

AUTHOR QUERY FORM

	Journal: J. Chem. Phys. Article Number: 018143JCP	Please provide your responses and any corrections by annotating this PDF and uploading it to AIP's eProof website as detailed in the Welcome email.
---	--	--

Dear Author,

Below are the queries associated with your article. Please answer all of these queries before sending the proof back to AIP.

Location in article	Query / Remark: click on the Q link to navigate to the appropriate spot in the proof. There, insert your comments as a PDF annotation.
Q1	AU: Please check the edits done in the sentence beginning with “Illustrative trajectories for....” if the intended meaning is preserved.
Q2	AU: Please check the definition of OPLS, PME, and GPU.
Q3	AU: Please check journal abbreviation of Refs. 27 and 115.
Q4	AU: Please provide article title and update Ref. 32.
Q5	AU: Please check Refs. 44, 84, 87, and 108 for content errors, or provide the DOI.
Q6	AU: Please check publisher location of Refs. 63, 86, and 104.

Thank you for your assistance.

Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing

John D. Chodera^{1,a)} and Michael R. Shirts^{2,b)}

¹Research Fellow, California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, 260J Stanley Hall, Berkeley, California 94720, USA

²Department of Chemical Engineering, University of Virginia, Virginia 22904, USA

(Received 13 June 2011; accepted 26 October 2011; published online XX XX XXXX)

The widespread popularity of replica exchange and expanded ensemble algorithms for simulating complex molecular systems in chemistry and biophysics has generated much interest in discovering new ways to enhance the phase space mixing of these protocols in order to improve sampling of uncorrelated configurations. Here, we demonstrate how both of these classes of algorithms can be considered as special cases of Gibbs sampling within a Markov chain Monte Carlo framework. Gibbs sampling is a well-studied scheme in the field of statistical inference in which different random variables are alternately updated from conditional distributions. While the update of the conformational degrees of freedom by Metropolis Monte Carlo or molecular dynamics unavoidably generates correlated samples, we show how judicious updating of the thermodynamic state indices—corresponding to thermodynamic parameters such as temperature or alchemical coupling variables—can substantially increase mixing while still sampling from the desired distributions. We show how state update methods in common use can lead to suboptimal mixing, and present some simple, inexpensive alternatives that can increase mixing of the overall Markov chain, reducing simulation times necessary to obtain estimates of the desired precision. These improved schemes are demonstrated for several common applications, including an alchemical expanded ensemble simulation, parallel tempering, and multidimensional replica exchange umbrella sampling. © 2011 American Institute of Physics. [doi:10.1063/1.3660669]

I. INTRODUCTION

A broad category of simulation methodologies known as *generalized ensemble*¹ or *extended ensemble*² algorithms have enjoyed increasing popularity in the field of biomolecular simulation over the last decade. The two most popular algorithmic classes within this category are undoubtedly *replica exchange*,³ which includes parallel tempering^{4–6} and Hamiltonian exchange,^{7–10} among others, and its serial equivalent, the method of *expanded ensembles*,¹¹ which includes simulated tempering^{12,13} and simulated scaling.¹⁴ In both classes of algorithms, a mixture of thermodynamic states are sampled within the same simulation, with each simulation walker able to access multiple thermodynamic states through a stochastic hopping process, which we will generically refer to as consisting of *swaps* or *exchanges*. In expanded ensemble simulations, the states are explored via a biased random walk in state space; in replica exchange simulations, multiple coupled walks are carried out in parallel without biasing factors. Both methods allow estimation of equilibrium expectations at each state as well as free energy differences between states. In both cases, stochastic transitions between different thermodynamic states can reduce correlation times and increase sampling efficiency relative to straightforward Monte Carlo

or molecular dynamics simulations by allowing the system to avoid barriers between important configuration substates.

Because of their popularity, these algorithms and their properties have been the subject of intense study over recent years. For example, given optimal weights, expanded ensemble simulations have been shown to have provably higher exchange acceptance rates than replica exchange simulations using the same set of thermodynamic states.¹⁵ Higher exchange attempt frequencies have been demonstrated to improve mixing for replica exchange simulations.^{16,17} Alternative velocity rescaling schemes have been suggested to improve exchange probabilities.¹⁸ Other work has examined the degree to which replica exchange simulations enhance sampling relative to straightforward molecular dynamics simulations.^{19–25} Numerous studies have examined the issue of how to optimally choose thermodynamic states to enhance sampling in systems with second-order phase transitions,^{26–32} though systems with strong first-order-like phase transitions (such as two-state protein systems) remain challenging.^{33,34} A number of combinations^{35,36} and elaborations^{19,37–40} of these algorithms have also been explored. A few publications have examined the mixing and convergence properties of replica exchange and expanded ensemble algorithms with mathematical rigor^{41–44} but there remain many unanswered questions about these sampling algorithms, both in terms of theoretical bounds and practical guidelines for how much these methods accelerate sampling for complex molecular systems.

^{a)}Electronic mail: jchodera@berkeleyu.edu.

^{b)}Author to whom correspondence should be addressed. Electronic mail: michael.shirts@virginia.edu.

Standard practice for expanded ensemble and replica exchange simulations is that exchanges are to be attempted only between “neighboring” thermodynamic states—for example, the states with temperatures immediately above or below the current temperature in a simulated or parallel tempering simulation.^{4–10} The rationale behind this choice is that states further away in state space will have low probability of acceptance due to diminished phase space overlap, and thus attempts should focus on the states for which exchange attempts are most likely to be accepted. Increasing the proximity of neighboring thermodynamic states in both kinds of simulations can further increase the probability that exchange attempts will be accepted. However, restricting exchange attempts to neighboring states can then result in slow overall diffusion in state space due to the larger number of replicas needed to span the thermodynamic range of interest.⁴⁵ Some exchange schemes have been proposed to improve this diffusion process, such as all-pairs exchange,⁴⁶ and optimized exchange moves¹⁸ but the problem of designing optimal exchange moves and overall sampling improvement is still very much a challenge; see Ref. 47 for a recent comparison of different methods. The problem of slow diffusion is exacerbated in “multidimensional” simulations that make use of a 2D or 3D grid of thermodynamic states,^{7,48,49} where diffusion times in state space increase greatly due to the increase in dimensionality.⁵⁰

Here, we show how the many varieties of expanded ensemble and replica exchange simulations can all be considered to be forms of *Gibbs sampling*, a sampling scheme well-known to the statistical inference literature,^{51,52} though unrelated to simulations in the “Gibbs ensemble” for determining phase equilibria.^{53–55} When the problem of sampling among a number of ensembles is viewed in this statistical context, a number of alternative schemes can readily be proposed for updating the thermodynamic state while preserving the distribution of configurations and thermodynamic states sampled by the algorithm. By making simple modifications to the exchange attempt schemes, we show that great gains in sampling efficiency can be achieved under certain conditions with little or no extra cost. There is essentially no drawback to implementing these algorithmic improvements, as the additional computational cost is negligible, their implementation sufficiently simple to encourage widespread adoption, and there appears to be no hindrance of sampling in cases where these schemes offer no great efficiency gain. Importantly, we also demonstrate that schemes that encourage mixing in state space can also encourage mixing of the overall Markov chain, reducing correlation times in coordinate space, leading to more uncorrelated samples being generated for a fixed amount of computer time.

This paper is organized as follows. In Sec. II, we describe expanded ensemble and replica exchange algorithms in a general way, casting them as a form of Gibbs sampling. In Sec. III, we propose multiple approaches to the state exchange process in both classes of algorithm with the aim of encouraging faster mixing in among the thermodynamic states accessible in the simulation, and hence the overall Markov chain. In Sec. IV, we illustrate how and why these modified schemes enhance mixing of the overall chain for a

simple one-dimensional model system. In Sec. V, we apply these algorithmic variants to some examples from physical chemistry, using several different common benchmark systems from biomolecular simulation, and examine several metrics of simulation efficiency. Finally, we make recommendations for the adoption of simple algorithmic variants that will improve efficiency in Sec. VI.

II. THEORY

Before describing our suggested algorithmic modifications (Sec. III), we first present some theoretical tools we will use to analyze expanded ensemble and replica exchange simulations in the context of Gibbs sampling.

A. Thermodynamic states and thermodynamic ensembles

To be as general as possible, we describe the expanded ensemble and replica exchange algorithms as sampling a mixture of K thermodynamic states. Here, a *thermodynamic state* is parameterized by a vector of time-independent thermodynamic parameters λ . For notational convenience and to make what follows general, we define the *reduced potential*⁵⁶ $u(x)$ of a physical system,

$$u(x) = \beta \left[H(x) + pV(x) + \sum_i \mu_i n_i(x) + \cdots \right], \quad (1)$$

corresponding to its thermodynamic state, where x denotes the configuration of the system specifying any physical variables allowed to change, including the volume $V(x)$ (in the case of a constant pressure ensemble) and $n_i(x)$ the number of molecules of each of $i = 1, \dots, M$ components of the system, in the case of a (semi)grand ensemble. The reduced potential is a function of the Hamiltonian H , the inverse temperature $\beta = (k_B T)^{-1}$, the pressure p , and the vector of chemical potentials for each of M components μ_i . Other thermodynamic parameters and their conjugate coordinates can be included in a similar manner, or some of these can be omitted, as required by the physics of the system. We denote the set of all thermodynamic parameters by $\lambda \equiv \{\beta, H, p, \vec{\mu}, \dots\}$.

We next denote a configuration of the molecular system by $x \in \Omega$, where Ω is allowed configuration space, which may be continuous or discrete. A choice of thermodynamic state gives rise to set of configurations of the system that are sampled by a given time-independent probability distribution at equilibrium. So each x will have associated unnormalized probability density $q(x)$, which is a function of λ , where $q(x) > 0$ for all $x \in \Omega$. If we define the normalization constant, or *partition function*, Z as

$$Z \equiv \int_{\Omega} dx q(x), \quad (2)$$

we can define a normalized probability density

$$\pi(x) = Z^{-1} q(x). \quad (3)$$

A physical system in equilibrium with its environment obeying classical statistical mechanics will sample

configurations distributed according to the Boltzmann distribution,

$$q(x) \equiv e^{-u(x)}. \quad (4)$$

In this paper, we consider a set of K thermodynamic states defined by their thermodynamic parameter vectors, $\lambda_k \equiv \{\beta_k, H_k, p_k, \bar{\mu}_k, \dots\}$, with $k = 1, \dots, K$, where H_k denotes any modifications of the Hamiltonian H as a function of k , including biasing potentials. Each new choice of k gives rise to a reduced potential $u_k(x)$, un-normalized and normalized probability distributions $q_k(x)$ and $\pi_k(x)$ (which we also sometimes write as $\pi(k, x)$), and a partition function Z_k . Although in this paper, we generally assume a Boltzmann distribution, there is nothing to prevent some or all of the states from being sampled using non-thermodynamic (non-Boltzmann) statistics using alternative choices of the un-normalized density $q_k(x)$, as in the case of multicanonical simulations⁵⁷ or Tsallis statistics.⁵⁸ To ensure that any configuration x has finite, nonzero density in all K thermodynamic states, we additionally require that the same thermodynamic parameters be specified for all thermodynamic states, though their values may of course differ.

B. Gibbs sampling

Suppose we wish to sample from the joint distribution of two random variables, x and y . We denote this joint distribution by $\pi(x, y)$. Often, it is not possible to generate directly the uncorrelated sample pairs (x, y) from the joint distribution due to the complexity of the function $\pi(x, y)$. In these cases, a standard approach to sampling is to use some form of Markov chain Monte Carlo (MCMC),⁵² such as the Metropolis-Hastings algorithm^{59,60} or hybrid Monte Carlo.⁶¹ While general in their applicability, MCMC algorithms suffer from the drawback that they often must generate *correlated* samples, potentially requiring long running times to produce a sufficient number of effectively uncorrelated samples to allow the computation of properties to the desired precision.^{62,63}

Assume we can draw samples, either independently or through some Markov chain Monte Carlo procedure, from the *conditional* distributions of one or more of the variables, $\pi(x|y)$ or $\pi(y|x)$, where the value of the second variable is fixed. To generate a set of sample pairs $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots\}$ from $\pi(x, y)$, we can iterate the update scheme

$$\begin{aligned} x^{(n+1)}|y^{(n)} &\sim \pi(x|y^{(n)}), \\ y^{(n+1)}|x^{(n+1)} &\sim \pi(y|x^{(n+1)}), \end{aligned}$$

where $x \sim \pi$ denotes that the random variable x is sampled or “updated” from the distribution $\pi(x)$.

This procedure is termed *Gibbs sampling* or the *Gibbs sampler* in the statistical literature, and has been employed and studied extensively.^{51,52} In many cases, it may be possible to draw uncorrelated samples from either or both distributions, but this is not required.⁶⁴ The algorithm still samples from the desired joint distribution correlated samples are gen-

erated instead. The choice of which variable to update—in this example, x or y —can be either deterministic (e.g., update x then y) or stochastic (e.g., a random number determines whether x or y is to be updated); both schemes sample from the desired joint distribution $\pi(x, y)$. However, each method has different dynamic properties and can introduce different correlation structure in the sequence of sample pairs. In particular, we note that a stochastic choice of which variable to update obeys detailed balance, while a deterministic choice obeys the weaker balance condition.⁶⁵ In both cases, the distribution $\pi(x, y)$ is preserved.

In the sections below, we describe how expanded ensemble and replica exchange simulations can be considered as special cases of Gibbs sampling on the probability distribution $\pi(x, k)$, which is now a function of both coordinates and thermodynamic states, and how this recognition allows us to consider simple variations of these techniques that will enhance mixing in phase space with little or no extra cost. In the algorithms we consider here, the thermodynamic state variable k is discrete, but continuous k are also completely valid in this formalism if an appropriate continuous sampler is used.

C. Expanded ensembles

In an expanded ensemble simulation,¹¹ a single replica (or “walker”) samples pairs (x, k) from a joint distribution of configurations $x \in \Gamma$ and state indices $k \in \{1, \dots, K\}$ given by

$$\pi(x, k) \propto \exp[-u_k(x) + g_k], \quad (5)$$

where g_k is an state-dependent weighting factor. This space is therefore a *mixed*, *generalized*, or *expanded* ensemble which samples from multiple thermodynamic ensembles simultaneously. g_k is chosen to give a specific weighting of each subensemble in the expanded ensemble, and is generally determined through some iterative procedure.^{11,12,14,31,66–68} The set of g_k is frequently chosen to give each thermodynamic ensemble equal probability, in which case $g_k = -\ln Z_k$, but they can be set to arbitrary values as desired.

In the context of Gibbs sampling, an expanded ensemble simulation proceeds by alternating between sampling from the two conditional distributions,

$$\pi(x|k) = \frac{q_k(x)}{\int_{\Omega} dx q_k(x)} = \frac{e^{-u_k(x)}}{\int_{\Omega} dx e^{-u_k(x)}}, \quad (6)$$

$$\pi(k|x) = \frac{e^{g_k} q_k(x)}{\sum_{k'=1}^K e^{g_{k'}} q_{k'}(x)} = \frac{e^{g_k - u_k(x)}}{\sum_{k'=1}^K e^{g_{k'} - u_{k'}(x)}}. \quad (7)$$

In all but trivial cases, sampling from the conditional distribution $\pi(x|k)$ must be done using some form of Markov chain Monte Carlo sampler that generates correlated samples, due to the complex form of $u_k(x)$ and the difficulty of computing the normalizing constant in the denominator.⁵² Typically, Metropolis-Hastings Monte Carlo^{59,60} or molecular dynamics is used,⁶⁹ generating an updated configura-

tion $x^{(n+1)}$ that is correlated with the previous configuration $x^{(n)}$, though some types of sampling, such as reservoir exchange methods⁷⁰ attempt to sample independently from a given ensemble through a reservoir. However, as we will see in Sec. III A, multiple choices for sampling from the conditional distribution $\pi(k|x)$ are possible due to the simplicity of its form.

D. Replica exchange ensembles

In a replica exchange, we consider K simulations, with one simulation in each of the thermodynamic K states. The current state of the replica exchange simulation is given by (X, S) , where X is a vector of the replica configurations, $X \equiv \{x_1, x_2, \dots, x_K\}$, and $S \equiv \{s_1, \dots, s_K\} \in \mathcal{S}_K$ is a permutation of the state indices $\{1, \dots, K\}$ associated with each of the replica configurations $\{x_1, \dots, x_K\}$. Then

$$\pi(X, S) \propto \prod_{i=1}^K q_{s_i}(x_i) \propto \exp \left[- \sum_{i=1}^K u_{s_i}(x_i) \right] \quad (8)$$

with the conditional densities given by

$$\pi(X|S) = \prod_{i=1}^K \left[\frac{e^{-u_{s_i}(x_i)}}{\int_{\Omega} dx e^{-u_{s_i}(x_i)}} \right], \quad (9)$$

$$\pi(S|X) = \frac{\exp \left[- \sum_{i=1}^K u_{s_i}(x_i) \right]}{\sum_{S' \in \mathcal{S}_K} \exp \left[- \sum_{i=1}^K u_{s'_i}(x_i) \right]}. \quad (10)$$

As in the case of expanded ensemble simulations, updating the vector of configurations X must be by some form Markov chain Monte Carlo or molecular dynamics simulation, invariably generating configurations with some degree of correlation. Unlike the case of expanded ensembles, generating independent samples in the conditional permutation space is very challenging for even moderate numbers of states because of the expense of computing the denominator of $\pi(S|X)$,⁷¹ which includes a sum over all permutations in the set $\mathcal{S}_K$. However, as we shall see in Sec. III B, there are still effective ways to generate *nearly* uncorrelated permutations that have improved mixing properties over traditional exchange attempt schemes.

III. ALGORITHMS

We now describe the *algorithms* used in sampling from the expanded ensemble and replica exchange ensembles described in Sec. II. We start with the typical neighbor exchange schemes commonly used in the literature, and then describe additional novel schemes based on Gibbs sampling that can encourage more rapid mixing among the accessible thermodynamic states.

A. Expanded ensemble simulation

For an expanded ensemble simulation, the *conditional* distribution of the state index k given x is, again

$$\pi(k|x) = \frac{e^{g_k - u_k(x)}}{\sum_{k'=1}^K e^{g_{k'} - u_{k'}(x)}}.$$

We can use any proposal/acceptance scheme that ensures this conditional distribution is sampled in the long run for any fixed x . We can choose at each step to sample in either k or x according to some fixed probability p , in which case detailed balance is obeyed. We can also alternate N_k and N_x steps of k and x sampling, respectively. Although this algorithm does not satisfy detailed balance, it does satisfy the weaker condition of *balance*⁶⁵ which is sufficient to preserve sampling from the joint stationary distribution $\pi(x, k)$. In the cases that proposal probabilities are based on past history, however, the algorithm will not preserve the equilibrium distribution,⁷² though in some cases the deviations caused by the history dependence can be mitigated with proper choices of parameters.⁷³

1. Neighbor exchange

In the neighbor exchange scheme, the proposed state index j given the current state index i is chosen randomly from one of the neighboring states, $i \pm 1$, with probability,

$$\alpha(j|x, i) = \begin{cases} \frac{1}{2} & \text{if } j = i - 1 \\ \frac{1}{2} & \text{if } j = i + 1 \\ 0 & \text{else} \end{cases} \quad (11)$$

and accepted with probability,

$$P_{\text{accept}}(j|x, i) = \begin{cases} 0 & \text{if } j \notin \{1, \dots, K\} \\ \min \left\{ 1, \frac{e^{g_j - u_j(x)}}{e^{g_i - u_i(x)}} \right\} & \text{else} \end{cases}. \quad (12)$$

This scheme was originally suggested by Marinari and Parisi¹² and has been used extensively in subsequent work.^{35,74} A slight variation of this scheme considers the set $\{1, \dots, K\}$ to lie on a torus, such that state $i + nK$ is equivalent to state i for integral n , with the proposal and acceptance probability otherwise left unchanged.

An alternative scheme avoids having to reject choices of j that lead to $j \notin \{1, \dots, K\}$ by modifying the proposal scheme,

$$\alpha(j|x, i) = \begin{cases} \frac{1}{2} & \text{if } k \in \{2, \dots, K-1\}, |j-i| = 1 \\ 1 & \text{if } i = 1, j = i + 1 \leq K \\ 1 & \text{if } i = K, j = i - 1 \geq 1 \\ 0 & \text{else} \end{cases} \quad (13)$$

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

and modifying the acceptance criteria for these two moves to be³⁵

$$P_{\text{accept}}(j|x, i) = \min \left\{ 1, \frac{1}{2} \frac{e^{g_j - u_j(x)}}{e^{g_i - u_i(x)}} \right\} \quad (14)$$

to include the correct Metropolis-Hastings ratio of proposal probabilities.

2. Independence sampling

The most straightforward way of generating an uncorrelated state index i from the conditional distribution $\pi(k|i)$ is by *independence sampling*, in which we propose an update of the state index i by drawing a new j from $\pi(i|x)$ with probability

$$\alpha(j|x, i) = \pi(i|x), \quad (15)$$

and always accepting this new j . While well known in the statistical inference literature⁵²—and the update scheme most closely associated with the use of the Gibbs sampler there—this scheme has been recently discovered independently in the context of expanded ensembles²⁵ and distributed replica sampling.⁷⁵ A straightforward way to implement this update scheme is to generate a uniform random number r on the interval $[0, 1)$, and select the smallest k where $r < \sum_{i=1}^k \pi(i|x)$.

3. Metropolized independence sampling

In what we term a *metropolized independence sampling* move,⁷⁶ a new state index k' is proposed from the distribution,

$$\alpha(j|x, i) = \begin{cases} \frac{\pi(j|x, i)}{1 - \pi(j|x, i)} & j \neq i \\ 0 & j = i \end{cases} \quad (16)$$

and accepted with probability,

$$P_{\text{accept}}(j|x, i) = \min \left\{ 1, \frac{1 - \pi(i|x, i)}{1 - \pi(j|x, i)} \right\}. \quad (17)$$

This scheme has the surprising property that, despite including a rejection step (unlike the independence sampling in Sec. III A 2 above), the mixing rate in $\pi(k|x)$ can be proven to be greater than that of independence sampling,⁷⁶ using the same arguments that Peskun used to demonstrate the optimality of the Metropolis-Hastings criteria over other criteria for swaps between two states. This can be rationalized by noting that metropolized independence sampling updates will always try move away from the current state, whereas standard independence sampling has some nonzero probability to propose to remain in the current state.

4. Restricted range sampling

In some situations, such as simulated scaling¹⁴ or other schemes in which the Hamiltonian differs in a non-trivial way among thermodynamic states, there may be a non-negligible cost in evaluating the un-normalized probability distributions $q_k(x)$ for all k . Because transitions to a states with minimal phase space overlap will have very low probability, prior knowledge of which states have the highest phase space overlap could reduce computational effort with little loss in sam-

pling efficiency if states with poor overlap are excluded from consideration for exchange.

One way to implement such a *restricted range sampling* scheme is to define a set of proposal states $\mathcal{S}_i$ for each state $i \in \{1, \dots, K\}$, with the requirement that $i \in \mathcal{S}_j$ if and only if $j \in \mathcal{S}_i$, and propose transitions from the current (x, i) to a new state j with probability,

$$\alpha(j|x, i) = \begin{cases} \frac{e^{g_j - u_j(x)}}{\sum_{k \in \mathcal{S}_i} e^{g_k - u_k(x)}} & j \in \mathcal{S}_i \\ 0 & j \notin \mathcal{S}_i \end{cases}. \quad (18)$$

This proposal is accepted with probability,

$$P_{\text{accept}}(j|x, i) = \min \left(1, \frac{\sum_{k \in \mathcal{S}_i} e^{g_k - u_k(x)}}{\sum_{k' \in \mathcal{S}_j} e^{g_{k'} - u_{k'}(x)}} \right). \quad (19)$$

We can easily see that this scheme satisfies detailed balance for fixed x . The probability the sampler is initially in $i \in \mathcal{S}_j$ and transitions to $j \in \mathcal{S}_i$, where $j \neq i$, is given by,

$$\begin{aligned} \pi(i|x) \alpha(j|x, i) P_{\text{accept}}(j|x, i) \\ = \left[\frac{e^{g_i - u_i(x)}}{Z(\mathcal{S}_{\text{all}})} \right] \left[\frac{e^{g_j - u_j(x)}}{Z(\mathcal{S}_i)} \right] \left[\min \left(1, \frac{Z(\mathcal{S}_i)}{Z(\mathcal{S}_j)} \right) \right] \end{aligned} \quad (20)$$

$$= \left[\frac{e^{g_j - u_j(x)} e^{g_i - u_i(x)}}{Z(\mathcal{S}_{\text{all}})} \right] [\min(Z^{-1}(\mathcal{S}_i), Z^{-1}(\mathcal{S}_j))] \quad (21)$$

$$= \left[\frac{e^{g_j - u_j(x)}}{Z(\mathcal{S}_{\text{all}})} \right] \left[\frac{e^{g_i - u_i(x)}}{Z(\mathcal{S}_j)} \right] \left[\min \left(1, \frac{Z(\mathcal{S}_j)}{Z(\mathcal{S}_i)} \right) \right] \quad (22)$$

$$= \pi(j|x) \alpha(i|x, j) P_{\text{accept}}(i|x, j), \quad (23)$$

where $Z(\mathcal{S}_i) = \sum_{k \in \mathcal{S}_i} e^{g_k - u_k(x)}$, and $\mathcal{S}_{\text{all}} = \{1, \dots, K\}$. This is simply the detailed balance condition, ensuring that this scheme will sample from the distribution $\pi(i|x)$. Therefore, this scheme samples from the stationary probability $\pi(j|x)$.

For example, we can define $\mathcal{S}_i = \{i - n, \dots, i + n\}$, with $n \ll K$, for all i , making appropriate adjustments to this range at $i < n$ and $i > K - n$. Then we only need to compute the reduced potentials for states $\{\min(1, i - 2n), \dots, \max(K, i + 2n)\}$, rather than all states $\{1, \dots, K\}$. The additional evaluations for $\{\min(1, i - 2n), \dots, i - n - 1\}$ and $\{\max(i + n + 1, K), \dots, \max(K, i + 2n)\}$ are required to ensure that we can calculate both sums in the acceptance criteria (Eq. (19)).

Restricted range sampling simply reduces to independence sampling, as presented in Sec. III A 2, when $\mathcal{S}_i = \{1, \dots, K\}$, and all proposals are therefore accepted. We also note that metropolized independence sampling, given in Sec. III A 3, is exactly equivalent to using the restricted range scheme with $\mathcal{S}_i = \{1, \dots, K\}$ excluding i , such that $\alpha(i|x, i) = 0$ for all i . Any other valid scheme of sets $\mathcal{S}_i$ can be metropolized by removing i from $\mathcal{S}_i$.

Clearly, other state decomposition schemes exist, though the efficiency of such schemes will almost certainly depend on the underlying nature of the thermodynamic states under study. It is possible to define state schemes that preserve

detailed balance, but that are not ergodic, such as $S_1 = S_3 = S_5 = \{1, 3, 5\}$ and $S_2 = S_4 = S_6 = \{2, 4, 6\}$ for $K = 6$, so some care must be taken. In most cases, users will likely use straightforward rules to find locally defined sets such as $S_i = \{i - n, \dots, i + n\}$ or the metropolized version $S_i = \{i - n, \dots, i - 1, i + 1, \dots, i + n\}$, and ergodicity as well as detailed balance will be satisfied. Further analysis of the performance tradeoffs involved in choices of the sets, situations where sets might be chosen stochastically, or more efficient choices of sets that satisfy only balance is beyond the scope of this study.

5. Other schemes

The list above is by no means intended to be exhaustive—many other schemes can be used for updating the state index k , provided they sample from $\pi(k|x)$. Compositions of different schemes are also permitted—even something simple as applying the neighbor exchange scheme a number of times, rather than just once, could potentially improve mixing properties at little or no additional computational cost.

B. Replica exchange simulation

1. Neighbor exchange

In standard replica exchange simulation algorithms, an update of the state permutation S of the (X, S) sampler state only considers exchanges between neighboring states.⁴⁻¹⁰ One such scheme involves attempting to exchange either the set of state index pairs $\{(1, 2), (3, 4), \dots\}$ or $\{(2, 3), (4, 5), \dots\}$, chosen with equal probability.⁷⁷

Each state index pair (i, j) exchange attempt is attempted independently, with the exchange of states i and j associated with configurations x_i and x_j , respectively, accepted with probability

$$P_{\text{accept}}(x_i, i, x_j, j) = \min \left\{ 1, \frac{e^{-[u_i(x_j) + u_j(x_i)]}}{e^{-[u_i(x_i) + u_j(x_j)]}} \right\}. \quad (24)$$

2. Independence sampling

Independence sampling in replica exchange would consist of generating an uncorrelated, independent sample from $\pi(S|X)$. The most straightforward scheme for doing so would require compiling a list of all possible $K!$ permutations of S , evaluating the un-normalized probability $\exp[-\sum_i u_{S_i}(x_i)]$ for each, normalizing by their sum, and then selecting a permutation S' according to this normalized probability. Even if the entire $K \times K$ matrix $\mathbf{U} \equiv (u_{ij})$ with $u_{ij} \equiv u_i(x_j)$ is precomputed, the cost of this sampling scheme becomes impractical even for modestly large K .

Instead, we note that an *effectively* uncorrelated sample from $\pi(S|X)$ can be generated by running a MCMC sampler scheme for a short time with trivial or small additional computational expense. For each step of the MCMC sampler, we pick a pair of state indices (i, j) , with $i \neq j$, uniformly from the set $\{1, \dots, K\}$. The state pair associated with the configurations x_i and x_j are swapped with the same replica exchange Metropolis-like criteria shown in Eq. (24), with the

labels of the states updated after each swap. If we precompute the matrix $\mathbf{U}$, then these updates are extremely inexpensive, and many Monte Carlo update steps of the state permutation vector S can be taken to decorrelate from the previous sample for a fixed set of configurations X , effectively generating an uncorrelated sample $S' \sim P(S|X)$.

In the case where all u_{ij} are equal, then the number of swaps required is $K \ln K$ —a well-known result due to Aldous and Diaconis.⁷⁸ Empirically, we have found that swapping K^3 to K^5 times each state update iteration appears to be sufficient for the molecular cases examined in this paper and in our own work without consuming a significant portion of the replica exchange iteration time, but further experimentation may be required for some systems. The improvement in efficiency over neighbor exchange algorithms stems from the fact that, for each state, there is a nonzero probability of moving to all other states in proportion to the Boltzmann weight difference, thereby allowing exchange with more remote states without significantly down-weighting exchange with neighboring states. While complete mixing in state space is not a requirement for validity of the algorithm in preserving the equilibrium distribution, increasing the number of swap attempts will lead to increased space sampling until the limit of independent sampling is reached.

For parallel tempering, in which only the inverse temperature β_k varies with state index k , computation of $\mathbf{U}$ is trivial if the potential energies of all K states are known. On the other hand, computation of all $u_i(x_j)$ for all $i, j = 1, \dots, K$ may be time-consuming if the potential energy must be recomputed for each state, such as in an alchemical simulation. If the Bennett acceptance ratio (BAR) (Ref. 79) or the improved multistate version MBAR (Ref. 56) are used to analyze data generated during the simulation, however, all such energies are required anyway, and so no extra work is needed if the state update interval matches the interval at which energies are written to disk. Alternatively, if the number of Monte Carlo or molecular dynamics time steps in between each state update is large compared to K , the overall impact on simulation time of the need to compute $\mathbf{U}$ will be minimal.

Instead of performing random pair selections, we could also apply multiple passes of the standard neighbor exchange algorithm (Sec. III B 1)—a method that may be more easily implemented in existing simulation codes. With a sufficient number of passes, independence sampling will also be achieved, though convergence to an independent permutation S will be at a different rate than the random swap algorithm suggested above. Use of multiple consecutive state swaps between configuration updates is not entirely novel—we have heard several anecdotal examples of people experimenting with multiple consecutive state swaps, with sparse mentions in the literature.^{80,81} However, we believe this is the first study to characterize the theory and properties of this particular modification of standard replica exchange.

3. Other schemes

The list of replica exchange methods above is by no means exhaustive—other schemes can be used for updating the state index k , provided they sample from the space of permutations $\pi(S|X)$ in a way that preserves the conditional

distribution. For example, it may be efficient for a node of a parallel computer to perform many exchanges only among replicas held in local memory, and to attempt few exchanges between nodes due to network constraints. Compositions of different schemes are again also permitted.

C. Metrics of efficiency

There is currently no universally accepted metric for assessing sampling efficiency in molecular simulation, and thus it is difficult to quantify exactly how much our proposed algorithmic modifications improve sampling efficiency. In the end, efficient algorithms will decrease the computational effort to achieve an estimate of the desired statistical precision for the expectations or free energy differences of interest. Unfortunately, this can depend strongly on property of interest, the thermodynamic states that are being sampled, and the dynamics of the system studied. While there exist metrics that describe the *worst case* convergence behavior by approximating the slowest eigenvalue of the Markov chain,^{82,83} the worst case behavior can often differ from practical behavior by orders of magnitude.⁸⁴ Here, we make use of a few metrics that will help us understand the time scale of these correlations in sampling under practical conditions.

Complex systems often get stuck in metastable states in configuration space with residence times a substantial fraction of the total available simulation time. This dynamical behavior hinders the sampling of uncorrelated configurations by molecular dynamics simulation or Metropolis Monte Carlo schemes.^{85,86} Systems can remain stuck in these metastable traps even as a replica in an expanded ensemble or replica exchange simulation travels through multiple thermodynamic states,⁸⁷ either because the trap exists in multiple thermodynamic states or because the system does not have enough time to escape the trap before returning to states where the trap exists. While approaches for detecting and characterizing the kinetics of these metastable states exist,^{87,88} the combination of error from discretizing the conformation space and statistical error makes these approaches to compute relaxation times in configuration space not ideal for our purposes.

Here, we instead consider three simple statistics of the observed state index of each replica trajectory as surrogates to assess the improvements in overall efficiency of sampling. Instead of considering the full expanded ensemble simulation trajectory $\{(x^{(0)}, k^{(0)}), (x^{(1)}, k^{(1)}), \dots\}$ or the replica exchange simulation trajectory $\{(X^{(0)}, S^{(0)}), (X^{(1)}, S^{(1)}), \dots\}$, we consider the trajectory of individual replicas projected onto the sequence of thermodynamic state indices $\mathbf{s} \equiv \{s_0, s_1, \dots\}$ visited during the simulation. In long replica exchange simulations, each replica executes a random walk, and statistics can be pooled.⁸⁹ If significant metastabilities in configuration space exist, we hypothesize that these configurational states will have different typical reduced potential $u(x)$ distributions, and therefore induce metastabilities in the state index trajectory $\mathbf{s}$ as well. This metastabilities will be detectable by the methods described below. Each of the measures provides a different way to interpret the mixing of the simulation in state space; we will refer to all of them in the rest of the paper as “mixing times.”

1. Relaxation time from empirical state transition matrix, τ_2

One way to characterize how rapidly the simulation is mixing in state space is to examine the *empirical transition matrix* among states, the $K \times K$ row-stochastic matrix $\mathbf{T}$. An individual element of this matrix, T_{ij} , is the probability that an expanded ensemble or replica exchange walker currently in state i will be found in state j the next iteration of state sampling. From a given expanded ensemble or replica exchange simulation, we can estimate $\mathbf{T}$ by examining the expanded ensemble trajectory history or pooled statistics from individual replicas,

$$T_{ij} \approx \frac{N_{ij} + N_{ji}}{\sum_{k=1}^K [N_{ik} + N_{ki}]}, \quad (25)$$

where N_{ij} is the number of times the replica is observed to be in state j one update interval after being in state i . To obtain a transition matrix $\mathbf{T}$ with purely real eigenvalues, we have assumed both forward and time-reversed transitions in state indexes are equally probable, which is true in the limit of infinite time for all methods described in this paper. To assess how quickly the simulation is transitioning between different thermodynamic states, we compute the eigenvalues $\{\mu_1, \mu_2, \dots, \mu_K\}$ of $\mathbf{T}$ and sort them in descending order, such that $1 = \mu_1 \geq \mu_2 \geq \dots \geq \mu_K$. If $\mu_2 = 1$, the Markov chain is *decomposable*, meaning that two more subsets of the thermodynamic states exist where *no* transitions have been observed between these sets, a clear indicator of very poor mixing in the simulation. In this case, the thermodynamic states characterized by $\{\lambda_1, \dots, \lambda_K\}$ should be adjusted, or additional thermodynamic states inserted to enhance overlap in problematic regions. Several schemes for optimizing the choice of these state vectors exist,^{26–32} but are beyond the scope of this work to discuss here.

If the second-largest eigenvalue μ_2 is such that $0 < \mu_2 < 1$ we can estimate a corresponding *relaxation time* τ_2 as

$$\tau_2 = \frac{\tau}{1 - \mu_2}, \quad (26)$$

where τ is the effective time between exchange attempts. τ_2 then provides an estimate of the total simulation time required for the autocorrelation function in the state index $k^{(n)}$ of a replica at iteration n of the simulation to decay to $1/e$ of the initial value. This estimate holds if the time scale of decorrelation in the configurational coordinate x is fast compared to the decorrelation of the state index k ; that is, if essentially uncorrelated samples could be drawn from $\pi(x|k)$ for each update of $x^{(n+1)}|k^{(n)}$. Because configuration updates for useful molecular problems generally have long correlation times, this τ_2 time represents a lower bound on the observed correlation time for both the state index $k^{(n)}$ and the configuration $x^{(n)}$.

2. Correlation time of the replica state index, τ_{ac}

As a more realistic estimate of how quickly correlations in the state index $k^{(n)}$ decay in a replica trajectory, we also directly compute the correlation time of the state index

history using the efficiency computation scheme described in Sec. 5.2 of Ref. 89, where τ_{ac} is equal to the integrated area under the autocorrelation function. For replica exchange simulations, where all replicas execute an equivalent walk in state space, the unnormalized autocorrelation functions were averaged over all replicas before computing the autocorrelation time by integrating the area under the autocorrelation function. This time, τ_{ac} , gives a practical estimate of how much simulation time must elapse for correlations in the state index to fall to $1/e$. The *statistical inefficiency* is the number of samples required to collect each uncorrelated sample, and can be estimated for a Markovian process by $2\tau_{ac} + 1$, with τ_{ac} in units of time between samples.

3. Average end-to-end transit time of the replica state index, τ_{end}

As an additional estimate of practical efficiency, we measure the average end-to-end transition time for the state index, τ_{end} . This is the average of the time elapsed between the first visit of the state index $k^{(n)}$ to one end point ($k = 1$ or $k = K$) after visiting the opposite end point ($k = K$ or $k = 1$, respectively). This metric of efficiency, or the related “round-trip” time, has seen common use in diagnosing efficiency for simulated-tempering and replica exchange simulations.^{18,28,90,91}

IV. MODEL ILLUSTRATION

To illustrate the motivation behind the idea that speeding up sampling in one coordinate—the state index or permutation—will enhance sampling of the overall Markov chain of (x, k) or (X, P) , we consider a simulated tempering simulation in a one-dimensional model potential,

$$U(x) = 10(x - 1)^2(x + 1)^2, \quad (27)$$

shown in the top panel of Figure 1, along with the corresponding stationary distribution $\pi(x)$ at several temperatures from $k_B T = 1$ to $k_B T = 10$. To simplify our illustration, we directly numerically compute the log-weight factors

$$g_k = -\ln \int_{-\infty}^{+\infty} dx e^{-\beta_k U(x)} \quad (28)$$

so that the simulation has an equal probability to be in each of the K states.

The K inverse temperatures β_k that can be visited during the simulated tempering simulation are chosen to be geometrically spaced,

$$\beta_k = 10^{-(k-1)/(K-1)} \quad \text{for } k = 1, \dots, K \quad (29)$$

Each iteration of the simulation consists of an update of the temperature index k using either neighbor exchange (Sec. III A 1) or independence sampling updates (Sec. III A 2), followed by 100 steps of Metropolis Monte Carlo^{59,60} using a Gaussian proposal with zero mean and standard deviation of 0.1 in the x -coordinate. Simulations are initiated from $(x_0, k_0) = (-1, 1)$.

Illustrative trajectories for $K = 16$ are shown in the middle and bottom panels of Figure 1, along with the correlation

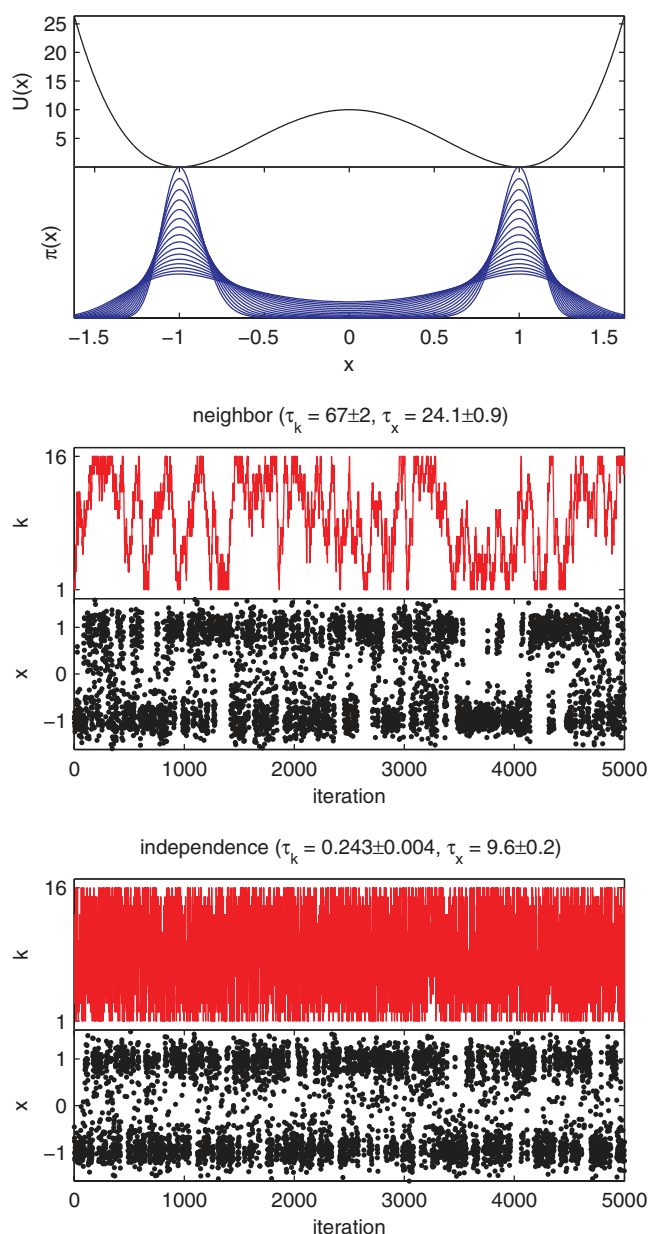


FIG. 1. Simulated tempering for a one-dimensional model. Top panel: Potential energy $U(x)$ and stationary probabilities $\pi(x)$ for one-dimensional two-well model potential at 16 temperatures spanning $k_B T = 1$, where barrier crossing is hindered, to $k_B T = 10$, where barrier crossing is rapid. Middle panel: Temperature index, k , and position, x , histories for a simulated tempering simulation where neighbor swap in temperature are attempted each iteration. Bottom panel: Temperature index, k , and position, x , histories for a simulated tempering simulation where independence sampling of the temperature index is performed each iteration. Only the first 5000 iterations are shown, though simulations of 10^6 iterations were conducted to estimate the correlation times τ_k and τ_x printed above each panel, shown in number of iterations required to produce an effectively uncorrelated sample in either k or x , respectively. Statistical uncertainties shown represent one standard error of the mean.

times τ_k and τ_x computed for the temperature index k and the configurational coordinate x , respectively, from a long trajectory of 10^6 iterations. Independence sampling in state space k greatly reduces the correlation time, and hence statistical inefficiency, in k compared to neighbor sampling. Importantly, because k and x are coupled, we clearly see that increasing

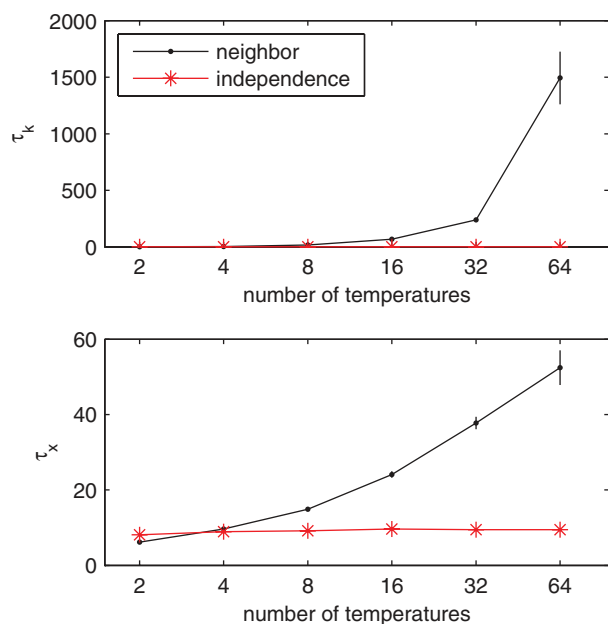


FIG. 2. Autocorrelation times as function of number of temperatures for one-dimensional model. The integrated autocorrelation time τ_k for state index k (top) and τ_x for position x (bottom) as a function of the number of exponentially-spaced temperatures spanning the range $k_B T \in [1, 10]$. The correlation times for neighbor swap (black points) and independence sampling updates (red stars) are shown for each. Error bars represent one standard error of the mean.

the mixing in the index k also substantially reduces the correlation time in the configurational coordinate x . We find that $\tau_x = 9.6 \pm 0.2$ for independence sampling, compared to 24.1 ± 0.9 for neighbor moves.

Figure 2 compares the correlation times for k and x estimated from simulations of length 10^6 for different numbers of temperatures spanning the same range of $k_B T \in [1, 10]$, with temperatures again geometrically spaced according to Eq. (29). As the number of temperatures spanning this range increases, the correlation time in the temperature coordinate k increases, as one would expect for a random walk on domains of increasing size. Notably, increasing the number of temperatures also has the effect of increasing the correlation time of the configuration coordinate x . When independence sampling is used to update the temperature index k instead, the mixing time in k is greatly reduced, and both correlation times τ_k and τ_x remain small even as the number of temperatures is increased.

V. APPLICATIONS

To demonstrate that the simple state update modifications we describe in Sec. III lead to real efficiency improvements in practical simulation problems, we consider three typical simulation problems: An alchemical expanded ensemble simulation of united atom (UA) methane in water to compute the free energy of transfer from gas to water; a parallel tempering simulation of terminally blocked alanine dipeptide in implicit solvent; and a two-dimensional replica exchange umbrella sampling simulation of alanine dipeptide in implicit solvent to compute the potential of mean force. These systems are small compared to modern applications of biophysical and

biochemical interest. However, they are realistic enough to demonstrate the fundamental issues in multiensemble simulations, but still sufficiently tractable that a large quantity of data can be collected to prove that the differences in efficiency of our proposed mixing schemes are highly significant.

A. Expanded ensemble alchemical simulations of Lennard-Jones spheres in water

1. United atom methane

We first compare different types of Gibbs sampling state space updates in an expanded ensemble alchemical simulation of the kind commonly used to compute the free energy of hydration of small molecules.^{35,90} If the state mixing schemes proposed here lead to more efficient sampling among alchemical states, a larger number of effectively uncorrelated samples will be generated for a simulation of a given duration, and thus require less computation effort to reach the desired degree of statistical precision.

An optimized potentials for liquid simulation-UA united atom methane particle ($\sigma = 0.373$ nm, $\epsilon = 1.230096$ kJ/mol) was solvated in a cubic simulation cell containing 893 TIP3P (Ref. 92) waters. For all simulations, a modified version of GROMACS 4.5.2 (Ref. 93) was used.⁹⁴ A velocity Verlet integrator⁹⁵ was used to propagate dynamics with a time step of 2 fs. A Nosé-Hoover chain of length 10 (Ref. 96) and time constant $\tau_T = 10.0$ ps was used to thermostat the system to 298 K. A measure-preserving barostat was used according to Tuckerman *et al.*^{97,98} to maintain the average system pressure at 1 atm, with $\tau_p = 10.0$ ps and compressibility 4.5×10^{-5} bar⁻¹. Rigid geometry was maintained for all waters using the analytical SETTLE scheme.⁹⁹ A neighborlist and photomagnetoelastic (PME) cutoff of 0.9 nm were used, with a PME order of 6, spacing of 0.1 nm and a relative tolerance of 10^{-6} at the cutoff. The Lennard-Jones potential was switched off, with the switch beginning at 0.85 nm and terminating at the cutoff of 0.9 nm. An analytical dispersion correction was applied beyond the Lennard-Jones cutoff to correct the energy and pressure computation.¹⁰⁰ The neighborlist was updated every 10 steps.

A set of $K = 6$ alchemically modified thermodynamic states were used in which the Lennard-Jones interactions between the methane and solvent were eliminated using a soft-core Lennard-Jones potential,¹⁰¹

$$U_{ij}(r; \lambda) = 4\epsilon_{ij}\lambda f(r; \lambda)[1 - f(r; \lambda)],$$

$$f(r; \lambda) \equiv [\alpha(1 - \lambda) + (r/\sigma_{ij})^6]^{-1} \quad (30)$$

with values of the alchemical coupling parameter λ_k chosen to be $\{0.0, 0.3, 0.6, 0.7, 0.8, 1.0\}$.

To simplify our analysis of efficiency, we fix the log-weights g_k to “perfect weights,” where all states are visited with equal probability. This also decouples the issue of efficiency of state updates with efficiency of different weight update schemes, of which many have been proposed.^{11,12,14,31,66–68} The “perfect” log-weights were estimated for this system as follows: A 1 ns expanded ensemble simulation using independence sampling was run, with weights g_k initialized to zero, then adjusted using a

TABLE I. Efficiency measures for expanded ensemble alchemical simulation of united atom methane in water. Times measuring mixing in state space are: τ_2 , estimated from second eigenvalue of the empirical state transition matrix; τ_{ac} , estimated from autocorrelation function of the alchemical state index; τ_{end} , estimated average end-to-end transit time for the alchemical state index. τ_N is a structural parameter, the autocorrelation function of the number of TIP3P oxygen molecules within 0.3 nm (87.5% of the Lennard-Jones σ_{ij} (0.3428 nm)) of the center of the united atom methane particle. The *relative speedup* in sampling efficiency is given relative to the standard neighbor exchange scheme.

	Mixing times (ps)				Relative speedup			
	τ_2	τ_{ac}	τ_{end}	τ_N	τ_2	τ_{ac}	τ_{end}	τ_N
1 state move attempted every 0.1 ps								
Neighbor exchange	1.693 $\pm$ 0.008	6.7 $\pm$ 0.4	11.9 $\pm$ 0.2	5.9 $\pm$ 0.4	1.0	1.0	1.0	1.0
Independence sampling	0.771 $\pm$ 0.004	6.2 $\pm$ 0.2	7.2 $\pm$ 0.1	5.4 $\pm$ 0.2	2.20 $\pm$ 0.02	1.08 $\pm$ 0.08	1.65 $\pm$ 0.04	1.10 $\pm$ 0.09
Metropolized independence	0.645 $\pm$ 0.003	4.6 $\pm$ 0.2	6.6 $\pm$ 0.1	4.4 $\pm$ 0.2	2.62 $\pm$ 0.02	1.5 $\pm$ 0.1	1.81 $\pm$ 0.04	1.3 $\pm$ 0.1
1 000 state moves attempted every 0.1 ps								
Neighbor exchange	0.764 $\pm$ 0.006	4.9 $\pm$ 0.3	7.2 $\pm$ 0.1	4.6 $\pm$ 0.3	1.0	1.0	1.0	1.0
Independence sampling	0.769 $\pm$ 0.005	4.8 $\pm$ 0.2	7.2 $\pm$ 0.1	4.5 $\pm$ 0.2	0.99 $\pm$ 0.01	1.01 $\pm$ 0.08	1.00 $\pm$ 0.02	1.01 $\pm$ 0.07
Metropolized independence	0.774 $\pm$ 0.005	5.0 $\pm$ 0.2	7.5 $\pm$ 0.1	4.7 $\pm$ 0.2	0.99 $\pm$ 0.01	0.98 $\pm$ 0.08	0.96 $\pm$ 0.02	0.98 $\pm$ 0.07
1 state move attempted every 5 ps								
Neighbor exchange	85.8 $\pm$ 2.3	177.7 $\pm$ 17.6	330.0 $\pm$ 16.1	105.3 $\pm$ 12.1	1.0	1.0	1.0	1.0
Independence sampling	39.0 $\pm$ 0.9	69.2 $\pm$ 6.1	141.1 $\pm$ 4.7	49.1 $\pm$ 3.8	2.20 $\pm$ 0.08	2.6 $\pm$ 0.3	2.3 $\pm$ 0.1	2.1 $\pm$ 0.3
Metropolized independence	31.8 $\pm$ 0.4	51.4 $\pm$ 1.9	115.7 $\pm$ 3.4	37.4 $\pm$ 1.4	2.70 $\pm$ 0.08	3.5 $\pm$ 0.4	2.9 $\pm$ 0.2	2.8 $\pm$ 0.3

Wang-Landau scheme,¹⁴ until occupancy of each state was roughly even to within statistical noise. With these approximate weights, a 2 ns expanded ensemble simulation using independence sampling with fixed weights was run, and the free energy of each state was estimated using MBAR.⁵⁶ The log-weights g_k were set to these estimated free energies, which were {0.0, 0.32, -0.46, -1.67, -2.83, -3.66}, in units of $k_B T$. Simulations using these weights deviated by an average of 5% from flat histogram occupancy in states, with an average maximum deviation over all simulations of less than 10%.

The state update procedure was carried out either every 0.1 ps (frequent update) or 5 ps (infrequent update), in order to test the effect of state updates that were much faster than, or on the order of, the conformational correlation times of molecular dynamics, as water orientational correlation times are a few picoseconds.¹⁰² Production simulations with fixed log-weights were run with for 25 ns (250 000 state updates), for frequent updates, or 100 ns (20 000 state updates), for infrequent updates. Three types of state moves were attempted: (1) neighbor exchange moves (described in Sec. III A 1), (2) independence sampling (Sec. III A 2, and (3) metropolized independence sampling (Sec. III A 3. In the case of frequent updates, we additionally performed 1000 trials of the state update every 0.1 ps, instead of a single update, before returning to coordinate update moves with molecular dynamics.

Statistics of the observed replica trajectories are shown in Table I. All three mixing efficiency measures of the state index trajectories described in Sec. III C were computed: relaxation time of the empirical state transition matrix (τ_2), autocorrelation of the state function (τ_{ac}), and average end-to-end distance (τ_{end}).

We additionally look at a measure of correlation in the coordinate direction. For each configuration, we examine the number of O atoms of the water molecules N that are found in the interior of the united atom methane, set to be 0.3 nm (or 87.5% of the Lennard-Jones $\sigma_{ij} = 0.3428$ nm) from the

center. We then compute the autocorrelation function of τ_N of this variable, which is affected both by the dynamics of the state and the dynamical response of the system to changes in state. Uncertainties in these time autocorrelation functions are computed by subdividing the trajectories into $N_S = 10$ subtrajectories, computing the standard error, and then dividing by $\sqrt{N_S}$ to obtain standard error of the $N_S \times$ longer trajectory. Uncertainties changed by less than 5% when computed with $N_S = 20$ for frequent update simulations, and less than 10% for infrequent update simulations.

The relaxation time τ_2 estimated from the second eigenvalue of the empirical state transition matrix (Sec. V B) does appear to provide a lower bound for the other estimated mixing times. For the infrequent state updates, it is only about 25% smaller than τ_N . This suggests that when transition times in state space are of the same order of magnitude as conformational rearrangements τ_2 is not only a lower bound, but is characteristic of sampling through the joint state-configuration space. We additionally note that mixing time τ_2 is empirically exactly proportional to the update frequency; the mixing times for the infrequent update state are exactly (5 ps/0.1 ps) = 50 times longer than the frequent state mixing times, a direct consequence of the fact that the probability of successful state transitions is directly proportional to the rate of attempted transitions.

For both the frequent and infrequent state updates, independence sampling and metropolized independence sampling yield a clear, statistically significant speedup by all sampling metrics. This speedup is accentuated for infrequent updates. For frequent updates, the speedup is between 1.3 and 2.6 for metropolized independence sampling, while for infrequent updates, it ranges between 2.7 and 3.5, as seen in Table I. As expected, attempting many state updates in a row (1000 state moves) using any of the state update schemes effectively recapitulates the independence sampling scheme. Repeated application of any method that obeys the balance condition

TABLE II. Efficiency measures for expanded ensemble alchemical simulation of large LJ sphere in water. Times measuring mixing in state space are: τ_2 , estimated from second eigenvalue of empirical state transition matrix; τ_{ac} , estimated from autocorrelation time of alchemical state index; τ_{end} , estimated average end-to-end transit time for alchemical state index. τ_N is a structural parameter, the autocorrelation function of the number of TIP3P oxygen molecules within 0.5 nm (85.3% of the Lennard-Jones σ_{ij} (0.5860 nm)) of the center of the large Lennard-Jones particle. The *relative speedup* in sampling efficiency is given relative to the standard neighbor exchange scheme.

	Mixing times (ps)				Relative speedup			
	τ_2	τ_{ac}	τ_{end}	τ_N	τ_2	τ_{ac}	τ_{end}	τ_N
1 state move attempted every 0.1 ps								
Neighbor exchange	9.51 $\pm$ 0.01	65.8 $\pm$ 4.2	126.3 $\pm$ 4.2	58.1 $\pm$ 4.3	1.0	1.0	1.0	1.0
Independence sampling	2.586 $\pm$ 0.009	42.9 $\pm$ 2.4	88.4 $\pm$ 2.7	41.5 $\pm$ 2.0	3.68 $\pm$ 0.01	1.5 $\pm$ 0.1	1.43 $\pm$ 0.06	1.4 $\pm$ 0.1
Metropolized independence	2.181 $\pm$ 0.006	48.6 $\pm$ 4.0	88.3 $\pm$ 3.0	46.7 $\pm$ 3.4	4.36 $\pm$ 0.01	1.4 $\pm$ 0.1	1.43 $\pm$ 0.07	1.2 $\pm$ 0.1
1 state move attempted every 1 ps								
Neighbor exchange	95.0 $\pm$ 0.2	211.1 $\pm$ 58.9	507.6 $\pm$ 19.3	167.6 $\pm$ 16.0	1.0	1.0	1.0	1.0
Independence sampling	25.8 $\pm$ 0.1	67.3 $\pm$ 3.6	196.0 $\pm$ 5.8	63.1 $\pm$ 3.3	3.69 $\pm$ 0.02	3.1 $\pm$ 0.9	2.6 $\pm$ 0.1	2.7 $\pm$ 0.3
Metropolized independence	21.6 $\pm$ 0.1	66.8 $\pm$ 2.4	169.2 $\pm$ 4.7	62.1 $\pm$ 2.5	4.40 $\pm$ 0.02	3.2 $\pm$ 0.9	3.0 $\pm$ 0.1	2.7 $\pm$ 0.3

will eventually converge to the same independent sampling distribution. If state updates are relatively inexpensive, then any state update scheme that ensures the correct distribution is sampled can be iterated many times, effectively resulting in an independence sampling scheme. Interestingly, this means that metropolized independence sampling becomes worse when repeated several times, as it eventually turns into simple independence sampling.

Although the acceleration of independence sampling over neighbor exchange is more dramatic with longer intervals between state updates, more frequent state updates appear to always be better than less frequent updates. For example, neighbor exchange with more frequent updates achieves shorter correlation times than either independence sampling scheme for infrequent updates, confirming earlier findings that frequent exchange attempts in neighbor exchange reduce correlation times.^{16,17} We also note that for this particular system, metropolized independence sampling is slightly but clearly better than independence sampling in all sampling measures, providing a strong incentive to use metropolized independence sampling when it is inexpensive to do so.

2. Larger Lennard-Jones spheres

As united atom methane is much smaller than typical biomolecules of interest, we additionally examined an alchemical expanded ensemble simulation of a much larger Lennard-Jones sphere. In this case, the sphere has $\sigma_{ii} = 1.09$ nm and $\epsilon_{ii} = 1.230096$ kJ/mol, again solvated in a cubic simulation cell containing 893 TIP3P (Ref. 92) waters. These parameters result in a sphere-water $\sigma_{ij} = 0.561$ nm, and therefore a particle 5.0 times as large in volume as the UA methane sphere. Because of the larger volume of the solute, $K = 18$ alchemically modified thermodynamic states were required, with $\lambda = [0, 0.15, 0.3, 0.45, 0.55, 0.6, 0.64, 0.66, 0.68, 0.70, 0.72, 0.75, 0.78, 0.81, 0.84, 0.87, 0.90, 1.0]$. All other simulation parameters (other than simulation length) were the same as the UA methane simulations. Log-weights g_k for the equilibrium expanded ensemble simulation were determined in the same manner as for united atom methane,

except that a 15 ns simulation was used to generate the data for MBAR, yielding weights $g_k = \{0.0, 1.74, 2.96, 3.39, 2.84, 2.01, 0.73, -0.34, -1.75, -3.35, -4.96, -7.19, -9.11, -10.70, -11.98, -12.98, -13.72, -14.65\}$. Frequent state updates were performed every 0.1 ps, but infrequent state moves were performed every 1 ps rather than 5 ps to obtain better statistics for the larger molecule. The production expanded ensemble simulations were run for a total of 100 ns for frequent exchange, and 250 ns for infrequent exchange. The same three types of moves in state space were attempted as with UA methane.

Statistics of the observed replica trajectories are shown in Table II. All three convergence rate diagnostics of the state index trajectories described in Sec. III C were computed. In general, the relaxation time estimated from the second eigenvalue of the empirical state transition matrix (Sec. V B) again provides a lower bound for the other computed relaxation times. For the infrequent sampling interval τ_2 is of the same order of magnitude (2 to 5 times less) than the other sampling measures. Again, for both the frequent (0.1 ps) and infrequent (1 ps) state update intervals, independence sampling and metropolized independence sampling yield a clear speedup over neighbor exchange. The improvement in sampling efficiency appears to be valid for both small and large particles.

B. Parallel tempering simulations of terminally blocked alanine peptide in implicit solvent

We next consider a parallel tempering simulation, a form of replica exchange in which the thermodynamic states differ only in inverse temperature β_k . A system containing terminally blocked alanine (sequence Ace-Ala-Nme) was constructed using the LEaP program¹⁰³ from the AMBER-TOOLS 1.2 package with bugfixes 1–4 applied. The AMBER parm96 forcefield was used¹⁰⁴ along with the Onufriev-Bashford-Case generalized Born-surface area (OBC GBSA) implicit solvent model (corresponding to model I of Ref. 105 equivalent to igb=2 in AMBER's sander program and using the mbondi2 radii selected within LEaP).

TABLE III. Efficiency measures for parallel tempering simulation of alanine dipeptide in implicit solvent. Mixing times listed are: τ_2 , estimated from second eigenvalue of empirical state transition matrix; τ_{ac} , estimated from autocorrelation time of alchemical state index; τ_{end} , estimated average end-to-end transit time for alchemical state index. Autocorrelation times of trigonometric functions of ϕ and ψ torsion angles are listed as $\tau_{\cos \phi}$, $\tau_{\sin \phi}$, $\tau_{\cos \psi}$, $\tau_{\sin \psi}$. The statistical error is given as one standard error of the mean.

	State mixing times (ps)			Structural correlation times (ps)			
	τ_2	τ_{ac}	τ_{end}	$\tau_{\cos \phi}$	$\tau_{\sin \phi}$	$\tau_{\cos \psi}$	$\tau_{\sin \psi}$
Neighbor exchange	91.8 ± 0.6	80 ± 2	360 ± 30	25 ± 2	110 ± 9	25 ± 2	66 ± 6
Independence sampling	2.62 ± 0.01	1.60 ± 0.06	28.7 ± 0.7	12.4 ± 0.5	8.7 ± 0.4	11.8 ± 0.6	9.1 ± 0.5

A custom Python code making use of the graphics processing unit (GPU)-accelerated OPENMM package^{106–108} and the PYOPENMM Python wrapper¹⁰⁹ was used to conduct the simulations. All forcefield terms are identical to those used in AMBER except for the surface area term, which was left as default in the OPENMM implementation through a GBSAOBC Force term. Parallel tempering simulations of 2000 iterations were run, with dynamics propagated by 500 steps each iteration using a 2 fs time step and the leapfrog Verlet integrator.^{110,111} Velocities were reassigned from the Maxwell-Boltzmann distribution each iteration. The Python scripts for simulation and data analysis used here are available online at <http://simtk.org/home/gibbs>.

For the replica-mixing phase, the simulation employed either neighbor exchange (Sec. III A 1) or independence sampling (Sec. III B 2), with K^3 attempted swaps of replica pairs selected at random. The efficiency was measured in several ways, shown in Table III. In addition to the standard mixing metrics described in Sec. III C, an estimate of the configurational relaxation times was also made; due to the circular nature of the torsional coordinates ϕ and ψ known to be slow degrees of freedom for this system,¹¹² we instead computed the autocorrelation times for $\sin \phi$, $\cos \phi$, $\sin \psi$, and $\cos \psi$. All replicas were treated as equivalent, and their raw statistics (e.g., autocorrelation functions before normalization) were averaged to produce these estimates. Statistical error was again estimated by blocking.

As expected, the various metrics indicate that the parallel tempering replicas mix in state space much more rapidly with independence sampling than when only neighbor exchanges are attempted. The amount by which mixing is accelerated depends on the metric used to quantify this, but it is roughly one to two orders of magnitude. The structural relaxation times also reflect a speedup, though much more modest than the acceleration in state space sampling—roughly a factor of two to ten, depending on the metric examined.

C. Two-dimensional replica exchange umbrella sampling of terminally-blocked alanine peptide in implicit solvent

Finally, we consider a two-dimensional replica exchange umbrella sampling situation, commonly used to compute potentials of mean force along two coordinates of interest. We again consider the alanine dipeptide in implicit solvent, and employ umbrella potentials to restrain the ϕ and ψ torsions near reference values (ϕ_k^0, ψ_k^0) for $K = 101$ replicas spaced evenly on a 10×10 toroidal grid, with the inclusion of one

replica without any bias potential for ease of post-simulation analysis.

Because harmonic constraints are not periodic, we employ periodic bias potential based on the von Mises circular normal distribution,

$$U'_k(x) \equiv -\kappa [\cos(\phi - \phi_k^0) + \cos(\psi - \psi_k^0)], \quad (31)$$

where κ has units of energy. For sufficiently large values of κ , this will localize the torsion angles in an approximately Gaussian distribution near the reference torsions (ϕ_k^0, ψ_k^0) with a standard deviation of $\sigma \equiv (\beta\kappa)^{1/2}$.

Here, we employ a κ of $(2\pi/30)^{-2}\beta^{-1}$ so that neighboring bias potentials are separated by 3σ . This was sufficient to localize sampling near the reference torsion values for most sterically unhindered regions. The simulation was run at 300 K, using a 2 fs time step with 5 ps between replica exchange attempts. A total of 2000 iterations were conducted, with each iteration consisting of mixing the replica state assignments via a state update phase, a new velocity assignment from the Maxwell-Boltzmann distribution, propagation of dynamics, and writing out the resulting configuration data. The first 100 iterations were discarded as equilibration.

The same mixing schemes examined in the parallel tempering simulation were evaluated here, and the results of the efficiency metrics are summarized in Table IV. Note that the end-to-end time does not have a clear interpretation in terms of the average transit time between a maximum and minimum thermodynamic parameter here—it simply reflects the average time between exchanges between a particular localized umbrella and the unbiased state.

As in the parallel tempering case, we find that both mixing times in state space and the structural correlation times are reduced by use of Gibbs sampling, albeit to a lesser degree than in the parallel tempering case. Here, state relaxation times are reduced by a factor of two to six, depending on the metric considered, while structural correlation times are reduced by a factor of four or five.

VI. DISCUSSION

We have presented the framework of Gibbs sampling on the joint set of state and coordinate variables to better understand different expanded ensemble and replica exchange schemes, and demonstrated how this framework can identify simple ways to enhance the efficiency of expanded ensemble and replica exchange simulations by modifying the

TABLE IV. Efficiency measures for two-dimensional replica exchange umbrella sampling for the alanine dipeptide in implicit solvent. Mixing times in state space listed are: τ_2 , estimated from second eigenvalue of empirical state transition matrix; τ_{ac} , estimated from autocorrelation time of alchemical state index; τ_{end} , estimated average end-to-end transit time for alchemical state index. Autocorrelation times of trigonometric functions of ϕ and ψ torsion angles are listed as $\tau_{\cos \phi}$, $\tau_{\sin \phi}$, $\tau_{\cos \psi}$, $\tau_{\sin \psi}$. The statistical error is given as one standard error of the mean.

	State mixing times (ps)			Structural correlation times (ps)			
	τ_2	τ_{ac}	τ_{end}	$\tau_{\cos \phi}$	$\tau_{\sin \phi}$	$\tau_{\cos \psi}$	$\tau_{\sin \psi}$
Neighbor exchange	82 ± 4	31.0 ± 0.9	350 ± 30	47 ± 2	57 ± 2	26.4 ± 0.8	27.1 ± 0.9
Independence sampling	24.2 ± 0.3	5.45 ± 0.06	175 ± 6	8.92 ± 0.09	9.9 ± 0.1	5.63 ± 0.04	6.09 ± 0.04

thermodynamic state update phase of the algorithms. While the actual efficiency improvement will depend on the system and simulation details, we believe there is likely little, if any, drawback to using these improvements in a broad range of situations.

For simulated and parallel tempering simulations, in which only the temperature is varied among the thermodynamic states, the recommended scheme (independence sampling updates, Secs. III A 2 and III B 2) is simple and inexpensive enough to be easily adopted by simulated and parallel tempering codes. Because calculation of exchange probability requires no additional energy evaluations, it is effectively free. Other expanded ensemble or replica exchange simulations where the potential does not vary between states (such as exchange among temperatures and pressures⁴⁸ or pH values¹¹³) are also effectively free, as no additional energy evaluations are required in these cases either. As long as state space evaluations are cheap compared to configuration updates, independence sampling will mix more rapidly than neighbor updates, though this advantage will be reduced as the interval spent between configuration updates by molecular dynamics or Monte Carlo simulation or the total time performing these coordinate updates becomes very small.

In some cases, exchange of information between processors during replica exchange in tightly coupled parallel codes may incur some cost, mainly in the form of latency. In many cases, however, the decrease in mixing times could more than offset any loss in parallel efficiency. If the recommended independence sampling schemes would consume a substantial fraction of the iteration time, or where the parallel implementation of state updates is already complex, it may still be relatively inexpensive to perform simply the same state update scheme *several times*, achieving enhanced mixing with little extra coding or computational overhead. Alternatively, the Gibbs sampling formalism could be used to design some other scheme that performs frequent state space sampling only on replicas that are local in the topology of the code.

For simulated scaling¹⁴ or Hamiltonian exchange simulations,^{7–10} independence sampling updates of state permutation vector S requires evaluation of the reduced potential $u_k(x)$ at all K states for the current configuration (in simulated scaling) or all replica configurations x_k (for Hamiltonian exchange), which requires more energy evaluations than the neighbor exchange scheme. However, if the intent is to make use of the MBAR estimator,⁵⁶ which produces optimal estimates of free energy differences and expectations, all of these energies are required for analysis anyway, and so the com-

putational impact on simulation time is negligible. It is more computationally efficient to evaluate these additional reduced potentials *during* the simulation, instead of post-processing simulation data, which is especially true if the additional reduced potential evaluations are done in parallel. Alternatively, if a simulated scaling simulation is run and one does not wish to use MBAR, restricted range state updates (Sec. III A 4) offer improved mixing behavior with minimal additional number of energy evaluations.

We have found that examining the exchange statistics, the empirical state transition matrix and its dominant eigenvalues, is extremely useful in diagnosing equilibration and convergence, as well as poor choices of thermodynamic states. It is often very easy to see, from the diagonally dominant structure of this matrix, where regions of poor state overlap occur. Poor overlap among sets of thermodynamic states observed early in simulations from the empirical state transition matrix are likely to also frustrate post-simulation analysis with techniques such as MBAR and histogram reweighting methods,^{56,89,114} making such metrics useful diagnostic tools.

For more complex state topologies in expanded ensemble or replica exchange simulations, where, for example, several different pressures or temperatures are included simultaneously, there may not exist a simple grid of values, or it may not be easy to identify which states are the most efficient neighbors. Using independence sampling eliminates the need to plan efficient exchange schemes among neighbors, or even to determine which states are neighbors. This may encourage the addition of states that aid in reducing the correlation time of the overall Markov chain solely by speeding decorrelation of conformational degrees of freedom, since they will automatically couple to states with reasonable phase space overlap.

It is important to stress, however, that expanded ensemble and replica exchange simulations are not a cure-all for all systems with poor sampling. In the presence of a first-order or pseudo-first-order phase transition, phase space mixing may still take an exponentially long time even when simulated or parallel tempering algorithms are used.⁴² Optimization of the state exchange scheme, as described here, can only help so much; further efficiency gains would require design of intermediate states that abolish the first-order phase transition. Schemes for optimal state selection are an area of active research.^{26–32}

Finally, we observe that the independence sampling scheme for a simulated tempering simulation or any

simulation where the contribution to the reduced potential is a thermodynamic parameter λ multiplying a conjugate configuration-dependent variable $h(x)$ naturally generalizes to a continuous limit. As the number K of thermodynamic states λ_k is increased between some fixed lower and upper limits, this process eventually results in the thermodynamic state index k effectively becoming a continuous variable λ .² Such a *continuous tempering* simulation would sample from the joint distribution $\pi(x, \lambda) \propto \exp[-\lambda h(x) + g(\lambda)]$, with the continuous log weighting function $g(\lambda)$ replacing the discrete g_k in simulated tempering simulations.

The Gibbs sampler and variations on it remain exciting areas for future exploration, and we hope that our conditional state space sampling formulation will make it much easier for other researchers to envision, develop, and implement new schemes for sampling from multiple thermodynamics states. We also hope it encourages exploration of further connections between the two deeply interrelated fields of statistical mechanics and statistical inference.

ACKNOWLEDGMENTS

The authors thank Sergio Bacallado (Stanford University), Jed Pitera (IBM Almaden), Adrian Roitberg (University of Florida), Scott Schmidler (Duke University), and William Swope (IBM Almaden) for insightful discussions on this topic, and Imran Haque (Stanford University), David Minh (Argonne National Labs), Victor Martin-Mayor (Universidad Complutense de Madrid), and Anna Schneider (University of California, Berkeley) for a critical reading of the manuscript. The authors especially thank David M. Rogers (Sandia National Labs) who for spotting a key error in an early version of the restricted range sampling method, and the anonymous referees for constructive comments. Additionally, the authors are grateful to OPENMM developers Peter Eastman, Mark Friedrichs, Randy Radmer, and Christopher Bruns, and project leader Vijay Pande (Stanford University and SimBios) for their generous help with the OPENMM GPU-accelerated computing platform and associated PYOPENMM Python wrappers. J.D.C. acknowledges support from a QB3-Berkeley Distinguished Postdoctoral Fellowship.

¹A. Mitsutake, Y. Sugita, and Y. Okamoto, *Biopolymers* **60**, 96 (2001).

²Y. Iba, *Int. J. Mod. Phys. C* **12**, 623 (2001).

³C. J. Geyer, in *Computing Science and Statistics: The 23rd Symposium on the Interface* (Interface Foundation, Fairfax, 1991), pp. 156–163.

⁴K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).

⁵U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).

⁶Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).

⁷Y. Sugita, A. Kitao, and Y. Okamoto, *J. Chem. Phys.* **113**, 6042 (2000).

⁸H. Fukunishi, O. Watanabe, and S. Takada, *J. Chem. Phys.* **116**, 9058 (2002).

⁹S. Jang, S. Shin, and Y. Pak, *Phys. Rev. Lett.* **91**, 058305 (2003).

¹⁰W. Kwak and U. H. E. Hansmann, *Phys. Rev. Lett.* **95**, 138102 (2005).

¹¹A. P. Lyubartsev, A. A. Martynov, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).

¹²E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).

¹³C. J. Geyer and E. A. Thompson, *J. Am. Stat. Assoc.* **90**, 909 (1995).

¹⁴H. Li, M. Fajer, and W. Yang, *J. Chem. Phys.* **126**, 024106 (2007).

¹⁵S. Park, *Phys. Rev. E* **77**, 016709 (2008).

¹⁶D. Sindhikara, Y. Meng, and A. E. Roitberg, *J. Chem. Phys.* **128**, 024103 (2008).

¹⁷D. Sindhikara, D. J. Emerson, and A. E. Roitberg, *J. Chem. Theory Comput.* **6**, 2804 (2010).

¹⁸W. Nadler and U. H. E. Hansmann, *Phys. Rev. E* **76**, 057102 (2007).

¹⁹Y. M. Rhee and V. S. Pande, *Biophys. J.* **84**, 775 (2003).

²⁰D. M. Zuckerman and E. Lyman, *J. Chem. Theory Comput.* **2**, 1200 (2006).

²¹W. Zheng, M. Andrec, E. Gallicchio, and R. M. Levy, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15340 (2007).

²²H. Nymeyer, *J. Chem. Theory Comput.* **4**, 626 (2008).

²³R. Denschlag, M. Lingenheil, and P. Tavan, *Chem. Phys. Lett.* **458**, 244 (2008).

²⁴E. Rosta and G. Hummer, *J. Chem. Phys.* **131**, 165102 (2009).

²⁵E. Rosta and G. Hummer, *J. Chem. Phys.* **132**, 034102 (2010).

²⁶D. A. Kofke, *J. Chem. Phys.* **117**, 6911 (2002).

²⁷H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, *J. Stat. Mech.: Theory Exp.* **2006**, 03018 (2006).

²⁸S. Trebst, M. Troyer, and U. H. E. Hansmann, *J. Chem. Phys.* **124**, 174903 (2006).

²⁹W. Nadler and U. H. E. Hansmann, *Phys. Rev. E* **75**, 026109 (2007).

³⁰D. Gront and A. Kolinski, *J. Phys.: Condens. Matter* **19**, 036225 (2007).

³¹S. Park and V. S. Pande, *Phys. Rev. E* **76**, 016703 (2007).

³²D. K. Shenfeld, H. Xu, M. P. Eastwood, R. O. Dror, and D. E. Shaw, “■,” *Phys. Rev. E* (to be published).

³³T. Neuhaus, M. P. Magiera, and U. H. E. Hansmann, *Phys. Rev. E* **76**, 045701 (2007).

³⁴J. Kim, T. Keyes, and J. E. Straub, *J. Chem. Phys.* **132**, 224107 (2010).

³⁵M. K. Fenwick and F. A. Escobedo, *J. Chem. Phys.* **119**, 11998 (2003).

³⁶A. Mitsutake and Y. Okamoto, *J. Chem. Phys.* **121**, 2491 (2004).

³⁷A. Mitsutake, Y. Sugita, and Y. Okamoto, *J. Chem. Phys.* **118**, 6664 (2003).

³⁸A. Okur, D. R. Roe, G. Cui, V. Hornak, and C. Simmerling, *J. Chem. Theory Comput.* **3**, 557 (2007).

³⁹E. Gallicchio, R. M. Levy, and M. Parashar, *J. Comput. Chem.* **29**, 288 (2008).

⁴⁰U. H. E. Hansmann, *Physica A* **389**, 1400 (2010).

⁴¹N. Madras and D. Randall, *Ann. Appl. Probab.* **12**, 581 (2002).

⁴²N. Bhatnagar and D. Randall, in *SODA 2004: Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004), pp. 478–487.

⁴³D. B. Woodard, *Ann. Appl. Probab.* **19**, 617 (2009).

⁴⁴D. B. Woodard, S. C. Schmidler, and M. Huber, *Electron. J. Probab.* **14**, 780 (2009).

⁴⁵J. Machta, *Phys. Rev. E* **80**, 056706 (2009).

⁴⁶P. Brenner, C. R. Sweet, D. VonHandorf, and Jesús A. Izaguirre, *J. Chem. Phys.* **126**, 074103 (2007).

⁴⁷M. Lingenheil, R. Denschlag, G. Mathias, and P. Tavan, *Chem. Phys. Lett.* **478**, 80 (2009).

⁴⁸D. Paschek and Angel E. García, *Phys. Rev. Lett.* **93**, 238105 (2004).

⁴⁹W. Jiang and B. Roux, *J. Chem. Theory Comput.* **6**, 2559 (2010).

⁵⁰J. Rudnick and G. Gaspari, *Elements of the Random Walk: An introduction for Advanced Students and Researchers*, 1st ed. (Cambridge University Press, Cambridge, England, 2010).

⁵¹S. Geman and D. Geman, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721 (1984).

⁵²J. S. Liu, *Monte Carlo strategies in Scientific Computing*, 2nd ed. (Springer, New York, 2002).

⁵³A. Panagiotopoulos, *Mol. Phys.* **61**, 813 (1987).

⁵⁴A. Z. Panagiotopoulos, N. Quirke, M. Stapleton, and D. J. Tildesley, *Mol. Phys.* **63**, 527 (1988).

⁵⁵It is ironic that although J. Willard Gibbs invented a large number of important thermodynamic concepts and methods, neither Gibbs sampling nor the Gibbs ensemble are among them.

⁵⁶M. R. Shirts and J. D. Chodera, *J. Chem. Phys.* **129**, 124105 (2008).

⁵⁷E. Mezei, *J. Comp. Phys.* **68**, 237 (1987).

⁵⁸C. Tsallis, *J. Stat. Phys.* **52**, 479 (1988).

⁵⁹N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

⁶⁰W. K. Hastings, *Biometrika* **57**, 97 (1970).

⁶¹S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Phys. Lett. B* **195**, 216 (1987).

⁶²H. Müller-Krumbhaar and K. Binder, *J. Stat. Phys.* **8**, 1 (1973).

- Q6⁶³W. Janke, in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, edited by J. Grotendorst, D. Marx, and A. Murmatsu (John von Neumann Institute for Computing, Jülich, Germany, 2002), Vol. 10, pp. 423–445.
- ⁶⁴An even more general class of samplers, the so-called *hit-and-run* schemes, explore this notion further. See also Ref. 115.
- ⁶⁵V. I. Manousiouthakis and M. W. Deem, *J. Chem. Phys.* **110**, 2753 (1999).
- ⁶⁶F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- ⁶⁷S. Park, D. L. Ensign, and V. S. Pande, *Phys. Rev. E* **74**, 066703 (2006).
- ⁶⁸R. Chelli, *J. Chem. Theory Comput.* **6**, 1935 (2010).
- ⁶⁹Caution must be exercised when using molecular dynamics to sample from $\pi(x|k)$, as the phase space distribution actually sampled will deviate from the distribution $\pi(x|k) \propto e^{-u_k(x)}$ in a timestep-dependent manner unless the integration scheme is wrapped in a Monte Carlo acceptance step, as in hybrid Monte Carlo (see also Ref. 61), metropolized Langevin, or generalized hybrid Monte Carlo (see also Refs. 116 and 117) integration schemes. If this deviation is significant, it could result in significant errors in sampling the overall (x, k) Markov chain that could exceed the statistical error in computed properties.
- ⁷⁰A. E. Roitberg, A. Okur, and C. Simmerling, *J. Phys. Chem. B* **111**, 2415 (2007).
- ⁷¹The denominator, in fact, is the *matrix permanent* of a $K \times K$ matrix $A \equiv (a_{ij})$ with $a_{ij} = e^{-u_i(x_j)}$, computation of which is known to be sharp-P-complete, and hence too inefficient for our purposes except for very small K .
- ⁷²M. A. Miller, L. M. Amon, and W. P. Reinhardt, *Chem. Phys. Lett.* **331**, 278 (2000).
- ⁷³G. Bussi, A. Laio, and M. Parrinello, *Phys. Rev. Lett.* **96**, 090601 (2006).
- ⁷⁴U. H. E. Hansmann and Y. Okamoto, *Phys. Rev. E* **54**, 5863 (1996).
- ⁷⁵T. Rodinger, P. L. Howell, and R. Pomès, *J. Chem. Phys.* **129**, 155102 (2008).
- ⁷⁶J. S. Liu, *Biometrika* **83**, 681 (1996).
- ⁷⁷It is again common to consider the set of K states to lie on a torus, such that exchanges of $(K, 1)$ are also attempted.
- ⁷⁸D. Aldous and P. Diaconis, *Am. Math. Monthly* **93**, 333 (1986).
- ⁷⁹C. H. Bennett, *J. Comput. Phys.* **22**, 245 (1976).
- ⁸⁰J. W. Pitera and W. Swope, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7587 (2003).
- ⁸¹L. A. Fernandez, V. Martin-Mayor, S. Perez-Gaviro, A. Tarancon, and A. P. Young, *Phys. Rev. B* **80**, 024422 (2009).
- ⁸²S. T. Garren and R. L. Smith, *Bernoulli* **6**, 215 (2000).
- ⁸³X. Zhang, D. Bhatt, and D. M. Zuckerman, *J. Chem. Theory Comput.* **6**, 3048 (2010).
- ⁸⁴P. Diaconis and P. Hanlon, *Contemp. Math.* **138**, 99 (1992).
- ⁸⁵C. Schütte, A. Fischer, W. Huisinga, and P. Deuffhard, *J. Comput. Phys.* **151**, 146 (1999).
- ⁸⁶C. Schütte and W. Huisinga, in *Handbook of Numerical Analysis - Special Volume on Computational Chemistry*, edited by P. G. Ciaret and J.-L. Lions (Elsevier, New York, 2002), Vol. 10.
- ⁸⁷X. Huang, G. R. Bowman, S. Bacallado, and V. S. Pande, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19765 (2009).
- ⁸⁸P. Deuffhard and M. Weber, *Linear Algebr. Appl.* **398**, 161 (2005).
- ⁸⁹J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, *J. Chem. Theory Comput.* **3**, 26 (2007).
- ⁹⁰F. A. Escobedo and F. J. Martinez-Veracoechea, *J. Chem. Phys.* **129**, 154107 (2009).
- ⁹¹R. Denschlag, M. Lingenheil, and P. Tavan, *Chem. Phys. Lett.* **473**, 193 (2009).
- ⁹²W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- ⁹³B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).
- ⁹⁴Please contact the corresponding author for instructions on accessing this version. Alternatively, this functionality is also being prepared for inclusion in GROMACS 4.6.
- ⁹⁵W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, *J. Chem. Phys.* **76**, 637 (1982).
- ⁹⁶G. J. Martyna, M. L. Klein, and M. Tuckerman, *J. Chem. Phys.* **97**, 2635 (1992).
- ⁹⁷T. Yu, J. Alejandre, R. López-Rendón, G. J. Martyna, and M. E. Tuckerman, *Chem. Phys.* **370**, 294 (2010).
- ⁹⁸M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna, *J. Phys. A* **39**, 5629 (2006).
- ⁹⁹S. Miyamoto and P. A. Kollman, *J. Comput. Chem.* **13**, 952 (1992).
- ¹⁰⁰M. R. Shirts, D. L. Mobley, J. D. Chodera, and V. S. Pande, *J. Phys. Chem. B* **111**, 13052 (2007).
- ¹⁰¹M. R. Shirts and V. S. Pande, *J. Chem. Phys.* **122**, 134508 (2005).
- ¹⁰²M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, *J. Chem. Phys.* **119**, 5740 (2003).
- ¹⁰³W. Zhang, T. Hou, C. Schafmeister, W. S. Ross, and D. A. Case, “AMBERTOOLS 1.2: LEaP” (2008).
- ¹⁰⁴P. A. Kollman, R. Dixon, W. Cornell, T. Vox, C. Chipot, and A. Pohorille, in *Computer Simulation of Biomolecular Systems*, edited by A. Wilkinson, P. Weiner, and W. F. van Gunsteren (Kluwer/Escom, The Netherlands/Virginia, 1997), Vol. 3, pp. 83–96.
- ¹⁰⁵A. Onufriev, D. Bashford, and D. A. Case, *Proteins: Struct., Funct., Genet.* **55**, 383 (2004).
- ¹⁰⁶M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. LeGrand, A. L. Beberg, D. L. Ensign, C. M. Bruns, and V. S. Pande, *J. Comput. Chem.* **30**, 864 (2009).
- ¹⁰⁷P. Eastman and V. S. Pande, *Comput. Sci. Eng.* **12**, 34 (2010).
- ¹⁰⁸P. Eastman and V. S. Pande, *J. Comput. Chem.* **31**, 1268 (2010).
- ¹⁰⁹C. M. Bruns, R. A. Radmer, J. D. Chodera, and V. S. Pande, “PyOpenMM” (2010).
- ¹¹⁰L. Verlet, *Phys. Rev.* **159**, 98 (1967).
- ¹¹¹L. Verlet, *Phys. Rev.* **165**, 201 (1968).
- ¹¹²J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, *Multiscale Model. Simul.* **5**, 1214 (2006).
- ¹¹³Y. Meng and A. E. Roitberg, *J. Chem. Theory Comput.* **6**, 1401 (2010).
- ¹¹⁴S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).
- ¹¹⁵H. C. Andersen and P. Diaconis, *Journal de la Société Française de Statistique* **148**, 5 (2007).
- ¹¹⁶A. M. Horowitz, *Phys. Lett. B* **268**, 247 (1991).
- ¹¹⁷T. Lelièvre, M. Rousset, and G. Stoltz, *Free Energy Computations* (Imperial College, London, 2010).