Optimal updating magnitude in adaptive flat-distribution sampling

Cheng Zhang,¹ Justin A. Drake,¹ Jianpeng Ma,^{2,3} and B. Montgomery Pettitt^{1,a)}

¹*Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, Galveston, Texas 77555-0304, USA*

²*Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA*

³*Department of Bioengineering, Rice University, Houston, Texas 77005, USA*

(Received 7 February 2017; accepted 8 October 2017; published online 3 November 2017)

We present a study on the optimization of the updating magnitude for a class of free energy methods based on flat-distribution sampling, including the Wang-Landau (WL) algorithm and metadynamics. These methods rely on adaptive construction of a bias potential that offsets the potential of mean force by histogram-based updates. The convergence of the bias potential can be improved by decreasing the updating magnitude with an optimal schedule. We show that while the asymptotically optimal schedule for the single-bin updating scheme (commonly used in the WL algorithm) is given by the known inverse-time formula, that for the Gaussian updating scheme (commonly used in metadynamics) is often more complex. We further show that the single-bin updating scheme is optimal for very long simulations, and it can be generalized to a class of bandpass updating schemes that are similarly optimal. These bandpass updating schemes target only a few long-range distribution modes and their optimal schedule is also given by the inverse-time formula. Constructed from orthogonal polynomials, the bandpass updating schemes generalize the WL and Langfeld-Lucini-Rago algorithms as an automatic parameter tuning scheme for umbrella sampling. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5008618>

I. INTRODUCTION

Free energy calculation^{1,2} is a central theme in computational physics and chemistry that can provide insight into an array of phenomena not easily studied with regular simulation methods. Given a system, we wish to compute a distribution, $p(z)$, along a collective variable, z , by sampling via either Monte Carlo (MC)^{1,2} or molecular dynamics (MD)^{1,3} simulations. The distribution is related to the dimensionless free energy, or potential of mean force (PMF), as $-\ln p(z)$. However, reliable estimation of the PMF is often difficult when the collective variable is energetically restricted to local regions of the complicated, multimodal free energy surface. We can overcome, at least partially, these energetic barriers in sampling by introducing a bias potential⁴ that offsets the PMF such that the distribution along z is flat.^{5–9}

Many efficient flat-distribution-sampling (FDS) techniques are based on adaptive construction of the bias potential, e.g., the Wang-Landau (WL) algorithm^{7,8} and metadynamics^{9,10} among others.^{11–18} These techniques discourage future visits to sampled configurations by incrementally elevating the bias potential around the current value of z . A key difference lies in the updating window function, which specifies, apart from an overall updating magnitude, the relative change to the bias potential upon each visit. The window function often takes the form of a discrete δ -function (a boxcar function one histogram bin wide) in the WL algorithm or that of a Gaussian in metadynamics.¹⁴ We will refer to the two updating schemes

as the single-bin and Gaussian updating schemes, respectively. While the former is popular for MC simulations,^{7,8,12} the latter is commonly used for MD simulations as it provides a smoother bias potential suitable for deriving the bias force.^{9,14} In both cases, regular updates to the bias potential disrupts the underlying equilibrium sampling,^{19–21} and the convergence of the bias potential often benefits from a gradually decreasing updating magnitude.

The schedule of reducing the updating magnitude affects the precision and accuracy of the bias potential and hence those of the PMF.^{20–30} While an overly slow reduction is ineffective in restricting the variance of the bias potential, a too rapid reduction often converges the bias potential to a wrong value. For the single-bin case, it is known that the optimal schedule is given by the inverse time^{20–24} or the inverse number of updating steps.

In this study, we first present a method of computing the optimal schedule for an adaptive FDS simulation under a general updating scheme, including the single-bin and Gaussian versions. We will map the optimization problem to a mechanical one, in which the schedule plays the role of the velocity of a free particle with a position-dependent mass, and the final error that of the action. This mapping allows the optimal schedule to be derived from the equation of motion that minimizes the action. We will show that while the optimal schedule in the single-bin case is given by the known inverse-time formula, that for the Gaussian case is more complex.

We will then show that among different updating schemes, the single-bin updating scheme is optimal in error minimization for very long simulations. A similar but partial asymptotic

^{a)}mpettitt@utmb.edu

optimality applies to a class of bandpass updating schemes that target only a few long-range distribution modes. Compared with the Gaussian updating scheme, bandpass updating schemes allow a simple protocol (borrowed from the single-bin updating scheme) to optimally control the updating magnitude.

Finally, we present an application of the bandpass updating schemes by a construction from a sequence of orthogonal polynomials. The bias potential in this case assumes a polynomial form, and we will adjust the coefficients on-the-fly to approximate the target distribution. This method serves as a generalization of the WL and Langfeld-Lucini-Rago (LLR) algorithms,^{15,16} and we show that a similar construction with a quadratic bias potential^{31,32} allows one to sample an array of Gaussian distributions with predefined means and variances.

The article is organized as follows. We present the methods in Sec. II, provide numerical examples in Sec. III, and conclude the article in Sec. IV.

II. METHODS

We present the methods in the following order. In Sec. II A, we briefly review the basics of FDS and introduce our notations. Then, in Sec. II B, we derive the method of computing the optimal schedule in the simple case of the single-bin updating scheme (as in the WL algorithm) by proving the optimality of the known inverse-time formula. We extend the method to the general case of a multiple-bin updating scheme that encompasses the Gaussian updating scheme used in metadynamics in Sec. II C. Then, we compare different updating schemes and show that the single-bin updating scheme and a class of generalized bandpass updating schemes are optimal for convergence in the long time limit in Sec. II D. Finally, we use the bandpass updating schemes to build a class of adaptive FDS algorithms with polynomial bias potentials in Sec. II E.

A. Flat-distribution sampling

Consider the problem of computing the distribution, p_i , along a discrete collective variable, i , for a given system. For a continuous collective variable, z , we can use the equivalent integer i to represent the index of a small interval or a bin (z , $z + \delta z$).

Sampling a multimodal distribution with deep basins is often challenging in regular MC and MD simulations. To promote transitions among the basins, we can apply a bias potential to alter the intrinsic distribution to a smoother one, ρ_i . We will broadly define a flat-distribution-sampling (FDS) simulation as a biased simulation targeting a flat ρ_i (i.e., $\rho_i \equiv 1/n$ for n bins) or a nearly flat one.^{27,33,34} Since the equilibrium distribution under a bias potential, u_i , is given by

$$\pi_i = \frac{p_i e^{-u_i}}{\sum_{j=1}^n p_j e^{-u_j}} \propto p_i e^{-u_i}, \quad (1)$$

we wish to find a bias potential that allows the resulting π_i to coincide with the desired ρ_i . Such a bias potential would satisfy

$$u_i \rightarrow \ln \frac{p_i}{\rho_i} + \text{const.}, \quad (2)$$

and the PMF, $-\ln p_i$, can be readily deduced from u_i .

Many methods⁵⁻⁹ have been developed to find the desired bias potential. In an earlier class of equilibrium FDS methods,^{5,6,35} the bias potential, u_i , is estimated beforehand and fixed during simulation, and a correction is derived from the normalized histogram, H_i , accumulated from the simulation as

$$\hat{u}_i = u_i + \ln \frac{H_i}{\rho_i}. \quad (3)$$

However, this formula often requires a long accumulation period for a precise histogram and thus disallows continuous improvement of the bias potential as the simulation lengthens.

Below we will focus on a more recent class of adaptive FDS methods, including the WL algorithm^{7,8} and metadynamics,⁹ which continuously and incrementally update the bias potential. In the WL algorithm,^{7,8} the bias potential, $u_i(t)$, is updated at each time step t as

$$u_i(t+1) = u_i(t) + \delta_{i,i(t)} \frac{\alpha(t)}{\rho_{i(t)}}, \quad (4)$$

where $i(t)$ is the bin at step t , $\delta_{i,i(t)}$ is the Kronecker delta function, which is 1 when $i = i(t)$ or 0 otherwise, and $\alpha(t)$ is the updating magnitude. We refer to this scheme as the *single-bin* updating scheme, for it applies only to the bias potential at the current bin, $i(t)$. In a more general *multiple-bin* updating scheme, the bias potential at neighboring bins is also updated,

$$u_i(t+1) = u_i(t) + w_{i,i(t)} \frac{\alpha(t)}{\rho_{i(t)}}. \quad (5)$$

A common example is the Gaussian updating scheme used in metadynamics, with w_{ij} being roughly a Gaussian centered on bin $j = i(t)$. Note that if updating occurs only every few sampling steps, then t denotes the number of updating steps.

While adaptive FDS simulations can often sample the desired distribution, the frequent updates cause the bias potential to fluctuate constantly and disrupt the underlying equilibrium sampling. To reduce the sampling error, one often needs to gradually decrease the updating magnitude.^{19-26,30,36}

In the original WL algorithm,^{7,8} the updating magnitude, $\alpha(t)$ (or $\ln f$ therein), is controlled in stages. In each stage, $\alpha(t)$ is kept constant, and the histogram along i is collected and monitored. Once the histogram is sufficiently flat, we reset the histogram and start a new stage with $\alpha(t)$ reduced by a factor of $1/2$.^{7,8} While this scheme works well for early stages, it tends to reduce the magnitude too quickly in late stages, making the asymptotic error saturate.²²

A more effective way of controlling the updating magnitude in late stages is to follow the inverse-time schedule,^{21,22,24,25}

$$\alpha(t) = \frac{1}{t}, \quad (6)$$

where t , referred to as the “time” below, is the number of updating steps from the beginning of the simulation. This is a general result,^{16,37} and it has an intuitive explanation:^{27,38} the bias potential undergoing the single-bin updating scheme can be mapped to a cumulative moving average over all data points collected so far; upon the addition of a new data point, the average absorbs its contribution by the due weight, the inverse of the current sample size, t (cf. Appendix A).

B. Single-bin updating scheme

In this section, we derive the optimal schedule, $\alpha(t)$, for the single-bin updating scheme, Eq. (4), as a preparation for the more general multiple-bin case. We will first express the error of the bias potential, $u_i(t)$, as a functional of $\alpha(t)$, and then minimize it by variation.

1. Error

We define the net error in the bias potential by deducting two contributions. First, we deduct the target value [Eq. (2)] by introducing a shifted bias potential

$$v_i(t) \equiv u_i(t) - \ln \frac{p_i}{\rho_i}, \quad (7)$$

which should tend to a constant of i upon convergence.

Next, we observe that the bias potential following Eq. (4) generally grows over time. Since a uniform increase of $u_i(t)$, or, equivalently, that of $v_i(t)$, does not affect the resulting distribution, $\pi_i(t)$, the real deviation of the bias potential depends only on the difference between $v_i(t)$ and a baseline value, $v_{av}(t)$, for the average growth.²⁹ In the second deduction, we define the bin average as $v_{av}(t) \equiv \sum_{i=1}^n \rho_i v_i(t)$ and the deviation of the bias potential as

$$v_{*i}(t) \equiv v_i(t) - v_{av}(t). \quad (8)$$

The total error is then given by

$$E(t) = \sum_{i=1}^n \rho_i \langle v_{*i}^2(t) \rangle. \quad (9)$$

Below we will use the notations “ $_{av}$ ” and “ $_{*i}$ ” for other variables.

2. Differential equation

To proceed, we approximate the finite difference equation, Eq. (4), by a differential equation³⁹

$$\dot{u}_i(t) \equiv \frac{du_i(t)}{dt} \approx \alpha(t) \frac{h_i(t)}{\rho_i} \equiv \alpha(t) f_i(t), \quad (10)$$

with $h_i(t) = \delta_{i,i(t)}$ being the instantaneous histogram, which is equal to 1 for the bin $i(t)$ at time t or zero otherwise, and we have defined $f_i(t) \equiv h_i(t)/\rho_i$. The shifted bias potential, $v_i(t)$, follows the same differential equation since the shift is a constant in time: $\dot{v}_i(t) \approx \alpha(t) f_i(t)$. Deducting the bin average [cf. Eq. (8)], we get

$$\dot{v}_{*i}(t) \approx \alpha(t) f_{*i}(t). \quad (11)$$

We can split $f_{*i}(t)$ into a deterministic averaged part, $\langle f_{*i}(t) \rangle$, and a random fluctuating part, $\xi_{*i}(t)$,

$$f_{*i}(t) = \langle f_{*i}(t) \rangle + \xi_{*i}(t).$$

The deterministic part can be related to an average in an ensemble consisting of many copies of similar simulations sharing the same schedule $\alpha(t)$ and the same bias potential at time t . The initial states and the stochastic forces during the process may differ, however. For a sufficiently small $\alpha(t)$, the bias potential remains roughly the same for a short period, and we may assume a quasi-equilibrium sampling process [cf. Eq. (1)] such that

$$\langle f_i(t) \rangle \approx \frac{\pi_i(t)}{\rho_i} = \frac{e^{-v_{*i}(t)}}{\sum_{j=1}^n \rho_j e^{-v_{*j}(t)}} \approx 1 - v_{*i}(t),$$

where we have assumed the smallness of $v_{*i}(t)$ in the linear expansion and used Eq. (8) in the last step. Deducting the bin average and noticing that $f_{av}(t) = \sum_{i=1}^n \rho_i [h_i(t)/\rho_i] = \sum_{i=1}^n \delta_{i,i(t)} = 1$, we get $\langle f_{*i}(t) \rangle = \langle f_i(t) \rangle - 1 \approx -v_{*i}(t)$ and

$$f_{*i}(t) \approx -v_{*i}(t) + \xi_{*i}(t). \quad (12)$$

We will approximate the histogram fluctuation part, $\xi_{*i}(t)$, as white noise such that

$$\langle \xi_{*i}(t) \xi_{*i}(t') \rangle = G_i \delta(t - t'), \quad (13)$$

for some constant G_i , and $\delta(t)$ is Dirac's delta function. We expect the approximation to hold for simulations much longer than the autocorrelation time.

From Eqs. (11) and (12), we get

$$\dot{v}_{*i}(t) \approx -\alpha(t) [v_{*i}(t) - \xi_{*i}(t)],$$

and the formal solution is

$$v_{*i}(T) = v_{*i}(0) e^{-q(T)} + \int_0^T \theta(q(t)) \xi_{*i}(t) dt, \quad (14)$$

where $q(t) \equiv \int_0^t \alpha(t') dt'$ and

$$\theta(q') \equiv e^{q' - q(T)}. \quad (15)$$

Thus, without the random part, $\xi_{*i}(t)$, the deviation of the bias potential, $v_{*i}(t)$, would tend to 0 at long times, reaching convergence, if $\alpha(t)$ decays sufficiently slowly such that $q(T) \rightarrow \infty$ as $T \rightarrow \infty$.

3. Optimal schedule

From Eqs. (9) and (14), we can write the total error at the end of period T as a sum of a residual error, $E_R(T)$ (from the decay of the initial error), and an asymptotic error, $E_A(T)$,

$$E(T) = E_R(T) + E_A(T), \quad (16)$$

where

$$E_R(T) = E(0) e^{-2q(T)}, \quad (17)$$

$$\begin{aligned} E_A(T) &= \sum_{i=1}^n \rho_i \int_0^T \int_0^T \langle \xi_{*i}(t) \xi_{*i}(t') \rangle \theta(q(t)) \theta(q(t')) dt dt' \\ &= \Gamma \int_0^T \theta^2(q(t)) dt. \end{aligned} \quad (18)$$

In Eq. (18), we have used Eq. (13) and introduced

$$\Gamma \equiv \sum_{i=1}^n \rho_i G_i \approx \sum_{i=1}^n \rho_i \int_{-\infty}^{\infty} \langle \xi_{*i}(t) \xi_{*i}(0) \rangle dt.$$

The error depends implicitly on the curve $\alpha(t)$, or, equivalently, the curve $q(t)$. We wish to find $q(t)$ that minimizes this error. It is convenient to vary the curve, $q(t)$ ($0 < t < T$), under a fixed value of $q(T)$, since it allows us to focus on the asymptotic error, $E_A(T)$. The contribution of the residual error will be considered later by adjusting the value of $q(T)$. Note that the asymptotic error has the form of the action of a free particle at $\theta(q(t))$, with the endpoints fixed at $\theta(q(0)) = e^{-q(T)}$ and $\theta(q(T)) = 1$. Thus, the velocity should be a constant

$$\frac{d}{dt} \theta(q(t)) = c. \quad (19)$$

Then, by using Eq. (15), we have

$$e^{q(t)-q(T)} = c(t+t_0) = \frac{t+t_0}{T+t_0}, \quad (20)$$

where t_0 is some constant, and the previous constant, c , has been determined from the $t = T$ case in the second step. Taking the logarithm and differentiating both sides with respect to t , we get the inverse-time schedule

$$\alpha(t) = \frac{1}{t+t_0}, \quad (21)$$

and Eq. (6) is the special case of $t_0 = 0$.

We now determine the optimal value of $q(T)$, which was previously held fixed. This is equivalent to choosing a value of t_0 because $q(T) = \ln(1 + \frac{T}{t_0})$ from the $t = 0$ case of Eq. (20). Using Eq. (21) in Eqs. (17) and (18) yields

$$E_R(T) = \frac{E(0)t_0^2}{(T+t_0)^2}, \quad (22)$$

$$E_A(T) = \frac{\Gamma T}{(T+t_0)^2}, \quad (23)$$

and the total has a minimum of

$$E(T) = \frac{\Gamma}{T+t_0}, \quad (24)$$

reachable at $t_0 = \Gamma/E(0)$.

C. Multiple-bin updating scheme

We now turn to the general multiple-bin updating scheme. By definition, Eq. (5), a visit to bin $j = i(t)$ in this case results in updates of the bias potential not only at bin j but also at a few neighboring bins i 's. Similar to Eq. (10), we can approximate the updating scheme as a differential equation,

$$\dot{u}_i(t) \approx \alpha(t) \sum_{j=1}^n w_{ij} f_j(t) = \alpha(t) \sum_{j=1}^n w_{ij} \frac{h_j(t)}{\rho_j}. \quad (25)$$

The optimization procedure is similar to the single-bin case. The key simplification comes from the projection of the bias potential to the eigenmodes of the updating matrix, $\mathbf{w}$, formed by the relative magnitudes, w_{ij} 's.

1. Updating matrix and spectral decomposition

We will first review the spectral decomposition of the updating matrix, $\mathbf{w}$, which helps establish the convergence criterion and the decomposition of error into independent eigenmodes. The matrix, $\mathbf{w}$, is subject to several conditions. First, we have a fixed-point condition.^{29,40} To sample the desired distribution, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_n)$, we require the growth rate of the bias potential, $\dot{u}_i(t)$ in Eq. (25), to be a constant of i if $h_j(t)$ were the same as ρ_j , i.e., $\sum_{j=1}^n w_{ij}$ should be a constant to allow the bias potential to grow uniformly in the asymptotic regime. If the growth rate is positive,⁴¹ we can, by a proper overall scaling of $\mathbf{w}$, write this condition as

$$\sum_{j=1}^n w_{ij} = 1. \quad (26)$$

In other words, $(1, \dots, 1)^T$ is a right eigenvector of $\mathbf{w}$ with eigenvalue 1. Thus, the transpose $\mathbf{w}^T$ resembles a transition matrix, except that some elements can be negative for cases

considered here. Below we will borrow techniques used in studying transition matrices,⁴² but we limit ourselves to properties that are unaffected by the admission of negative elements into $\mathbf{w}$.

Second, we limit ourselves to matrices that satisfy the detailed balance condition for $\boldsymbol{\rho}$,

$$\rho_i w_{ij} = \rho_j w_{ji}.$$

This requires the scaled updating matrix, $\sqrt{\rho_i/\rho_j} w_{ij}$, to be symmetric, and the symmetry allows the diagonalization of $\mathbf{w}$ with a set of eigenvectors, ϕ_{ki} , as

$$\sum_{i=1}^n \phi_{ki} w_{ij} = \lambda_k \phi_{kj}, \quad (27)$$

with ϕ_{ki} 's satisfying the orthonormal conditions,⁴²

$$\sum_{k=0}^{n-1} \phi_{ki} \phi_{kj} = \delta_{ij} \rho_i, \quad (28)$$

$$\sum_{i=1}^n \frac{\phi_{ki} \phi_{li}}{\rho_i} = \delta_{kl}. \quad (29)$$

Here we have started the index of the eigenmodes at 0 instead of 1 for later convenience.

The above symmetry implies the existence of a left eigenvector, $\boldsymbol{\rho}$, corresponding to the uniform right eigenvector associated with Eq. (26), with eigenvalue 1,

$$\sum_{i=1}^n \rho_i w_{ij} = \sum_{i=1}^n \rho_j w_{ji} = \rho_j.$$

We will label this eigenvector by index $k = 0$, and

$$\phi_{0i} = \rho_i, \quad (30)$$

with $\lambda_0 = 1$. Clearly, Eq. (30) satisfies the normalization condition given by Eq. (29) for $k = l = 0$, while the orthogonality between this eigenvector and others demands

$$\sum_{i=1}^n \phi_{ki} = \delta_{k0}. \quad (31)$$

We can further define a generalized Fourier transform, $\mathcal{F}$, for variable, x_i , as

$$\tilde{x}_k \equiv \mathcal{F}[x_i]_k \equiv \sum_{i=1}^n \phi_{ki} x_i, \quad (32)$$

and the inverse transform as

$$x_i = \frac{1}{\rho_i} \sum_{k=0}^{n-1} \phi_{ki} \tilde{x}_k. \quad (33)$$

Then, we can rewrite Eq. (31) as $\mathcal{F}[1]_k = \delta_{k0}$, and Eq. (30) shows that the bin average $x_{av} = \mathcal{F}[x_i]_0 = \tilde{x}_0$, so that

$$\tilde{x}_{*k} = \tilde{x}_k - x_{av} \delta_{k0} = \begin{cases} 0 & k = 0, \\ \tilde{x}_k & k > 0. \end{cases} \quad (34)$$

Thus, the only effect of the bin-average deduction on $\tilde{x}_k$ is to annihilate the $k = 0$ mode. From Eqs. (32) and (33), we have Parseval's theorem,

$$\sum_{k=1}^{n-1} \tilde{x}_k \tilde{y}_k = \sum_{k=0}^{n-1} \tilde{x}_{*k} \tilde{y}_{*k} = \sum_{i=1}^n \rho_i x_{*i} y_{*i}. \quad (35)$$

We can now use Eq. (27) to diagonalize the multiple-bin updating scheme, Eq. (5) [using $v_i(t)$ instead of $u_i(t)$], as

$$\tilde{v}_k(t+1) = \tilde{v}_k(t) + \alpha(t) \lambda_k \tilde{f}_k(t). \quad (36)$$

In continuous time, we have, for $k > 0$,

$$\dot{\tilde{v}}_k(t) = \alpha(t) \lambda_k \tilde{f}_k(t) \approx -\alpha(t) \lambda_k [\tilde{v}_k(t) - \tilde{\xi}_k(t)], \quad (37)$$

where we have Fourier transformed Eq. (12) and used Eq. (34) in the second step.

Equation (37) shows that in the absence of the fluctuating term, $\tilde{\xi}_k(t)$, the deviation of the bias potential would tend to zero at a rate of $\alpha(t) \lambda_k$. Thus, the updating scheme is convergent only if no eigenvalue, λ_k , is negative for $k > 0$. The formal solution of Eq. (37) is given by

$$\tilde{v}_k(T) = \tilde{v}_k(0) e^{-\lambda_k q(T)} + \int_0^T \dot{\theta}_k(q(t)) \tilde{\xi}_k(t) dt, \quad (38)$$

where

$$\theta_k(q') \equiv e^{\lambda_k [q' - q(T)]}. \quad (39)$$

2. Error

By Eq. (35), we can rewrite the error defined in Eq. (9) as a sum of the errors in individual modes,⁴³

$$E(T) = \sum_{k=1}^{n-1} \langle \tilde{v}_k^2(T) \rangle.$$

As in Eq. (16), we can decompose the error as the sum of the residual error and the asymptotic error, with

$$E_R(T) = \sum_{k=1}^{n-1} \langle \tilde{v}_k^2(0) \rangle e^{-2\lambda_k q(T)}, \quad (40)$$

$$\begin{aligned} E_A(T) &= \sum_{k=1}^{n-1} \int_0^T \int_0^T \langle \tilde{\xi}_k(t) \tilde{\xi}_k(t') \rangle \theta_k(q(t)) \theta_k(q(t')) dt dt' \\ &= \int_0^T \sum_{k=1}^{n-1} \Gamma_k \dot{\theta}_k^2(q(t)) dt, \end{aligned} \quad (41)$$

where

$$\Gamma_k \approx \int_{-\infty}^{\infty} \langle \tilde{\xi}_k(t) \tilde{\xi}_k(0) \rangle dt \rightarrow \sum_{t=-\infty}^{\infty} \langle \tilde{\xi}_k(t) \tilde{\xi}_k(0) \rangle, \quad (42)$$

with the second expression being the counterpart in the discrete-time setting. Equation (42) represents a first approximation for modeling a general fluctuation, $\tilde{\xi}_k(t)$, as an equivalent white noise. In this way, the underlying sampling process affects the error only through the few numbers, Γ_k 's. Note that by Eq. (35), we have $\sum_{k=1}^{n-1} \langle \tilde{\xi}_k(t) \tilde{\xi}_k(0) \rangle = \sum_{i=1}^n \rho_i \langle \xi_{*i}(t) \xi_{*i}(0) \rangle$, and summing t from $-\infty$ to ∞ yields

$$\sum_{k=1}^{n-1} \Gamma_k = \sum_{i=1}^n \rho_i G_i = \Gamma. \quad (43)$$

This shows the equivalence of Eqs. (18) and (41) in the case of the single-bin updating scheme, where $\lambda_k \equiv 1$.

3. Optimal schedule

As in the single-bin case, by varying the schedule, $\alpha(t)$, under a fixed value of $q(T) = \int_0^T \alpha(t) dt$, we can focus on the asymptotic error, $E_A(T)$. The value of $q(T)$ will be adjusted later in the minimization of the total error. We can rewrite Eq. (41) much like the action of a particle whose position is given by $q(t)$,

$$E_A(T) = \int_0^T \mathcal{L}[q(t)] dt,$$

where the Lagrangian is

$$\mathcal{L}[q(t)] = \sum_{k=1}^{n-1} \Gamma_k \dot{\theta}_k^2[q(t)] = M^2(q(T) - q(t)) \dot{q}^2(t),$$

where we have used Eq. (39) and defined the (square-root) mass function as

$$M(\bar{q}) \equiv \sqrt{\sum_{k=1}^{n-1} \Gamma_k \lambda_k^2 e^{-2\lambda_k \bar{q}}}. \quad (44)$$

In this mechanical analogy, the schedule, $\dot{q}(t) = \alpha(t)$, corresponds to the velocity of the particle, and the above Lagrangian, containing only the kinetic energy, describes a free particle with a position-dependent mass. Here, the Hamiltonian coincides with the Lagrangian,

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial \dot{q}} \dot{q} - \mathcal{L} = 2\mathcal{L} - \mathcal{L} = \mathcal{L}.$$

Since the Lagrangian does not explicitly depend on time, t , the Hamiltonian is conserved,⁴⁴ i.e., the asymptotic error grows linearly with time at a rate of $\mathcal{L} = \mathcal{H}$. We may set

$$\sqrt{\mathcal{H}} = \sqrt{\mathcal{L}} = M(q(T) - q(t)) \dot{q}(t) = \frac{C_M}{T}, \quad (45)$$

for some positive C_M . Equation (45) serves as a generalization of Eq. (19). By integrating this equation of motion, we get

$$\int_{q(T)-q(t)}^{q(T)} M(\bar{q}) d\bar{q} = C_M \frac{t}{T}, \quad (46)$$

and C_M can be determined from the $t = T$ case as

$$C_M = \int_0^{q(T)} M(\bar{q}) d\bar{q}. \quad (47)$$

Thus, Eq. (45) gives an implicit equation for $q(t)$ and hence the optimal schedule, $\alpha(t) = \dot{q}(t)$. This solution can be characterized geometrically, as shown in Appendix B. The minimal asymptotic error is

$$E_A(T) = \mathcal{L} T = \frac{C_M^2}{T} = \frac{1}{T} \left(\int_0^{q(T)} M(\bar{q}) d\bar{q} \right)^2. \quad (48)$$

4. Optimal initial updating magnitude

We now determine the optimal value of $q(T)$ that minimizes the total error defined in Eq. (16). To model the initial error for Eq. (40), we will assume that at $t = 0$ the system has undergone a preliminary FDS simulation under a constant updating magnitude, a_0 , for T_0 steps. If the initial bias potential before the preliminary run is zero, we have, from Eq. (38) [with $T \rightarrow T_0, q(t) \rightarrow a_0 t$],

$$\langle \tilde{v}_k(0) \rangle = \tilde{u}_k e^{-\lambda_k a_0 T_0}, \quad (49)$$

$$\langle \tilde{v}_k^2(0) \rangle = \frac{1}{2} \Gamma_k \lambda_k a_0 + \epsilon_k e^{-2\lambda_k a_0 T_0}, \quad (50)$$

where $\epsilon_k \equiv \left(\tilde{u}_k^2 - \frac{1}{2}\Gamma_k \lambda_k a_0\right) e^{-2\lambda_k a_0 T_0}$.

From Eqs. (40), (50), and (48), we find that the optimal $q(T)$ that minimizes $E_R(T) + E_A(T)$ satisfies

$$\Delta(q(T)) \equiv \frac{M(q(T))}{T} \int_0^{q(T)} M(\bar{q}) d\bar{q} - \frac{a_0}{2} M^2(q(T)) - \sum_{k=1}^{n-1} \lambda_k \epsilon_k e^{-2\lambda_k q(T)} = 0. \quad (51)$$

We can rewrite the condition in terms of the initial updating magnitude by Eqs. (45) and (47),

$$\alpha(t=0) = \frac{a_0}{2} + \sum_{k=1}^{n-1} \frac{\lambda_k \epsilon_k e^{-2\lambda_k q(T)}}{M^2(q(T))} \xrightarrow{\epsilon_k \rightarrow 0} \frac{a_0}{2}. \quad (52)$$

If the initial bias, ϵ_k , is negligible, the optimal initial updating magnitude will reach the ideal limit given by half of the updating magnitude in the preliminary run.

5. Procedure for computing the optimal schedule and error

Based on the above results, we give a procedure for computing the optimal schedule and the error for a given adaptive FDS simulation. Basically, we need to extract λ_k 's, Γ_k 's, and ϵ_k 's from the given system and updating scheme (input) and then apply them to Eq. (46) for the schedule, as shown in Fig. 1. While the updating scheme alone determines λ_k 's, Γ_k 's depend also on the system and the underlying sampling method. The procedure is given below.

1. Compute the eigenvalues, λ_k 's, from the updating matrix, $\mathbf{w}$. Note that no eigenvalue can be negative in a convergent updating scheme.
2. Run a preliminary adaptive FDS simulation under a constant updating magnitude, a_0 .
3. Estimate Γ_k 's (cf. Appendix C) and the initial bias ϵ_k .⁴⁵
4. Compute the optimal value of $q(T)$ by solving Eq. (51).⁴⁶
5. Compute the optimal schedule by numerically integrating Eqs. (46) and (47).⁴⁷
6. Compute the total error $E(T)$ from Eqs. (16), (40), (50), and (48).

6. Gaussian updating scheme

A commonly used multiple-bin updating scheme in metadynamics is the Gaussian updating scheme, whose updating matrix can be characterized by a rigid Gaussian function centered on the current bin (with some boundary adjustment). In

FIG. 1. Computation of the optimal schedule and the error.

terms of the continuous collective variable, z , the eigenvalues of the updating matrix are given by⁴⁰

$$\lambda_k = \exp\left(-\frac{\pi^2 \sigma_z^2 k^2}{2 \Delta z^2}\right), \quad (53)$$

where σ_z is the width of the Gaussian and $\Delta z = n \delta z g/2$ is the effective span with $g = 1$ for a periodic variable or 2 otherwise (cf. Appendix D 2).

The optimal schedule of the Gaussian updating scheme generally has no simple closed-form expression. Nonetheless, we can show that it always lies above the inverse-time schedule and may assume a roughly exponentially decaying form initially (cf. Appendix B).

7. Homogeneous updating schemes

More generally, we can replace the rigid Gaussian function in the Gaussian updating scheme by a window function of different shape. We call the resulting updating scheme *homogeneous* or translationally invariant.

By morphing the window function but preserving the translational invariance, we can design a homogeneous updating scheme that is more convenient for error control than the Gaussian updating scheme. As shown in Appendix D, the homogeneous updating schemes share the same set of orthonormal eigenvectors, ϕ_{ki} , which are cosine and sine functions, and differ only by the eigenvalues. Thus, the window function is completely determined by the eigenvalues, and the updating matrix is given by⁴⁰

$$w_{ij} = \frac{1}{\rho_i} \sum_{k=0}^{n-1} \lambda_k \phi_{ki} \phi_{kj}. \quad (54)$$

As shown below, the optimized updating scheme allows the optimal schedule to coincide with the inverse-time formula and thus spare the need to go through the general procedure in Sec. II C 5.

D. Optimization of updating schemes

Now that we can compute the optimal schedule for a given updating scheme, we turn to the problem of optimizing the updating scheme.

1. Asymptotic optimality of the single-bin updating scheme

We first show that the single-bin updating scheme is asymptotically optimal. We wish to find a lower bound of the asymptotic error, $E_A(T)$ [given by Eq. (41)], for a class of updating matrices that share the same set of eigenvectors, hence the same Γ_k 's, but not the eigenvalues, λ_k 's. If all eigenvalues are positive, we can use the Cauchy-Schwarz inequality,⁴⁸

$$\left(\int_0^T dt \sum_{k=1}^{n-1} \Gamma_k \dot{\theta}_k^2(q(t))\right) \left(\int_0^T dt \sum_{k=1}^{n-1} \Gamma_k\right) \geq \left[\int_0^T dt \sum_{k=1}^{n-1} \Gamma_k \dot{\theta}_k(q(t))\right]^2 = \left[\sum_{k=1}^{n-1} \Gamma_k (1 - e^{-\lambda_k q(T)})\right]^2.$$

In the long time limit, $q(T) \rightarrow \infty$, the factor $e^{-\lambda_k q(T)}$ is negligible, and the lower bound is independent of the λ_k 's:

$E_A(T) \geq \sum_{k=1}^{n-1} \Gamma_k / T$. This inequality applies to any schedule and any updating scheme in the class, and the equality is achieved if $\theta_k(q(t)) = \text{constant}$ for all $k > 0$ at any t . Solving these equations yields a set of identical eigenvalues, $\lambda_1 = \dots = \lambda_{n-1}$.

We can show that the above optimal solution represents the single-bin updating scheme. Since the $k = 0$ eigenmode represents a uniform shift of the bias potential, we can set $\lambda_0 = \lambda_1$ without changing the nature of the updating scheme. By Eqs. (54) and (28), we have

$$w_{ij} = \frac{1}{\rho_i} \sum_{k=0}^{n-1} \lambda_k \phi_{ki} \phi_{kj} = \lambda_1 \delta_{ij},$$

which is a multiple of the identity updating matrix that defines the single-bin updating scheme.

We can further construct an optimal *time-dependent* updating scheme by a superposition of single-bin updating schemes. Since each eigenmode evolves independently with an effective updating magnitude, $\lambda_k \alpha(t)$, as shown in Eq. (37), the ideal error reduction requires a time-dependent eigenvalue, $\lambda_k(t)$, (hence a time-dependent updating scheme) that allows the effective updating magnitude to follow the eigenmode-specific optimal schedule, $\lambda_k(t) \alpha(t) = 1/(t + t_{0k})$, with $t_{0k} = \Gamma_k / \langle \dot{v}_k^2(0) \rangle$, which is borrowed from Eq. (21) for the single-bin case. In long times, however, this time-dependent optimal solution is reduced to the single-bin updating scheme under the inverse-time schedule as the differences among the t_{0k} 's become negligible.

2. Bandpass updating schemes

We can generalize the single-bin updating scheme to a class of perfect (brick-wall) *bandpass* updating schemes that discard certain noise modes of short wavelengths. If the eigenmodes are sorted in descending order of the typical wavelength, and the PMF is sufficiently smooth to be represented by the first K eigenmodes, we may assign the eigenvalues of the first K eigenmodes to 1 and the remaining eigenvalues to 0,

$$\lambda_0 = \dots = \lambda_{K-1} = 1 \quad \text{and} \quad \lambda_K = \dots = \lambda_{n-1} = 0. \quad (55)$$

The updating matrix can then be reconstructed from the eigenvalues by Eq. (54). In effect, each updating step of the bandpass updating scheme is equivalent to a single-bin updating step, Eq. (4), followed by a filtering step that eliminates the components of the noise modes of $k \geq K$ from the bias potential. Like the single-bin updating scheme, the bandpass updating scheme is also asymptotically optimal, but only up to the first K fluctuation modes.

A distinct advantage of the bandpass updating schemes is that they share with the single-bin updating scheme the asymptotically optimal schedule, the inverse-time schedule, Eq. (21). Thus, it is unnecessary to go through the general procedure described in Sec. II C 5 for these updating schemes.

We will consider two special classes of bandpass updating schemes: homogeneous and polynomial bandpass updating schemes. First, by applying this construction to the basis of cosine and sine functions, we get a homogeneous bandpass updating scheme (cf. Appendix D 3), which resembles the

Gaussian updating scheme, but with long-range wiggles. In Sec. III A, we will further compare the homogeneous bandpass updating scheme with the Gaussian updating scheme in an example. The other class of polynomial bandpass updating schemes will be discussed below.

E. Polynomial bandpass updating schemes

Here we consider a class of bandpass updating schemes constructed from a sequence of orthogonal polynomials. The resulting polynomial bandpass updating schemes are more intuitive than the homogeneous bandpass updating scheme in that the former express the bias potential as a simple polynomial of the collective variable.

1. Orthogonal polynomial basis

For convenience, we replace the discrete bin index, i , by the continuous variable, z , ρ_i by $\rho(z) \delta z$, ϕ_{ki} by $\phi_k(z) \delta z$, w_{ij} by $w(z, z') \delta z$, etc. We define the eigenvectors by $\phi_k(z) = R_k(z) \rho(z)$, where $R_k(z)$ is the k th member of a sequence of orthogonal polynomials that satisfy

$$\int R_k(z) R_{k'}(z) \rho(z) dz = \overline{R_k(z) R_{k'}(z)} = \delta_{k,k'},$$

where the overline denotes the average for the distribution, $\rho(z)$. The first few members are⁴⁹

$$R_0(z) = 1, \quad R_1(z) = \frac{z - \bar{z}}{\sigma_z}, \quad R_2(z) = \frac{(z - \bar{z})^2 - \sigma_z^2}{\kappa_z},$$

where $\sigma_z \equiv \sqrt{\overline{(z - \bar{z})^2}}$ and $\kappa_z \equiv \sqrt{\overline{[(z - \bar{z})^2 - \sigma_z^2]^2}}$. We can readily verify that the eigenvectors satisfy Eqs. (28)–(31). The eigenvectors allow us to construct the updating matrix from Eq. (54) with the eigenvalues given by Eq. (55):

$$w(z, z') = \sum_{k=0}^{K-1} \frac{\phi_k(z) \phi_k(z')}{\rho(z)} = \sum_{k=0}^{K-1} R_k(z) R_k(z') \rho(z').$$

Then, by Eq. (5), the bias potential is updated as

$$u(z, t+1) = u(z, t) + \sum_{k=0}^{K-1} R_k(z) R_k(z(t)) \alpha(t).$$

We can write the bias potential as a superposition of the orthogonal polynomials

$$u(z, t) = \sum_{k=0}^{K-1} c_k(t) R_k(z), \quad (56)$$

with the variable coefficients updated as

$$c_k(t+1) = c_k(t) + \alpha(t) R_k(z(t)). \quad (57)$$

Note that although the target distribution, $\rho(z)$, does not explicitly appear in the bias potential, Eq. (56), or in the updating scheme, Eq. (57), it affects both via the parameters, $\bar{z}$, σ_z , κ_z , etc. For $k > 0$, $R_k(z)$ is a polynomial of degree k with zero expectation, and $R_k(z(t))$ roughly represents the deviation of the k th statistical moment of the instantaneous distribution from its expectation. As high-order moments are generally unreliable, this method works better with a small K for a narrow distribution. However, by constructing a mixture or expanded ensemble,^{13,31,35,50–53} we can simultaneously

sample multiple target distributions to cover a wider range, as exemplified below.

2. Langfeld-Lucini-Rago (LLR) algorithm

For example, to sample a flat distribution, $\rho(z) = 1/\Delta z$, over $(z_c - \frac{1}{2}\Delta z, z_c + \frac{1}{2}\Delta z)$, we can use the Legendre polynomials.⁵⁴ If the interval is sufficiently small, we can use $K = 2$ to limit the basis to the first two modes, $R_0(z) = 1$ and $R_1(z) = \sqrt{12}(z - z_c)/\Delta z$ (with $\sigma_z = \Delta z/\sqrt{12}$). If we drop the $k = 0$ mode that corresponds to a uniform shift of the bias potential, the updating scheme is essentially the LLR algorithm¹⁵ in the single-interval case, and the inverse-time schedule has been recommended for the optimal convergence of c_1 .¹⁶

In practice, we often apply the above algorithm simultaneously to an array of adjacent intervals to cover a wide range of the collective variable, z .^{15,16} Note that in this expanded ensemble setup, the updating equation for c_0 is no longer trivial, but it corresponds to a single-bin updating scheme of constructing the bias potential for transitions among different intervals. The WL algorithm is in fact recovered by the $K = 1$ case. However, the LLR algorithm estimates c_0 from an approximate relation similar to Eq. (63) below.

3. Adaptive Gaussian ensemble

In analogy to the above algorithm, we will consider the case of sampling a Gaussian distribution

$$\rho(z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left[-\frac{(z - z_c)^2}{2\sigma_z^2}\right], \quad (58)$$

using a basis of Hermite polynomials.⁵⁴ As discussed above, the method is effective only for a local region of a multimodal distribution, and thus we will assume a small width, σ_z . Because the Gaussian distribution has no hard boundaries as in the previous case, we need to use, in addition to the linear mode, $R_1(z) = (z - z_c)/\sigma_z$, a quadratic one, $R_2(z) = [(z - z_c)^2 - \sigma_z^2]/(\sqrt{2}\sigma_z^2)$, to restrict the sampling region. The bias potential is thus quadratic, which is common in umbrella sampling^{17,32} and the Gaussian ensemble method.^{31,55}

4. Expanded ensemble

We can expand the sampling range by constructing an expanded ensemble from a superposition of S Gaussian umbrellas centered at different places. The expanded ensemble treats the umbrella index, $s = 1, \dots, S$, also as a random variable, and the biased sampling targets a joint distribution of the umbrella index and the continuous variable, z , $\rho(s, z) = \rho_s \rho^{(s)}(z)$. Here, the marginal inter-umbrella distribution is flat, $\rho_s = 1/S$, and the conditional intra-umbrella distribution, $\rho^{(s)}(z)$, is the Gaussian given in Eq. (58).⁵⁶ To sample this distribution, we give each umbrella an independent bias potential of the form of Eq. (56),

$$u^{(s)}(z) = \hat{c}_0^{(s)} + \frac{c_1^{(s)}}{\sigma_z^{(s)}}(z - z_c^{(s)}) + \sqrt{2}c_2^{(s)}\frac{(z - z_c^{(s)})^2}{2(\sigma_z^{(s)})^2}, \quad (59)$$

with $\hat{c}_0^{(s)} \equiv c_0^{(s)} - c_2^{(s)}/\sqrt{2}$, and update the coefficients as

$$c_k^{(s)}(t+1) = c_k^{(s)}(t) + \frac{\alpha(t)}{\rho_s} R_k^{(s)}(z(t)) \delta_{s,s(t)}, \quad (60)$$

where $s(t)$ denotes the umbrella index at time t .

In the expanded ensemble, inter-umbrella transitions can be handled via the frameworks of parallel^{13,31,50} and/or simulated tempering.^{35,51,52} In the latter framework, the transition probability from the s th umbrella to the s' th umbrella is given by

$$A(s \rightarrow s') = \min\left\{1, \frac{\exp[-u^{(s')}(z(t), t)]}{\exp[-u^{(s)}(z(t), t)]}\right\}, \quad (61)$$

where $z(t)$ is the current value of z . In this case, the parameter, $c_0^{(s)}$, serves as the bias potential for inter-umbrella transitions, and it is updated by Eq. (60), much like in the single-bin updating scheme.⁵¹

Since the bandpass updating schemes are a straightforward extension of the single-bin updating scheme, we can borrow the existing protocol²² to optimally control the updating magnitude. We can adopt the WL stage-by-stage strategy early in simulation and switch to the inverse-time schedule once the updating magnitude from the WL strategy falls under the value given by the inverse-time formula.⁵⁷ However, for this application, it is natural to redefine the measure of histogram fluctuations used in the WL strategy for stage transitions by Eq. (E3) to include the intra-umbrella fluctuations.

Upon convergence, the parameters, $c_k^{(s)}$, can be related to the PMF. By Eqs. (2) and (58), we have

$$u^{(s)}(z) \approx \ln p(z) + \frac{(z - z_c^{(s)})^2}{2(\sigma_z^{(s)})^2} + \text{const.}$$

Comparing this to Eq. (59), we get

$$\hat{c}_0^{(s)} \approx \ln p(z_c^{(s)}) + \text{const.}, \quad (62a)$$

$$\frac{c_1^{(s)}}{\sigma_z^{(s)}} \approx \left. \frac{d \ln p(z)}{dz} \right|_{z=z_c^{(s)}}, \quad (62b)$$

$$\frac{\sqrt{2}c_2^{(s)} - 1}{(\sigma_z^{(s)})^2} \approx \left. \frac{d^2 \ln p(z)}{dz^2} \right|_{z=z_c^{(s)}}. \quad (62c)$$

Thus, $\hat{c}_0^{(s)}$ as a function of $z_c^{(s)}$ approximates the negative PMF, $\ln p(z)$,¹⁷ $c_1^{(s)}/\sigma_z^{(s)}$ approximates the mean force at $z = z_c^{(s)}$, and $\sqrt{2}c_2^{(s)}$, serving as the relative “pressure” to restrain the width of the distribution, would exceed 1 in a barrier region where the PMF is locally concave, i.e., $-d^2 \ln p(z)/dz^2 < 0$. From Eqs. (62a) and (62b), we also have approximately⁵²

$$\hat{c}_0^{(s+1)} - \hat{c}_0^{(s)} \approx \frac{1}{2} \left(\frac{c_1^{(s+1)}}{\sigma_z^{(s+1)}} + \frac{c_1^{(s)}}{\sigma_z^{(s)}} \right) (z_c^{(s+1)} - z_c^{(s)}). \quad (63)$$

Note that the LLR algorithm (cf. Sec. II E 2) implicitly uses a similar relation for the estimation of $c_0^{(s)}$ [which is equivalent to $\hat{c}_0^{(s)}$ here for lack of $c_2^{(s)}$ in that case].

We summarize the method as the following procedure.

1. Set up an array of Gaussian umbrellas covering the desired range, e.g., we may set, for $s = 1, \dots, S$, the mean, $z_c^{(s)} = z_{\min} + \frac{s-1}{S-1}(z_{\max} - z_{\min})$, and the width, $\sigma_z^{(s)}$, to be a constant, σ_z . For the polynomial coefficients, we can set $c_1^{(s)}/\sigma_z^{(s)}$ to some estimated mean force, η_z ; $c_2^{(s)} = 1/\sqrt{2}$; and $c_0^{(s)} = \eta_z[z_c^{(s)} - \frac{1}{2}(z_{\min} + z_{\max})]$ by Eq. (63).
2. Equilibrate the system at umbrella $s = 1$. Repeat the following steps until the simulation ends.

3. Perform a regular MC or MD step in the current umbrella, $s = s(t)$, under the bias potential $u^{(s)}(z, t)$.
4. Attempt a transition to another umbrella with the acceptance probability given by Eq. (61).⁵⁸
5. Update $c_k^{(s)}$'s according to Eq. (60) and increase t by 1.
6. Change the updating magnitude, $\alpha(t)$.

After simulation, we can use the multiple histogram method^{1,2,32,59} to recover the unbiased distribution,

$$p(z) \delta z \propto \frac{\sum_{s=1}^S n^{(s)}(z)}{\sum_{s=1}^S N^{(s)} e^{-u^{(s)}(z)}}, \quad (64)$$

where $n^{(s)}(z)$ is the number of visits to the bin $(z - \frac{1}{2}\delta z, z + \frac{1}{2}\delta z)$ in umbrella s and $N^{(s)}$ is the total number of visits to the umbrella s .

III. NUMERICAL RESULTS

A. Comparison of homogeneous updating schemes

In this example, we will study the optimal schedule and compare the single-bin, Gaussian (Sec. II C 6), and homogeneous bandpass (Sec. II D 2) updating schemes.

1. System and simulation details

Our test system is a Lennard-Jones (LJ) fluid of $N = 100$ particles and we wish to compute the PMF along the distance, r , between two particular particles. We will assume reduced units in the calculation. The two particular particles, labeled as 1 and 2, are mutually non-interacting but do interact with other particles via the pair potential $\varphi(r) = 4(r^{-12} - r^{-6})$ (truncated at half of the box size). The PMF is the negative logarithm of the intrinsic distribution of the distance, $p(r)$, which is proportional to the cavity distribution function.⁶⁰

We used biased sampling to calculate the PMF from $r = 0$ to half of the box size on a grid of size $\delta r = 0.01$ in reduced units. The reduced density was $\rho = 0.1$. The number of bins was $n = 500$. The reduced temperature was $T = 3$. The bias potential, $u(r)$, was initially set to zero and underwent the test updating scheme and schedule. Each simulation started with an “equilibration” phase of 10^7 steps under a constant updating magnitude, $a_0 = 10^{-4}$ (corresponding to $\ln f = n a_0 = 0.05$ in the WL algorithm), followed by a production phase of $T = 10^8$ steps with the updating magnitude given by either the inverse-time or the optimal schedule. The deviation of the bias potential was calculated from Eqs. (7) and (8), and the error from Eq. (9), with the reference bias potential pre-calculated from a longer simulation under the single-bin updating scheme. We repeated each simulation multiple times and report the averages. For the Gaussian updating scheme, we used a set of fixed reference values of Γ_k 's⁶¹ to ensure the same optimal schedule for multiple runs.

We used the Metropolis algorithm^{1,62} for configuration sampling. Each step consisted of one MC trial for displacing one of the two tagged particles and another four trials for displacing the other particles. In each trial, a random particle was displaced in each spatial dimension according to a uniform distribution over $(-\varepsilon, \varepsilon)$. The two tagged particles were fixed on the x -axis and only displaced in this direction. We used $\varepsilon = 0.1/\rho$ to ensure a reasonable acceptance ratio.

TABLE I. Errors of the bias potential under the single-bin, Gaussian (with width $\sigma = 0.28$) and homogeneous bandpass (with mode cutoff $K = 26$) updating schemes on the LJ system. The numbers have been averaged over multiple runs, and the predicted values are shown in parentheses for comparison; E_{initial} and E_{final} denote the errors at the beginning and end of the production phase, respectively.

Scheme	Schedule	E_{initial}	E_{final}
Single-bin	$1/t$	$5.7(4.5) \times 10^{-2}$	$9.1(9.1) \times 10^{-6}$
Gaussian	$1/t$	$6.2(6.2) \times 10^{-3}$	$1.46(1.44) \times 10^{-5}$
Gaussian	Optimal	$6.4(6.2) \times 10^{-3}$	$2.72(2.74) \times 10^{-6}$
Bandpass	$1/t$	$8.3(8.0) \times 10^{-3}$	$1.69(1.67) \times 10^{-6}$

2. Results

Both the optimal and potentially sub-optimal inverse-time schedules reduced the error significantly, as shown in Table I. While the final errors agreed well with the predicted values (which were calculated using the reference Γ_k 's and $\tilde{u}_k$'s), there were considerable deviations in the initial errors. This shows that the assumptions used in our methods, including those on quasi-equilibrium sampling, could fail with a large updating magnitude.

As shown in Fig. 2, the optimal schedule for the Gaussian updating scheme with width $\sigma = 0.28$ was more complex than the inverse-time one. We chose the Gaussian updating scheme here because the optimal and inverse-time schedules were identical in the single-bin and bandpass updating schemes. The optimal schedule lay above the inverse-time schedule, and its initial decay was roughly exponential, in agreement with the discussion in Appendix B. Since our equilibration phase was sufficiently long, the initial updating magnitude, $\alpha(0)$, was around the ideal value of $a_0/2$ [cf. Eq. (52)].

The optimal schedule more effectively reduced the error than the inverse-time schedule for the Gaussian updating scheme, as shown in Table I. By decomposing the error into the components of the eigenmodes, $\langle \tilde{v}_k^2 \rangle$'s, using Eqs. (32) and (D2) (Fig. 3), we found that the inverse-time schedule failed to reduce the errors in the $k = 7$ to 21 modes adequately. By a delayed reduction, the optimal schedule removed more errors in these modes and produced a flatter profile. However, the gain in error reduction of using the optimal schedule diminished for shorter simulations, e.g., when the production phase was shortened to $T = 10^7$ steps, the final errors under the optimal and inverse-time schedules became 2.2×10^{-5} and 3.4×10^{-5} , respectively, which were more similar.

FIG. 2. Optimal schedule of a Gaussian updating scheme.

FIG. 3. Error components, $\langle \tilde{v}_k^2 \rangle$'s, of the PMF. The optimal schedule more effectively reduced the errors of middle-range eigenmodes than the inverse-time schedule ($1/t$). The “corrected” values were obtained from Eq. (A6). The lines are a guide for the eye.

For the Gaussian updating scheme, the error depended not only on the schedule but also on the width of the Gaussian. Although widening the Gaussian can suppress local noise, it can also mask finer features of the PMF.⁶³ Thus, there is an optimal width that balances the two factors. According to the analytical prediction, the optimal width (assuming the optimal schedule) was $\sigma = 0.28$, which we used in the simulations, and it would decrease with the simulation length and approach the limit of the single-bin updating scheme, $\sigma \rightarrow 0$, asymptotically.

The homogeneous bandpass updating scheme had a similar behavior. The mode cutoff, K , also has an optimal value, which was 26 according to the prediction. The minimal error of the bandpass case was slightly less than the corresponding value in the Gaussian case, as shown in Table I. The optimal cutoff would increase with the simulation length and approach the limit of the single-bin updating scheme, $K \rightarrow n$, asymptotically.

Although the Gaussian and homogeneous bandpass updating schemes helped reduce the error in this case, it was only because the sampling along the collective variable was free from major barriers such that the histogram fluctuation was largely local in nature. For similar simulations performed at a higher density, $\rho = 0.8$, we found that the above two updating schemes did not improve the convergence of the bias potential significantly compared to the single-bin updating scheme. However, the multiple-bin updating schemes have advantages in MD simulations when a smooth bias potential is needed for calculating the force.

For simulations using sub-optimal updating schemes and schedules, we can use the histogram-based correction formula, Eq. (A6), to improve the estimated PMF. In Fig. 3, we show that for a simulation with a sub-optimal schedule, this correction produced the flatter profile seen in the single-bin case. While this correction also introduces some local noise in the large k modes, the noise is likely harmless for a post-simulation analysis, as the bias potential no longer needs to be differentiated for force calculation.

B. Adaptive Gaussian ensemble

As another proof of principle, we tested the adaptive Gaussian ensemble method discussed in Secs. II E 3 and II E 4 on the two-dimensional $L \times L$ ten-state Potts model^{2,7,8,64}

in the energy space. In this case, the intrinsic distribution is proportional to the density of states, $g(E)$. We spread the Gaussian umbrellas of the same width, $\sigma_E^{(s)} = L$, over $[-1.8L^2, -0.8L^2]$ with an even spacing of L . For simplicity, only transitions between neighboring umbrellas were performed.⁶⁵ Initially, we set $c_1^{(s)} = \tilde{\beta}_c \sigma_E^{(s)}$ for all umbrellas with $\tilde{\beta}_c = 1.4$ being the approximate inverse critical temperature. For configuration sampling in each umbrella, we used the Wolff algorithm^{2,66} (which was more efficient than the Metropolis algorithm) at the inverse temperature, $\beta = c_1^{(s)}/\sigma_E^{(s)}$, followed by a Metropolis step with acceptance probability $\min\{1, e^{c_2^{(s)}[R_2(E) - R_2(E')]} \}$, where E and E' are the energies of the old and new configurations, respectively. The acceptance ratio was no less than 91% in any umbrella. In the initial WL stages, the initial updating magnitude was 0.01, the fluctuation threshold for stage transitions was $F = 0.2$ [using the definition given by Eq. (E3)], and the reduction factor was $1/2$.

The results for the $L = 64$ case after $10^5 L^2$ steps are shown in Fig. 4. The Gaussian histograms were evenly spaced with similar widths as intended, as shown in Fig. 4(a). We computed the density of states by Eq. (64) and found the critical temperature that makes the two distribution peaks of equal height to be $T_c = 0.70161$, in agreement with a previous study.⁸ Since Gaussian umbrellas were sufficiently narrow, we could readily verify Eqs. (62): the logarithmic density of states, $\ln g(E)$, was roughly captured by $\hat{c}_0^{(s)}$ as a function of $E_c^{(s)}$, as shown Fig. 4(b); $c_1^{(s)}/\sigma_E^{(s)}$, roughly representing the local or statistical

FIG. 4. Sampling using adaptive Gaussian ensembles on the 64×64 ten-state Potts model. (a) Normalized histograms. (b) $\hat{c}_0^{(s)} = c_0^{(s)} - c_2^{(s)}/\sqrt{2}$ against $E_c^{(s)}$ (circles) roughly agreed with the logarithmic density of states, $\ln g(E)$ (dashed line). In either case, we have deducted the linear contribution, $\beta_c E$, and applied a shift so as to normalize the distribution at the critical temperature, β_c . (c) $c_1^{(s)}/\sigma_E$ (circles) as an approximation of the statistical temperature, $\beta(E) = d \ln g(E)/dE; \beta_c$ (dashed line). (d) $\sqrt{2} c_2^{(s)}$ exceeded 1.0 in the unstable region, where $\beta(E)$ is increasing.

temperature, $\beta(E) = d \ln g(E)/dE$, manifested the backbending behavior^{12,13} typical in a first-order phase transition, as shown in Fig. 4(c); $\sqrt{2} c_2^{(s)}$ was indeed above 1.0 in the unstable region, where $\beta(E)$ is increasing or, equivalently, $\ln g(E)$ is locally convex, as shown in Fig. 4(d).

IV. CONCLUSIONS AND DISCUSSIONS

In conclusion, we have studied the optimization of the updating magnitude for a general class of free energy calculation methods based on flat-distribution sampling, including the Wang-Landau (WL) algorithm and metadynamics. These methods incrementally update a bias potential along a collective variable to offset the potential of mean force (PMF) and thereby sample a flat distribution. The optimal updating magnitude can improve the convergence of the bias potential and thus reduce the effort of computing the PMF.

The optimal schedule of the overall updating magnitude depends on the updating scheme, which is mathematically characterized by an updating matrix, with the element, w_{ij} , representing the relative updating magnitude at bin i when bin j is visited. Our method computes the optimal schedule from the eigenvalues of the updating matrix, which reflect the relative rates of error reduction, and the intrinsic histogram fluctuations of the eigenmodes. For the single-bin updating scheme (as often used in the WL case), whose updating matrix is the identity matrix with equal eigenvalues, we confirmed that the optimal schedule is given by the known inverse-time formula, Eq. (21). For a general multiple-bin scheme, including the Gaussian one often used in metadynamics, the optimal schedule, having to balance the different rates of error reduction of the eigenmodes, usually has no simple closed-form expression, but it is implicitly given by an equation of motion of a free particle with a position-dependent mass, Eq. (46).

In comparing different updating schemes, we found the single-bin updating scheme to be optimal for error control in the long time limit. The single-bin updating scheme can be generalized to a class of bandpass updating schemes, which discard a few noise eigenmodes and are similarly optimal for the non-noise eigenmodes. The bandpass updating schemes allow simpler control for the updating magnitude as the asymptotically optimal schedule is given by the inverse-time formula.

As an application, we studied a particular class of the bandpass updating schemes constructed from a family of orthogonal polynomials. This construction, as a generalization of the WL and LLR^{15,16} algorithms, delegates the biased sampling to a local umbrella with a simple polynomial bias potential and applies the flat-distribution sampling to the umbrellas located at different places. While our example is limited to a quadratic bias potential,^{31,32} this approach may also be used for adaptive parameter adjustment for bias potentials of different forms.^{13,67}

This study has several limitations. First, we have ignored the *systematic* error^{19–21} caused by the frequent updates to the bias potential. While this is justified in the asymptotic regime, where the updating magnitude is small and the sampling is a quasi-equilibrium one,^{19–21,27,29} it is inadequate in early stages of a simulation. The white noise approximation can also be

inaccurate for relatively short simulations. For the Gaussian updating scheme, the optimal schedule requires a long preliminary run for parameter estimation, while the gain in error reduction is sometimes rather modest for a relatively short simulation. Finally, the adaptive Gaussian ensemble method introduced in Sec. II E 3 can be refined to improve its general applicability and stability.

ACKNOWLEDGMENTS

We thank Dr. Y. Mei, Dr. W. Tang, J. Wang, Dr. M. W. Deem, Dr. K.-Y. Wong, O. Nassar, Dr. C. Lai, Dr. S. Ou, D. Stuhlsatz, Dr. C. Myers, and Dr. O. Samoylova for helpful discussions and the anonymous reviewers for critical reading of earlier versions of the manuscript. We gratefully acknowledge the Robert A. Welch Foundation (H-0037), the National Science Foundation (No. CHE-1152876), and the National Institutes of Health (No. GM-037657) for partial support of this work. J.M. thanks the support from the National Institutes of Health (Nos. R01-GM067801 and R01-GM116280), the Welch Foundation (Q-1512), and National Science Foundation (Grant No. PHY-1427654). This research used computing resources of the National Science Foundation XSEDE grid.

APPENDIX A: MOVING AVERAGE INTERPRETATION

We can understand the optimality of the inverse-time schedule for the single-bin updating scheme by a moving average interpretation. Here, the bias potential undergoing the recurrence updating relation, Eq. (4), is mapped to an equivalent moving average with a dynamic weight determined by the updating schedule. For the single-bin updating scheme, the inverse-time schedule is optimal because it corresponds to an equally weighted average.

1. Moving average and recurrence relation

Consider the weighted cumulative moving average (CMA),

$$X_\omega(t) = \frac{\omega(0)x(0) + \cdots + \omega(t)x(t)}{\Omega(t)}, \quad (\text{A1})$$

where $\Omega(t) \equiv \omega(0) + \cdots + \omega(t)$ is the cumulative weight. We can write the average as a recurrence relation

$$X_\omega(t) = X_\omega(t-1) + \alpha(t) [x(t) - X_\omega(t-1)], \quad (\text{A2})$$

where the updating magnitude is defined as

$$\alpha(t) \equiv \frac{\omega(t)}{\Omega(t)} = 1 - \frac{\Omega(t-1)}{\Omega(t)}, \quad (\text{A3})$$

and $x(t) - X_\omega(t-1)$ is the change from the previous average proposed by the data point at step t .

Conversely, we can revert a recurrence relation in the form of Eq. (A2) with a given $\alpha(t)$ to an average in the form of Eq. (A1) with the weight given by

$$\frac{\omega(t)}{\omega(0)} = \left(\prod_{\tau=1}^{t-1} \frac{1}{1 - \alpha(\tau)} \right) \frac{\alpha(t)}{1 - \alpha(t)}. \quad (\text{A4})$$

In particular, the equally weighted average, $\omega(t) = \omega(0)/t_0$ (for $t > 0$), corresponds to the inverse-time updating magnitude, Eq. (21).

2. Application to updating schemes

To apply the above formalism to the updating schemes, we will need a correction formula for the bias potential. We will use a linearized version of Eq. (3),

$$\hat{u}_i = u_i + \frac{H_i}{\rho_i} - 1.$$

Using this formula on the instantaneous histogram, $h_i(t)$, yields an independent estimate of the ideal bias potential from a single data point at step t ,

$$\hat{u}_i(t) \approx u_i(t) + \frac{h_i(t)}{\rho_i} - 1 = u_i(t) + f_i(t) - 1. \quad (\text{A5})$$

For the single-bin updating scheme, Eq. (4), we have

$$\begin{aligned} u_i(t+1) &= u_i(t) + \alpha(t)f_i(t) \\ &\approx u_i(t) + \alpha(t)[\hat{u}_i(t) - u_i(t)] + \text{const.}, \end{aligned}$$

where we have used Eq. (A5) and ignored a uniform shift in the second step. Comparing this to Eq. (A2), we find that $u_i(t+1)$ follows the same time evolution as the CMA $X_\omega(t)$ of $x(t) = \hat{u}_i(t)$. Thus, the bias potential can be thought as a CMA with the weight given by Eq. (A4). The inverse-time schedule is optimal because it corresponds to the equally weighted average, which is statistically most efficient, at least asymptotically.

Similarly, for the multiple-bin updating scheme, we have, from Eqs. (7) and (36), $\tilde{u}_k(t+1) = \tilde{u}_k(t) + \lambda_k \alpha(t)\tilde{f}_k(t)$, and from Eq. (A5), $\hat{\tilde{u}}_k(t) \approx \tilde{u}_k(t) + \tilde{f}_k(t)$, for $k > 0$. Thus,

$$\tilde{u}_k(t+1) = \tilde{u}_k(t) + \lambda_k \alpha(t) [\hat{\tilde{u}}_k(t) - \tilde{u}_k(t)],$$

which means that $\tilde{u}_k(t+1)$ is equivalent to a CMA of $\hat{\tilde{u}}_k(t)$, and the weight is given by Eq. (A4), with $\alpha(t)$ replaced by the effective updating magnitude, $\lambda_k \alpha(t)$.

3. Upper bound of the updating magnitude

The moving average interpretation imposes an upper bound on the updating magnitude. By Eq. (A3), the effective updating magnitude, $\lambda_k \alpha(t)$, of any mode $k > 0$ should not exceed 1.0. But this is not necessarily satisfied for the updating schedule derived from the continuous-time formalism. An expedient fix, which we used in Sec. II C 5, is to derive the effective updating magnitude from Eq. (A3) for the $k = 1$ mode, which has the largest eigenvalue, i.e., $\alpha(t) = \lambda_1^{-1}[1 - \Omega(t-1)/\Omega(t)]$, with

$$\begin{aligned} \ln \frac{\Omega(t-1)}{\Omega(t)} &\approx \frac{1}{t_+ - t_-} \ln \frac{\Omega(t_-)}{\Omega(t_+)} = \sum_{\tau=t_-+1}^{t_+} \frac{\ln[1 - \lambda_1 \alpha(\tau)]}{t_+ - t_-} \\ &\approx -\frac{\lambda_1 \int_{t_-}^{t_+} \alpha(\tau) d\tau}{t_+ - t_-} = -\lambda_1 \frac{q(t_+) - q(t_-)}{t_+ - t_-}, \end{aligned}$$

where t_- and t_+ are two neighboring grid points around t . This guarantees $\lambda_k \alpha(t) \leq 1$ for all nontrivial modes.

4. Histogram-based correction

For a general multiple-bin updating scheme, each mode of the bias potential corresponds to a CMA with a dynamic weight, which is not necessarily the optimal equal weight.

We can, however, recover the equally weighted average by averaging both sides of Eq. (A5),

$$\hat{U}_i(t) \approx U_i(t) + F_i(t) - 1 \approx U_i(t) + \ln \frac{H_i(t)}{\rho_i}, \quad (\text{A6})$$

where the capital letters denote the equally weighted averages. This correction serves as an approximate generalization of Eq. (3) for FDS simulations under a variable bias potential.⁶⁸ However, it does not remove the systematic bias caused by the adaptive updates and thus still requires the updating magnitude to be sufficiently small.

APPENDIX B: CHARACTERIZATION OF OPTIMAL SCHEDULES

Here we give a geometric characterization of the optimal schedule. First, we define a normalized mass distribution as

$$m(\bar{q}) = \frac{M(\bar{q})}{\int_0^{q(T)} M(q') dq'} = \frac{M(\bar{q})}{C_M}, \quad (\text{B1})$$

such that $\int_0^{q(T)} m(\bar{q}) d\bar{q} = 1$. Then, from Eqs. (45) and (46), we find that for $\bar{q}(t) \equiv q(T) - q(t)$,

$$m(\bar{q}(t)) = \frac{1}{T \alpha(t)}, \quad (\text{B2})$$

$$\int_{\bar{q}(t)}^{q(T)} m(\bar{q}) d\bar{q} = \frac{t}{T}. \quad (\text{B3})$$

The above two equations map each point on the schedule, $(t, \alpha(t))$, to a point, $(\bar{q}(t), m(\bar{q}(t)))$, on the mass distribution such that the ordinate of the latter curve is $1/(T\alpha)$, and the area under the curve between $\bar{q}(t)$ and $q(T)$, or the complementary cumulative distribution function of $m(\bar{q})$, is equal to t/T , as shown in Fig. 5(a).

We can now show that for simulations with equal lengths and equal initial updating magnitudes, the optimal schedule for a Gaussian updating scheme always lies above the inverse-time schedule, which is optimal for the single-bin updating scheme. From Eqs. (B2), (B3), (B1), and (44), we have

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{\alpha(t)} \right) &= \frac{dm}{d(t/T)} = -\frac{dm}{m d\bar{q}} = -\frac{1}{2} \frac{dM^2(\bar{q})/d\bar{q}}{M^2(\bar{q})} \\ &= \frac{\sum_{k=1}^{n-1} \Gamma_k \lambda_k^3 e^{-2\lambda_k \bar{q}}}{\sum_{k=1}^{n-1} \Gamma_k \lambda_k^2 e^{-2\lambda_k \bar{q}}} \equiv \bar{\lambda}(\bar{q}), \end{aligned}$$

which is a weighted average of λ_k 's with the weight being $\Gamma_k \lambda_k^2 e^{-2\lambda_k \bar{q}}$. For a Gaussian updating scheme, we have $\bar{\lambda}(\bar{q})$

FIG. 5. (a) Geometric construction of the optimal schedule, $\alpha(t)$, from the normalized mass distribution, $m(\bar{q})$. (b) Inverse construction of $m(\bar{q})$ from the optimal schedule, $\alpha(t)$.

TABLE II. Examples of the variable transformation between the reduced schedule, $\alpha^*(t^*)$, and the reduced mass distribution, $m^*(\bar{q}^*)$, as defined in Eqs. (B2) and (B3). Here, $\gamma > 0$ and we have defined $s_\gamma \equiv 1/(e^\gamma - 1)$, $t_0^{*-1} \equiv (1 + q_0^{*-1})^a - 1$, and $c \equiv q_0^{*a}(1 + t_0^*)$.

$m^*(\bar{q}^*)$	$\alpha^*(t^*)$	$\bar{q}^*(t^*)$
$(1 + s_\gamma)\gamma e^{-\gamma \bar{q}^*}$	$\frac{1}{\gamma(t^* + s_\gamma)}$	$\frac{1}{\gamma} \ln\left(\frac{1 + s_\gamma}{t^* + s_\gamma}\right)$
$\frac{1}{\gamma(\bar{q}^* + s_\gamma)}$	$(1 + s_\gamma)\gamma e^{-\gamma t^*}$	$(1 + s_\gamma)e^{-\gamma t^*} - s_\gamma$
$\frac{c a}{(\bar{q}^* + q_0^*)^{a+1}}$	$\frac{c^{\frac{1}{a}}/a}{(t^* + t_0^*)^{\frac{1}{a}+1}}$	$q_0^* \left[\left(\frac{1 + t_0^*}{t^* + t_0^*} \right)^{\frac{1}{a}} - 1 \right]$

< 1 for $\bar{q} > 0$ by Eq. (53). Thus the inverse updating magnitude increases more slowly in the Gaussian updating scheme than in the single-bin updating scheme, resulting in a larger updating magnitude in the former case.

The schedule, $\alpha(t)$, also defines a variable transformation from the mass distribution, $m(\bar{q})$. If we define $\bar{q}^* \equiv \bar{q}/q(T)$ and $t^* \equiv t/T$ such that both are limited to the interval $[0, 1]$, as well as

$$m^* \equiv q(T) m(\bar{q}) = -\frac{dt^*}{d\bar{q}^*},$$

$$\alpha^* \equiv \frac{T}{q(T)} \alpha(t) = -\frac{d\bar{q}^*}{dt^*},$$

then it is clear that both $m^*(\bar{q}^*)$ and $\alpha^*(t^*)$ can describe the differential relation between $\bar{q}^*$ and t^* , but with the natural variables being $\bar{q}^*$ and t^* , respectively. Further, the transformation is an involution: if we apply the same transform on $\alpha^*(t^*)$, we would recover $m^*(\bar{q}^*)$, as shown in Fig. 5(b). A few examples are shown in Table II. For example, in the single-bin updating scheme, the mass distribution decays exponentially and we would have the inverse-time schedule. On the other hand, the eigenvalues of the Gaussian updating scheme can differ greatly in magnitude [cf. Eq. (53)]. Then, at a large $\bar{q}$, the sum in Eq. (44) will be dominated by the largest term, where $\lambda_k \approx 1/\bar{q}$, and $M(\bar{q}) \propto \sqrt{\Gamma_k}/\bar{q}$. In a local sampling process, Γ_k often rapidly tends to a constant with increasing k . Thus, we have approximately $m(\bar{q}) \propto M(\bar{q}) \propto 1/\bar{q}$, which gives a roughly exponentially decaying schedule.

APPENDIX C: INTEGRALS OF THE AUTOCORRELATION FUNCTIONS OF THE EIGENMODES

Here we give two methods of estimating the integrals of the autocorrelation functions, Γ_k 's, in an adaptive FDS simulation under a constant updating magnitude. They have the advantage of not requiring explicit computation of the autocorrelation functions of the eigenmodes.

First, because the difference between u_i and v_i is a constant of time [cf. Eq. (7)], $\tilde{u}_k$ and $\tilde{v}_k$ share the same variance. Then, from Eqs. (49) and (50), we have

$$\text{var } \tilde{u}_k = \text{var } \tilde{v}_k \approx \frac{1}{2} \Gamma_k \lambda_k a_0, \quad (\text{C1})$$

for a sufficiently long period, T_0 . We can compute the variance, $\text{var } \tilde{u}_k$, by trajectory averaging in a long simulation, and use

Eq. (C1) to estimate Γ_k . However, this method would fail for a near zero, λ_k , and it is best used for the single-bin updating scheme, where $\lambda_k \equiv 1$.

To use the method in a simulation under a multiple-bin updating scheme, we can construct a virtual bias potential, u_i^{vir} , that emulates the bias potential in the single-bin case as (cf. Appendix A 2)

$$u_i^{\text{vir}}(t+1) = u_i^{\text{vir}}(t) + a_0 [\hat{u}_i(t) - u_i^{\text{vir}}(t)],$$

where $\hat{u}_i(t)$ is given by Eq. (A5). We can then apply Eq. (C1) to u_i^{vir} with the effective $\lambda_k \equiv 1$.

The above method requires a sufficiently small updating magnitude, $a_0 \ll 2/(\max_k \Gamma_k)$, and it tends to overestimate Γ_k (possibly because the sampling often deviates from the presumed quasi-equilibrium limit).

Alternatively, we can estimate Γ_k using Eq. (E2). By breaking the trajectory into blocks of T_{blk} steps, we can compute the histogram fluctuation, $\tilde{F}_k(T_{\text{blk}})$, and estimate Γ_k from the average of $T_{\text{blk}} \tilde{F}_k^2(T_{\text{blk}})$ over the blocks. The trajectory block should be long enough to cover at least a few autocorrelation times (which is roughly $\Gamma_k/2$), but not too long when dynamic updates to the bias potential exist so that the histogram flattening is not largely caused by the frequent non-equilibrium updates. Thus, we require $\max_k \Gamma_k/2 \ll T_{\text{blk}} \ll 1/a_0$.

APPENDIX D: HOMOGENEOUS UPDATING SCHEMES

Here, we discuss some mathematical details on the homogeneous updating schemes. We can characterize the updating matrix, w_{ij} , of a homogeneous updating scheme by a rigid window function that shifts with the current bin j (ignoring the boundary adjustment). The unadjusted window function is defined by the relative updating magnitude, μ_l , on the l th bin from the central bin. By the translational invariance, the updating matrices of homogeneous updating schemes share the same set of eigenvectors, and the eigenvalues are related to the window function as a cosine transform.

1. Eigenvalues and window function

For a periodic variable,²⁹ the element of the updating matrix, w_{ij} , can only be a function of the difference of indices, $i - j$, modulo n , and we may write

$$w_{ij} = \mu_{i-j} + \mu_{i-j+n} + \mu_{i-j-n},$$

for some numbers, μ_l 's, which define the unadjusted window function. For simplicity, we will assume that the μ_l 's are symmetric, $\mu_l = \mu_{-l}$, and can be cutoff at $l = b$ (i.e., $\mu_l = 0$ for $l > b$), where b is the integral part of $(n-1)/2$. To satisfy Eq. (26), we further impose the normalization, $\sum_{l=-b}^b \mu_l = 1$.

To find the eigenvectors, we define a periodic variable, ϕ_i , with the out-of-boundary values given by $\phi_i = \phi_{i \pm n}$, such that $i \pm n$ lies in between 1 and n . Then, the multiplication of the matrix, $\mathbf{w}$, is equivalent to a convolution with the window function,

$$\sum_{j=1}^n w_{ij} \phi_j = \sum_{j=1-n}^{2n} \mu_{i-j} \phi_j = \sum_{l=-b}^b \mu_l \phi_{i-l}.$$

The characteristic equation, $\mathbf{w} \phi = \lambda \phi$, can then be solved by discrete Fourier transform, with the orthogonal eigenvectors,

$\phi^{(1)}, \dots, \phi^{(n)}$, given by

$$\phi_i^{(k)} = \phi_{ki} = \frac{\sqrt{2 - \delta_{k,0}}}{n} \cos\left(\frac{k i 2 \pi}{n}\right),$$

where the function takes the cosine form for $k \leq n/2$ or the sine form otherwise. The eigenvalues are given by

$$\lambda_k = \mu_0 + 2 \sum_{l=1}^b \mu_l \cos\left(\frac{k l 2 \pi}{g n}\right), \quad (\text{D1})$$

with $g = 1$ here. Note also the two-fold degeneracy, $\lambda_{n-k} = \lambda_k$.

For a non-periodic variable, we will use the reflective boundary condition,⁴⁰ which sets up two reflective mirrors at $i = 1/2$ and $i = n + 1/2$. Then the variable and the images in the two mirrors can be mapped to an equivalent periodic variable of period $2n$. Thus, we can borrow the eigenvalues from Eq. (D1) with a doubled period, i.e., $g = 2$, and an increased cutoff $b = n - 1$ (so $\mu_n = 0$). The eigenvectors sharing the same periodicity and being even about $i = 1/2$ are given by

$$\phi_i^{(k)} = \phi_{ki} = \frac{\sqrt{2 - \delta_{k,0}}}{n} \cos\left[\frac{k \left(i - \frac{1}{2}\right) 2 \pi}{2 n}\right]. \quad (\text{D2})$$

The updating matrix can be reconstructed by Eq. (54),

$$w_{ij} = \mu_{i-j} + \mu_{i+j-1} + \mu_{i+j-2n-1}.$$

We can recover the window function from the eigenvalues by inversely transforming Eq. (D1),

$$\mu_i = \frac{1}{g n} \sum_{k=0}^{g n-1} \lambda_k \cos\left(\frac{k i 2 \pi}{g n}\right). \quad (\text{D3})$$

In the non-periodic case, we have defined $\lambda_k \equiv \lambda_{2n-k}$ for $k = n + 1, \dots, 2n - 1$, as well as

$$\lambda_n = (-1)^{n-1} \lambda_0 + 2 \sum_{k=1}^{n-1} (-1)^{n-k-1} \lambda_k, \quad (\text{D4})$$

to satisfy the constraint, $\mu_n = 0$.

2. Gaussian updating scheme

The Gaussian updating scheme (cf. in Sec. II C 6) is associated with a Gaussian-shaped window function, μ_l . Below we show its convergence in the continuous limit and how to maintain the convergence on a finite grid.

We recall that the bin index, i , can be related to the continuous variable z as $z = z_0 + i \delta z$, where δz is the bin size. We further define $\mu(z) \equiv \mu_i / \delta z$ and $\Delta z \equiv (n \delta z) g / 2$. For $n \gg 1$, we can approximate Eq. (D1) as an integral,

$$\lambda_k = \int_{-\Delta z}^{\Delta z} \mu(z) \cos\left(\frac{\pi k z}{\Delta z}\right) dz, \quad (\text{D5})$$

with the normalization, $\int_{-\Delta z}^{\Delta z} \mu(z) dz = 1$.

If the width of the Gaussian, σ_z , is much less than Δz , then we can extend the limits of the integrals to infinity and normalize $\mu(z)$ as

$$\mu(z) \approx \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left(-\frac{z^2}{2\sigma_z^2}\right).$$

Then, from Eq. (D5), we get Eq. (53), which shows that all eigenvalues are positive. However, when discretized on a finite

grid, some eigenvalues can turn negative. To keep the eigenvalues positive, we can redefine the Gaussian updating scheme from the eigenvalues given by Eq. (53), and then inversely compute the window function using Eq. (D3) or the updating matrix using Eq. (54).

3. Homogeneous bandpass updating schemes

In a homogeneous bandpass updating scheme, the eigenvalues are either 0 or 1. We can use Eq. (D3) to find the window function. For a periodic variable, we will take into account the two-fold degeneracy and modify Eq. (55) by setting $\lambda_{n-K+1}, \dots, \lambda_{n-1}$ also to 1. Then, $\mu_l = \text{sincr}(K, l, n)$ for $l \leq b$ and $w_{ij} = \text{sincr}(K, i - j, n)$, where we have defined

$$\text{sincr}(K, l, n) \equiv \begin{cases} \frac{\sin\left(\frac{(2K-1)l\pi}{n}\right)}{n \sin\frac{l\pi}{n}} & \text{for } l \neq 0, \\ \frac{2K-1}{n} & \text{for } l = 0. \end{cases}$$

For a non-periodic variable, we have $\lambda_n = (-1)^{n-K}$ from Eqs. (55) and (D4), and $\mu_l = (-1)^{n-K+l} / (2n) + \text{sincr}(K, l, 2n)$ for $l \leq b$. The updating matrix is given by $w_{ij} = \text{sincr}(K, i - j, 2n) + \text{sincr}(K, i + j - 1, 2n)$.

APPENDIX E: HISTOGRAM FLUCTUATION

Here we first relate the error in a long FDS simulation under the single-bin updating scheme to a measure of histogram fluctuation, which can be used in the WL criterion of stage transitions. We then modify the measure to include intra-umbrella components for the polynomial bandpass updating schemes.

Consider an FDS simulation under a nearly exact bias potential. By Eq. (12), the fluctuation of the instantaneous histogram is dominated by the random part, $f_{*i}(t) \approx \xi_{*i}(t)$, and

$$\tilde{f}_k(t) \approx \tilde{\xi}_k(t). \quad (\text{E1})$$

If we define the time average,

$$\tilde{F}_k(T) \equiv \frac{1}{T} \sum_{t=1}^T \tilde{f}_k(t) = \mathcal{F} \left[\frac{H_i(T)}{\rho_i} \right]_k,$$

for the histogram fluctuation of mode k , the total fluctuation is given by

$$F^2(T) \equiv \sum_{k=1}^{n-1} \tilde{F}_k^2(T) = \sum_{i=1}^n \rho_i F_{*i}^2(T) = \sum_{i=1}^n \frac{[H_i(T) - \rho_i]^2}{\rho_i},$$

where we have used Eq. (35) in the second step. From Eq. (E1), we have

$$\begin{aligned} \langle \tilde{F}_k^2(T) \rangle &\approx \frac{1}{T} \sum_{t=0}^{\infty} \langle \tilde{f}_k(t) \tilde{f}_k(0) \rangle \\ &\approx \frac{1}{T} \sum_{t=-\infty}^{\infty} \langle \tilde{\xi}_k(t) \tilde{\xi}_k(0) \rangle = \frac{\Gamma_k}{T}, \end{aligned} \quad (\text{E2})$$

and the total histogram fluctuation is

$$\langle F^2(T) \rangle = \sum_{k=1}^{n-1} \langle \tilde{F}_k^2(T) \rangle = \sum_{k=1}^{n-1} \frac{\Gamma_k}{T} = \frac{\Gamma}{T},$$

where we have used Eq. (43) in the last step. We may relate the histogram fluctuation to the minimal error under the single-bin updating scheme, Eq. (24), as $E(T) \approx \langle F^2(T) \rangle$.

As an estimate of the error, the above total histogram fluctuation can be used in the WL criterion for stage transitions. However, for a bandpass updating scheme, we should restrict the sum to the first K modes, where the histogram flattening applies, i.e., $F^2(T) \equiv \sum_{k=1}^{K-1} \tilde{F}_k^2(T)$.

In particular, in sampling of the bivariate distribution, $\rho(s, z)$, in Sec. II E 4, the above definition becomes

$$\begin{aligned} F^2 &= \sum_{\nu=0}^{S-1} \sum_{k=0}^{K-1'} \tilde{F}_{\nu,k}^2 \\ &= \sum_{s=1}^S \left[\frac{(H_s - \rho_s)^2}{\rho_s} + \sum_{k=1}^{K-1} \frac{H_s^2}{\rho_s} |\tilde{F}_k^{(s)}|^2 \right], \quad (\text{E3}) \end{aligned}$$

where ν and k are the indices of inter- and intra-umbrella fluctuation modes and the primed sum excludes the term of $\nu = k = 0$; on the next line, H_s is the normalized inter-umbrella histogram, and $\tilde{F}_k^{(s)}$ is the histogram fluctuation of the k th mode in umbrella s , which is the average of $R_k^{(s)}(z(t))$ for the data points in umbrella s .

¹D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Computational Science, 2nd ed. (Academic Press, San Diego, 2001).

²E. Newman and G. Barkema, *Monte Carlo Methods in Statistical Physics*, 1st ed. (Clarendon Press, Oxford, 1999); D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, 4th ed. (Cambridge University Press, Cambridge, 2015).

³M. Karplus and J. A. McCammon, *Nat. Struct. Mol. Biol.* **9**, 646 (2002).

⁴G. M. Torrie and J. P. Valleau, *Chem. Phys. Lett.* **28**, 578 (1974); *J. Comput. Phys.* **23**, 187 (1977).

⁵M. Mezei, *J. Comput. Phys.* **68**, 237 (1987).

⁶B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992); J. Lee, *ibid.* **71**, 211 (1993).

⁷F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).

⁸F. Wang and D. P. Landau, *Phys. Rev. E* **64**, 056101 (2001).

⁹T. Huber, A. E. Torda, and W. F. van Gunsteren, *J. Comput.-Aided Mol. Des.* **8**, 695 (1994); A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562 (2002); A. Laio and F. L. Gervasio, *Rep. Prog. Phys.* **71**, 126601 (2008); A. Barducci, M. Bonomi, and M. Parrinello, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **1**, 826 (2011); L. Sutto, S. Marsili, and F. L. Gervasio, *ibid.* **2**, 771 (2012).

¹⁰C. Micheletti, A. Laio, and M. Parrinello, *Phys. Rev. Lett.* **92**, 170601 (2004).

¹¹Q. Yan, T. S. Jain, and J. J. de Pablo, *Phys. Rev. Lett.* **92**, 235701 (2004).

¹²J. Kim, J. E. Straub, and T. Keyes, *Phys. Rev. Lett.* **97**, 050601 (2006); *J. Chem. Phys.* **126**, 135101 (2007).

¹³J. Kim, T. Keyes, and J. E. Straub, *J. Chem. Phys.* **132**, 224107 (2010).

¹⁴C. Junghans, D. Perez, and T. Vogel, *J. Chem. Theory Comput.* **10**, 1843 (2014).

¹⁵K. Langfeld, B. Lucini, and A. Rago, *Phys. Rev. Lett.* **109**, 111601 (2012).

¹⁶R. Pellegrini, K. Langfeld, B. Lucini, and A. Rago, *PoS LATTICE2014*, 229 (2014).

¹⁷L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.* **426**, 168 (2006); J. B. Abrams and M. E. Tuckerman, *J. Phys. Chem. B* **112**, 15742 (2008).

¹⁸L. Zheng, M. Chen, and W. Yang, *J. Chem. Phys.* **130**, 234105 (2009).

¹⁹C. Zhou and R. N. Bhatt, *Phys. Rev. E* **72**, 025701 (2005).

²⁰A. N. Morozov and S. H. Lin, *Phys. Rev. E* **76**, 026701 (2007).

²¹C. Zhou and J. Su, *Phys. Rev. E* **78**, 046705 (2008).

²²R. E. Belardinelli and V. D. Pereyra, *Phys. Rev. E* **75**, 046701 (2007); *J. Chem. Phys.* **127**, 184105 (2007); R. E. Belardinelli, S. Manzi, and V. D. Pereyra, *Phys. Rev. E* **78**, 067701 (2008); R. E. Belardinelli and V. D. Pereyra, *ibid.* **93**, 053306 (2016).

²³F. Liang, C. Liu, and R. J. Carroll, *J. Am. Stat. Assoc.* **102**, 305 (2007).

²⁴A. N. Morozov and S. H. Lin, *J. Chem. Phys.* **130**, 074903 (2009).

²⁵Y. Komura and Y. Okabe, *Phys. Rev. E* **85**, 010102 (2012); A. A. Caparica and A. G. Cunha-Netto, *ibid.* **85**, 046702 (2012); A. A. Caparica, *ibid.* **89**, 043301 (2014).

²⁶D. Min, Y. Liu, I. Carbone, and W. Yang, *J. Chem. Phys.* **126**, 194104 (2007).

²⁷A. Barducci, G. Bussi, and M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).

²⁸B. M. Dickson, *Phys. Rev. E* **84**, 037701 (2011).

²⁹J. F. Dama, M. Parrinello, and G. A. Voth, *Phys. Rev. Lett.* **112**, 240602 (2014).

³⁰P. Poulain, F. Calvo, R. Antoine, M. Broyer, and P. Dugourd, *Phys. Rev. E* **73**, 056704 (2006).

³¹T. Neuhaus and J. S. Hager, *Phys. Rev. E* **74**, 036702 (2006); T. Neuhaus, M. P. Magiera, and U. H. E. Hansmann, *ibid.* **76**, 045701 (2007).

³²F. Zhu and G. Hummer, *J. Comput. Chem.* **33**, 453 (2012).

³³P. Dayal, S. Trebst, S. Wessel, D. Würtz, M. Troyer, S. Sabhapandit, and S. N. Coppersmith, *Phys. Rev. Lett.* **92**, 097201 (2004); S. Trebst, D. A. Huse, and M. Troyer, *Phys. Rev. E* **70**, 046701 (2004).

³⁴S. Singh, C.-c. Chiu, and J. J. de Pablo, *J. Stat. Phys.* **145**, 932 (2011).

³⁵E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992); A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).

³⁶Y. Crespo, F. Marinelli, F. Pietrucci, and A. Laio, *Phys. Rev. E* **81**, 055701 (2010); Y. F. Atchadé and J. S. Liu, *Stat. Sin.* **20**, 209 (2010), see <http://www3.stat.sinica.edu.tw/statistica/j20n1/J20N16/J20N16.html>; G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre, and G. Stoltz, *Math. Comput.* **84**, 2297 (2015).

³⁷H. Robbins and S. Monro, *Ann. Math. Stat.* **22**, 400 (1951).

³⁸S. Marsili, A. Barducci, R. Chelli, P. Procacci, and V. Schettino, *J. Phys. Chem. B* **110**, 14011 (2006).

³⁹In switching to the continuous-time setup, we find it convenient to shift the origin of time by -1 , e.g., the sum $\sum_{i=1}^T$ is mapped to the integral $\int_0^T dt$.

⁴⁰G. Bussi, A. Laio, and M. Parrinello, *Phys. Rev. Lett.* **96**, 090601 (2006).

⁴¹Although in general the growth rate is not necessarily positive, it is often the largest, hence positive, eigenvalue for “natural” updating matrices considered in this study.

⁴²N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland Personal Library, 3rd ed. (Elsevier Science, 2007).

⁴³If the error weighting distribution differs from ρ_i in Eq. (9), the error would contain quadratic cross terms among different eigenmodes. While the solution for the optimal schedule retains the same form, the mass function defined in Eq. (44) will include the corresponding cross terms.

⁴⁴H. Goldstein, J. L. Safko, and C. P. Poole, *Classical Mechanics*, 3rd ed. (Pearson, Harlow, Essex, 2001), p. 680; L. D. Landau and E. M. Lifshitz, *Mechanics*, Course of Theoretical Physics, 3rd ed. (Elsevier Butterworth-Heinemann, Oxford, 1976); V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed. (Springer, New York, 1997); J. V. José and E. J. Saletan, *Classical Dynamics: A Contemporary Approach*, 1st ed. (Cambridge University Press, Cambridge, 1998).

⁴⁵The value of $\tilde{u}_k$ can be estimated from the histogram-corrected bias potential by Eq. (A6) although the approach often overestimates the ϵ_k 's for large- k modes.

⁴⁶We used a variant of the Newton-Raphson method [see, e.g., W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, 2007)] with the bisection method as a fallback. We solved the equation for $\ln q(T)$ to avoid negative values of $q(T)$ and imposed a lower bound, $|\Delta(q(T))/q(T)|$, for $\Delta'(q(T))$ in computing its change. The initial value was given by $q(T) = \ln(1 + \frac{1}{2} T a_0)$.

⁴⁷We used the trapezoidal rule for the integration. The grid points were $q_{[i]} = (q(T) + 1) - (q(T) + 1)^{1-i/N_g}$, where $i = 0, \dots, N_g$ with $N_g = 20\,000$. The grid points were designed for long simulations under the Gaussian updating scheme: for a mass function, $M(\tilde{q})$, that is roughly proportional to $1/\tilde{q}$ (cf. Appendix B), this grid makes the integral, $t_{[i+1]} - t_{[i]}$, approximately a constant. The resulting values, $(q_{[i]}, t_{[i]})$, were used to compute the updating magnitude as (cf. Appendix A 3) $\alpha_{[i]} = \frac{1}{\lambda_1} \left[1 - \exp\left(-\lambda_1 \frac{q_{[i+1]} - q_{[i]}}{t_{[i+1]} - t_{[i]}}\right) \right]$, with $q_{[-1]} \equiv t_{[-1]} \equiv 0$, $q_{[N_g+1]} \equiv q(T)$, and $t_{[N_g+1]} \equiv T$. The values of $(t_{[i]}, \alpha_{[i]})$ were linearly interpolated for the updating magnitude, $\alpha(t)$, at any t .

⁴⁸This inequality follows from the non-negativity of the quadratic polynomial of y , $\int_0^T dt \sum_{k=1}^{N-1} \Gamma_k (y - \theta_k)^2 \equiv A_2 y^2 + A_1 y + A_0$, which implies that the discriminant is nonpositive, $A_1^2 - 4A_0 A_2 \leq 0$.

- ⁴⁹The orthogonal polynomials share the same forms for different target distributions, but the parameters, $\bar{z}$, σ_z , κ_z , etc., are determined by the target distribution, $\rho(z)$.
- ⁵⁰R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986); C. J. Geyer, in *Computing science and statistics: Proceedings of the 23rd Symposium on the Interface* (American Statistical Association, New York, 1991), pp. 156–163; K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996); U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997); Y. Sugita and Y. Okamoto, *ibid.* **314**, 141 (1999); D. J. Earl and M. W. Deem, *Phys. Chem. Chem. Phys.* **7**, 3910 (2005); D. M. Zuckerman, *Annu. Rev. Biophys.* **40**, 41 (2011); S. Rauscher, C. Neale, and R. Pomès, *J. Chem. Theory Comput.* **5**, 2640 (2009).
- ⁵¹H. Li, M. Fajer, and W. Yang, *J. Chem. Phys.* **126**, 024106 (2007).
- ⁵²S. Park and V. S. Pande, *Phys. Rev. E* **76**, 016703 (2007); P. H. Nguyen, Y. Okamoto, and P. Derreumaux, *J. Chem. Phys.* **138**, 061102 (2013); T. Zhang, P. H. Nguyen, J. Nasica-Labouze, Y. Mu, and P. Derreumaux, *J. Phys. Chem. B* **119**, 6941 (2015).
- ⁵³M. R. Shirts, e-print [arXiv:1704.00891](https://arxiv.org/abs/1704.00891) [cond-mat], 2017.
- ⁵⁴G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 5th ed. (Academic Press, San Diego, 2001).
- ⁵⁵J. H. Hetherington, *J. Low Temp. Phys.* **66**, 145 (1987); M. S. S. Challa and J. H. Hetherington, *Phys. Rev. Lett.* **60**, 77 (1988); M. Costeniuc, R. S. Ellis, H. Touchette, and B. Turkington, *Phys. Rev. E* **73**, 026105 (2006).
- ⁵⁶This method should not be confused with the Gaussian updating scheme. In this case, the bias potential, $u^{(s)}(z)$, is quadratic and the conditional intra-umbrella target distribution, $\rho^{(s)}(z)$, is Gaussian although the marginal inter-umbrella distribution, ρ_s , is flat. In the Gaussian updating scheme, the bias potential is a sum of Gaussian functions, and the target distribution is flat.
- ⁵⁷We slightly modified the transition criterion from the WL stages to the inverse-time schedule. We allowed such a switch only upon the completion of a WL stage and interpreted t as the number of updating steps in this just completed WL stage. If the switch was successful, the above t was used as t_0 in the inverse-time regime.
- ⁵⁸The intra-umbrella sampling step and the inter-umbrella transition step can be executed at different frequencies without affecting the correctness of the algorithm. Although a random mixing of the two steps is sometimes preferred, the sequential execution adopted here is also valid, as discussed in V. I. Manousiouthakis and M. W. Deem, *J. Chem. Phys.* **110**, 2753 (1999).
- ⁵⁹A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988); **63**, 1195 (1989); S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, *J. Comput. Chem.* **13**, 1011 (1992); B. Roux, *Comput. Phys. Commun.* **91**, 275 (1995); C. Bartels and M. Karplus, *J. Comput. Chem.* **18**, 1450 (1997); M. Souaille and B. Roux, *Comput. Phys. Commun.* **135**, 40 (2001); M. N. Kobrak, *J. Comput. Chem.* **24**, 1437 (2003); E. Gallicchio, M. Andrec, A. K. Felts, and R. M. Levy, *J. Phys. Chem. B* **109**, 6722 (2005); J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, *J. Chem. Theory Comput.* **3**, 26 (2007); M. K. Fenwick, *J. Chem. Phys.* **129**, 125106 (2008); J. Kim, T. Keyes, and J. E. Straub, *ibid.* **135**, 061103 (2011); M. R. Shirts and J. D. Chodera, *ibid.* **129**, 124105 (2008).
- ⁶⁰J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, Amsterdam, 2006).
- ⁶¹For accuracy, we derived the reference Γ_k 's from a long equilibrium simulation under the reference bias potential using the block histogram method with $T_{\text{blk}} = 10^5$ steps (cf. [Appendix C](#)). We also assumed the reference bias potential in calculating ϵ_k 's.
- ⁶²N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- ⁶³A. Laio, A. Rodriguez-Forte, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, *J. Phys. Chem. B* **109**, 6714 (2005).
- ⁶⁴F. Y. Wu, *Rev. Mod. Phys.* **54**, 235 (1982).
- ⁶⁵For wider umbrellas, one may need to mix the nearest-neighbor inter-umbrella transitions with some longer-range transitions to improve the convergence.
- ⁶⁶U. Wolff, *Phys. Rev. Lett.* **62**, 361 (1989).
- ⁶⁷V. Martin-Mayor, *Phys. Rev. Lett.* **98**, 137207 (2007); S. Davis, *Phys. Rev. E* **84**, 050101 (2011); R. A. X. Persson, *ibid.* **87**, 054101 (2013).
- ⁶⁸We can also derive Eq. (A6) by forming an independent estimator of the unbiased distribution,³⁸ which is, by Eq. (2), $\hat{p}_i(t) \approx h_i(t) e^{u_i(t)}$, assuming a proper shift of $u_i(t)$. Time averaging the equation yields $\rho_i e^{\tilde{U}_i(t)} = \hat{P}_i(t) \approx H_i e^{U_i(t)}$.