_1 and ρ_2 is required.

B. Parallel GEM

In parallel GEM, exchange attempts are accepted with the probability given in Eq. 11. Therefore, the average acceptance rate for the $1 \leftrightarrow 2$ exchange is

$$\begin{aligned}\langle A^P\rangle &= \int dx_1\, dx_2\, \frac{e^{-h_1(x_1)}}{Z_1}\, \frac{e^{-h_2(x_2)}}{Z_2}\, \min\left\{1, \frac{e^{\Delta h(x_2)}}{e^{\Delta h(x_1)}}\right\} \\ &= \int_* dx_1\, dx_2\, \frac{e^{-h_1(x_1)}}{Z_1}\, \frac{e^{-h_2(x_2)}}{Z_2} \\ &\quad + \int_\dagger dx_1\, dx_2\, \frac{e^{-h_1(x_2)}}{Z_1}\, \frac{e^{-h_2(x_1)}}{Z_2}\,.\end{aligned} \tag{25}$$

The subscript $1 \leftrightarrow 2$ has been dropped because it is the only exchange possible when we consider two ensembles. The integrals $\int_*$ and $\int_\dagger$ are restricted to the regions $\Delta h(x_1) < \Delta h(x_2)$ and $\Delta h(x_1) > \Delta h(x_2)$, respectively. The two integrals are in fact identical, as we can verify by swapping the dummy variables x_1 and x_2:

$$\langle A^P\rangle = 2\int_* dx_1\, dx_2\, \frac{e^{-h_1(x_1)}}{Z_1}\, \frac{e^{-h_2(x_2)}}{Z_2}\,. \tag{26}$$

Using ρ_1 and ρ_2, we rewrite this as

$$\langle A^P\rangle = 2\int_{-\infty}^{\infty} d\epsilon_2 \int_{-\infty}^{\epsilon_2} d\epsilon_1\, \rho_1(\epsilon_1)\, \rho_2(\epsilon_2)\,. \tag{27}$$

Geometrical interpretation of the acceptance rate is not as straightforward as in serial GEM, but it is clear that parallel GEM also requires significant overlap between ρ_1 and ρ_2 for reasonable acceptance rates.

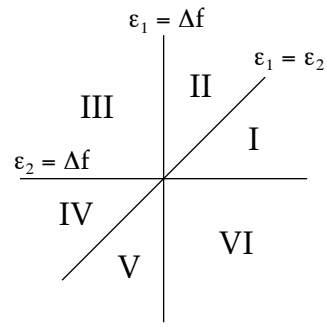

FIG. 2: Three lines, $\epsilon_1 = \Delta f$, $\epsilon_2 = \Delta f$, and $\epsilon_1 = \epsilon_2$, divide the (ϵ_1, ϵ_2) plane into six regions.

C. Comparison

We now prove $\langle \hat{A}^S\rangle \geq \langle A^P\rangle$. To compare Eq. 24 (a single-integral form) with Eq. 27 (a double-integral form), we convert Eq. 24 into a double-integral form:

$$\begin{aligned}\langle \hat{A}^S\rangle &= \int_{-\infty}^{\Delta f} d\epsilon_1\, \rho_1(\epsilon_1) + \int_{\Delta f}^{\infty} d\epsilon_2\, \rho_2(\epsilon_2) \\ &= \int_{-\infty}^{\infty} d\epsilon_2 \int_{-\infty}^{\Delta f} d\epsilon_1\, \rho_1(\epsilon_1)\, \rho_2(\epsilon_2) \\ &\quad + \int_{\Delta f}^{\infty} d\epsilon_2 \int_{-\infty}^{\infty} d\epsilon_1\, \rho_1(\epsilon_1)\, \rho_2(\epsilon_2)\,.\end{aligned} \tag{28}$$

We have thus expressed both quantities in terms of double integrals of $\rho_1(\epsilon_1)\, \rho_2(\epsilon_2)$. Let us examine what regions of the (ϵ_1, ϵ_2) plane each quantity covers. As illustrated in Fig. 2, three lines, $\epsilon_1 = \Delta f$, $\epsilon_2 = \Delta f$, and $\epsilon_1 = \epsilon_2$, divide the plane into six regions, labeled I through VI. The first term of Eq. 28 covers III, IV, and V, and the second term covers I, II, and III:

$$\langle \hat{A}^S\rangle = \left(\int_{\mathrm{I}} + \int_{\mathrm{II}} + 2\int_{\mathrm{III}} + \int_{\mathrm{IV}} + \int_{\mathrm{V}}\right) dR\,, \tag{29}$$

where $dR := d\epsilon_1\, d\epsilon_2\, \rho_1(\epsilon_1)\, \rho_2(\epsilon_2)$. And, Eq. 27 can be written as

$$\langle A^P\rangle = 2\left(\int_{\mathrm{II}} + \int_{\mathrm{III}} + \int_{\mathrm{IV}}\right) dR\,. \tag{30}$$

The difference between the two quantities is

$$\langle \hat{A}^S\rangle - \langle A^P\rangle = \left(\int_{\mathrm{I}} - \int_{\mathrm{II}} - \int_{\mathrm{IV}} + \int_{\mathrm{V}}\right) dR\,. \tag{31}$$

Taking advantage of the symmetry between regions I and II (and regions IV and V) with respect to the reflection in the line $\epsilon_1 = \epsilon_2$, we compare $\int_{\mathrm{I}} dR$ and $\int_{\mathrm{II}} dR$ (and $\int_{\mathrm{IV}} dR$ and $\int_{\mathrm{V}} dR$). By swapping the dummy variables ϵ_1 and ϵ_2, $\int_{\mathrm{I}} dR$ can be turned into an integral over II:

$$\int_{\mathrm{I}} dR = \int_{\mathrm{II}} d\epsilon_1\, d\epsilon_2\, \rho_1(\epsilon_2)\, \rho_2(\epsilon_1)\,. \tag{32}$$

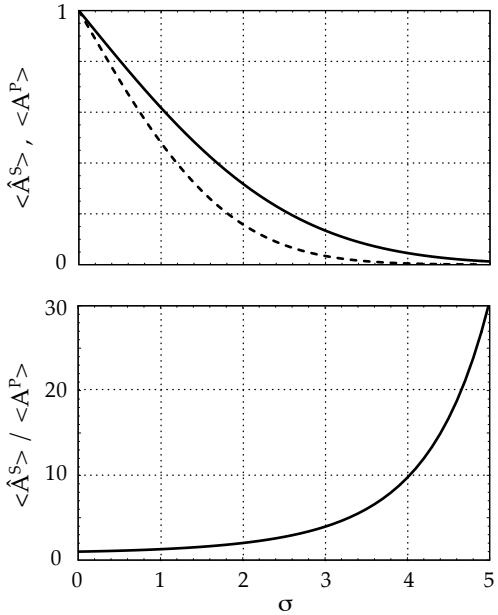

FIG. 3: Acceptance rates, assuming Gaussian distributions. Top: $\langle \hat{A}^S \rangle$ for the serial GEM (solid line) and $\langle A^P \rangle$ for the parallel GEM (dashed line). Bottom: the rate between the two acceptance rates.

Then, we use the duality of Eq. 19 to obtain

$$\begin{aligned} \int_I dR &= \int_{II} d\epsilon_1 d\epsilon_2 \, \rho_1(\epsilon_1) \, \rho_2(\epsilon_2) \, e^{\epsilon_2 - \epsilon_1} \\ &\geq \int_{II} d\epsilon_1 d\epsilon_2 \, \rho_1(\epsilon_1) \, \rho_2(\epsilon_2) = \int_{II} dR \, , \end{aligned} \tag{33}$$

where the inequality holds because $e^{\epsilon_2 - \epsilon_1} \geq 1$ in region II. It can be shown similarly that

$$\int_{IV} dR \leq \int_V dR \, . \tag{34}$$

From Eqs. 31, 33, and 34, we find

$$\langle \hat{A}^S \rangle \geq \langle A^P \rangle \, , \tag{35}$$

which completes the proof.

The actual difference between $\langle \hat{A}^S \rangle$ and $\langle A^P \rangle$ depends on the specific forms of ρ_1 and ρ_2. As a typical example, let us consider the cases where they are Gaussian. Suppose ρ_1 is given as

$$\rho_1(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\epsilon - \mu)^2}{2\sigma^2} \right] \, . \tag{36}$$

Then, from Eqs. 14 and 19 we find

$$\Delta f = \mu - \sigma^2/2 \tag{37}$$

and

$$\rho_2(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\epsilon - \mu + \sigma^2)^2}{2\sigma^2} \right] \, . \tag{38}$$

That is, if ρ_1 is a Gaussian distribution with mean μ and variance σ^2, then ρ_2 must be another Gaussian distribution with the mean shifted by $-\sigma^2$ and the same variance.

Plotted in Fig. 3 are the acceptance rates, $\langle \hat{A}^S \rangle$ (Eq. 24) and $\langle A^P \rangle$ (Eq. 27), calculated with these Gaussian distributions. Notice that since μ has no effect on the acceptance rates, σ is the only relevant parameter. When $\sigma = 0$, two ensembles are identical, and both acceptance rates are unity. As σ increases, the overlap between ρ_1 and ρ_2 decreases; both acceptance rates fall toward zero, but their ratio $\langle \hat{A}^S \rangle / \langle A^P \rangle$ diverges to infinity. In practical applications of GEM, one chooses temperatures (or ensembles in general) such that the acceptance rate is not too low, typically between 30 and 50%. If σ in this Gaussian example is chosen according to this criterion, serial GEM has about 20 to 50% higher acceptance rates. This may appear to be only a moderate increase. If we consider, however, the time scale of mixing on the ensemble space such as the first passage time between the lowest and the highest temperatures, even a moderate increase in the acceptance rate can be a significant advantage. Having higher acceptance rates also means that a certain desired acceptance rate can be achieved with fewer temperatures (or fewer ensembles in general).

V. CONCLUSIONS

In this paper, we have derived a cumulant approximation for free energy weights in serial GEM and explained why such a simple method can be so effective, especially in simulated tempering, as was demonstrated by Park and Pande [11] in an all-atom simulation of a peptide in an explicit solvent. We have also derived general formulas for acceptance rates of serial and parallel GEM and proved that, with free energy weights, serial GEM always has higher acceptance rates. This solidifies the empirical findings of Mitsutake and Okamoto [9]. The duality between forward and backward transitions, which plays a crucial role in these derivations, seems to be a key to deeper understanding of the statistical mechanics of GEM.

In contrast to the recent popularity of parallel GEM (e.g., parallel tempering), serial GEM (e.g., simulated tempering) has gained relatively little attention, because of the difficulty of weight determination. This difficulty seems to be greatly reduced now because free energy weights can be readily obtained by the cumulant approximation combined with adaptive weighting schemes. In addition to being robust in various computing environments, serial GEM has the advantage of having higher acceptance rates, as we have proved here. The acceptance rate is an important criterion of efficiency, but certainly not the only one. It remains to be seen how the increase in the acceptance rate affects the sampling of microstates and the convergence of various estimates. The answers are likely to be system-

dependent and require further comparative study of the serial and parallel algorithms.

Acknowledgments

I thank my dear friend Seung Yeop Lee, whose insight was crucial for the proof of Eq. 35. I also thank John Chodera, Xuhui Huang, Christopher Jarzynski, and Vijay Pande for helpful comments on the manuscript. I gratefully acknowledge the support of the Argonne National Laboratory through the Director's Fellowship under the U.S. Dept. of Energy Contract DE-AC02-06CH11357.

[1] E. Marinari and G. Parisi, *Simulated tempering: A new Monte-Carlo scheme*, Europhys. Lett. **19**, 451 (1992).
[2] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles*, J. Chem. Phys. **96**, 1776 (1992).
[3] H. Li, M. Fajer, and W. Yang, *Simulated scaling method for localized enhanced sampling and simultaneous alchemical free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations*, J. Chem. Phys. **126**, 024106 (2007).
[4] C. J. Geyer and E. A. Thompson, *Annealing Markov chain Monte Carlo with applications to ancestral inference*, J. Am. Stat. Assoc. **90**, 909 (1995).
[5] K. Hukushima and K. Nemoto, *Exchange Monte Carlo method and application to spin glass simulations*, J. Phys. Soc. Jpn **65**, 1604 (1996).
[6] U. H. E. Hansmann, *Parallel tempering algorithm for conformational studies of biological molecules*, Chem. Phys. Lett. **281**, 140 (1997).
[7] Y. Sugita and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*, Chem. Phys. Lett. **314**, 141 (1999).
[8] M. Hagen, B. Kim, P. Liu, R. A. Friesner, and B. J. Berne, *Serial replica exchange*, J. Phys. Chem. B **111**, 1416 (2007).
[9] A. Mitsutake and Y. Okamoto, *Replica-exchange simulated tempering method for simulations of frustrated systems*, Chem. Phys. Lett. **332**, 131 (2000).
[10] A. Mitsutake and Y. Okamoto, *Replica-exchange extensions of simulated tempering method*, J. Chem. Phys. **121**, 2491 (2004).
[11] S. Park and V. S. Pande, *Choosing weights for simulated tempering*, Phys. Rev. E **76**, 016703 (2007).
[12] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equations of state calculations by fast computing machines*, J. Chem. Phys. **21**, 1087 (1953).
[13] R. W. Zwanzig, *High-temperature equation of state by a perturbation method. I. Nonpolar gases*, J. Chem. Phys. **22**, 1420 (1954).
[14] C. Jarzynski, *Nonequilibrium equality for free energy differences*, Phys. Rev. Lett. **78**, 2690 (1997).
[15] C. H. Bennett, *Efficient estimation of free energy differences from Monte Carlo data*, J. Comp. Phys. **22**, 245 (1976).
[16] G. Hummer, *Fast-growth thermodynamic integration: Error and efficiency analysis*, J. Chem. Phys. **114**, 7330 (2001).
[17] S. Park and K. Schulten, *Calculating potentials of mean force from steered molecular dynamics simulations*, J. Chem. Phys. **120**, 5946 (2004).
[18] G. E. Crooks, *Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems*, J. Stat. Phys. **90**, 1481 (1998).
[19] C. Jarzynski, *Rare events and the convergence of exponentially averaged work values*, Phys. Rev. E **73**, 046105 (2006).
[20] S. Park, D. L. Ensign, and V. S. Pande, *Bayesian update method for adaptive weighted sampling*, Phys. Rev. E **74**, 066703 (2006).

The following government license should be removed before publication:

The submitted manuscript has been created by UChicago Argonne, LLC, Operation of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.