The Journal of Chemical Physics

Brownian bridges for stochastic chemical processes—An approximation method based on the asymptotic behavior of the backward Fokker–Planck equation

Cite as: J. Chem. Phys. **156**, 184108 (2022); https://doi.org/10.1063/5.0080540
Submitted: 01 December 2021 • Accepted: 18 April 2022 • Accepted Manuscript Online: 19 April 2022 • Published Online: 10 May 2022

Shiyan Wang, Anirudh Venkatesh, Doraiswami Ramkrishna, et al.

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Brownian bridges for stochastic chemical processes—An approximation method based on the asymptotic behavior of the backward Fokker–Planck equation

Cite as: J. Chem. Phys. 156, 184108 (2022); doi: 10.1063/5.0080540
Submitted: 1 December 2021 • Accepted: 18 April 2022 •
Published Online: 10 May 2022

Shiyan Wang,[a] Anirudh Venkatesh, Doraiswami Ramkrishna,[b] and Vivek Narsimhan[c]

AFFILIATIONS

Davidson School of Chemical Engineering, Purdue University, West Lafayette, Indiana 47907, USA

[a]E-mail: wang2502@purdue.edu
[b]E-mail: ramkrish@purdue.edu
[c]Author to whom correspondence should be addressed: vnarsim@purdue.edu

ABSTRACT

A Brownian bridge is a continuous random walk conditioned to end in a given region by adding an effective drift to guide paths toward the desired region of phase space. This idea has many applications in chemical science where one wants to control the endpoint of a stochastic process—e.g., polymer physics, chemical reaction pathways, heat/mass transfer, and Brownian dynamics simulations. Despite its broad applicability, the biggest limitation of the Brownian bridge technique is that it is often difficult to determine the effective drift as it comes from a solution of a Backward Fokker–Planck (BFP) equation that is infeasible to compute for complex or high-dimensional systems. This paper introduces a fast approximation method to generate a Brownian bridge process without solving the BFP equation explicitly. Specifically, this paper uses the asymptotic properties of the BFP equation to generate an approximate drift and determine ways to correct (i.e., re-weight) any errors incurred from this approximation. Because such a procedure avoids the solution of the BFP equation, we show that it drastically accelerates the generation of conditioned random walks. We also show that this approach offers reasonable improvement compared to other sampling approaches using simple bias potentials.

I. INTRODUCTION

Continuous random walks—i.e., diffusion with drift—are omnipresent in a wide range of chemical fields, such as sampling polymer conformations,[1,2] studying chemical reaction pathways,[3,4] and molecular/Brownian dynamics simulations.[5] In many practical situations, one seeks to quantify random walks whose paths are constrained to stay in a given region, hit a region before another one, or end in a given region.[6] Such ideas allow one to quantify rare events in a chemical process or conversely quantify most probable configurations and reaction pathways. Because of these reasons, there has been much interest in developing robust techniques that can sample and control random paths efficiently.

We study a concept known as a Brownian bridge, which is a random process whose start and end points are known.[7] A Brownian bridge is created by adding an effective drift to a continuous random walk such that the paths end in the desired region of phase space with the correct conditional statistics (Fig. 1). This idea falls into the broader class of "driven" processes that was introduced by Doob in the 1950s (referred in the mathematics community as the Doob *h*-transform[8,9]) and has only recently experienced interest in the chemical physics community to study rare events and understand non-equilibrium thermodynamics under large deviation theory.[10–14] In fact, the effective drift for a Brownian bridge can be generalized to create conditioned processes that stay in a given region for its entire path or create random walks that hit one region of phase space before another.[15,16]

The Brownian bridge can condition any continuous random walk to end in a given region and in many situations offer advantages over other techniques. For example, when chemical physicists

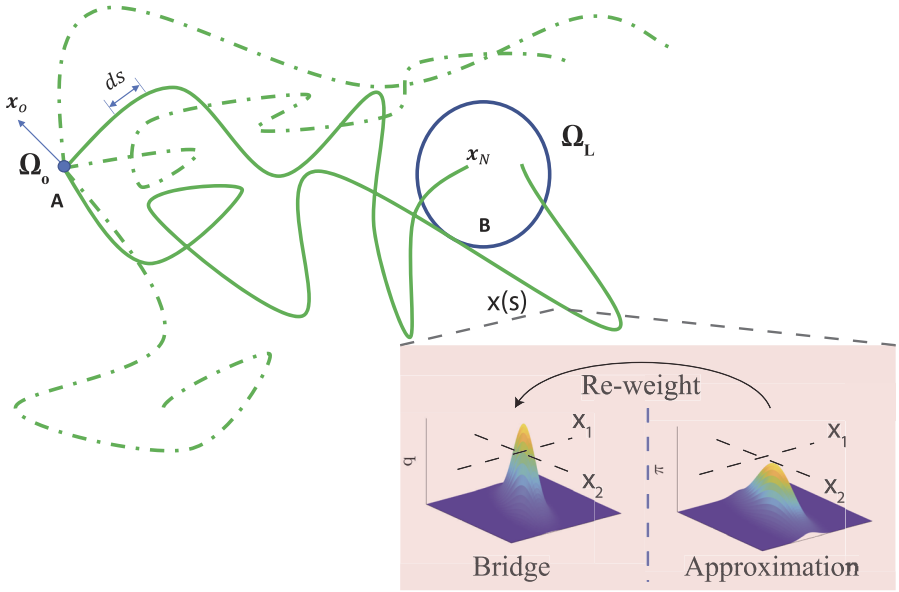

FIG. 1. Brownian bridge samples the stochastic paths that start from point $\boldsymbol{x}_o \in \Omega_o$ (at $s = 0$) and are conditioned to end in domain $\boldsymbol{x}_N \in \Omega_L$ (at $s = L$). In other words, out of all paths that a continuous random walk can realize (solid and dashed), the Brownian bridge will only sample the subset that ends in region Ω_L (solid lines). For a random walk at a given intermediate position $\boldsymbol{x}(s)$, the Brownian bridge is implemented by adding a drift velocity $\boldsymbol{u} = 2\boldsymbol{D} \cdot \frac{\partial \ln q}{\partial \boldsymbol{x}}$ to bias the path, where $\boldsymbol{D}$ is the diffusion matrix of the random walk and $q(\boldsymbol{x}, s) = P(\boldsymbol{x}(L) \in \Omega_L | \boldsymbol{x}(s))$ is a hitting probability that satisfies the Backward Fokker–Planck (BFP) equation (3). This paper introduces a method to approximate the drift velocity $\boldsymbol{u}$ without solving the BFP equation and how to correct (reweight) any errors incurred from such an approximation.

examine paths with two known endpoints, they often use a path-integral formulation.[17,18] This formulation does not work well if one wants the process to end in a region rather than a point, but the Brownian bridge allows for this flexibility. Similarly, many molecular dynamics simulations at constant temperature use a Langevin equation to describe the dynamics of all molecules involved. To sample the transition between two possibly rare states, one often uses sampling techniques, such as umbrella/blue moon sampling,[19,20] metadynamics,[21] or adaptive forcing approaches,[22] to add a biasing force/potential along a reaction coordinate. Although useful, these techniques often struggle to sample rare events over a prescribed length/time interval, which the Brownian bridge can implement straightforwardly. Finally, Brownian bridges can be useful in barrier crossing problems where one wants to examine the short-time barrier crossing trajectories and ignore the uninteresting dynamics that occur, while the system waits in a potential energy well.[23]

In our manuscripts,[16,24] we demonstrated how the Brownian bridge idea could be used in polymer physics by examining a canonical problem in polymer field theory—the conformation of an ideal polymer in an external field. We demonstrated that one can use the bridge methodology to efficiently grow polymer chains to end with a given topology (e.g., rings or finite winding number) or end with a given total energy. The latter is useful for sampling rare events (high total energy) or most probable configurations (low energy). In the situations we examined, the bridge methodology vastly outperforms traditional sampling techniques, such as Rosenbluth techniques and Metropolis Monte Carlo methods, in some cases simulating exceedingly rare conformations with an order of magnitude fewer samples than before.[16] Furthermore, we showed that this methodology can be extended to constrain paths to lie in a given region for its entire path.[16]

Despite the applications of Brownian bridges and other "driven" processes with a Doob h-transform, there is one glaring limitation of these techniques that hamper their implementation in the chemistry community. The drift used in the Brownian bridge to drive paths to the correct regions in phase space is determined using the solution of a Backward Fokker–Planck (BFP) equation, which is infeasible for complicated systems. The solution to the BFP equation becomes problematic due to the "curse of dimensionality"—i.e., it scales unfavorably as the number of dimensions increases. Thus, there is a need to develop and leverage approximation methods to calculate the drift in order to scale up the methods for higher dimensions. This work will discuss a novel approach to address this issue, and an approximation method will be developed based on the asymptotic properties of the BFP equation.

This work will show how to systematically generate approximations to the drift without solving the BFP equation and how to correct (i.e., reweight) any errors incurred from this approximation. In Sec. II, we begin by first defining the stochastic system and the concept of a Brownian bridge. Next, we introduce the approximation method based on ideas from path integrals and importance sampling. There are two key elements in our scheme: (a) an approximate solution to the BFP equation based on its asymptotic properties and (b) a reweighting procedure. In Sec. III, we use our approximation scheme to sample several 1D Wiener processes that are conditioned to end at a given point or a region. These ideas are then extended to higher dimensional systems. In Sec. IV, we show that the approximation scheme offers an improvement in efficacy compared to alternate importance sampling approaches where (a) one flattens the potential energy landscape, (b) one adds a constant drift toward the end region, and (c) one adds a confining, harmonic potential centered around the end region. In Sec. V, we perform a time and spatial complexity analysis to show that the approximation method outlined here gains a significant computational edge over obtaining an exact Brownian bridge via solution to the BFP equation from a finite element method. In Sec. VI, we summarize our findings and discuss the future developments.

II. PROBLEM SETUP AND METHODS

A. Stochastic differential equation for continuous random walk

A stochastic differential equation is a mathematical description of a continuous random walk. Let s be a coordinate that represents the arc length of a random walk, where $0 < s < L$, where L is the total length. Let $\boldsymbol{x}_o$ be the initial position of the walk and $\boldsymbol{x}(s)$ be the position at arc length s. If the random walk has a drift $\boldsymbol{A}$ and a diffusion tensor $\boldsymbol{D} = \frac{1}{2}\boldsymbol{B}\cdot\boldsymbol{B}^T$, the stochastic differential equation in Itô form that describes this process is

$$d\boldsymbol{x} = \boldsymbol{A}(s,\boldsymbol{x}(s))ds + \boldsymbol{B}(s,\boldsymbol{x}(s))\cdot d\mathscr{B}, \quad \boldsymbol{x}(0) = \boldsymbol{x}_o, \ s \in (0,L). \tag{1}$$

In the above equation, we note that the drift $\boldsymbol{A}$ and tensor $\boldsymbol{B}$ can be functions of the arc length s and phase space coordinate $\boldsymbol{x}$. For instance, when $\boldsymbol{A} = -\frac{1}{2}(\boldsymbol{B}\cdot\boldsymbol{B}^T)\cdot\nabla U(\boldsymbol{x}) = -\boldsymbol{D}\cdot\nabla U$, the system represents an over-damped Langevin equation, where $U(\boldsymbol{x})$ is the dimensionless external potential. The above notation is shorthand for the following: if we want to compute the random walk position $\boldsymbol{x}(s+ds)$ given the known position $\boldsymbol{x}(s)$, we perform a forward Euler–Maruyama integration:[25] $\boldsymbol{x}(s+ds) = \boldsymbol{x}(s) + \boldsymbol{A}(s,\boldsymbol{x}(s))ds + \sqrt{ds}\boldsymbol{B}(s,\boldsymbol{x}(s))\cdot\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a normal, independently distributed Gaussian random variable with mean zero and covariance $\boldsymbol{I}$ (the identity matrix). The symbol $d\mathscr{B} = \sqrt{ds}\boldsymbol{Z}$ denotes a Wiener process.[25] For example, many polymer chains are represented using such random walks. A flexible polymer with no interactions is simply $d\boldsymbol{x} = d\mathscr{B}$ if the coordinates are scaled properly. This ideal random walk—i.e., a Gaussian chain—accurately describes polymer conformations over distances larger than the persistence length of the backbone.[26–27] Ideal polymers that have backbone stiffness can be represented by more complicated stochastic differential equations (SDEs)—e.g., worm-like chain model.[28] Other examples of continuous random walks include an Ornstein–Uhlenbeck (UO) process[29,30] ($d\boldsymbol{x} = -k\boldsymbol{x}ds + \sigma d\mathscr{B}$) and geometric Brownian motion[31] ($d\boldsymbol{x} = k\boldsymbol{x}ds + \sigma\boldsymbol{x}d\mathscr{B}$), where constants k represent the drift constants of both systems, and constants σ are the volatility of the random fluctuations that are modeled as Brownian motions.

B. Brownian bridge

Suppose that we have the random walk described in Eq. (1), but now we want to condition the paths to end in a region Ω_L at $s = L$ (Fig. 1). We note that for this paper, the total length of the walks (L) is a specified input. For example, in simulating a polymer conformation, L represents the length of the polymer backbone. The stochastic differential equation that samples the conditional probability $P(\boldsymbol{x}(s)|\boldsymbol{x}(0) = \boldsymbol{x}_0, \boldsymbol{x}(L) \in \Omega_L)$ is a bridge $\boldsymbol{x}^{Br}$, which satisfies[16,24]

$$d\boldsymbol{x}^{Br} = d\boldsymbol{x} + \boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\frac{\partial}{\partial\boldsymbol{x}}(\ln q)ds \quad \text{with} \quad \boldsymbol{x}(0) = \boldsymbol{x}_o. \tag{2}$$

The bridge is the same as the original random walk $\boldsymbol{x}$ but with an additional drift term $\boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\frac{\partial}{\partial\boldsymbol{x}}(\ln q)$. The term $q(\boldsymbol{x},s)$ is a hitting probability (Fig. 1), which is a probability that the original random walk reaches the desired end region Ω_L at $s = L$, given that its earlier position is $\boldsymbol{x}(s)$ at arc length s. From a physics standpoint, one can think of the extra drift as an "entropic drift" that drives a path toward the end region[16] since the entropic free-energy barrier for reaching the end-state is proportional to $T\Delta S \sim -\beta\ln(q)$. To compute the hitting probability q, one can invoke classic theories from continuous Markov chains to show that it satisfies a backward Fokker–Planck equation[32] (Ω_L is the endpoint or designated region) as follows:

$$\mathscr{L}q = \frac{\partial q}{\partial s} + \boldsymbol{A}\cdot\frac{\partial q}{\partial\boldsymbol{x}} + \frac{1}{2}\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right) : \frac{\partial^2 q}{\partial\boldsymbol{x}\partial\boldsymbol{x}} = 0,$$
$$q(\boldsymbol{x}, s = L) = \begin{cases} 1 & \boldsymbol{x} \in \Omega_L, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The above equation is integrated backward from $s = L$ to $s = 0$. Furthermore, if one wants to confine the whole path to a region rather than just its endpoint (Brownian excursion[16]), then one can add boundary conditions to the above equation. Note that a numerical solution of a Backward Fokker–Planck (BFP) equation becomes infeasible for high dimensional systems. In Sec. II D, we will discuss an approximation method that allows one to obtain the statistics of the Brownian bridge without an explicit solution to the BFP equation.

C. Probability of paths: Path integral approach

The probability of a specific path for the SDE $d\boldsymbol{x} = \boldsymbol{A}ds + \boldsymbol{B}\cdot d\mathscr{B}$ is expressed in a discrete manner ($L = N\cdot\Delta s$),

$$p\{\boldsymbol{x}\} \sim \exp\left(-\frac{1}{2\Delta s}\sum_{k=0}^{N-1}(\Delta\boldsymbol{x}_k - \boldsymbol{A}(\boldsymbol{x}_k)\Delta s) \cdot \left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1}\cdot(\Delta\boldsymbol{x}_k - \boldsymbol{A}(\boldsymbol{x}_k)\Delta s)\right), \tag{4}$$

where $\Delta\boldsymbol{x}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$. For a Brownian bridge, the corresponding probability is

$$p^{Br}\{\boldsymbol{x}\} \sim \exp\left(-\frac{1}{2\Delta s}\sum_{k=0}^{N-1}\left(\Delta\boldsymbol{x}_k - (\boldsymbol{A}(\boldsymbol{x}_k) + \boldsymbol{u}^d)\Delta s\right) \cdot \left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1}\cdot\left(\Delta\boldsymbol{x}_k - (\boldsymbol{A}(\boldsymbol{x}_k) + \boldsymbol{u}^d)\Delta s\right)\right), \tag{5}$$

where the exact drift is expressed as $\boldsymbol{u}^d = (\boldsymbol{B}\cdot\boldsymbol{B}^T)\cdot\nabla_{\boldsymbol{x}}\ln(q)$. The relationship between the probability of the Brownian bridge and the original stochastic process can also be written as $p^{Br}\{\boldsymbol{x}\} \sim \mathbb{1}(\boldsymbol{x}(L) \in \Omega_L)\cdot p\{\boldsymbol{x}\}$, where $\mathbb{1}$ is the indicator function.

D. Method of approximation: Reweighting

Instead of solving the hitting probability q via the Backward Fokker–Planck (BFP) equation (3), let us use an approximate hitting probability ψ that satisfies the same boundary and end conditions as q in the backward Fokker–Planck equation(3). Suppose that we now add a drift to the original SDE that approximates the bridge: $d\boldsymbol{x} = (\boldsymbol{A} + \boldsymbol{u}^{ap})dt + \boldsymbol{B}\cdot d\mathscr{B}$, where the approximated drift is $\boldsymbol{u}^{ap} = (\boldsymbol{B}\,\boldsymbol{B}^T)\cdot\nabla_{\boldsymbol{x}}\ln(\psi)$.

The probability of a path under this approximation is

$$p^{ap}\{\boldsymbol{x}\} \sim \exp\Bigg(-\frac{1}{2\Delta s}\sum_{k=0}^{N-1}\left(\Delta \boldsymbol{x}_k - (\boldsymbol{A}(\boldsymbol{x}_k)+\boldsymbol{u}^{ap})\Delta s\right) \cdot \left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1}\cdot\left(\Delta \boldsymbol{x}_k - (\boldsymbol{A}(\boldsymbol{x}_k)+\boldsymbol{u}^{ap})\Delta s\right)\Bigg)$$
$$\sim p\{\boldsymbol{x}\}\exp\Bigg(\frac{1}{2}\sum_{k=0}^{N-1}\Big(\Delta \boldsymbol{x}_k\cdot\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1}\cdot\boldsymbol{u}^{ap} + \boldsymbol{u}^{ap}\cdot\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1}\cdot\Delta \boldsymbol{x}_k - \boldsymbol{A}\Delta s\cdot\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1}\cdot\boldsymbol{u}^{ap} - \boldsymbol{u}^{ap}\cdot\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1}\cdot\boldsymbol{A}\Delta s - \boldsymbol{u}^{ap}\cdot\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1}\cdot\boldsymbol{u}^{ap}\Delta s\Big)\Bigg). \tag{6}$$

If one uses Itô's formula $\Delta \ln(\psi) = \frac{\partial \ln\psi}{\partial s}\Delta s + \frac{\partial \ln\psi}{\partial \boldsymbol{x}}\Delta \boldsymbol{x} + \frac{1}{2}(\boldsymbol{B}\cdot\boldsymbol{B}^T):\frac{\partial^2 \ln\psi}{\partial \boldsymbol{x}\partial \boldsymbol{x}}\Delta s$ as $\Delta s \to 0$ and uses the fact that $\exp(\sum \Delta \ln\psi) \sim \mathbb{1}(\boldsymbol{x}(L)\in\Omega_L)$, one can obtain the following expression:

$$p^{ap}\{\boldsymbol{x}\} \sim p^{Br}\{\boldsymbol{x}\}\exp\Bigg(-\sum_{k=0}^{N-1}\Bigg(\frac{\partial \ln\psi}{\partial s} + \boldsymbol{A}\cdot\frac{\partial \ln\psi}{\partial \boldsymbol{x}} + \frac{1}{2}\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right):\left(\frac{\partial^2 \ln\psi}{\partial \boldsymbol{x}\partial \boldsymbol{x}} + \frac{\partial \ln\psi}{\partial \boldsymbol{x}}\frac{\partial \ln\psi}{\partial \boldsymbol{x}}\right)\Bigg)\Delta s\Bigg). \tag{7}$$

In the limit $\Delta s \to 0$, the above expression can be written as

$$p^{ap}\{\boldsymbol{x}\} \sim p^{Br}\{\boldsymbol{x}\}\exp\left(-\int_0^L \mathscr{G}[\ln\psi]ds\right), \tag{8}$$

where

$$\mathscr{G}[\ln\psi] = \frac{\partial \ln\psi}{\partial s} + \boldsymbol{A}\cdot\frac{\partial \ln\psi}{\partial \boldsymbol{x}} + \frac{1}{2}\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right):\left(\frac{\partial^2 \ln\psi}{\partial \boldsymbol{x}\partial \boldsymbol{x}} + \frac{\partial \ln\psi}{\partial \boldsymbol{x}}\frac{\partial \ln\psi}{\partial \boldsymbol{x}}\right) = \frac{1}{\psi}\mathscr{L}\psi. \tag{9}$$

In the above equation, $\mathscr{L}$ is the Backward Fokker–Planck (BFP) operator described in Eq. (3). Note that Eq. (9) shows that, for the reweighting procedure, one needs to calculate the Hessian matrix, $\frac{\partial^2 \ln\psi}{\partial \boldsymbol{x}\partial \boldsymbol{x}}$. Alternatively, one can avoid the second derivative calculations by considering the following equation instead of Eq. (7):

$$p^{ap}\{\boldsymbol{x}\} \sim p^{Br}\{\boldsymbol{x}\}\exp\Bigg(-\sum_{k=0}^{N-1}\Bigg(\Delta\ln\psi - \frac{\partial \ln\psi}{\partial \boldsymbol{x}}\cdot\Delta \boldsymbol{x}_k + \boldsymbol{A}\cdot\frac{\partial \ln\psi}{\partial \boldsymbol{x}}\Delta s + \frac{1}{2}\boldsymbol{B}\cdot\boldsymbol{B}^T:\frac{\partial \ln\psi}{\partial \boldsymbol{x}}\frac{\partial \ln\psi}{\partial \boldsymbol{x}}\Delta s\Bigg)\Bigg). \tag{10}$$

As $\Delta s \to 0$, the new reweighting function is expressed as

$$p^{ap}\{\boldsymbol{x}\} \sim p^{Br}\{\boldsymbol{x}\}\exp\Bigg(\int_0^L \frac{1}{2}\boldsymbol{B}\cdot\boldsymbol{B}^T:\frac{\partial \ln\psi}{\partial \boldsymbol{x}}\frac{\partial \ln\psi}{\partial \boldsymbol{x}}ds + \int_0^L \frac{\partial \ln\psi}{\partial \boldsymbol{x}}\cdot\boldsymbol{B}\cdot d\mathscr{B} - \ln\psi(L) + \ln\psi(0)\Bigg), \tag{11}$$

which does not need the repetitive calculation of the Hessian matrix. However, one should be cautious about the sampling efficiency of this alternative formalism compared to the original equations (8) and (9) as the uncertainty of ensembles is amplified by avoiding the calculation of the Hessian matrix.

Equations (8) and (9) show that the process with the approximate hitting probability ψ has the same statistics as the bridge if we reweight each generated path by $W = \exp\left[\int_0^L \frac{1}{\psi}\mathscr{L}\psi ds\right] = \exp\left[\int_0^L \mathscr{G}[\ln\psi]ds\right]$. There are many consequences to this reweighting scheme.

- One can now obtain the statistics of the bridge without knowing the exact hitting probability.
- If $\psi = q$ (the exact hitting probability), the quantity $\int_0^L \mathscr{G}[\ln\psi]ds = \int_0^L \frac{1}{\psi}\mathscr{L}\psi ds = 0$. The weight $W = 1$, and we obtain optimal sampling.
- If $\psi \neq q$, the quantity $\int_0^L \frac{1}{\psi}\mathscr{L}\psi ds$ gives an idea of how good or poor the approximation is. If there is a wide variation in the weights $W = \exp(\int_0^L \frac{1}{\psi}\mathscr{L}\psi ds)$, then one needs a better estimate.

In Sec. II E, we will design a good choice of ψ for various continuous random walks. We first design ψ for endpoint control and then extend the idea to controlling the end region. We will look at 1D stochastic processes first and then extend to $\mathcal{N}$ dimensions.

E. Design principle for approximation

A good approximation of a hitting probability should allow $\int_0^L \frac{1}{\psi}\mathscr{L}\psi ds \approx 0$. Here, we use the asymptotic properties of this operator to generate guidelines for an approximate hitting probability ψ.

Let us look at the simple problem of a pure Wiener process $d\boldsymbol{x} = d\mathscr{B}$ conditioned to end at a particular point $\boldsymbol{x}(s=L) = \boldsymbol{x}_N$. The hitting probability is Gaussian—i.e.,

$$q(\boldsymbol{x},s) = \mathbb{P}(\boldsymbol{x}(L) = \boldsymbol{x}_N|\boldsymbol{x},s) = \frac{1}{\sqrt{2\pi(L-s)}}\exp\left[-\frac{(\boldsymbol{x}-\boldsymbol{x}_N)^2}{2(L-s)}\right], \tag{12a}$$

$$\ln q(\boldsymbol{x},s) = -\frac{(\boldsymbol{x}-\boldsymbol{x}_N)^2}{2(L-s)} - \frac{1}{2}\ln(L-s) - \frac{1}{2}\ln(2\pi). \tag{12b}$$

In this case, we see that $\ln q \sim O(\frac{1}{L-s})$ as $s \to L$, which exhibits singular behavior. In fact, this scaling will hold for any SDE $d\boldsymbol{x} = \boldsymbol{A}ds + \boldsymbol{B}\cdot d\mathscr{B}$ as long as $\boldsymbol{B}\cdot\boldsymbol{B}^T$ is full rank since the hitting probability will always be diffusion dominant near the endpoint. Therefore, one criterion we can learn is that any approximation $\ln\psi$ to the hitting probability must recover the asymptotic property $\ln\psi \sim O(\frac{1}{L-s})$ as $s \to L$, as long as $\boldsymbol{B}\cdot\boldsymbol{B}^T$ is full-rank.

Another restriction on $\ln\psi$ comes from the fact that the weight $W = \exp\left(\int_0^L \mathscr{G}[\ln\psi]ds\right)$ cannot have too large of a dynamical range for different paths $\boldsymbol{x}(s)$ that are sampled. At minimum, we need to ensure that $\mathscr{G}[\ln\psi] \sim O(1)$ or smaller as $s \to L$. Thus, if we want to guide process to an end region, the approximate hitting probability ψ should at minimum satisfy the following two properties:

$$1.\ \ln\psi \sim O\left(\frac{1}{L-s}\right) \quad \text{as } s \to L, \tag{13}$$

2. $\mathcal{G}[\ln\psi] \sim O(1)$ or less as $s \to L$. (14)

If both conditions hold, we can safely reweight the paths to obtain the correct statistics for the bridge. Note that this approximate principle is expected to be accurate as long as the stochastic process is diffusion dominant. In Sec. V, we will show that it is much faster than the conventional procedure that solves the BFP equation directly.

F. Summary of the approximation method for the Brownian bridge

Below, we summarize the procedure for the approximation method.

1. Generate an approximate hitting probability ψ that satisfies conditions (13) and (14).
2. Generate an approximate drift $\boldsymbol{u}^{ap} = \boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\frac{\partial \ln\psi}{\partial \boldsymbol{x}}$ and add it to the SDE to guide paths to the endpoint or end region: $d\boldsymbol{x} = (\boldsymbol{A} + \boldsymbol{u}^{ap})ds + \boldsymbol{B}\cdot d\mathcal{B}$.
3. For each path $\{\boldsymbol{x}\}(s)$ generated in this process, calculate a weight $W = \exp(\int_0^L \mathcal{G}[\ln\psi]ds)$ using Eq. (9).
4. Paths reweighted by W now have the statistics of the conditioned random walk. Thus, if we want to obtain an ensemble average of an observable α, we compute $\langle\alpha\rangle = \sum_i \alpha_i W_i / \sum_i W_i$, where W_i is the weight of sample i.

III. RESULTS AND DISCUSSION

A. 1D Wiener process with a fixed endpoint

Let us look at a generic, one-dimensional SDE $dx = A(x)ds + B(x)d\mathcal{B}$ conditioned to end at a particular point $x(s = L) = x_N$. Because we know that the hitting probability will be diffusion dominated near the endpoint, we can expect it to have similar asymptotic form as Eq. (12b) when $s \to L$. Thus, we can state that the approximate solution takes the form

$$\ln\psi = -\frac{1}{2(L-s)}f(x) + g(x) - \frac{1}{2}\ln(L-s), \tag{15}$$

where $f(x)$ and $g(x)$ are unknown functions. Using the condition that $\mathcal{G}[\ln\psi] \sim O(1)$ as $s \to L$ allows one to solve for the functions $f(x)$ and $g(x)$, which gives the approximate hitting probability to be

$$\ln\psi = -\frac{1}{2(L-s)}\left(\int_{x_N}^{x}\frac{dy}{B(y)}\right)^2 - \int_{x_N}^{x}\frac{A(y)}{B^2(y)}dy + \frac{1}{2}\ln B - \frac{1}{2}\ln(L-s). \tag{16}$$

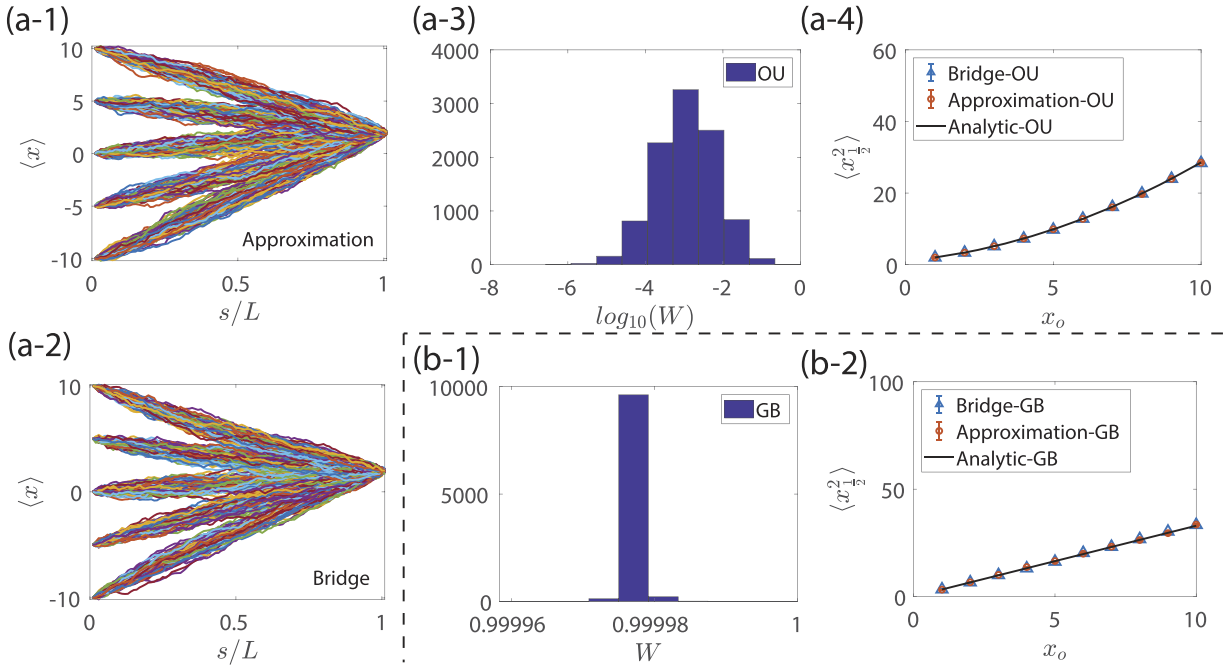

FIG. 2. Comparison of the exact Brownian bridge to our proposed approximate solution for 1D Ornstein–Uhlenbeck (OU) and geometric Brownian (GB) processes. (a) For an OU process, we present paths for different starting positions $x_o = -10, -5, 0, 5$, and 10 and endpoint $x_N = 2$. (a-1) Approximate bridge and (a-2) exact Brownian bridge. The parameters are $\sigma = k = 1$, $L = 1$, and the number of runs for each instance = 100. (a-3) Distribution of normalized weight W for the approximate OU process ($x_o = 10$; number of runs = 10^4). (a-4) Comparison of the ensemble of $x^2_{\frac{1}{2}}$, which is $\langle x^2\rangle$ at $s/L = 1/2$, for different starting positions x_o (number of runs = 10^4). (b) For the GB process, we present (b-1) the distribution of normalized weight W from the approximate GB ($x_o = 10$) and (b-2) the ensemble of $x^2_{\frac{1}{2}}$. For these runs, the endpoint is $x_N = 2$. The other parameters are $L = 1$, $\sigma = k = 1$, and the number of runs = 10^4. In (a-4) and (b-2), the error bars indicate the standard deviation of the sample mean.

We will use this approximate hitting probability to sample random walks that finish at the point $x = x_N$. To do this, we first generate an approximate entropic term $u^{ap} = B^2 \frac{\partial \ln \psi}{\partial x}$ and add it to the SDE to guide paths to the endpoint: $dx = (A + u^{ap})ds + Bd\mathscr{B}$. For each path $x(s)$ generated in this process, we calculate a weight $W = \exp(\int_0^L \mathscr{G}[\ln \psi]ds)$. Paths reweighted by W will now have the statistics of the conditioned random walk.

We will demonstrate this methodology for two different stochastic processes: 1D Ornstein–Uhlenbeck (OU) process[29,33] and geometric Brownian (GB) motion.[31] For the OU process, the SDE is written as $dx = -kxds + \sigma d\mathscr{B}$. According to Eq. (16), the approximate hitting probability and the corresponding $\mathscr{G}[\ln \psi]$ value are

$$\ln \psi = -\frac{1}{2(L-s)\sigma^2}(x-x_N)^2 + \frac{1}{2}\frac{k}{\sigma^2}(x^2 - x_N^2) + \frac{1}{2}\ln\left(\frac{\sigma}{L-s}\right),$$
$$\mathscr{G}\ln[\psi] = -\frac{1}{2}\frac{k^2x^2}{\sigma^2} + \frac{k}{2}. \tag{17}$$

If the SDE is a geometric Brownian motion ($dx = kxds + \sigma xd\mathscr{B}$), we have

$$\ln \psi = -\frac{1}{2(L-s)\sigma^2}\left[\ln\left(\frac{x}{x_N}\right)\right]^2 - \frac{k}{\sigma^2}\ln\left(\frac{x}{x_N}\right) + \frac{1}{2}\ln\left(\frac{\sigma x}{L-s}\right),$$
$$\mathscr{G}[\ln \psi] = \text{const.} \tag{18}$$

Figures 2(a-1) and 2(a-2) show sample paths for the OU process starting at different points x_o and conditioned to end at $x_N = 2$. The paths denoted "approximation" are paths constructed from the approximate hitting probability discussed above, while the paths denoted "bridge" are constructed using the exact hitting probability q (known analytically for the OU process). The simulations from the Brownian bridge and the approximate solution show different paths, yet both reach the end point in Fig. 2(a). The weight distribution for the approximate OU process is in the range of $W \in (10^{-4}, 10^0)$, showing that reweighting is necessary to capture the correct statistics. For the GB process, the approximate solution is identical to the exact bridge solution [see weight distribution for GB in Fig. 2(b-1)].

In Figs. 2(a-4) and 2(b-2), we graph the quantity $\langle x_{\frac{1}{2}}^2 \rangle$, which is the ensemble average of the second moment x^2 at the midpoint of the path ($s = L/2$). The data points "bridge" are sampled from the exact bridge process, while data points "approximation" are sampled from the approximate bridge and reweighted. For both the OU and GM processes, an analytic solution to $\langle x_{\frac{1}{2}}^2 \rangle$ is known using the following formula:

$$\langle x_{\frac{1}{2}}^2 \rangle = \frac{\int_{-\infty}^{\infty} x^2 p\left(x, s = \frac{L}{2} | x_o, s_o\right) p\left(x_N, L | x, s = \frac{L}{2}\right) dx}{\int_{-\infty}^{\infty} p\left(x, s = \frac{L}{2} | x_o, s_o\right) p\left(x_N, L | x, s = \frac{L}{2}\right) dx}. \tag{19}$$

In the above expression, $p(x, s|x_o, s_o)$ is the probability at position x and arc length s given a previous position x_o and arc length s_o. The analytic solutions are

$$\text{OU}: \langle x_{\frac{1}{2}}^2 \rangle = \frac{\sigma^2(1-e^{-k})}{2k(1+e^{-k})} + \frac{(x_o + x_N)^2 e^{-k}}{(1+e^{-k})^2},$$
$$\text{GB}: \langle x_{\frac{1}{2}}^2 \rangle = \exp\left(\frac{2\ln(x_o x_N) + 1}{2\sigma^2}\right). \tag{20}$$

The ensemble statistics from the approximation method confirms the correct solution [see analytic solutions for OU and GB in Figs. 2(a-4) and 2(b-2)]. We observe excellent agreement even though we did not compute an exact hitting probability from a backward Fokker–Planck equation.[16]

B. 1D Wiener process that ends in a given region

Section III A demonstrates a systematic method to approximate hitting probabilities for one-dimensional random walks conditioned to end at a particular point. This approximate hitting probability allows one to sample the conditioned random walk exactly, as long as one reweights each path accordingly. The advantage of this approach is that one no longer needs to solve a backward Fokker–Planck equation, which allows one to sample constrained random walks for more complicated systems where determining the Fokker–Planck solution is difficult or infeasible.

Here, we would like to develop approximations for paths that are constrained to end in a particular region rather than a point. Similar to before, let us examine the hitting probability $q(x, s)$ for a pure diffusion ($dx = d\mathscr{B}$) to gain intuition for its asymptotic properties near the end of the path ($s \to L$). If we examine paths conditioned to end in the region $x_N \in [a, \infty)$, the hitting probability is

$$\mathbb{P}(x(L) \geq a | x(s)) = \int_a^{\infty} \frac{dx_N}{\sqrt{2\pi(L-s)}} \exp\left[-\frac{(x_N - x)^2}{2(L-s)}\right]$$
$$= \frac{1}{2}\text{erfc}\left[\frac{a-x}{\sqrt{2(L-s)}}\right], \tag{21}$$

where $\text{erfc}(\cdot)$ is a complimentary error function. When $s \to L$, Eq. (21) behaves like $\ln(q) \approx -\frac{1}{2(L-s)}(x-a)^2 - \ln(a-x) + \frac{1}{2}\ln(L-s)$ when x is far outside the hitting region ($\frac{a-x}{\sqrt{L-s}} \gg 1$) and $\ln(q) \sim 0$ when x is far inside the hitting region ($\frac{x-a}{\sqrt{L-s}} \gg 1$). There is a smooth variation in between these two regimes.

These expressions allow one to derive an approximate hitting probability for a general 1D SDE: $dx = Ads + Bd\mathscr{B}$. Since any conditioned continuous random walk is dominated by diffusion near the end of the path ($s \to L$), one would expect that the hitting probability would take a similar asymptotic form as the pure random walk. We propose the form for the approximate hitting probability to be $\ln(\psi) = -\frac{1}{2(L-s)}f(x) + g(x) + \frac{1}{2}\ln(L-s)$ for x far outside the hitting region ($\frac{a-x}{|B|\sqrt{L-s}} \gg 1$) and $\ln(\psi) = 0$ for x far inside the hitting region ($\frac{x-a}{|B|\sqrt{L-s}} \gg 1$). To make sure that the these expressions do not create large variations in sample weights $W = \exp(\int_0^L \mathscr{G}[\ln \psi]ds)$, we enforce the condition $\mathscr{G}[\ln \psi] \sim O(1)$ as $s \to L$ in each of these regions. This allows one to solve for functions $f(x)$ and $g(x)$ and hence gives the approximate hitting probability in these regions,

$$\ln\psi \approx \begin{cases} \frac{1}{2}\ln[B(L-s)] - \ln\left(\left|\int_a^x \frac{dx}{B}\right|\right) - \int_a^s \frac{A}{B^2}dx - \frac{1}{2(L-s)}\left(\int_a^x \frac{dx}{B}\right)^2, & \frac{a-x}{|B|\sqrt{L-s}} \gg 1, \\ 0, & \frac{x-a}{|B|\sqrt{L-s}} \gg 1. \end{cases} \tag{22}$$

In between these two regions, there should be a smooth variation in $\ln(\psi)$. Below is an analytic expression that interpolates between these two regions such that when x is close to the boundary, $\frac{a-x}{|B|\sqrt{L-s}} \sim O(1)$. There are other interpolation expressions that one can make, but we found that the following equation works fairly well for our applications:

$$\ln\psi = \ln\phi - (1-\phi)\left[\int_a^x \frac{A}{B^2}dx - \frac{1}{2}\ln B\right]. \tag{23}$$

In the above formula, $\phi = \frac{1}{2}\mathrm{erfc}\left(-\frac{\int_a^x \frac{dx}{B}}{\sqrt{2(L-s)}}\right)$.

In Fig. 3(a), we examine an Ornstein–Uhlenbeck (OU) process starting at different initial conditions x_0 and conditioned to end in a region $x_N \in [4, \infty)$. Figure 3(a-1) generates paths from the exact Brownian bridge using the exact hitting probability, and Fig. 3(a-2) generates paths from the approximate bridge using the estimated hitting probability in Eq. (23). Figure 3(b) shows the ensemble average of the first and second moments at the midpoint of the path ($s = L/2$). This figure shows that the approximate solution after reweighting gives the same results as the exact bridge even though the approximate approach did not explicitly compute a full solution of the backward Fokker–Planck equation. Next, we derive the theoretical solution for both first and second moments at the midpoint. For the OU process, both solutions can be generally written as

$$\begin{aligned} \langle x_{\frac{1}{2}}\rangle &= \frac{\int_a^\infty dx_N \int_{-\infty}^\infty x p\left(x, s=\frac{L}{2}|x_o, s_o\right) p\left(x_N, L|x, s=\frac{L}{2}\right)dx}{\int_a^\infty dx_N \int_{-\infty}^\infty p\left(x, s=\frac{L}{2}|x_o, s_o\right) p\left(x_N, L|x, s=\frac{L}{2}\right)dx}, \\ \langle x^2_{\frac{1}{2}}\rangle &= \frac{\int_a^\infty dx_N \int_{-\infty}^\infty x^2 p\left(x, s=\frac{L}{2}|x_o, s_o\right) p\left(x_N, L|x, s=\frac{L}{2}\right)dx}{\int_a^\infty dx_N \int_{-\infty}^\infty p\left(x, s=\frac{L}{2}|x_o, s_o\right) p\left(x_N, L|x, s=\frac{L}{2}\right)dx}, \end{aligned} \tag{24}$$

where a is the lower integral limit. For Eq. (24), the analytical forward probability $p(x, s|x_o, s_o)$ is known, and thus, direct numerical integration is possible. In Fig. 3(b), sampling from the exact bridge and approximate bridge compares well with the theoretical result from Eq. (24). Furthermore, since the error bars are roughly the same for the bridge and the approximation, there is little loss in

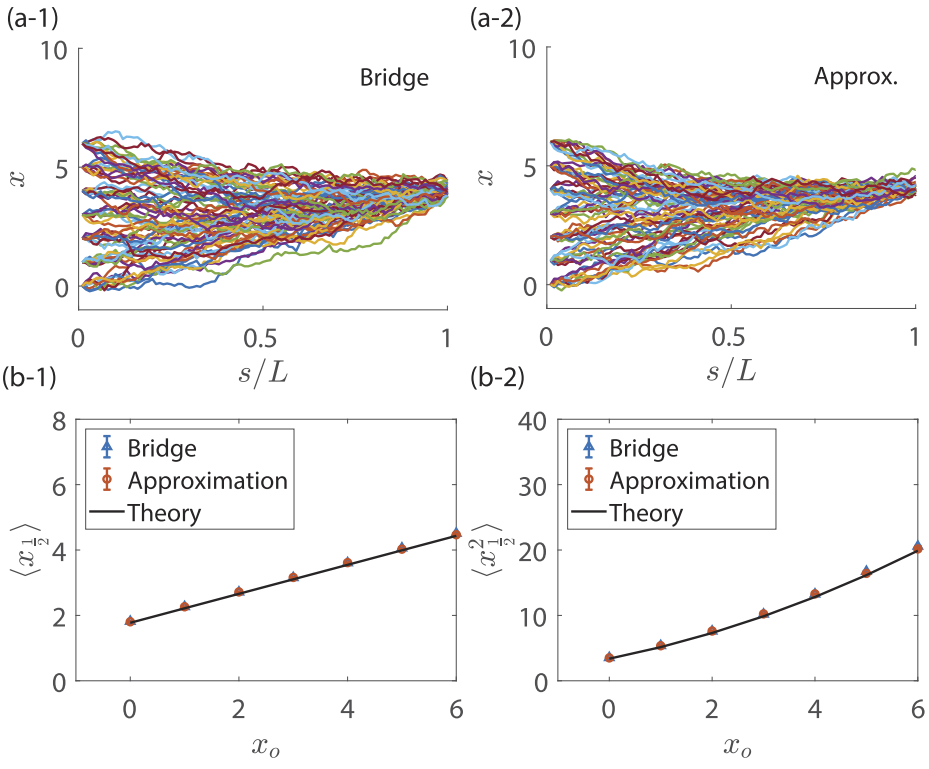

FIG. 3. Application of the Brownian bridge to the 1D Ornstein–Uhlenbeck (OU) process using the approximation method for ending in a region. (a) Showcase of OU paths (number of runs = 10 for each starting position x_o) where the target region is $[4, \infty)$: (a-1) exact Brownian bridge and (a-2) approximation method. (b) Comparison of the ensemble of (b-1) $\langle x\rangle$ and (b-2) $\langle x^2\rangle$ at $s/L = 1/2$ (number of runs, 10^4) for different starting positions x_o. For all runs, $L = 1$ and $x_o = 0, 1, 2, 3, 4, 5,$ and 6. Parameter specifications: $\sigma = 1$ and $k = 1$. In (b-1) and (b-2), the uncertainties indicate one standard deviation of the sample mean.

sampling precision. We should note that, for longer simulation trajectories (larger L), the distribution of reweighting could become wider, which could result in poorer sampling precision. If L becomes larger (longer trajectory), the system would be drift dominant. Since our approximate scheme is particularly suitable for the diffusion dominant problems, the approximation method would likely show a slower convergence rate as the trajectories become longer.

C. $\mathcal{N}$-dimensional Langevin equation with constant diffusion

Let us examine an $\mathcal{N}$ dimensional Langevin equation as follows:

$$d\boldsymbol{x} = -\boldsymbol{D}\cdot\nabla_{\boldsymbol{x}}U(\boldsymbol{x})ds + \sqrt{2\boldsymbol{D}}\cdot d\mathcal{B}, \tag{25}$$

where $\boldsymbol{D} = \frac{1}{2}\boldsymbol{B}\cdot\boldsymbol{B}^T$ is a constant diffusion matrix and $U(\boldsymbol{x})$ is the external potential energy. This type of equation is commonly employed in Brownian dynamics simulations, where $\boldsymbol{x}$ represents the coordinates of all particles. If one wants to condition the coordinates to end at $\boldsymbol{x} = \boldsymbol{x}_N$, one can develop an approximate hitting probability ψ for this process using the same ideas discussed earlier. If one lets $\ln(\psi) \sim O(\frac{1}{L-s})$ and $\mathcal{G}[\ln\psi] \sim O(1)$ as $s \to L$, one obtains

$$\begin{aligned}\ln\psi &= -\frac{1}{4(L-s)}(\boldsymbol{x}-\boldsymbol{x}_N)\cdot\boldsymbol{D}^{-1}\cdot(\boldsymbol{x}-\boldsymbol{x}_N)\\ &\quad -\frac{n}{2}\ln(L-s) + \frac{U(\boldsymbol{x})}{2},\\ \mathcal{G}[\ln\psi] &= -\frac{1}{2}\boldsymbol{D}:(\nabla U(\boldsymbol{x})\nabla U(\boldsymbol{x}) - \nabla\nabla U(\boldsymbol{x})).\end{aligned} \tag{26}$$

The approximate bridge process is

$$d\boldsymbol{x}^{Approx} = (-\boldsymbol{D}\cdot\nabla U(\boldsymbol{x}) + 2\boldsymbol{D}\cdot\nabla\ln\psi)ds + \sqrt{2\boldsymbol{D}}\cdot d\mathcal{B}. \tag{27}$$

We note that, in Eq. (26), one needs to calculate the Hessian of potential energy for the weight estimation (i.e., $\mathcal{G}[\ln\psi]$). For a large system (e.g., a folding polymer described by an empirical force-field), the repetitive calculation of the Hessian matrix may be expensive.

Next, we move to the case where we condition the paths to end in a region Ω_N rather than a single point. Given the expression in Eq. (26), the approximate hitting probability for an endpoint is written as

$$\begin{aligned}\mathbb{P}(\boldsymbol{x}_N|\boldsymbol{x}(s)) &\approx \exp(\ln\psi)\\ &\approx \exp\left(-\frac{1}{4(L-s)}(\boldsymbol{x}-\boldsymbol{x}_N)^T\cdot\boldsymbol{D}^{-1}\cdot(\boldsymbol{x}-\boldsymbol{x}_N)\right)\\ &\quad\times\exp\left(\frac{U(\boldsymbol{x})}{2}\right)(L-s)^{-n/2}.\end{aligned} \tag{28}$$

Therefore, the approximate hitting probability for a region is

$$\begin{aligned}\mathbb{P}(\boldsymbol{x}_N\in\Omega_N|\boldsymbol{x}(s)) \approx \int_{\Omega_N}d\boldsymbol{x}_N\,&\exp\left(-\frac{1}{4(L-s)}(\boldsymbol{x}-\boldsymbol{x}_N)^T\right.\\ &\left.\cdot\boldsymbol{D}^{-1}\cdot(\boldsymbol{x}-\boldsymbol{x}_N)\right)\exp\left(\frac{U(\boldsymbol{x})}{2}\right)(L-s)^{-n/2}.\end{aligned} \tag{29}$$

We perform the method of steepest descent (Bender and Orsag[34]) to obtain the asymptotic properties of this integral as $s \to L$. Let us consider the following quantity, which represents a minimum, non-dimensional mean-squared distance to the boundary $\partial\Omega_N$ of the hitting region:

$$d^2 = \min_{\boldsymbol{x}_N}\left(\frac{1}{4(L-s)}(\boldsymbol{x}-\boldsymbol{x}_N)^T\cdot\boldsymbol{D}^{-1}\cdot(\boldsymbol{x}-\boldsymbol{x}_N)\right). \tag{30}$$

When the position $\boldsymbol{x}(s)$ is far inside the boundary—i.e., $\boldsymbol{x}\in\Omega_N$ and $d \gg 1$—the approximate hitting probability [Eq. (29)] satisfies $\ln\psi \approx 0$. When the position $\boldsymbol{x}(s)$ is far outside the boundary—i.e., $\boldsymbol{x}\notin\Omega_N$ and $d \gg 1$—the approximate hitting probability becomes

$$\ln\psi \sim \frac{U(\boldsymbol{x})}{2} - \frac{1}{2}\ln(d^2) - \frac{1}{4}d^2. \tag{31}$$

Just like before, we choose a function that interpolates smoothly between these two limits and let $\ln(\psi) \sim O(1)$ near the boundary. One function we found that works well is

$$\ln\psi = \ln\phi + (1-\phi)\left(\frac{U(\boldsymbol{x})}{2}\right), \tag{32}$$

where

$$\phi = \begin{cases}\frac{1}{2}\mathrm{erfc}\left[\frac{d}{2}\right] & \text{when}\quad \boldsymbol{x}\notin\Omega_N,\\ \frac{1}{2}\mathrm{erfc}\left[-\frac{d}{2}\right] & \text{when}\quad \boldsymbol{x}\in\Omega_N.\end{cases} \tag{33}$$

D. Application of the approximation method for the $\mathcal{N}$-dimensional Langevin equation for rare events

Here, we illustrate how the approximation method improves the efficiency of sampling rare events compared to a conventional approach. To examine rare events, we study the Langevin equation (25) in two dimensions with a barrier crossing potential.[35] Here, we consider an asymmetric double well potential energy $U(x,y) = k(x^4 - 11/3x^3 - 2x^2 + 11x + 25/3 + 7.5y^2)$ and a diffusion matrix $\boldsymbol{D} = \frac{1}{2}\sigma^2\boldsymbol{I}$. In Fig. 4(a), the potential has two minima of $(x=-1, y=0)$ and $(x=2.75, y=0)$ with a transition state at $(x=1, y=0)$. We examine paths that start at one of the minima $(x_o, y_o) = (-1, 0)$ and are conditioned to end in a circle centered at the other minimum $(x_c, y_c) = (2.75, 0)$ with radius $R = 2$. The potential energy $U(x,y)$ guides paths toward the initial state, preventing barrier crossing [Fig. 4(a)]. The barrier height is modulated by the value of k: a large value of k leads to a stronger barrier.

The approximation method involves adding a biasing drift $u^{ap} = 2\boldsymbol{D}\cdot\frac{\partial\ln(\psi)}{\partial\boldsymbol{x}}$ to the Langevin equation, where $\ln(\psi)$ is the approximate hitting probability in Eq. (32). The generated paths are reweighted accordingly using weights $W = \exp(\int_0^L\mathcal{G}[\ln\psi]ds)$ described previously in Eq. (9). The conventional sampling approach involves generating paths from the Langevin equation (25) without a biasing drift and collecting the subset of samples that satisfy the end condition.

In Fig. 4(b), we see that the approximation method will guide all paths to the desired end region, while the conventional approach will only have a subset of the paths satisfying the end condition. In Fig. 4(c), we compare the statistics of the approximate method to

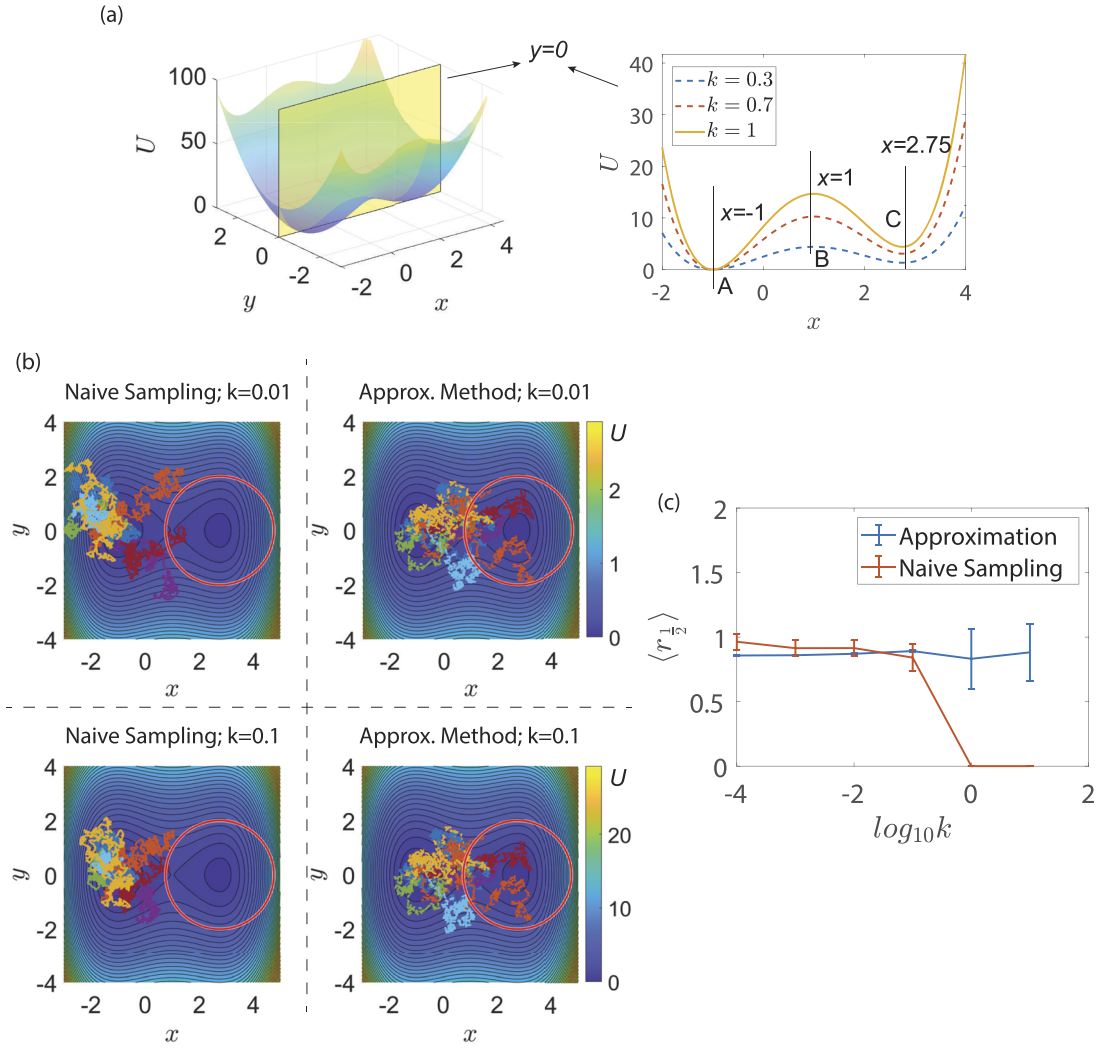

FIG. 4. Comparison between the approximate Brownian bridge and conventional sampling for the Langevin equation conditioned to end in a region for barrier crossing problems. The above fact examines a 2D Langevin equation (25) with the asymmetric double well potential energy $U(x,y) = k(x^4 - 11/3x^3 - 2x^2 + 11x + 25/3 + 7.5y^2)$, starting at position $(x_o, y_o) = (-1, 0)$ and conditioned to end in the circular region $(x_c, y_c) = (2.75, 0)$ with radius $R = 2$. (a) Showcase of the asymmetric 2D double well potential in the 1D slice ($y = 0$) with two minima (A: $x = -1$ and B: $x = 2.75$) and the transition state (C: $x = 1$). The barrier height is modulated by values of k. (b) Comparison of paths from the approximate Brownian bridge and conventional sampling for different values of k ($k = 0.01, 0.1$) with color maps for potential energy $U(x,y)$. (c) Ensemble average of $\langle r_{\frac{1}{2}} \rangle$, i.e., the distance of the path's midpoint from the origin, as value k increases (strong barrier). The approximation and conventional sampling give the same statistics although the approximation has lower error and does not fail from sampling inefficiency for larger k values. For (b) and (c), $\sigma = \sqrt{2}$, $L = 1$, and the number of runs is 10^4.

the conventional approach, as we increase the barrier height. We see that the both methods give the same ensemble average of $\langle r_{\frac{1}{2}} \rangle$, which is the average distance of the path's midpoint from the origin. However, the approximation method yields much less uncertainty in the ensemble value since the sampling efficiency is 100%. Furthermore, the approximate method allows one to sample rare events with strong barriers where the conventional approach will fail.

To further illustrate these points, we repeat the above calculation for an $\mathcal{N}$-dimensional Langevin equation to show how the approximate method allows one to sample conditional probabilities in higher dimensions. We examine the $\mathcal{N}$-dimensional Langevin equation (25) with potential energy $U(\boldsymbol{x}) = k(x_1^4 - 11/3x_1^3 - 2x_1^2 + 11x_1 + 25/3) + 7.5k(\sum_{i=2}^{\mathcal{N}} x_i)^2$ and diffusion matrix $\boldsymbol{D} = \frac{1}{2}\sigma^2 \boldsymbol{I}$. We study paths that start at position $\boldsymbol{x}_0 = (-1, 0, 0, ..., 0)$ and are conditioned to end in the region of the $\mathcal{N}$-sphere centered at $\boldsymbol{x}_c = (2.75, 0, 0, ..., 0)$ with radius $R = 3$. As the dimension of the system increases, the conventional sampling approach again fails to collect useful statistics. In Fig. 5(a), the conventional approach cannot collect sufficient data to sample $\langle r_{\frac{1}{2}} \rangle$ for $\mathcal{N} > 3$. On the other hand, the Brownian bridge using the approximation method enables

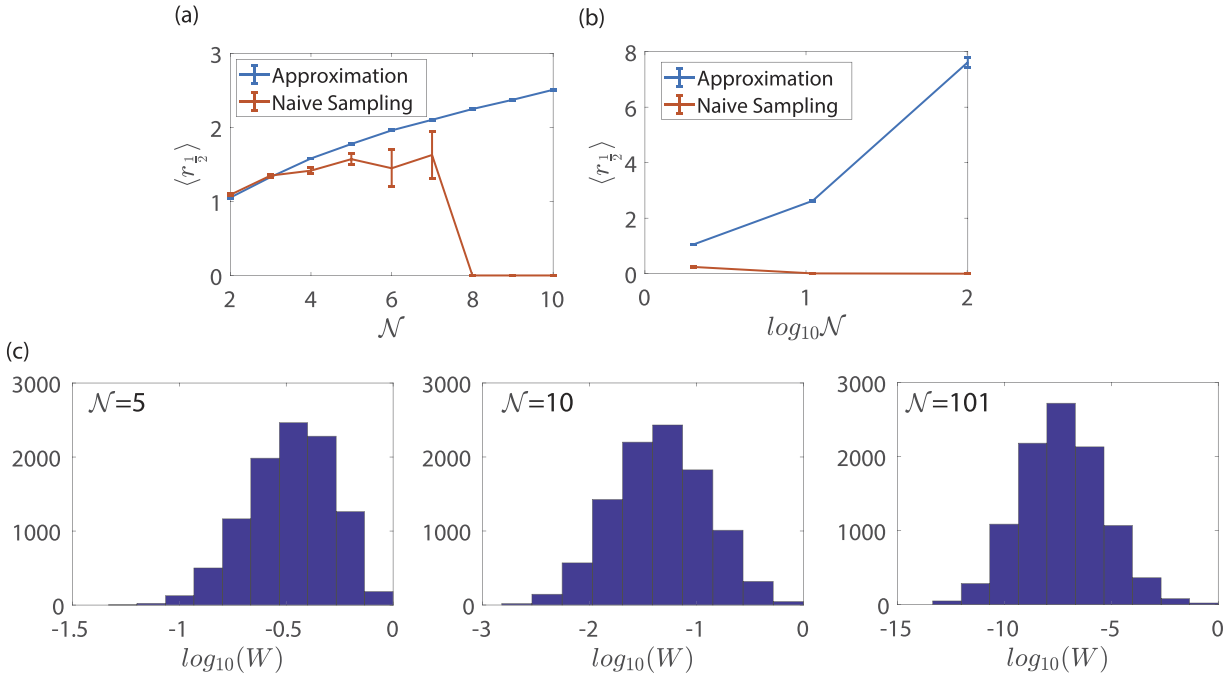

FIG. 5. Comparison between the approximate Brownian bridge and conventional sampling for the $\mathcal{N}$-dimensional Langevin equation. The above examines an $\mathcal{N}$-dimensional Langevin equation with diffusion matrix $\boldsymbol{D} = \frac{1}{2}\sigma^2\boldsymbol{I}$ and potential energy $U(\boldsymbol{x}) = k(x_1^4 - 11/3x_1^3 - 2x_1^2 + 11x_1 + 25/3) + 7.5k(\sum_{i=2}^{\mathcal{N}} x_i)^2$, starting at position $\boldsymbol{x}_o = (-1, 0, 0, ..., 0)$ and conditioned to end in the $\mathcal{N}$-sphere centered at $\boldsymbol{x}_c = (2.75, 0, 0, \ldots, 0)$ with radius $R = 3$. (a) Ensemble average of $\langle r_{\frac{1}{2}} \rangle$—i.e., the distance of the path's midpoint from the origin for $\mathcal{N}$: 2 – 10, as the number of dimensions increases. The approximation gives the same statistics as conventional sampling but can sample a larger dimensional space where traditional sampling fails ($\mathcal{N} > 5$). (b) Ensemble average of $\langle r_{\frac{1}{2}} \rangle$ for $\mathcal{N} \sim O(100)$, where naive sampling spectacularly fails. (c) Distribution of normalized weight from the approximate solution: $\mathcal{N} = 5, 10, 101$. Parameter specifications: $k = 0.001$ and $\sigma = \sqrt{2}$. The number of runs is 10^4.

prediction for $\mathcal{N} = 10$ or even higher [see $\mathcal{N} = 101$ in Fig. 5(b)]. Since all the runs are counted in the approximation method, the associated uncertainty of the ensemble of $\langle r \rangle$ from this technique is significantly smaller than that of conventional sampling [Fig. 5(a)]. Next, we examine the weight distribution for different dimensions. In Fig. 5(c), when the dimension $\mathcal{N}$ increases from 5, 10 to 101, the weight distribution broadens. This result indicates that as the dimension increases, the degradation of the sample weight will decrease the convergence rate of the sampling method (i.e., more samples are needed to obtain statistics within a given tolerance). In the future, it will be interesting to improve sampling at higher dimensions through a more accurate design of the drift based on the criteria discussed in this paper.

IV. PERFORMANCE AGAINST OTHER BIASED SAMPLING APPROACHES

It is important to compare the approximate bridge methodology to alternate sampling approaches. The alternate approaches we will consider are situations where (a) one flattens the potential energy landscape, (b) one adds a constant drift toward the end region, and (c) one adds a confining, harmonic potential centered around the end region. In these alternate approaches, one generates sample paths along the biased energy landscape and reweights the trajectories via importance sampling. First, we will provide the derivation of the sample weight when one uses a biasing potential. Next, we will compare the biased approaches to the approximate bridge methodology.

Suppose that we add an extra drift to $\boldsymbol{u}^d$ to the stochastic differential equation (SDE) as follows:

$$d\boldsymbol{x}^d = \left[\boldsymbol{A}(s, \boldsymbol{x}(s)) + \boldsymbol{u}^d(\boldsymbol{x}(s))\right]ds + \boldsymbol{B}(s, \boldsymbol{x}(s)) \cdot d\mathscr{B}, \quad \boldsymbol{x}(0) = \boldsymbol{x}_o, \ s \in (0, L). \tag{34}$$

This extra drift will bias all paths toward the desired end region but will not guarantee that the paths will end there. We also note that $\boldsymbol{u}^d$ depends only on the spatial coordinate $\boldsymbol{x}$. During the simulations, the probability of the sample paths satisfies

$$\frac{p\{\boldsymbol{x}\}}{p^d\{\boldsymbol{x}\}} \sim \exp\Bigg(-\frac{1}{2}\sum_{k=0}^{N-1}\Big(\Delta\boldsymbol{x}_k \cdot \left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1} \cdot \boldsymbol{u}^d + \boldsymbol{u}^d \cdot \left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1} \cdot \Delta\boldsymbol{x}_k - \boldsymbol{A}\Delta s \cdot \left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1} \cdot \boldsymbol{u}^d - \boldsymbol{u}^d \cdot \left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1} \cdot \boldsymbol{A}\Delta s - \boldsymbol{u}^d \cdot \left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)^{-1} \cdot \boldsymbol{u}^d \Delta s\Big)\Bigg), \tag{35}$$

where $p\{\boldsymbol{x}\}$ is the probability of the unbiased path (if $\boldsymbol{u}^d = 0$). If we let $\boldsymbol{u}^d(\boldsymbol{x}) = -\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)\cdot\nabla U^d(\boldsymbol{x})$, where $U^d(\boldsymbol{x})$ is the bias potential, Eq. (35) becomes

$$\frac{p\{\boldsymbol{x}\}}{p^d\{\boldsymbol{x}\}} \sim \exp\left(\sum_{k=0}^{N-1}\left(\Delta\boldsymbol{x}_k\cdot\nabla U^d - \boldsymbol{A}\Delta s\cdot\nabla U^d - \frac{1}{2}\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)\cdot\nabla U^d\cdot\nabla U^d\Delta s\right)\right). \tag{36}$$

We note that the bridge process $p^{Br}\{\boldsymbol{x}\} \sim p\{\boldsymbol{x}\}\mathbb{1}(\boldsymbol{x}(L)\in\Omega_L)$, where the latter is the indicator function for the end region. Multiplying both sides by this function gives us the weighting factor as follows:

$$W^d = \exp\left(\sum_{k=0}^{N-1}\left(\Delta\boldsymbol{x}_k\cdot\nabla U^d - \boldsymbol{A}\Delta s\cdot\nabla U^d - \frac{1}{2}\left(\boldsymbol{B}\cdot\boldsymbol{B}^T\right)\cdot\nabla U^d\cdot\nabla U^d\Delta s\right)\right)\mathbb{1}(\boldsymbol{x}(L)\in\Omega_L). \tag{37}$$

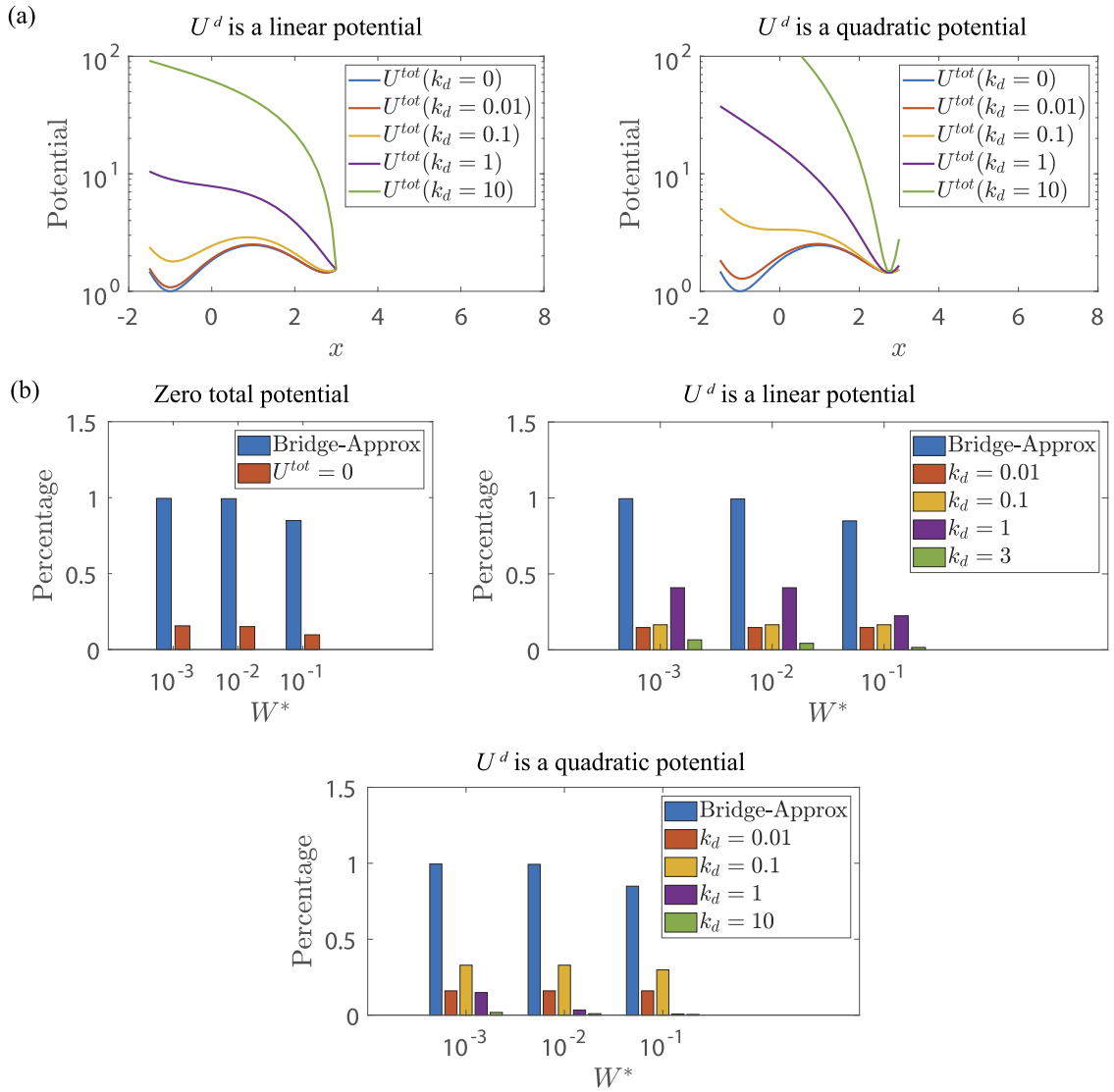

FIG. 6. Comparison between the approximate Brownian bridge and the following biased sampling approaches: zero total potential ($U + 2U^d = 0$), linear bias potential, and quadratic bias potential. The above examines the biased sampling of a 2D Langevin equation (25) with the asymmetric double well potential energy $U(x, y) = k(x^4 - 11/3x^3 - 2x^2 + 11x + 25/3 + 7.5y^2)$, starting at position $(x_o, y_o) = (-1, 0)$ and conditioned to end in the circular region $(x_c, y_c) = (2.75, 0)$ with radius $R = 3$. The parameters are $L = 1$, $k = 0.1$, and $\sigma = \sqrt{2}$, with the number of runs being 10^3. (a) 1D slice (at $y = 0$) of the total potential $U^{tot} = U + 2U^d$, where U^d is either a linear biasing potential [Eq. (38)] or a quadratic biasing potential [Eq. (39)]. Adjusting the value k_d in the bias potential adjusts its steepness. (b) Percentage of runs with normalized sample weights $W > W^*$, where the threshold weight takes the values $W^* = (10^{-3}, 10^{-2}, 10^{-1})$.

If the sample path does not reach the desired end region $\boldsymbol{x}(L) \in \Omega_L$, the weight is identically equal to zero. Otherwise, the weight will be the exponential term listed above. An ideal biasing potential $U^d(\boldsymbol{x})$ is one that will give a weight distribution that is not very broad.

To compare the performance of the bridge from our approximation scheme to the above approach of adding a biased potential, we consider the $\mathcal{N}$-dimensional Langevin equation with potential $U(\boldsymbol{x}) = k(x_1^4 - 11/3x_1^3 - 2x_1^2 + 11x_1 + 25/3) + 7.5k(\sum_{i=2}^{\mathcal{N}} x_i)^2$ defined in Sec. III, where there is a barrier in the direction of x_1. We are interested in examining paths that start at one minimum $\boldsymbol{x}_0 = (-1, 0, 0, \ldots, 0)$ and end in a circular ball of radius $R = 3$ centered around the other minimum $\boldsymbol{x}_c = (2.75, 0, 0, \ldots, 0)$.

When adding a biased drift, we will consider the following three potentials: (1) zero total potential $(U^{tot} = U + 2U^d = 0)$; (2) linear potential

$$U^d(\boldsymbol{x}; k_d) = -k_d(x_1 - c), \tag{38}$$

where k_d controls the strength of the biasing and c is a constant; and (3) quadratic potential

$$U^d(\boldsymbol{x}; k_d, x_b) = k_d\left[(x_1 - x_b)^2 + \sum_{i=2}^{\mathcal{N}} x_i^2\right], \tag{39}$$

where k_d again controls the strength of the biasing and $(x_b, 0, 0, \ldots, 0)$ is the center of the target region. Figure 6(a) shows the total potential energy $U^{tot} = U + 2U^d$ when the biasing drift comes from the linear and quadratic potentials. For the linear potential, the resultant constant force will push all trajectories toward the end region; however, one should note that too large of a forcing (k_d) could cause the random walks to be pushed past the end region. On the other hand, for the quadratic potential, the trajectories will always be confined around the end region.

Figure 6(b) quantifies the sample weights W from these biased simulations. In particular, we quantify the percentage of runs with normalized weight larger than a threshold value $W^* = [10^{-1}, 10^{-2}, 10^{-3}]$ since these trajectories will be the ones that dominate any ensemble average. For the first method, the total potential is flattened $(U^{tot} = 0)$ such that trajectories undergo pure diffusion. The quality of runs by such an approach is found to be less efficient than the approximate bridge method proposed in this paper. This observation intuitively makes sense as the approximate bridge method [Eq. (27)] is equivalent to flattening the energy landscape followed by adding a time-dependent bias that pushes the trajectories toward the end region.

For the linear potential, we find that the best performance occurs when $k_d = 1$ based on Fig. 6(b). The performance of the weight deteriorates when further increasing the value of k_d. For instance, if one set $k_d = 10$, there are zero trajectories reaching the end region. For the quadratic potential, the best performance occurs when $k_d = 0.1$. We should point out that, for a linear or quadratic biasing potential, the selection of k_d is highly dependent on the specific problem. Overall, for the problem described in Fig. 6(b), the Brownian bridge using the approximation method outperforms all three approaches.

Next, we consider the effect of dimension $\mathcal{N}$. Figure 7 compares the method of the added drift using a quadratic potential $(k_d = 0.1)$ to the approximate Brownian bridge approach. When $k_d = 0.1$ is fixed, the method using an added drift fails when $\mathcal{N} > 6$ [Fig. 7(a)]. Furthermore, we examine the weight distribution and find that both approaches become less efficient as the dimension increases. However, the approximate bridge performs better than the added drift method [Fig. 7(b)].

V. TIME AND SPACE COMPLEXITY ANALYSIS

This paper discusses an approximation method to generate statistics of a continuous random walk conditioned to end in a given region of phase space. The main advantage of this approach is that it avoids an explicit solution to the Backward Fokker–Planck (BFP) equation that is often needed to generate such random walks. Here, we perform a time and space complexity analysis of the approximation scheme and compare these results to the standard Brownian bridge technique. We will examine the $\mathcal{N}$-dimensional Langevin Eq. (25), where $U(\boldsymbol{x})$ is the potential energy function and

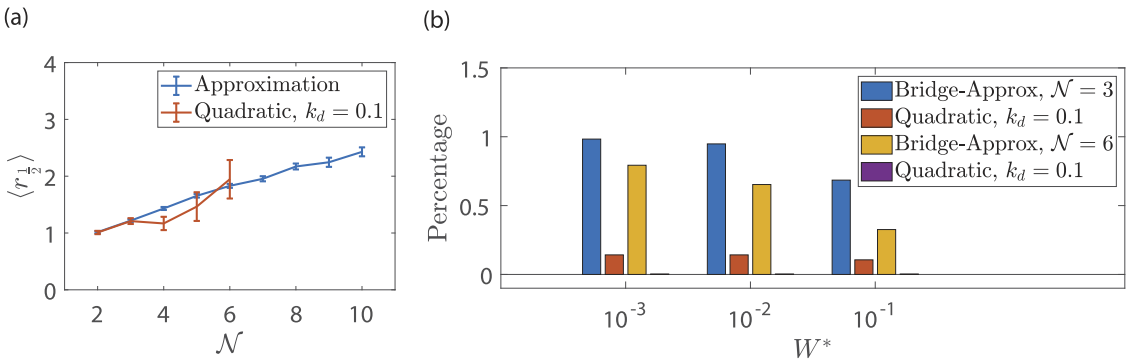

FIG. 7. Effect of dimension $\mathcal{N}$ on the approximate Brownian bridge and biased sampling with a quadratic potential [Eq. (39)]. The above examines an $\mathcal{N}$-dimensional Langevin equation with diffusion matrix $\boldsymbol{D} = \frac{1}{2}\sigma^2\boldsymbol{I}$ and potential energy $U(\boldsymbol{x}) = k(x_1^4 - 11/3x_1^3 - 2x_1^2 + 11x_1 + 25/3) + 7.5k(\sum_{i=2}^{\mathcal{N}} x_i)^2$, starting at position $\boldsymbol{x}_o = (-1, 0, 0, \ldots, 0)$ and conditioned to end in the $\mathcal{N}$-sphere centered at $\boldsymbol{x}_c = (2.75, 0, 0, \ldots, 0)$ with radius $R = 3$. The parameters are $L = 1$, $k = 0.1$, and $\sigma = \sqrt{2}$, and the number of runs is 10^3. (a) Ensemble average of $\langle r_{\frac{1}{2}} \rangle$—i.e., the distance of the path's midpoint from the origin for $\mathcal{N}$: 2 – 10. (b) Percentage of runs with normalized sample weights $W > W^*$, where the threshold weight takes the values $W^* = (10^{-3}, 10^{-2}, 10^{-1})$. We examine dimensions $\mathcal{N} = 3, 6$.

$\boldsymbol{D} = \frac{1}{2}\boldsymbol{B}\cdot\boldsymbol{B}^T$ is a constant diffusion matrix. We will examine paths that start at position $\boldsymbol{x} = \boldsymbol{x}_o$ at $s = 0$ and are conditioned to end in region $\boldsymbol{x} \in \Omega_N$ at $s = L$. It is assumed that the cost of computing the gradient of any function is $O(N_{dim})$ and the double gradient is $O(N_{dim}^2)$, where N_{dim} is the dimension of the phase space.

A. Standard Brownian Bridge

In the standard Brownian bridge technique, one adds a drift to the SDE to bias all paths to end in the desired region,

$$d\boldsymbol{x} = \left(-\frac{1}{2}\boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\nabla U(\boldsymbol{x}) + \boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\frac{\partial \ln(q)}{\partial \boldsymbol{x}}\right)ds + \boldsymbol{B}\cdot d\mathscr{B}, \quad (40)$$

where q is the solution of the Backward Fokker–Planck (BFP) equation (3). The procedure for generating the conditioned random walk is as follows:

1. Solve the BFP equation: Solve the BFP equation on a set of meshpoints in space $\boldsymbol{x} \in R^{(N_{dim})}$ and along the arc-length coordinate $0 < s < L$. Use the stored solution to compute the drift $\boldsymbol{u} = \boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\frac{\partial \ln q}{\partial \boldsymbol{x}}$ on the meshpoints. There are typically N_{time} meshpoints along the s-coordinate and N meshpoints along each dimension for the x coordinates. Thus, there are $N^{N_{dim}}$ meshpoints in space.
2. Generate Ito paths with modified drift: Generate N_{run} paths for the SDE using Euler–Maruyama integration. The integration will have N_{time} quadrature points over the coordinate $0 < s < L$. At any time step during integration, the additional drift $\boldsymbol{u} = \boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\frac{\partial \ln q}{\partial \boldsymbol{x}}$ is computed by interpolating the solution from the meshpoints in step 1.
3. Ensemble average: For an observable α, we compute the ensemble average of this quantity via $\langle\alpha\rangle = \frac{1}{N_{run}}\sum_i^{N_{run}}\alpha_i$, where α_i is the observable from run i.

From the above, the cost to perform the simulation is the cost for each of the steps. The time and space complexity of each step is described as follows:

1. Solving the BFP equation: A standard finite difference scheme will yield a computational cost of $(N_{time}N^{N_{dim}})(N_{dim} + N_{dim}^2) \sim O((N_{time}N^{N_{dim}})N_{dim}^2)$ as there are $N_{time}N^{N_{dim}}$ meshpoints (space + time), and one needs to compute the gradient ∇U and double derivative $\frac{\partial^2 q}{\partial \boldsymbol{x}\partial \boldsymbol{x}}$ at the cost of $O(N_{dim})$ and $O(N_{dim}^2)$ at each meshpoint, respectively. The space complexity is the order of the number of meshpoints, i.e., $O(N^{N_{dim}}N_{time})$. If instead one uses a state-of-the art technique to accelerate the solution of the Fokker–Planck equation, one can reduce the computational load considerably. For example, Proper Generalized Decomposition (PGD) approximates the solution to the BFP equation as a series of modes that are separable.[36] The number of modes $N_{mode} \leq 10$ for low dimensional problems $(N_{dim} \leq 2)$ will increase as the size of the dimensions N_{dim} increases. Here, solving the BFP equation has a time complexity of $O(N_{mode}^2 N N_{time} N_{dim}^2)$, and the space complexity is $O(N_{mode}N N_{time}N_{dim})$.
2. Generating Ito paths: The time complexity for generating paths is $O(2^{N_{dim}}N_{run}N_{time})$ when performing finite differences in step 1. For each of the N_{run} paths generated, there are N_{time} steps in the Euler–Maruyama integration scheme with each step taking $O(2^{N_{dim}})$ operations to perform the interpolation of computed hitting probability at a given position $\boldsymbol{x}$. The space complexity is $O(N_{dim}N_{run}N_{time})$ for storing all paths. When performing the PGD method, the time complexity is $O(2^{N_{dim}}N_{mode}N_{run}N_{time})$ and the space complexity is still $O(N_{dim}N_{run}N_{time})$.
3. Ensemble average: The time complexity is $O(N_{run})$, and the space complexity is $O(1)$.

B. Approximating scheme

In the approximating scheme, we add an approximate drift to the original random walk as follows:

$$d\boldsymbol{x} = \left(-\frac{1}{2}\boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\nabla U(\boldsymbol{x}) + \boldsymbol{B}\cdot\boldsymbol{B}^T\cdot\frac{\partial \ln(\psi)}{\partial \boldsymbol{x}}\right)ds + \boldsymbol{B}\cdot d\mathscr{B}, \quad (41)$$

where ψ is the approximate hitting probability computed from Eqs. (26) and (32) using the asymptotic properties of the BFP equation for the endpoint or end region, respectively. Since $q \neq \psi$, one reweights each path according to

$$W = \exp\left(\int_0^L \frac{1}{\psi}\mathscr{L}\psi ds\right), \quad (42)$$

where $\mathscr{L}$ is the backward Fokker–Planck operator [Eq. (3)]. The steps in performing this process are as follows:

1. Generate Ito paths: Generate N_{run} paths for the above SDE using the Euler–Maruyama integration with N_{time} quadrature points over the s-coordinate $0 < s < L$. The time complexity for performing this process is $O(N_{run}N_{time}N_{dim})$ as stated previously, with the space complexity being $O(N_{time}N_{dim}N_{run})$. Note that during the Ito integration, the time complexities for standard Brownian bridge and Approximate Brownian bridge are different, which is due to the fact that we are calculating the effective drift instead of performing interpolation.
2. Generate weight: Compute the weight W for each path using Eq. (9). The total cost is $O(N_{run} * N_{time}N_{dim}^2)$ as we have N_{run} paths, the integral in Eq. (42) is over N_{time} quadrature points, and the evaluation of operator $\mathscr{L}\psi$ involves the computation of double derivative $\frac{\partial^2\psi}{\partial \boldsymbol{x}\partial \boldsymbol{x}}$ at cost $O(N_{dim}^2)$. The space complexity is $O(N_{run})$.
3. Ensemble average: For an observable α, we compute $\langle\alpha\rangle = \frac{\sum_i^{N_{run}}\alpha_i W_i}{\sum_i^{N_{run}} W_i}$ over all runs with a time complexity of $O(N_{run})$ and space complexity $O(1)$.

C. Summary

Table I summarizes the time and space complexity of the approximating method and the standard Brownian bridge using the proper generalized decomposition (PGD) and the finite difference method to solve the Backward Fokker–Planck (BFP) equation. Generally, in most applications, the number of meshpoints along the s-coordinate and each spatial dimension are comparable $(N \sim N_{time})$. Figure 8 shows that if this is the case, the approximation method has a time complexity that scales as $O(N)$, while the proper generalized decomposition method (PGD)—the best case scenario

TABLE I. Time and space complexity of the approximation method, finite element algorithm with proper generalized decomposition (PGD),[16,36] and traditional finite difference method to solve the Brownian bridge problem.

Algorithm	Time complexity	Space complexity
Approximation method	$O(N_{dim}^2 N_{runs} N_{time})$	$O(N_{time} N_{dim} N_{run})$
PGD method	$O\left(N_{mode}^2 N N_{time} N_{dim}^2 + 2^{N_{dim}} N_{mode} N_{time} N_{run}\right)$	$O(N_{dim} N_{time} N_{mode} N + N_{dim} N_{run} N_{time})$
Finite difference method	$O\left(N_{time} N^{N_{dim}} N_{dim}^2 + 2^{N_{dim}} N_{time} N_{run}\right)$	$O(N^{N_{dim}} N_{time} + N_{dim} N_{run} N_{time})$
Algorithm ($N_{time} \approx N \gg 1$)	Time complexity	Space complexity
Approximation method	$O(N_{dim}^2 N_{runs} N)$	$O(N N_{dim} N_{run})$
Finite element method with PGD	$O(N_{mode}^2 N^2 N_{dim}^2)$	$O(N_{dim} N^2 N_{mode})$
Finite difference method	$O(N^{N_{dim}+1} N_{dim}^2)$	$O(N^{N_{dim}+1})$

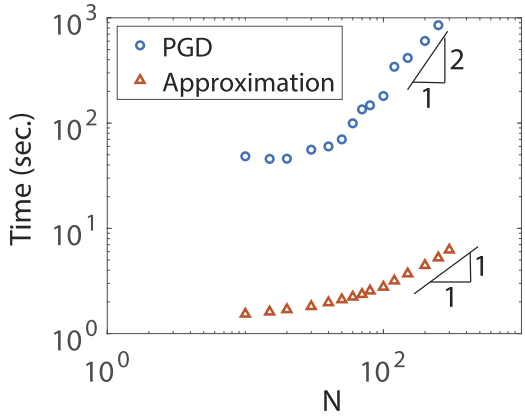

FIG. 8. Running time for solving the 1D Langevin equation with $U(x) = x^2$ using the Brownian bridge with two different solution methods: (1) approximation method and (2) finite element method with Proper Generalized Decomposition (PGD).[16,36] The running time is plotted with respect to the phase spatial resolution, N. Note that, in this case, the number of time steps is equal to the number of space steps ($N_{time} = N$). For the finite element method with PGD, the simulation time is dominated by the solution of the Backward Fokker–Planck (BFP) equation.[16]

for numerically solving the BFP—has a complexity of $O(N^2)$. The approximation method clearly has advantages when scaling to larger dimensions and more complex problems.

VI. CONCLUSION

Simulating stochastic processes with end control is critical in a variety of chemistry applications, such as understanding polymer conformations and characterizing reaction pathways. The Brownian bridge is a useful technique to generate such processes by adding a drift velocity to guide paths toward the correct regions in phase space. This drift is calculated from a hitting probability that satisfies a Backward Fokker–Planck (BFP) equation (a PDE) in terms of collective variables.

In this work, we continue to improve such bridge processes by proposing an approximation scheme instead of directly solving hitting probability from the BFP equation. If one biases a continuous random walk by an approximate drift $\boldsymbol{u}^{ap} = 2\boldsymbol{D} \cdot \frac{\partial \ln(\psi)}{\partial \boldsymbol{x}}$ and reweights each path by weight $W = \exp\left[\int_0^L \frac{1}{\psi} \mathscr{L} \psi ds\right]$, one can recover the exact statistics of the conditioned random walk. In these expressions, $\boldsymbol{D}$ is the diffusion matrix, ψ is an approximate solution to the BFP equation, and $\mathscr{L}$ is the BFP operator. We find that in many situations, most of the variation in the sample weight W occurs near the endpoint, and hence, only an accurate solution of ψ near the endpoint is required to obtain a reasonable approximation scheme. We show that the approximation scheme accurately captures the statistics of many conditioned random walks while drastically accelerating the sampling speed and efficiency. We also show that this approach offers reasonable improvement compared to other approaches using simple bias potentials. Finally, a time and space complexity analysis demonstrates that the approximation scheme scales well for higher dimensions and offers significant advantages compared to the traditional Brownian bridge approach. Finally, although we observe that the method can examine certain problems up to dimension $\mathcal{N} \sim O(100)$, an examination of the weight distribution shows that the weight generated by the approximation method becomes wider with the increasing dimension and increasing trajectory length. Future research would focus on refining such approximation methods for very high dimensional problems.[37–39]

ACKNOWLEDGMENTS

The authors acknowledge the financial support from the National Science Foundation (NSF Award No. CBET-2126230). The authors also thank the anonymous reviewers whose comments greatly improved this manuscript.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

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

REFERENCES

[1] M. P. Taylor, W. Paul, and K. Binder, "Phase transitions of a single polymer chain: A Wang–Landau simulation study," J. Chem. Phys. **131**(11), 114907 (2009).
[2] N. Bou-Rabee and E. Vanden-Eijnden, *Continuous-time Random Walks for the Numerical Solution of Stochastic Differential Equations* (American Mathematical Society, 2018), Vol. 256.
[3] L. Maragliano and E. Vanden-Eijnden, "A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations," Chem. Phys. Lett. **426**(1-3), 168–175 (2006).
[4] S. Habershon, "Sampling reactive pathways with random walks in chemical space: Applications to molecular dissociation and catalysis," J. Chem. Phys. **143**(9), 094106 (2015).
[5] I. Sharony, R. Chen, and A. Nitzan, "Stochastic simulation of nonequilibrium heat conduction in extended molecular junctions," J. Chem. Phys. **153**(14), 144113 (2020).
[6] W. E and E. Vanden-Eijnden, "Transition-path theory and path-finding algorithms for the study of rare events," Annu. Rev. Phys. Chem. **61**, 391–420 (2010).
[7] W. Vervaat, "A relation between Brownian bridge and Brownian excursion," Ann. Probab. **7**, 143–149 (1979).
[8] D. Joseph, "Conditional Brownian motion and the boundary limits of harmonic functions," Bull. Soc. Math. France **85**, 431–458 (1957).
[9] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales: Itô Calculus* (Cambridge University Press, 2000), Vol. 2.
[10] H. Touchette, "The large deviation approach to statistical mechanics," Phys. Rep. **478**(1-3), 1–69 (2009).
[11] H. Touchette, "Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels," J. Stat. Phys. **159**(5), 987–1016 (2015).
[12] R. Chetrite and H. Touchette, "Nonequilibrium microcanonical and canonical ensembles and their equivalence," Phys. Rev. Lett. **111**(12), 120601 (2013).
[13] R. Chetrite and H. Touchette, "Nonequilibrium Markov processes conditioned on large deviations," Ann. Henri Poincare **16**, 2005–2057 (2015).
[14] U. Ray, G. K.-L. Chan, and D. T. Limmer, "Exact fluctuations of nonequilibrium steady states from approximate auxiliary dynamics," Phys. Rev. Lett. **120**(21), 210602 (2018).
[15] S. N. Majumdar and H. Orland, "Effective Langevin equations for constrained stochastic processes," J. Stat. Mech.: Theory Exp. **2015**(6), P06039 (2015).
[16] S. Wang, D. Ramkrishna, and V. Narsimhan, "Exact sampling of polymer conformations using Brownian bridges," J. Chem. Phys. **153**(3), 034901 (2020).
[17] F. W. Wiegel, *Introduction to Path-Integral Methods in Physics and Polymer Science* (World Scientific, 1986).
[18] R. Quhe, M. Nava, P. Tiwary, and M. Parrinello, "Path integral metadynamics," J. Chem. Theory Comput. **11**(4), 1383–1388 (2015).
[19] G. Ciccotti and M. Ferrario, "Blue moon approach to rare events," Mol. Simul. **30**(11-12), 787–793 (2004).
[20] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, "Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics," ChemPhysChem **6**(9), 1809–1814 (2005).
[21] L. Sutto, S. Marsili, and F. L. Gervasio, "New advances in metadynamics," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**(5), 771–779 (2012).
[22] W. Shi and E. J. Maginn, "Continuous fractional component Monte Carlo: An adaptive biasing method for open system atomistic simulations," J. Chem. Theory Comput. **3**(4), 1451–1463 (2007).
[23] H. Orland, "Generating transition paths by Langevin bridges," J. Chem. Phys. **134**(17), 174114 (2011).
[24] S. Krishnaswami, D. Ramkrishna, and J. M. Caruthers, "Statistical-mechanically exact simulation of polymer conformation in an external field," J. Chem. Phys. **107**(15), 5929–5944 (1997).
[25] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer Science & Business Media, 2013), Vol. 23.
[26] M. Rubinstein, R. H. Colby *et al.*, *Polymer physics* (Oxford University Press, New York, 2003), Vol. 23.
[27] C. F. Abrams, N.-K. Lee, and S. P. Obukhov, "Collapse dynamics of a polymer chain: Theory and simulation," Europhys. Lett. **59**(3), 391 (2002).
[28] M. Bohdanecky, "New method for estimating the parameters of the wormlike chain model from the intrinsic viscosity of stiff-chain polymers," Macromolecules **16**(9), 1483–1492 (1983).
[29] R. Schöbel and J. Zhu, "Stochastic volatility with an Ornstein–Uhlenbeck process: An extension," Rev. Finance **3**(1), 23–46 (1999).
[30] G. Ascione, Y. Mishura, and E. Pirozzi, "Fractional Ornstein–Uhlenbeck process with stochastic forcing, and its applications," Methodol. Comput. Appl. Probab. **23**, 53 (2021).
[31] D. Dufresne, "The integral of geometric Brownian motion," Adv. Appl. Probab. **33**, 223–241 (2001).
[32] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, 1992), Vol. 1.
[33] D. T. Gillespie, "Exact numerical simulation of the Ornstein–Uhlenbeck process and its integral," Phys. Rev. E **54**(2), 2084 (1996).
[34] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory* (Springer Science & Business Media, 2013).
[35] J. A. Green and D. V. Shalashilin, "Benchmark calculation for tunnelling through a multidimensional asymmetric double well potential," Chem. Phys. Lett. **641**, 173–180 (2015).
[36] A. Amine, B. Mokdad, F. Chinesta, and K. Roland, "A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids," J. Non-Newtonian Fluid Mech. **139**(3), 153–176 (2006).
[37] J. P. Nilmeier, G. E. Crooks, D. D. Minh, and J. D. Chodera, "Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation," Proc. Natl. Acad. Sci. **108**(45), E1009–E1018 (2011).
[38] A. Das, *et al.* "Reinforcement learning of rare diffusive dynamics," J. Phys. Chem. **155**(13), 134105 (2021).
[39] J. Yan, H. Touchette, and G. M. Rotskoff, "Learning nonequilibrium control forces to characterize dynamical phase transitions," Physical Review E **105**(2), 024115 (2022).