Weighted ensemble: Recent mathematical developments

D. Aristoff, J. Copperman, G. Simpson, et al.

View Online   Export Citation   CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Weighted ensemble: Recent mathematical developments

D. Aristoff,[1,a] J. Copperman,[2] G. Simpson,[3] R. J. Webber,[4] and D. M. Zuckerman[2]

AFFILIATIONS

[1] Mathematics, Colorado State University, Fort Collins, CO 80521 USA
[2] Biomedical Engineering, Oregon Health and Science University, Portland, OR 97239 USA
[3] Mathematics, Drexel University, Philadelphia, Pennsylvania 19104 USA
[4] Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California 91125 USA

[a] Author to whom correspondence should be addressed: aristoff@rams.colostate.edu, copperma@ohsu.edu, grs53@drexel.edu, robertjacobswebber@gmail.com, and zuckermd@ohsu.edu

ABSTRACT

Weighted ensemble (WE) is an enhanced sampling method based on periodically replicating and pruning trajectories generated in parallel. WE has grown increasingly popular for computational biochemistry problems due, in part, to improved hardware and accessible software implementations. Algorithmic and analytical improvements have played an important role, and progress has accelerated in recent years. Here, we discuss and elaborate on the WE method from a mathematical perspective, highlighting recent results that enhance the computational efficiency. The mathematical theory reveals a new strategy for optimizing trajectory management that approaches the best possible variance while generalizing to systems of arbitrary dimension.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0110873

I. INTRODUCTION

Weighted ensemble (WE)[1] is an enhanced sampling method employing multiple trajectories of a stochastic process to estimate *mean first passage times* (MFPTs) and related statistics. WE can be applied to any stochastic dynamics model,[2] such as Langevin dynamics in a molecular system[1] or continuous-time jump dynamics in a reaction network.[3] In biomolecular systems, WE has enabled the estimation of MFPTs that are orders of magnitude larger than the combined lengths of the individual WE trajectories.[4–7] WE has also recently been used to elucidate the spike opening dynamics in the SARS CoV-2 virus.[8]

WE is based on *splitting* and *merging*, as shown in Fig. 1. During splitting, "favorable" or "interesting" trajectories, according to a user definition, are replicated. During merging, the "less favorable" or "redundant" trajectories are randomly eliminated from the ensemble. Trajectories are then re-weighted to preserve the statistics of the path ensemble.[2]

The first splitting and merging algorithm was reported in the 1950s and attributed to von Neumann.[9] In 1996, Huber and Kim[1] modified von Neumann's approach by grouping the trajectories into bins and applying splitting and merging in each of the bins

separately, while preserving the bin weights. It was recently shown that WE with fixed bin weights leads to convergent estimates; in contrast, von Neumann's original method can become unstable in the large time limit.[10,11]

Applications of WE have blossomed, and the method is now implemented in the widely used WESTPA software package[12,13] and the more recent wepy[14] and WeightedEnsemble.jl[15] packages. Meanwhile, a growing community of researchers is promoting the method and developing it in different directions.[4,5,16–24]

In modern biochemical applications, WE is often used to estimate the MFPT for a stochastic process to transition from a source state A into a target state B.[25] In addition, WE has been used to estimate other transition path statistics, including the distribution of reaction times from A to B,[4,5] the distribution of entry points into B,[16,26] and the characteristics of paths leading from A to B.[4,17,27] Here, we focus on the estimation of MFPTs, which is a significant and challenging application because the inverse of the MFPT is the reaction rate constant.[28]

Alternative methods for estimating MFPTs include Markov state models,[29] forward flux sampling,[30] adaptive multilevel splitting,[31–34] diffusion Monte Carlo,[35–37] exact milestoning,[38,39] non-equilibrium umbrella sampling,[40] and transition interface
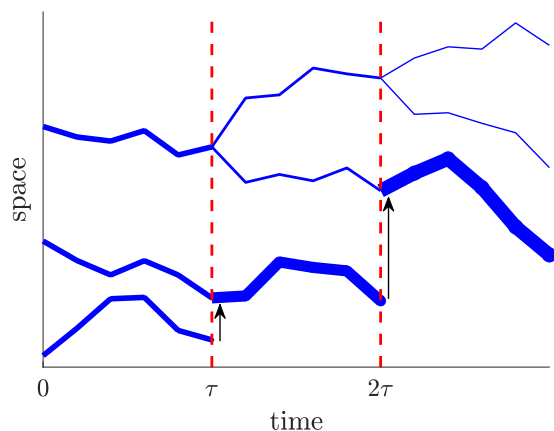
FIG. 1. An illustration of splitting and merging of weighted trajectories. It shows two splitting and merging steps in which the top trajectory is split and the bottom trajectories are merged, with merges indicated by arrows. Merging combines the individual trajectory weights, while splitting divides the weights. The weights are represented by the widths of trajectories.

TABLE I. Definitions of symbols used in this work.

Symbol	Definition
X_t	Underlying Markovian dynamics
x	Position in phase space
τ	Evolution time interval or lag time
$\xi_{s\tau}^i$	Position of the ith WE trajectory at time $s\tau$
$w_{s\tau}^i$	Weight of the ith WE trajectory at time $s\tau$
β	Inverse thermal energy ($1/k_B T$)
U	Potential energy
A	Initial (source) set
B	Target (sink) set
ρ_A	Initial (source) distribution inside A
T_B	First passage time to B
N_t	Number of arrivals in B by time t
q	Committor function from A to B
J	Steady-state flux into B
π	Steady state of source-sink dynamics
$\langle \rangle$	Mean or average for source-sink dynamics
$\langle \rangle_x$	Ensemble average starting at point x
$\langle \rangle_\pi$	Ensemble average starting at distribution π
h	(flux) discrepancy function
v^2	(flux) variance function
$\hat{J}_t$	WE estimate of steady-state flux up to time t
$\tilde{\pi}, \tilde{h}, \tilde{v}$	$\tau \to 0$ limits of π, h, v^2

sampling.[41] Like WE, these approaches use large numbers of short unbiased trajectories to compute the MFPT. However, the following combination of features makes WE especially attractive:

- WE only requires simulation of the stochastic model forward in time, never backward. Thus, WE can be applied to a wide variety of stochastic models arising in chemistry[1,3] and other fields, including astronomy,[42] climate science,[43,44] and systems biology.[3,45]
- WE is fully parallelizable over the time intervals of length τ between splitting/merging events.[12,13,46] Since the trajectories are simulated for the same time interval, the parallelization is straightforward.
- WE provides asymptotically unbiased, convergent estimates in the limit of many time steps for any choice of parameters.[2,10,11]
- WE has been observed to provide more accurate MFPT estimates than Markov state models[47] and similar or better convergence than non-equilibrium umbrella sampling.[48]

Despite progress over the last three decades, theoretical questions about WE have remained unanswered until recently. WE's performance is highly dependent on the choice of parameters,[4] including the definition of the bins and the desired number of trajectories in each bin. For a long time, it was unclear what the optimal parameter choices would be.

Recent work in the mathematical literature sheds new light on the optimal parameter choices for WE.[10,11,49,50] Indeed, the optimal merging coordinate is the local MFPT to B, given the current state. The optimal splitting coordinate is the local variance of the MFPT to B, given the current state and the evolution time interval τ. There is a theoretical limit on the variance reduction achievable through WE, and bins based on the optimal merging and splitting coordinates ensure the optimal variance in the limit of many trajectories and many time steps.

In this article, we aim to communicate these recent mathematical advances in a brief, accessible way. We identify the optimal splitting and merging coordinates in one- and two-dimensional examples. We propose optimized bins based on these coordinates, which differ from more traditional WE bins that employ a user-defined distance to some target structure, e.g., the (minimum) root-mean-squared distance in atomic coordinates. Our examples show that binning based on the distance to a target can encourage transitions along a physically inconsistent pathway and produce catastrophically wrong results. In such cases, more effective bins are needed, and optimized bins can help WE to approach the minimal possible variance. Table I summarizes notation used throughout.

The rest of this article is organized as follows. Section II discusses the computation of mean first passage times, Sec. III introduces the WE method, Sec. IV identifies the optimal merging and splitting coordinates for WE, Sec. V evaluates variance reduction strategies for WE, and Sec. VI concludes.

II. MEAN FIRST PASSAGE TIMES AND THE HILL RELATION

In this work, we study the MFPT of a Markovian stochastic process X_t from a source state A to a sink state B. In the biochemistry context, X_t could represent Langevin dynamics or constant-temperature dynamics generated with a stochastic thermostat. The MFPT could correspond to the characteristic time for the folding, binding, or conformational change of a simulated protein. Throughout, we write x for a particular location in phase space or a particular position of the stochastic process X_t.

To define the MFPT precisely, we must specify the distribution of starting points ρ_A within the source state A. The distribution ρ_A and the dynamics of X_t fully determine the MFPT, which is defined as the averaged length of the trajectories initiated from the source distribution ρ_A and absorbed upon reaching B. We do not address here the issue of choosing the "correct" ρ_A; for details on this, see Refs. 51 and 52.

When computing the MFPT, we "recycle" X_t according to the source distribution ρ_A upon arrival at the sink state B. This means that the position immediately changes from B to A, but the time index t continues to increase as usual. We assume that the distribution of positions for the source–sink dynamics converges as $t \to \infty$ to a unique steady-state distribution, denoted by π.

Under these recycling boundary conditions, the *Hill relation*[53] expresses the MFPT as the inverse of the steady-state flux into B. To state the Hill relation more precisely, write J for the steady-state flux into B, T_B for the first passage time to B, and N_t for the number of arrivals in B by time t. The Hill relation then dictates that the MFPT of X_t from A to B is the inverse of the steady-state flux into B, i.e.,

$$\langle T_B \rangle_A = J^{-1} = \left(\frac{d}{dt} \langle N_t \rangle_\pi \right)^{-1}. \quad (1)$$

Here and elsewhere, all averages are ensemble averages with respect to the source–sink dynamics, and we use subscripts to indicate the starting distribution for the dynamics at time 0. For example, $\langle \cdot \rangle_\pi$ denotes an ensemble average with starting distribution π and $\langle \cdot \rangle_A$ denotes an ensemble average with starting distribution ρ_A. Equation (1) holds for any $t \geq 0$ since π is the steady state.

The application of the Hill relation transforms the computation of a long expected time (the MFPT) into the computation of a small rate (the steady state flux into B). This transformation makes it possible, in principle, to compute the MFPT using trajectories with lengths much shorter than the MFPT. However, while these short
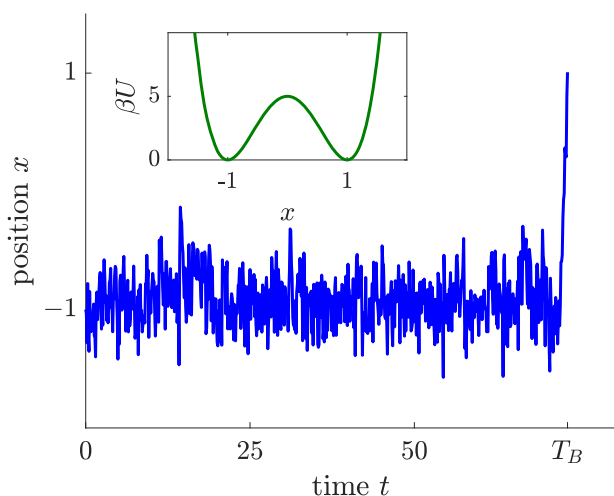

FIG. 2. A first passage event out of a local minimum of U. The first entry into the set $B = \{x > 1\}$ occurs at time T_B after a long waiting period. The dynamics satisfies the Smoluchowski equation [Eq. (2)] with $\beta U(x) = 5(x-1)^2(x+1)^2$ and $D = 1/5$ (inverse time units).

trajectories can be run in parallel to reduce the wall-clock time, a large number of trajectories may be required for estimating small fluxes.

Fortunately, the Hill relation can be combined with enhanced sampling approaches, such as WE, to accurately estimate the small flux into B. A requirement is that the trajectories approximate the steady-state distribution π, but this requirement can frequently be satisfied as the timescale for relaxation to steady state can be much shorter than the MFPT in problems of interest.[54]

Figure 2 shows a simple system for which the steady-state flux from A into B is small. Like all the examples and Figs. 1–10 presented in this paper, the data come from the overdamped Langevin ("Brownian") dynamics,

$$dX_t = [\nabla D(X_t) - \beta D(X_t)\nabla U(X_t)]dt + \sqrt{2D(X_t)}dW_t, \quad (2)$$

associated with the Smoluchowski equation,

$$\partial_t p(x,t) = \nabla \cdot [D(x)(\beta\nabla U(x)p(x,t) + \nabla p(x,t))]. \quad (3)$$

Here, $\beta > 0$ is an inverse temperature constant, $U : \mathbb{R}^n \to \mathbb{R}$ is a scalar potential, $D : \mathbb{R}^n \to \mathbb{R}$ is a diffusion coefficient, W is a standard Brownian motion, and $p(x,t)$ is the probability density at location x at time t. For simplicity, we assume a dimensionless configuration coordinate x, giving the diffusion coefficient D units of inverse time.

For exposition, many of our examples are in one or two spatial dimensions. We emphasize, however, that the WE method and the core mathematical analysis (Secs. IV A, IV B, and V A) apply to any Markovian stochastic dynamics in any spatial dimension.

The continuous-time Smoluchowski equation needs to be discretized for numerical simulation. Therefore, we assume that X_t is evolved as a discrete time series $X_0, X_\tau, X_{2\tau}, \ldots$, and X_t is only recycled when occupying state B at a multiple of the time interval τ. The difference between discrete- and continuous-time flux into B will be small if τ is small or the trajectories in B are slow to escape. We reserve variables t and τ for the "physical time," i.e., the time index for the original continuous-time dynamics X_t.

III. WEIGHTED ENSEMBLE: SPLITTING AND MERGING

Throughout this article, we consider a simple version of WE in which the splitting and merging of trajectories are formulated as *resampling*.[2,10,11,50] For a more general discussion of WE, see Ref. 25. The simple WE algorithm is described below and illustrated in Fig. 3.

1. **Initialization.** Select starting points in phase space $\xi_0^1, \ldots, \xi_0^N$ for N trajectories, and assign weights $w_0^1, \ldots, w_0^N$ to the trajectories with $\sum_{i=1}^N w_0^i = 1$. Here, the subscripts are physical time, and the superscripts are trajectory labels.

2. **Resampling.** At each multiple $s\tau$ of the lag time, partition the trajectories $\xi_{s\tau}^1, \ldots, \xi_{s\tau}^N$ into *bins*, or groupings of trajectories. Select the desired number of trajectories within each bin, called the *allocation*. The bins and allocation can change with time, but we constrain the total number of trajectories to be N. Next, *resample* trajectories within each bin, i.e., sample trajectories $\xi_{s\tau}^i$ with replacement using selection probabilities
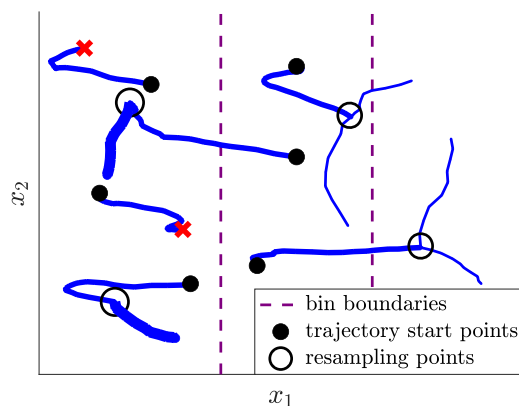
FIG. 3. An example of WE resampling and dynamics. The illustration shows six WE trajectories, which undergo evolution, resampling, and another round of evolution. During evolution, the trajectories travel between bins and the resampling enforces two trajectories in each bin. Circles indicate trajectories that are continued and possibly replicated during resampling; crosses indicate trajectories that are pruned during resampling. Here, trajectory weight is represented by linewidth, and the bins are based on the coordinate x_1.

proportional to the weights $w_{s\tau}^i$. Continue until reaching the desired number of trajectories for each bin.

Last, divide the bin weights evenly among the resampled trajectories within each bin.[10] Each resampled trajectory receives weight equal to the total weight in the bin (before resampling) divided by the number of resampled trajectories.

3. **Dynamics.** Evolve the trajectories independently for time τ according to the underlying dynamics, and label the trajectories $\xi_{(s+1)\tau}^1, \ldots, \xi_{(s+1)\tau}^N$. The trajectories keep the same weights $w_{(s+1)\tau}^1, \ldots, w_{(s+1)\tau}^N$ that were assigned during resampling.

4. **Convergence.** Repeat steps 2–3 as long as desired or possible. Estimate the steady-state flux J as the average WE probability flux into B from a burn-in time t_0 up to a final time t,

$$\hat{J}_{\text{WE}} = \frac{1}{t - t_0} \sum_{s=t_0/\tau}^{t/\tau - 1} \sum_{i=1}^{N} w_{s\tau}^i \mathbb{1}_B(\xi_{s\tau}^i), \tag{4}$$

where $\mathbb{1}_B$ is the characteristic function of B. The estimate $J \approx \hat{J}_{\text{WE}}$ in turn yields a MFPT estimate via the Hill relation (1).

The crucial step in WE is the resampling step, which leads to the splitting and merging of trajectories. Splitting replicates trajectories that are "valuable," while merging combines trajectories that are "similar." Splitting and merging were traditionally applied separately, but in the simple WE algorithm above, these steps are combined via resampling.[2] Resampling is easier to formulate and implement than the traditional splitting and merging. Moreover, resampling is statistically *optimal* when trajectories within a given bin are aside from their weights, indistinguishable.[46]

Splitting and merging must be carefully balanced, and a main contribution of our recent work[10,11,49,50] has been identifying the regions of state space where either splitting or merging should be encouraged. Identifying such regions naturally leads to variance reduction strategies for WE. We discuss these developments in Secs. IV and V.

Given sufficient resources, the WE method terminates when the estimated flux converges to a near-constant value. However, if the steady state cannot be reached during the available simulation time, enhanced WE initialization methods are needed.[55] For example, adjusting the burn-in time t_0 can reduce transient relaxation effects,[56] and adjusting the initial weights can also improve convergence.[57]

WE can be used to estimate other transition path statistics, in addition to the steady-state flux. For instance, the WE-generated paths and the associated weights can be used to estimate the distribution of reaction times (i.e., event durations) and the distribution of mechanistic pathways.[4] Our optimality theory extends to a wide range of transition path statistics. See Appendix 1 for a discussion.

In the subsequent analysis, we compare WE with a brute-force approach involving only dynamics steps with no resampling, which leads to a flux estimator,

$$\hat{J}_{\text{BF}} = \frac{1}{N(t - t_0)} \sum_{s=t_0/\tau}^{t/\tau - 1} \sum_{i=1}^{N} \mathbb{1}_B(\xi_{s\tau}^i). \tag{5}$$

The flux estimators $\hat{J}_{\text{WE}}$ and $\hat{J}_{\text{BF}}$ implicitly depend on the number of trajectories N, the time index t, and the burn-in time t_0.

IV. OPTIMAL SPLITTING AND MERGING COORDINATES

Recent mathematical analysis has revealed the existence of optimal reaction coordinates for merging and splitting in WE.[10,11,50] Sections IV A and IV B provide general formulas for these optimal reaction coordinates, independent of both the dynamics and dimension, while Sec. IV C provides more explicit formulas in the case of Smoluchowski dynamics.

A. Optimal merging and splitting coordinates

Optimal merging. Merging trajectories, i.e., pruning some members of a chosen set, is most beneficial and least harmful when the groups of trajectories to be merged are suitably "similar." For the MFPT problem, the scalar reaction coordinate that characterizes this similarity is the *flux discrepancy function*,[50]

$$h(x) = \lim_{t \to \infty} \left[\langle N_t \rangle_x - \langle N_t \rangle_\pi \right]. \tag{6}$$

Here, N_t counts the number of crossings into B at times $\tau, 2\tau, \ldots, t$, while $\langle \rangle_\pi$ and $\langle \rangle_x$ indicate ensemble averages starting from the distribution π or a particular location x. The flux discrepancy function $h(x)$ is the difference in expected future flux between trajectories started at x and trajectories started from π. Trajectories with similar h values make similar expected contributions to the flux estimate, so merging trajectories with similar h values is appropriate.[10,50]

Using the Hill relation, the flux discrepancy function can be rewritten in terms of MFPTs initiated from two different starting distributions (Appendix 2). The flux discrepancy function satisfies

$$h(x) = \frac{\langle T_B \rangle_\pi - \langle T_B \rangle_x}{\langle T_B \rangle_A}, \tag{7}$$

where we recall that T_B is the first passage time into B, excluding the time $t = 0$. Equation (7) allows us to reinterpret the optimal merging function as a normalized difference between the local MFPT to B

starting from x and the MFPT for trajectories initiated from π. Identity (7) also shows that the flux discrepancy function and the local MFPT are equivalent optimal coordinates for merging.

Optimal splitting. Splitting—the duplication of certain WE trajectories—increases local sampling and is of high benefit in specific regions of state space. Allocating a greater proportion of trajectories to such regions reduces the variance in the flux and MFPT estimates.

We now introduce the scalar reaction coordinate that describes optimal splitting behavior for the MFPT calculation, the *flux variance function*,[50]

$$v(x)^2 = \tau^{-1} \mathrm{Var}_x[h_0(X_\tau)]. \tag{8}$$

Here, $h_0(x) = \mathbb{1}_B(x) + h(x)$ is a modified flux discrepancy function that includes the flux at time $t = 0$, and Var_x denotes the variance for the source–sink dynamics started at x. The flux variance function quantifies the changes in the expected flux over a time interval τ for trajectories initiated at x. It is a local function that can guide WE trajectory allocation.

A heuristic strategy to optimize the MFPT estimate is to balance the variance contributed from each region, so a favorable allocation should satisfy[10,49,50]

$$\# \text{ trajectories near } x \propto \pi(x) v(x). \tag{9}$$

Compared to the steady state π, the proposed WE allocation (9) involves upsampling near x by a factor proportional to $v(x)$ to bring down the flux variance accordingly. Allocating in this way minimizes the contribution to the flux variance, assuming that WE is in the steady state.[10,49,50] We refer to this as the *optimal allocation distribution* and provide a brief sketch of the derivation in Appendix 3.

Optimality theorem. The following theorem shows that bins based on h and v minimize the variance in the WE flux estimate, up to an arbitrarily small tolerance $\varepsilon > 0$.

Theorem (Ref. 11). Assume that the recycled process X_t is geometrically ergodic. Then, the WE method has the following properties:

1. For any choice of bins and allocations, the WE flux estimate $\hat{J}_{\mathrm{WE}}$ given in (4) converges with probability one to the inverse MFPT,

$$\lim_{t \to \infty} \hat{J}_{\mathrm{WE}} = \frac{1}{\langle T_B \rangle_A}. \tag{10}$$

2. For any tolerance $\varepsilon > 0$ and any choice of bins and allocations, the WE variance satisfies

$$\mathrm{Var}\left(\hat{J}_{\mathrm{WE}}\right) \geq \frac{1-\varepsilon}{Nt} \left(\int v(x)\pi(x)dx\right)^2 \tag{11}$$

when t is sufficiently large.

3. For any tolerance $\varepsilon > 0$, if the bins are sufficiently small rectangles in h and v coordinates and each bin allocation is proportional to $\int_{\mathrm{bin}} v(x)\pi(x)dx$, the WE variance satisfies

$$\mathrm{Var}\left(\hat{J}_{\mathrm{WE}}\right) \leq \frac{1+\varepsilon}{Nt} \left(\int v(x)\pi(x)dx\right)^2 \tag{12}$$

for sufficiently large t and N. □

We emphasize that the theorem holds regardless of dimension, temperature, evolution time τ, and type of Markovian dynamics. It highlights the following WE strategy that is optimal in the limit of large N and t: we merge trajectories with similar h and v values and split to generate $\mathrm{Const} \times \pi(x)v(x)$ trajectories near x. The theorem supports the interpretation (9) of $\pi(x)v(x)$ as the optimal distribution for trajectory allocation.

B. Maximum gain of WE over direct dynamics

The above mathematical theory enables a quantitative comparison between WE and direct or "brute force" (BF) sampling of first-passage events using the system's underlying dynamics. The variance for N brute-force trajectories satisfies[11]

$$\mathrm{Var}\left(\hat{J}_{\mathrm{BF}}\right) \sim \frac{1}{Nt} \int v(x)^2 \pi(x)dx \tag{13}$$

in the limit $t \to \infty$, where $\hat{J}_{\mathrm{BF}}$ is defined in (5). In this formula, the flux variance function v^2 quantifies the variance in the flux estimates as trajectories are evolved forward over a time interval τ.

Dividing (13) by (12) and taking the tolerance $\varepsilon \to 0$, the ratio of the brute-force variance to the optimal WE variance, which we call the *gain over brute force sampling*, is then given by[11]

$$\frac{\mathrm{Var}\left(\hat{J}_{\mathrm{BF}}\right)}{\mathrm{Var}\left(\hat{J}_{\mathrm{WE}}\right)} = \frac{\int v^2(x)\pi(x)\,dx}{\left(\int v(x)\pi(x)\,dx\right)^2}. \tag{14}$$

This ratio quantifies the maximal possible variance reduction achievable by WE for any number of trajectories in the large time limit.

The optimal gain over direct, brute-force sampling has important implications. Analytical and numerical investigation of this quantity can yield insights into how much benefit is possible from using WE in a given problem. Additionally, knowing the optimal gain over direct simulation as a reference enables quantitative comparisons of practical WE binning strategies against the theoretically optimal performance.

C. Optimal WE for one-dimensional Smoluchowski dynamics

To make the preceding mathematical theory more explicit, we consider the Smoluchowski (overdamped Langevin) dynamics in a one-dimensional domain with a source state $A = \{x \leq a\}$ and a sink state $B = \{x \geq b\}$, where $a < b$. The recycling distribution is a delta function at the edge of A, i.e., $\rho_A = \delta(x - a)$. We consider the continuous-time limit, $\tau \to 0$, which leads to simpler, more interpretable mathematical expressions.

In the limit $\tau \to 0$, the process X_t is observed at all times and recycling occurs immediately upon entry into B. The steady-state distribution $\tilde{\pi}$ can be calculated using the relation[58]

$$\tilde{\pi}(x) \propto [1 - \tilde{q}(x)]e^{-\beta U(x)}. \tag{15}$$

Here and elsewhere, we use symbols with tildes to express the $\tau \to 0$ limit. $\tilde{q}$ is the *committor* function, the probability for X_t to reach B before A starting from x, which is given by[59]

$$\tilde{q}(x) = \begin{cases} 0, & x \le a, \\ \dfrac{\int_a^x e^{\beta U(y)}/D(y)\, dy}{\int_a^b e^{\beta U(y)}/D(y)\, dy}, & x < a < b, \\ 1, & x \ge b. \end{cases} \quad (16)$$

Using (15) and (16), the steady-state distribution is supported on $\{x < b\}$, and it satisfies

$$\tilde{\pi}(x) \propto e^{-\beta U(x)} \int_{\max\{x,a\}}^b \frac{e^{\beta U(y)}}{D(y)}\, dy. \quad (17)$$

See Fig. 4 for an illustration.

The flux discrepancy function $h = h_\tau$ and the flux variance function $v = v_\tau$ depend implicitly on the evolution interval τ, and they converge to well-defined limits as $\tau \to 0$,

$$\tilde{h}(x) = \lim_{\tau \to 0} h_\tau(x), \qquad \tilde{v}^2(x) = \lim_{\tau \to 0} v_\tau(x)^2. \quad (18)$$

As a result of (7), the discrepancy function $\tilde{h}$ satisfies

$$\tilde{h}(x) = \frac{\langle \tilde{T}_B \rangle_{\tilde{\pi}} - \langle \tilde{T}_B \rangle_x}{\langle \tilde{T}_B \rangle_A}. \quad (19)$$

Here, $\tilde{T}_B$ is the exact MFPT function to the target state B, which is given for $x \le b$ by[59]

$$\langle \tilde{T}_B \rangle_x = \int_x^b \frac{e^{\beta U(z)}}{D(z)} \int_{-\infty}^z e^{-\beta U(y)}\, dy\, dz. \quad (20)$$

As a result of (8), the variance function $\tilde{v}$ satisfies

$$\tilde{v}^2(x) = \lim_{\tau \to 0} \frac{\mathrm{Var}_x[\tilde{h}(X_\tau)]}{\tau} = 2D(x)\left| \frac{d}{dx}\tilde{h}(x) \right|^2, \quad (21)$$

where the second equality comes from applying Itô's lemma[59] for the drift and diffusion of a smooth function of Brownian dynamics. Formula (21) suggests that there should be more splitting in regions of high $\tilde{h}$ variability or, equivalently, high variability in the local MFPT $\langle \tilde{T}_B \rangle_x$.

Note that the committor $\tilde{q}$ is usually understood to be the relevant coordinate for computing the MFPT.[60] We have shown, however, that two different scalar coordinates, namely, $\tilde{h}$ and $\tilde{v}^2$, are the relevant ones for merging and splitting in the WE method. The difference between these coordinates is exhibited in the model problem with an asymmetric two-barrier system in Fig. 5. In this problem, the committor $\tilde{q}$ would not be an ideal reaction coordinate since it poorly resolves the largest barrier on the forward path from A to B. In contrast, $\tilde{h}$ and $\tilde{v}$ resolve the largest forward barrier, making them appropriate for WE.

A simple formula for the optimal WE allocation holds in the low-temperature limit as $\beta \to \infty$. We assume that the largest energy increase on the forward path to B occurs over an interval $[x_-, x_+]$, that is,

$$U(x_+) - U(x_-) = \Delta U = \max_{a \le y \le b,\, x < y} (U(y) - U(x)). \quad (22)$$

Note that the interval $[x_-, x_+]$ can include multiple energy barriers. Then, as $\beta \to \infty$, an application of Laplace's method (see Appendix 4) yields

$$\frac{\tilde{\pi}(x)\tilde{v}(x)}{\int \tilde{\pi}(y)\tilde{v}(y)dy} \to \begin{cases} 0, & x \le x_-, \\ \dfrac{D^{-1/2}(x)}{\int_{x_-}^{x_+} D^{-1/2}(y)dy}, & x_- < x < x_+, \\ 0, & x \ge x_+. \end{cases} \quad (23)$$

The optimal allocation is proportional to $1/\sqrt{D}$ over the interval $[x_-, x_+]$, and it is vanishingly small elsewhere, as illustrated in Fig. 6.
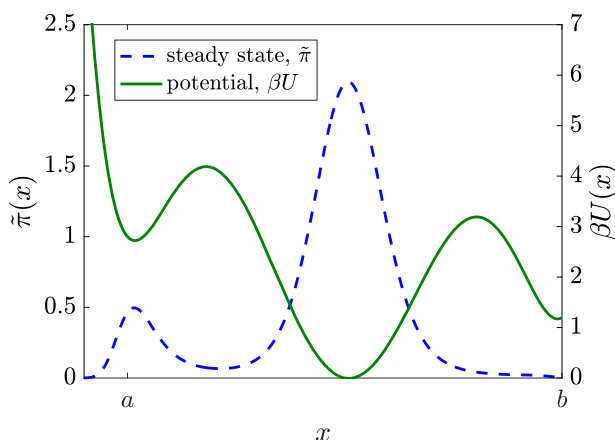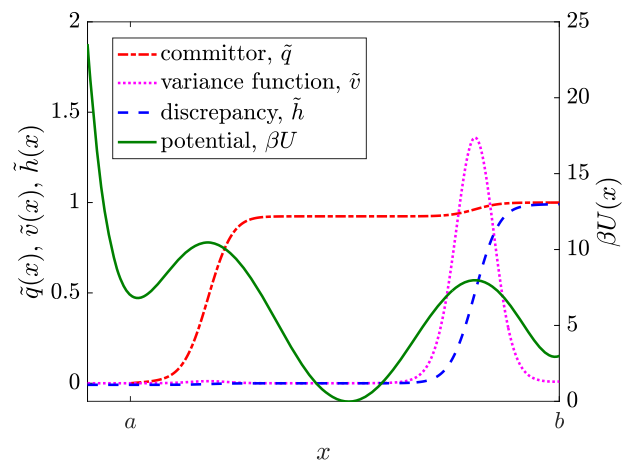

FIG. 4. *The steady state of the source-sink dynamics.* Pictured are the steady state $\tilde{\pi}$ of the Smoluchowski dynamics with recycling from b to a, a diffusion coefficient $D = 1$, and $\beta = 4$. In Figs. 4–7, $U(x) = 5(x-1)^2x^2(x+1)^2 + 0.5x^2 - 0.2x$.

FIG. 5. The ideal reaction coordinates for WE. Pictured are the committor $\tilde{q}$, the variance function $\tilde{v}$, and the discrepancy function $\tilde{h}$ for the Smoluchowski dynamics with recycling from b to a. The discrepancy and variance functions, but not the committor, resolve the largest forward barrier to b. Here, $D = 1$ and $\beta = 10$.
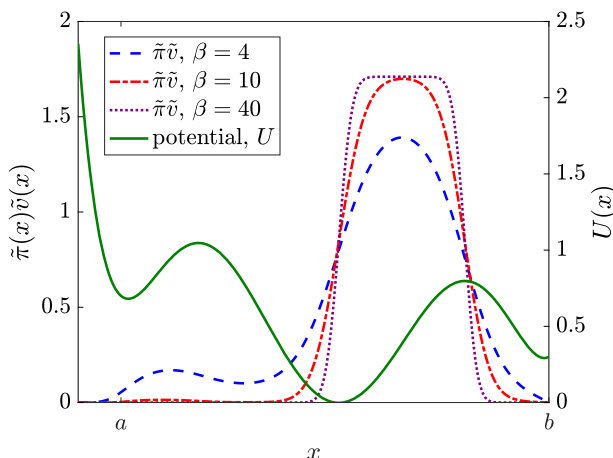
FIG. 6. The optimal allocation distribution at low temperature. Pictured is the (normalized) optimal allocation distribution $\tilde{\pi}\tilde{v}$ for the Smoluchowski dynamics with recycling from b to a and with increasing values of β when $D = 1$. In the limit $\beta \to \infty$, the optimal allocation distribution is constant along the interval of the largest energy increase; see (23).
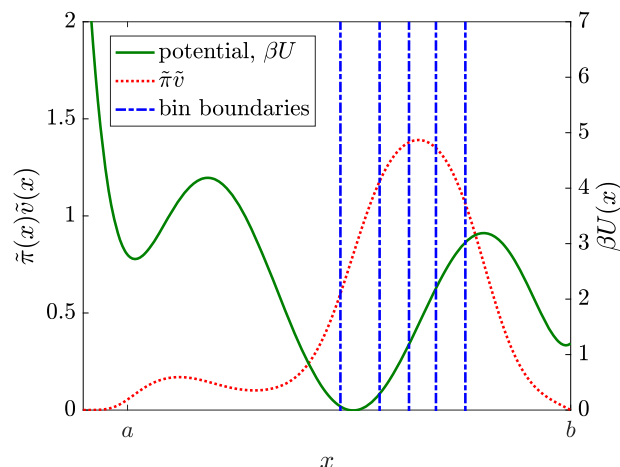

FIG. 7. Illustration of the MFPT binning strategy. Pictured are six bins, each one consisting of equal mass of πv; see (25). The model system is the same one pictured in Fig. 4.

In the low-temperature limit $\beta \to \infty$, the optimal gain over direct sampling grows exponentially with the size of the largest energy increase (see Appendix 4),

Gain over brute force sampling

$$\overset{\beta \to \infty}{\sim} \frac{\pi/\beta}{\left(\int_{x_-}^{x_+}\sqrt{D(x_+)/D(x)}\,dx\right)^2} \frac{e^{\beta\Delta U}}{\sqrt{|U''(x_-)U''(x_+)|}}. \quad (24)$$

This provides a formal explanation of earlier numerical findings[4,61] that WE's advantage over direct simulation grows dramatically as $\beta\Delta U$ increases.

V. AN OPTIMIZED WE STRATEGY

The theoretical results above can guide the optimization of WE. We describe one optimized WE strategy called "MFPT binning" in Sec. V A, and we compare it to a more traditional WE binning strategy in Sec. V B.

A. MFPT binning

From the theory in Sec. IV A, merging has a low cost when applied to trajectories with similar h values. This makes h a natural coordinate for WE binning. We propose the following *MFPT binning* strategy, in which bins are comprised of similar h values or equivalently based on (7), similar values of the local MFPT to B.

1. Define the bins to be intervals in the h coordinate, as follows: If there are K bins, choose the endpoints $\mathfrak{h}_0 < \mathfrak{h}_1 < \ldots < \mathfrak{h}_K$ such that

$$\int_{\mathfrak{h}_k \leq h(x) \leq \mathfrak{h}_{k+1}} \pi(x)v(x)\,dx = \text{constant}. \quad (25)$$

2. Following the optimal allocation rule (9), allocate trajectories uniformly so that approximately N/K trajectories are assigned to each bin.

In the MFPT binning strategy, the bins are intervals in the optimal merging coordinate h so that merging is low cost. The bin boundaries are based on the optimal splitting coordinate v^2 so that each bin makes an equivalent small contribution to the WE variance. We allocate the same number of trajectories to each bin, which is consistent with the optimal allocation rule (9). See Fig. 7 for an illustration.

The MFPT binning strategy requires an approximation to functions h and v^2. As a concrete strategy,[50,57] we can partition the state space of X_t into a discrete set of "microbins." Then, we can estimate the Markov transition matrix between microbins using trajectories evolved over a lag time τ. Conceptually, this is similar to using Markov state models[29,47] to estimate a coarse-grained transition matrix. Based on this transition matrix, we can obtain estimates for the flux discrepancy function h and the flux variance function v^2 in each microbin i using the formulas in Appendix 5.

B. Numerical comparison

In this section, we numerically compare two strategies for binning and allocation within WE:

1. The MFPT binning strategy discussed in Sec. V A.
2. WE bins based on the distance to B, with allocations proportional to $\int_{\text{bin}} v(x)\pi(x)dx$.

The latter strategy, which we call "radial binning," is based on the common WE approach[25] of defining bins based on the distance to B in some feature space. However, our results below show that radial binning can fail when the direct transition from the source set A to the target set B is not the same as the physical transition.
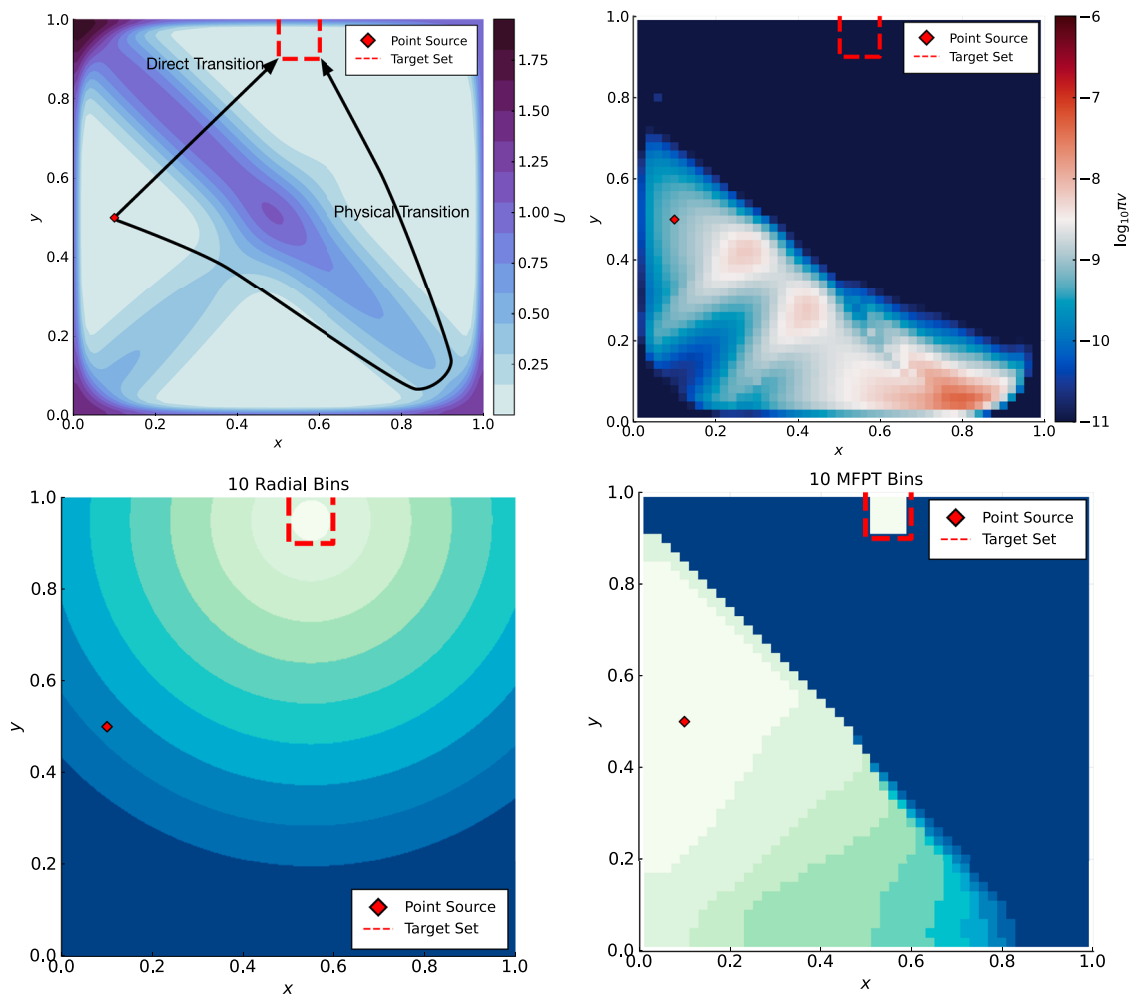
FIG. 8. Different binning strategies in a two-dimensional model. Clockwise from top left: the potential U, the optimal allocation distribution πv from (9), MFPT bins, and radial bins. Different shades indicate different bins in the bottom panels. Radial bins encourage a direct transition from the source set A to the target set B, while MFPT bins encourage the physical transition pictured at top left. The microbins used to define the MFPT bins are the small rectangles visible in the panels on the right.

We apply the MFPT and radial binning strategies to the two-dimensional Smoluchowski model system pictured in Fig. 8. The landscape is given by the expression

$$U(x, y) = U_1(x, y) + U_2(x, y) + 0.5U_3(x, y), \quad (26)$$

where

$$\log U_1(x, y) = -c_1(x - 0.25)^2 - c_1(y - 0.75)^2 \\ - 2c_2(x - 0.25)(y - 0.75), \quad (27)$$

$$\log U_2(x, y) = -c_3 x^2 (1 - x)^2 y^2 (1 - y)^2, \quad (28)$$

$$\log U_3(x, y) = -c_4 x^2 - c_4 y^2 + 2c_5 xy, \quad (29)$$

and the constants are given by

$$(c_1, c_2, c_3, c_4, c_5) = (50.5, 49.5, 10^5, 51, 49). \quad (30)$$

We set the diffusivity to $D = 1$ and the inverse temperature to $\beta = 30$.

To simulate the Smoluchowski dynamics, we apply the Euler–Maruyama integration with an integration step of size 0.001. Each evolution step is composed of ten integration steps so that $\tau = 0.01$. We constrain the trajectories to reside in $[0, 1]^2$. When a trajectory attempts to exit this region, we project it back into the box by setting the position as $x = \max(\min(x, 1), 0)$ and $y = \max(\min(y, 1), 0)$. We perform these computations using the WeightedEnsemble.jl package.[15]

We begin our study of this system by constructing a fine-grained approximation of functions π, h, and v, following the procedures given in Appendix 5. We define microbins using Voronoi cell centers that are spaced $\Delta x = \Delta y = 0.02$ units apart on a regular Cartesian grid. This spacing results in a total of $49 \times 49 = 2401$

microbins, which are the small rectangles visible in Fig. 8. We emphasize that the MFPT bins pictured in the bottom right panel resolve the largest energy barriers on the forward path from A to B, whereas the radial bins pictured in the bottom left have concentric circle bin boundaries that do not account for the energy landscape.

To evaluate the precision—and hence, efficiency—of the flux estimates, we define the normalized variance as

$$\text{Normalized Variance} = Nt \times \text{Var}\big(\hat{J}\big). \qquad (31)$$

Here, the variance of the estimated flux has been scaled to account for total simulation cost Nt. We compare the normalized variance for MFPT and radial binning in Fig. 9. We also show the normalized variance for direct (brute force) sampling without any splitting or killing, and we compare against the lowest possible normalized variance for WE, $\left(\int \pi v\right)^2$, evaluated using Eq. (11) together with the fine-grained model. We use $N = 10^4$ trajectories and average over 10^2 independent runs to produce the normalized variance estimates.

Figure 9 reveals a dramatic difference between MFPT and radial binning. Radial binning leads to worse-quality results than brute-force simulation for this non-trivial model. Meanwhile, MFPT

binning leads to a major improvement over BF simulation, particularly, when the number of bins is large (20–40 bins). With only five bins, we already see a modest gain with the MFPT binning strategy. Increasing the number of bins with MFPT binning systematically improves the output such that with 40 bins, we are within an order of magnitude of the estimated optimal constant. We achieve two and a half orders of magnitude of variance reduction over direct BF simulation.

In both radial and MFPT binning strategies, the allocations enforce the heuristic strategy (9) of having $\propto \pi(x)v(x)$ trajectories near x, which is derived mathematically in Sec. IV. However, analogous numerical experiments (not shown) demonstrate that radial binning with uniform allocation results in substantially higher variance. This comparison suggests that it is preferable to use the optimal allocation rule (9), even given an imperfect choice of WE bins. The MFPT binning method described here, with bins defined by (25), also outperforms the variance reduction strategies introduced in our previous work.[50]

We have generated the data in Fig. 9 using $N = 10^4$ trajectories, which may exceed the computational resources available for problems of interest. Figure 10 represents comparisons for 10^2, 10^3, and 10^4 trajectories, indicating that the normalized variance has no more
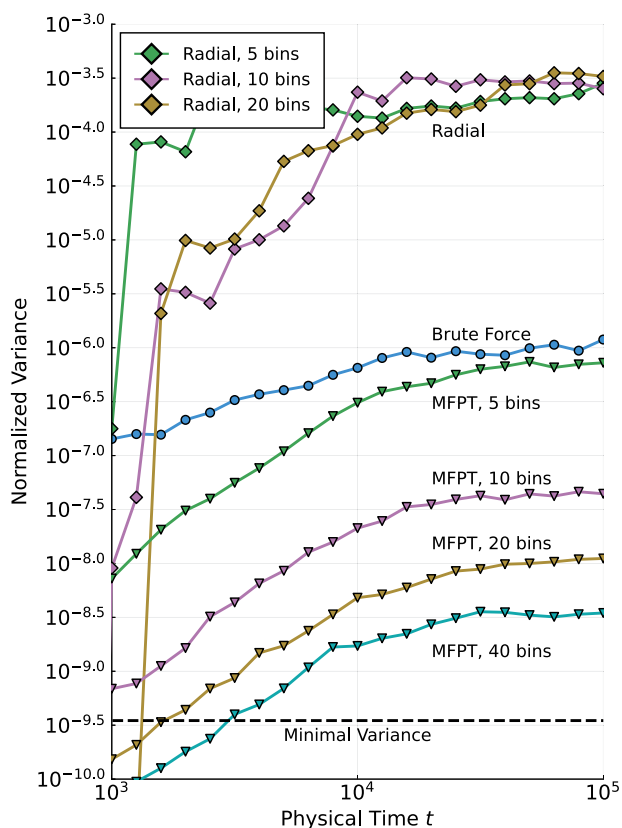
FIG. 9. Comparison of binning strategies for MFPT estimation. It shows estimates of the normalized variance (31) using MFPT and radial binning strategies. The normalized variance is 3–5 orders of magnitude smaller with MFPT binning. For this problem, the flux value is $\sim 10^{-6}$.
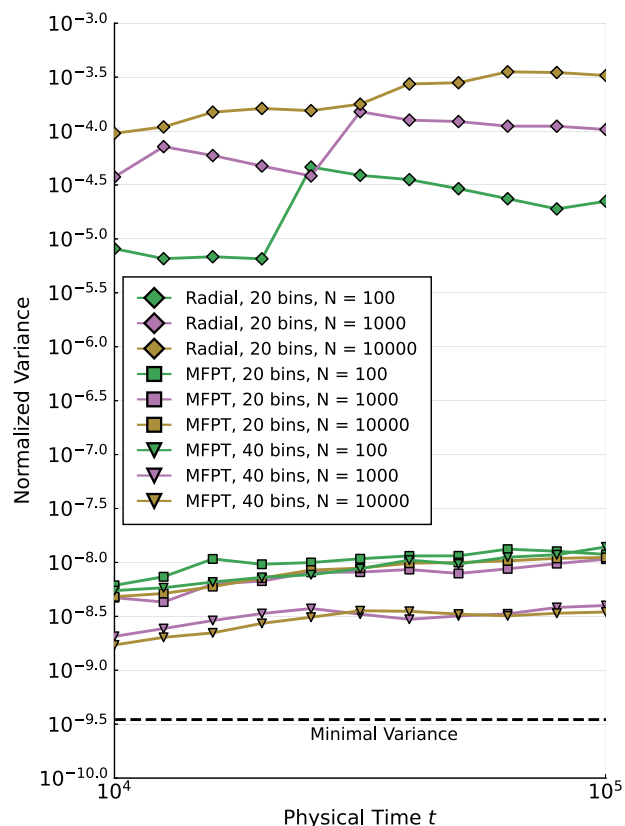
FIG. 10. Normalized variance for different numbers of trajectories. The six curves on the bottom show that the normalized variance of MFPT binning is not highly dependent on the number of WE trajectories.

than an order of magnitude in variability. In particular, the results for MFPT binning with 10^3 trajectories are nearly indistinguishable from those obtained with 10^4 trajectories.

VI. CONCLUSION

This work has attempted to assemble, expand upon, and make accessible to a chemical physics audience recent mathematical results of fundamental importance to weighted ensemble simulation. The mathematical theory exposes the fundamental capabilities of WE and guiding principles behind WE optimization. It leads to the identification of optimal reaction coordinates for merging and splitting, as well as the optimal gain over direct brute force sampling, which can be used to measure the benefit of WE for a given problem.

To illustrate these results in a physically intuitive way, we have derived explicit formulas in the case of the Smoluchowski dynamics (3) in one dimension. These formulas, presented for the first time here, demonstrate that the variance reduction from WE can be exponentially large in the size of the largest forward energy barrier from A to B.

The mathematical theory can guide variance reduction strategies for WE. We have introduced a simple binning strategy called MFPT binning, which approaches the minimal WE variance in a two-dimensional model problem. After pilot runs that estimate h and v (or equivalently, the local MFPT to B and its variance), we define the bins to be intervals in the h coordinate, with endpoints chosen so that each bin contributes an equal share of the variance. In numerical experiments, MFPT binning significantly outperforms a naive WE approach that constructs bins using the radial distance to B, and these results remain robust even for relatively small N.

A number of theoretical and practical challenges remain. A practical priority involves testing the MFPT binning approach in complex systems under limited-data conditions, given the present study's restriction to one- and two-dimensional model problems. We have begun to estimate h and v on the fly in complex systems using space-discretized Markov models built from available data and hope to report on these efforts in the future. On the theoretical side, our analysis has focused on reducing variance in MFPT computations, but the optimization needs to be extended other observables, e.g., correlation functions and pathway distributions. When targeting a single observable, we can optimize using a local discrepancy function and a local variance function, similar to the construction presented here.[11,50] Yet, it remains unclear how to optimize when targeting multiple observables simultaneously. Last, a theoretical and practical goal involves optimizing WE during the transient relaxation to the steady state (rather than within the steady state as was done here), given that many complex systems may not reach the steady state using traditional WE without interventions.[57]

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

D. Aristoff: Conceptualization (lead); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (lead); Project administration (equal); Resources (equal); Supervision (equal); Validation (supporting); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **J. Copperman**: Conceptualization (supporting); Methodology (supporting). **G. Simpson**: Conceptualization (supporting); Data curation (lead); Formal analysis (supporting); Funding acquisition (equal); Investigation (supporting); Methodology (supporting); Project administration (equal); Resources (equal); Software (lead); Supervision (equal); Validation (lead); Visualization (lead); Writing – original draft (supporting); Writing – review & editing (supporting). **R. J. Webber**: Conceptualization (supporting); Formal analysis (equal); Investigation (equal); Methodology (supporting); Validation (supporting); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **D. M. Zuckerman**: Conceptualization (supporting); Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: DETAILS AND DERIVATIONS

1. Computing other statistics using WE

WE can be used to compute other statistics, in addition to the MFPT from A to B. Namely, any integral $\int f(x)\pi(x)dx$ involving an observable f can be estimated using WE trajectories. In the long time limit, the trajectories satisfy

$$\int f(x)\pi(x)dx = \lim_{t\to\infty} \frac{\tau}{t} \sum_{s=0}^{t/\tau-1} \sum_{i=1}^{N} w_{s\tau}^i f(\xi_{s\tau}^i), \qquad (A1)$$

where $(\xi_{s\tau}^i, w_{s\tau}^i)_{1\le i\le N}$ denotes the position and weight of trajectory i at a multiple s of lag time τ. Therefore, as a practical estimator, we can terminate the simulations at a finite time t and approximate

$$\int f(x)\pi(x)dx \approx \frac{\tau}{t-t_0} \sum_{s=t_0/\tau}^{t/\tau-1} \sum_{i=1}^{N} w_{s\tau}^i f(\xi_{s\tau}^i), \qquad (A2)$$

where $t_0 > 0$ is a suitable burn-in time. Here, we have assumed that f is recorded only at τ intervals although WE permits saving points more frequently.

Estimator (A2) is readily extended to history-dependent observables, such as transition-path times or mechanistic pathways from A to B. This requires defining the trajectories as paths starting from A, which are recycled upon reaching B, and defining π as a stationary measure on path space. The recent mathematical literature[10,11,49,50] provides a set of error bounds and convergence guarantees for estimator (A2).

2. Relationship between flux discrepancy and MFPT

To establish the relationship

$$\lim_{t \to \infty}\left[\langle N_t \rangle_x - \langle N_t \rangle_\pi\right] = \frac{\langle T_B \rangle_\pi - \langle T_B \rangle_x}{\langle T_B \rangle_A}, \tag{A3}$$

we first introduce the following terminology:

- T_B^x is the first passage time into B for a process started at $X_0 = x$.
- T_B^π is the first passage time into B for a process started at $X_0 \sim \pi$, i.e., started at a point chosen randomly from the steady-state distribution.
- N_t^A is the number of arrivals into B by time t for a process started at $X_0 \sim \rho_A$.

Next, we recall the *renewal theorem* from classical probability,[62] which states that for any $s > 0$,

$$\lim_{t \to \infty}\left[\langle N_t^A \rangle - \langle N_{t-s}^A \rangle\right] = \frac{s}{\langle T_B \rangle_A}. \tag{A4}$$

Last, we verify

$$
\begin{aligned}
\lim_{t \to \infty}\left[\langle N_t \rangle_x - \langle N_t \rangle_\pi\right] &= \lim_{t \to \infty}\left[\langle N_{t-T_B^x}^A \rangle - \langle N_{t-T_B^\pi}^A \rangle\right] \\
&= \lim_{t \to \infty}\left[\langle N_t^A \rangle - \langle N_{t-T_B^\pi}^A \rangle\right] - \lim_{t \to \infty}\left[\langle N_t^A \rangle - \langle N_{t-T_B^x}^A \rangle\right] \\
&= \frac{\langle T_B^\pi \rangle - \langle T_B^x \rangle}{\langle T_B \rangle_A},
\end{aligned}
\tag{A5}
$$

where we have applied the renewal theorem after conditioning on $s = T_B^\pi$ and $s = T_B^x$.

3. Derivation of optimal allocation rule

In this section, we sketch the derivation of the optimal allocation rule,

$$\text{\# trajectories near } x \propto \pi(x)\nu(x). \tag{A6}$$

We use the following notation:

- $\text{weight}_{\text{bin}}$ and $\text{alloc}_{\text{bin}}$ indicate a bin weight and bin allocation.
- $\langle \rangle_{\text{bin}}$ and Var_{bin} denote the mean and variance with respect to the weighted trajectory distribution in a bin, so $\text{Var}_{\text{bin}}(f) = \langle f^2 \rangle_{\text{bin}} - \langle f \rangle_{\text{bin}}^2$.
- Finally, $\langle \rangle_t$ denotes the average over the WE ensemble up to time t.

In the limit as $t \to \infty$, the following formula[10,11] describes the WE variance:

$$t\,\text{Var}(\hat{f}_{\text{WE}}) \sim \left\langle \sum_{\text{bins}} \frac{\text{weight}_{\text{bin}}^2}{\text{alloc}_{\text{bin}}}\left[\text{Var}_{\text{bin}}(h) + \text{Var}_{\text{bin}}(\nu) + \langle \nu \rangle_{\text{bin}}^2\right] \right\rangle_t. \tag{A7}$$

We consider a greedy minimization strategy[10,50] for minimizing the variance formula (A7). The variance terms $\text{Var}_{\text{bin}}(h)$ and $\text{Var}_{\text{bin}}(\nu)$ are minimized by choosing bins in which h and ν values are nearly constant. The other term, namely,

$$\frac{\text{weight}_{\text{bin}}^2}{\text{alloc}_{\text{bin}}}\langle \nu \rangle_{\text{bin}}^2,$$

is minimized when the allocation is chosen using[10,49,50]

$$\text{alloc}_{\text{bin}} \propto \text{weight}_{\text{bin}} \times \langle \nu \rangle_{\text{bin}}, \tag{A8}$$

which reduces to (A6) in the limit of many trajectories.

4. Low temperature limit

Here, we derive the optimal splitting and merging coordinates in the low-temperature limit $\beta \to \infty$. First, we combine the expression for the steady state,

$$\tilde{\pi}(x) \propto e^{-\beta U(x)}\int_{\max\{x,a\}}^b \frac{e^{\beta U(y)}dy}{D(y)}, \tag{A9}$$

with the expression for the flux variance function,

$$\tilde{\nu}(x) = \frac{\sqrt{2}e^{\beta U(x)}}{D(x)^{1/2}\langle \tilde{T}_B \rangle_A}\int_{-\infty}^x \frac{dy}{e^{\beta U(y)}}, \tag{A10}$$

to obtain the optimal allocation rule,

$$\tilde{\pi}(x)\tilde{\nu}(x) \propto \frac{1}{D(x)^{1/2}}\int_{\max\{x,a\}}^b \frac{e^{\beta U(y)}dy}{D(y)}\int_{-\infty}^x \frac{dy}{e^{\beta U(y)}}. \tag{A11}$$

Next, we assume that the pair of points (x_-, x_+) is the unique solution to the maximization

$$\max_{a \le y \le b,\, x < y}(U(y) - U(x)), \tag{A12}$$

and $U''(x_+) < 0 < U''(x_-)$. Using Laplace's method[63] on (A11), we obtain, for any $x_- < x < x_+$,

$$\frac{1}{D(x)^{1/2}}\int_{\max\{x,a\}}^b \frac{e^{\beta U(y)}dy}{D(y)}\int_{-\infty}^x \frac{dy}{e^{\beta U(y)}} \overset{\beta \to \infty}{\sim} \frac{C}{D(x)^{1/2}}\frac{e^{\beta U(x_+)}}{\beta e^{\beta U(x_-)}}, \tag{A13}$$

where the prefactor C is explicitly given by

$$C = \frac{2\pi}{D(x_+)U''(x_-)^{1/2}|U''(x_+)|^{1/2}}. \tag{A14}$$

For x outside of the interval $[x_-, x_+]$, Laplace's method dictates that the optimal allocation $\pi(x)v(x)$ is exponentially smaller as $\beta \to \infty$, whence we recover the optimal allocation (23).

From the above calculations, we find that the optimal gain from using WE takes the form

$$\text{Gain over brute force sampling} = \frac{QR}{S^2}, \qquad (A15)$$

where

$$Q = \int_{-\infty}^{b} \frac{e^{\beta U(x)}}{D(x)} \int_{\max\{x,a\}}^{b} \frac{e^{\beta U(y)} dy}{D(y)} \left(\int_{-\infty}^{x} \frac{dy}{e^{\beta U(y)}} \right)^2 dx, \qquad (A16)$$

$$R = \int_{-\infty}^{b} e^{-\beta U(x)} \int_{\max\{x,a\}}^{b} \frac{e^{\beta U(y)} dy}{D(y)} dx, \qquad (A17)$$

$$S = \int_{-\infty}^{b} \frac{1}{D(x)^{1/2}} \int_{\max\{x,a\}}^{b} \frac{e^{\beta U(y)} dy}{D(y)} \int_{-\infty}^{x} \frac{dy}{e^{\beta U(y)}} dx. \qquad (A18)$$

Applying the Laplace principle to (A16)–(A18) verifies the expression

Gain over brute force sampling

$$\sim \frac{\pi/\beta}{\left(\int_{x_-}^{x_+} \sqrt{D(x_+)/D(x)}\, dx \right)^2} \frac{e^{\beta(U(x_+)-U(x_-))}}{\sqrt{|U''(x_-)U''(x_+)|}}. \qquad (A19)$$

5. Estimation of h and v^2

Here, we explain how to infer estimates for π, h, and v^2 based on a Markov transition matrix P between microbins. The steady-state distribution π is invariant under the dynamics, and it integrates to 1. Therefore, to obtain a coarse-grained estimate π_i of the probability mass for each microbin, we solve the eigenvalue problem

$$\pi^T P = \pi^T, \qquad \sum_i \pi_i = 1. \qquad (A20)$$

From definition (6), the flux discrepancy function h associated with the transition matrix P satisfies

$$h = \sum_{t=1}^{\infty} P^t (f - (\pi^T f)1), \qquad (A21)$$

where 1 is the vector of all ones, and

$$f_i = \begin{cases} 0, & \text{microbin } i \text{ is not in } B, \\ 1, & \text{microbin } i \text{ is in } B. \end{cases} \qquad (A22)$$

Multiplying both sides of (A21) by $I - P + 1\pi^T$ and invoking the relation $\pi^T P = \pi^T$, we find that h solves the linear system

$$(I - P + 1\pi^T)h = Pf - (\pi^T f)1. \qquad (A23)$$

The left-hand side of (A23) is invertible, so h can be obtained using a direct linear solver. Last, from definition (8), the flux variance function v^2 associated with the transition matrix P satisfies

$$v^2 = P(h_0^2) - (Ph_0)^2, \qquad h_0 = h + f, \qquad (A24)$$

where the squares are defined entrywise. The amount of resolution needed in this microbinning approach is a topic for future investigation.

REFERENCES

[1] G. A. Huber and S. Kim, "Weighted-ensemble Brownian dynamics simulations for protein association reactions," Biophys. J. **70**, 97–110 (1996).

[2] B. W. Zhang, D. Jasnow, and D. M. Zuckerman, "The 'weighted ensemble' path sampling method is statistically exact for a broad class of stochastic processes and binning procedures," J. Chem. Phys. **132**, 054107 (2010).

[3] R. M. Donovan, A. J. Sedgewick, J. R. Faeder, and D. M. Zuckerman, "Efficient stochastic simulation of chemical kinetics networks using a weighted ensemble of trajectories," J. Chem. Phys. **139**, 115105 (2013).

[4] B. W. Zhang, D. Jasnow, and D. M. Zuckerman, "Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin," Proc. Natl. Acad. Sci. U. S. A. **104**, 18043–18048 (2007).

[5] M. C. Zwier, J. W. Kaus, and L. T. Chong, "Efficient explicit-solvent molecular dynamics simulations of molecular association kinetics: Methane/methane, Na(+)/Cl(−), methane/benzene, and k(+)/18-crown-6 ether," J. Chem. Theory Comput. **7**, 1189–1197 (2011).

[6] E. Suárez, S. Lettieri, M. C. Zwier, C. A. Stringer, S. R. Subramanian, L. T. Chong, and D. M. Zuckerman, "Simultaneous computation of dynamical and equilibrium information using a weighted ensemble of trajectories," J. Chem. Theory Comput. **10**, 2658–2667 (2014).

[7] S. D. Lotz and A. Dickson, "Unbiased molecular dynamics of 11 min timescale drug unbinding reveals transition state stabilizing interactions," J. Am. Chem. Soc. **140**, 618–628 (2018).

[8] T. Sztain, S.-H. Ahn, A. T. Bogetti, L. Casalino, J. A. Goldsmith, E. Seitz, R. S. McCool, F. L. Kearns, F. Acosta-Reyes, S. Maji et al., "A glycan gate controls opening of the SARS-CoV-2 spike protein," Nat. Chem. **13**, 963–968 (2021).

[9] H. Kahn and T. E. Harris, Estimation of Particle Transmission by Random Sampling (National Bureau of Standards Applied Mathematics Series, 1951), Vol. 12, pp. 27–30.

[10] D. Aristoff, "An ergodic theorem for the weighted ensemble method," J. Appl. Probab. **59**, 152–166 (2022).

[11] R. J. Webber, D. Aristoff, and G. Simpson, "A splitting method to reduce MCMC variance," arXiv:2011.13899 (2020).

[12] M. C. Zwier, J. L. Adelman, J. W. Kaus, A. J. Pratt, K. F. Wong, N. B. Rego, E. Suárez, S. Lettieri, D. W. Wang, M. Grabe, D. M. Zuckerman, and L. T. Chong, "WESTPA: An interoperable, highly scalable software package for weighted ensemble simulation and analysis," J. Chem. Theory Comput. **11**, 800–809 (2015).

[13] J. D. Russo, S. Zhang, J. M. Leung, A. T. Bogetti, J. P. Thompson, A. J. DeGrave, P. A. Torrillo, A. Pratt, K. F. Wong, J. Xia et al., "WESTPA 2.0: High-performance upgrades for weighted ensemble simulations and analysis of longer-timescale applications," J. Chem. Theory Comput. **18**, 638 (2021).

[14] S. D. Lotz and A. Dickson, "Wepy: A flexible software framework for simulating rare events with weighted ensemble resampling," ACS Omega **5**, 31608–31623 (2020).

[15] D. Aristoff, F. G. Jones, R. J. Webber, G. Simpson, and D. M. Zuckerman, Weightedensemble.jl, Julia package, 2020.

[16] A. Dickson and C. L. Brooks III, "WExplore: Hierarchical exploration of high-dimensional spaces using the weighted ensemble algorithm," J. Phys. Chem. B **118**, 3532–3542 (2014).

[17] J. L. Adelman, A. L. Dale, M. C. Zwier, D. Bhatt, L. T. Chong, D. M. Zuckerman, and M. Grabe, "Simulations of the alternating access mechanism of the sodium symporter Mhp1," Biophys. J. **101**, 2399–2407 (2011).

[18] B. Abdul-Wahid, H. Feng, D. Rajan, R. Costaouec, E. Darve, D. Thain, and J. A. Izaguirre, "AWE-WQ: Fast-forwarding molecular dynamics using the accelerated weighted ensemble," J. Chem. Inf. Model. **54**, 3033–3043 (2014).

[19] S.-H. Ahn, A. A. Ojha, R. E. Amaro, and J. A. McCammon, "Gaussian-Accelerated molecular dynamics with the weighted ensemble method: A hybrid method improves thermodynamic and kinetic sampling," J. Chem. Theory Comput. **17**, 7938–7951 (2021).

[20] D. Ray and I. Andricioaei, "Weighted ensemble milestoning (WEM): A combined approach for rare event simulations," J. Chem. Phys. **152**, 234114 (2020).

[21] D. Ray, S. E. Stone, and I. Andricioaei, "Markovian weighted ensemble milestoning (M-WEM): Long-time kinetics from short trajectories," J. Chem. Theory Comput. **18**, 79 (2021).

[22] A. Ojha, S. Thakur, S.-H. Ahn, and R. Amaro, "DeepWEST: Deep learning of kinetic models with the weighted ensemble simulation toolkit for enhanced kinetic and thermodynamic sampling," ChemRxiv (2022), https://chemrxiv.org/engage/chemrxiv/article-details/6238beac13d478049d96d707.

[23] M. J. Tse, B. K. Chu, C. P. Gallivan, and E. L. Read, "Rare-event sampling of epigenetic landscapes and phenotype transitions," PLoS Comput. Biol. **14**, e1006336 (2018).

[24] R. Taylor, J. F. Allard, and E. Read, "Weighted ensemble simulation of kinetic segregation shows role of oligomers and close contacts," Biophys. J. **121**, 149a–150a (2022).

[25] D. M. Zuckerman and L. T. Chong, "Weighted ensemble simulation: Review of methodology, applications, and software," Annu. Rev. Biophys. **46**, 43–57 (2017).

[26] D. Bhatt and D. M. Zuckerman, "Beyond microscopic reversibility: Are observable nonequilibrium processes precisely reversible?" J. Chem. Theory Comput. **7**, 2520–2527 (2011).

[27] A. S. Saglam and L. T. Chong, "Protein–protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations," Chem. Sci. **10**, 2360–2372 (2019).

[28] D. Bhatt, B. W. Zhang, and D. M. Zuckerman, "Steady-state simulations using weighted ensemble path sampling," J. Chem. Phys. **133**, 014110 (2010).

[29] W. C. Swope, J. W. Pitera, and F. Suits, "Describing protein folding kinetics by molecular dynamics simulations. 1. Theory," J. Phys. Chem. B **108**, 6571–6581 (2004).

[30] R. J. Allen, P. B. Warren, and P. R. Ten Wolde, "Sampling rare switching events in biochemical networks," Phys. Rev. Lett. **94**, 018104 (2005).

[31] F. Cérou and A. Guyader, "Adaptive multilevel splitting for rare event analysis," Stochastic Anal. Appl. **25**, 417–443 (2007).

[32] F. Cérou, A. Guyader, T. Lelièvre, and D. Pommier, "A multiple replica approach to simulate reactive trajectories," J. Chem. Phys. **134**, 054108 (2011).

[33] C.-E. Bréhier, M. Gazeau, L. Goudenege, T. Lelièvre, and M. Rousset, "Unbiasedness of some generalized adaptive multilevel splitting algorithms," Ann. Appl. Prob. **26**, 3559–3601 (2016).

[34] I. Teo, C. G. Mayne, K. Schulten, and T. Lelièvre, "Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time," J. Chem. Theory Comput. **12**, 2983–2989 (2016).

[35] C. Giardina, J. Kurchan, V. Lecomte, and J. Tailleur, "Simulating rare events in dynamical processes," J. Stat. Phys. **145**, 787–811 (2011).

[36] E. Guevara Hidalgo, T. Nemoto, and V. Lecomte, "Finite-time and finite-size scalings in the evaluation of large-deviation functions: Numerical approach in continuous time," Phys. Rev. E **95**, 062134 (2017).

[37] U. Ray, G. K.-L. Chan, and D. T. Limmer, "Importance sampling large deviations in nonequilibrium steady states. I," J. Chem. Phys. **148**, 124120 (2018).

[38] J. M. Bello-Rivas and R. Elber, "Exact milestoning," J. Chem. Phys. **142**, 094102 (2015).

[39] D. Aristoff, J. M. Bello-Rivas, and R. Elber, "A mathematical framework for exact milestoning," Multiscale Model. Simul. **14**, 301–322 (2016).

[40] A. Warmflash, P. Bhimalapuram, and A. R. Dinner, "Umbrella sampling for nonequilibrium processes," J. Chem. Phys. **127**, 154112 (2007).

[41] T. S. van Erp, D. Moroni, and P. G. Bolhuis, "A novel path sampling method for the calculation of rate constants," J. Chem. Phys. **118**, 7762–7774 (2003).

[42] D. S. Abbot, R. J. Webber, S. Hadden, D. Seligman, and J. Weare, "Rare event sampling improves mercury instability statistics," Astrophys. J. **923**, 236 (2021).

[43] J. Finkel, R. J. Webber, E. P. Gerber, D. S. Abbot, and J. Weare, "Learning forecasts of rare stratospheric transitions from short simulations," Mon. Weather Rev. **149**, 3647–3669 (2021).

[44] J. Finkel, R. J. Webber, E. P. Gerber, D. S. Abbot, and J. Weare, "Exploring stratospheric rare events with transition path theory and short simulations," arXiv:2108.12727 (2021).

[45] R. M. Donovan, J.-J. Tapia, D. P. Sullivan, J. R. Faeder, R. F. Murphy, M. Dittrich, and D. M. Zuckerman, "Unbiased rare event sampling in spatial stochastic systems biology models using a weighted ensemble of trajectories," PLoS Comput. Biol. **12**, e1004611 (2016).

[46] E. Darve and E. Ryu, "Computing reaction rates in bio-molecular systems using discrete macro-states," in *Innovations in Biomolecular Modeling and Simulations* (Royal Society of Chemistry, 2012), pp. 138–206.

[47] H. Feng, R. Costaouec, E. Darve, and J. A. Izaguirre, "A comparison of weighted ensemble and Markov state model methodologies," J. Chem. Phys. **142**, 214113 (2015).

[48] J. L. Adelman and M. Grabe, "Simulating rare events using a weighted ensemble-based string method," J. Chem. Phys. **138**, 044105 (2013).

[49] D. Aristoff, "Analysis and optimization of weighted ensemble sampling," ESAIM: Math. Modell. Numer. Anal. **52**, 1219–1238 (2018).

[50] D. Aristoff and D. M. Zuckerman, "Optimizing weighted ensemble sampling of steady states," Multiscale Model. Simul. **18**, 646–673 (2020).

[51] M. Baudel, A. Guyader, and T. Lelièvre, "On the Hill relation and the mean reaction time for metastable processes," arXiv:2008.09790 (2020).

[52] T. Lelièvre, M. Ramil, and J. Reygner, "Estimation of statistics of transitions and hill relation for Langevin dynamics," arXiv:2206.13264 (2022).

[53] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics* (Courier Corporation, 2005).

[54] I. Ben-Ari and R. G. Pinsky, "Spectral analysis of a family of second-order elliptic operators with nonlocal boundary condition indexed by a probability measure," J. Funct. Anal. **251**, 122–140 (2007).

[55] A. J. DeGrave, A. T. Bogetti, and L. T. Chong, "The red scheme: Rate-constant estimation from pre-steady state weighted ensemble simulations," J. Chem. Phys. **154**, 114111 (2021).

[56] U. Adhikari, B. Mostofian, J. Copperman, S. R. Subramanian, A. A. Petersen, and D. M. Zuckerman, "Computational estimation of microsecond to second atomistic folding times," J. Am. Chem. Soc. **141**, 6519–6526 (2019).

[57] J. Copperman and D. M. Zuckerman, "Accelerated estimation of long-timescale kinetics from weighted ensemble simulation via non-Markovian "microbin" analysis," J. Chem. Theory Comput. **16**, 6763–6775 (2020).

[58] J. Lu and J. Nolen, "Reactive trajectories and the transition path process," Prob. Theory Relat. Fields **161**, 195–244 (2015).

[59] C. W. Gardiner *et al.*, *Handbook of Stochastic Methods*, 2nd ed. (Springer, Berlin, 1985).

[60] P. Ma, R. Elber, and D. E. Makarov, "The value of temporal information when analyzing reaction coordinates," J. Chem. Theory Comput. **16**, 6077–6090 (2020).

[61] A. J. DeGrave, J.-H. Ha, S. N. Loh, and L. T. Chong, "Large enhancement of response times of a protein conformational switch by computational design," Nat. Commun. **9**, 1013–1019 (2018).

[62] W. Feller, *An Introduction to Probability Theory and its Applications* (John Wiley and Sons, 2008), Vol. 2.

[63] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I* (Springer, New York, NY, 1999).