Efficient generation of barrier crossing trajectories using approximate Brownian bridges

George Curtis ⦿, Doraiswami Ramkrishna, and Vivek Narsimhan ⦿

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

We examine continuous random walks that are conditioned to reach one region before another. These conditioned processes are used to generate stochastic trajectories for barrier crossing events, which are generally rare and difficult to sample. The processes are generated using a Brownian bridge technique, resulting in near perfect sampling efficiency without accruing error in the conditional statistics of the process. The construction requires the hitting probability or committer function, which is a solution to the backward Fokker-Planck equation, a partial-differential equation that can be difficult to solve through general means. Therefore, we derive a one-dimensional approximation through asymptotic methods for barrier crossing trajectories. We show that this approximation has a simple analytical representation and approaches the true solution as the barrier height increases. Brownian bridge trajectories generated with this approximate solution are then shown to result in accurate conditional statistics when used in conjunction with importance sampling, even in the case when potential energy barriers are not large. We show this idea's effectiveness by simulating rare events in a stochastic chemical reaction network (Schögl reaction) with multiple steady states. This methodology shows great promise for future implementation to simulate rare barrier crossing events for a wide variety of physical processes.

I. INTRODUCTION

Continuous random walks (CRWs), processes characterized by diffusion and drift, are fundamental in modeling a broad spectrum of physical phenomena across many fields. These models provide valuable insight into the dynamics found in heat and mass transfer [1], polymer physics [2–4], nucleation [5], and chemical reaction processes [6]. In many of these processes, the most intriguing events are rare transitions between metastable states; however, these events (as the name implies) occur rarely and are therefore particularly difficult to sample directly. This paper proposes an efficient method to generate such events, specifically paths that reach one region before another.

Consider a stochastic process $\mathbf{X}_t$ whose probability density function $p(\mathbf{x}, t)$ satisfies the Fokker-Planck equation (FP) [7],

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}[A_i p] + \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial}{\partial x_i}\left[D_{ij}\frac{\partial p}{\partial x_j}\right], \quad (1)$$

where D_{ij} are the components of the diffusion tensor $\mathbf{D}$, and $\mathbf{A} = -\mathbf{D}\nabla_{\mathbf{x}}V$ represents the drift from a potential energy surface (PES) $V(\mathbf{x}, t)$. The FP equation (1) is associated with an Itö stochastic differential equation (SDE) [8,9],

$$d\mathbf{x}(t) = [\mathbf{A}(\mathbf{x}, t) + \nabla_{\mathbf{x}} \cdot \mathbf{D}]dt$$
$$+ \mathbf{B}(\mathbf{x}, t)d\mathbf{W}; \; \mathbf{x}(0) = \mathbf{x}_0, \quad t \geq 0, \quad (2)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ represents the state at time t, and $\mathbf{B}$ is a matrix related to the diffusion tensor as $\mathbf{D} = \frac{1}{2}(\mathbf{BB}^T)$ [10], often obtained by Cholesky decomposition [11]. The symbol $d\mathbf{W} \in \mathbb{R}^n$ is a Wiener process, where $\mathbf{W}(t)$ is an n-dimensional standard Brownian motion (BM) with independent Gaussian increments, $\mathbf{W}(t + s) - \mathbf{W}(s) \sim \mathcal{N}(0, t)$ [12].

Let Ξ be the path ensemble for (2), i.e., the set of all possible trajectories. We are interested only in the subset $\Gamma \subset \Xi$, the set of paths that reach one region of phase space before another. That is, if we define two regions $\Omega_1, \Omega_2 \subset \Omega$ with $\Omega_1 \cap \Omega_2 = \emptyset$, then we are interested only in the trajectories for which $\mathbf{X}_t \in \{\mathbf{x}(t) \in \Omega_1, \mathbf{x}(s) \notin \Omega_2 \mid \forall \, s \in [0, t]\}$, i.e., the first hitting time to region Ω_1 is less than for region Ω_2, $T_1 < T_2$. These types of paths occur often in reaction dynamics; for example, a set of reactants produces two products A and B, and only the reaction pathway to A is of interest. Specifically, we are interested in barrier crossing trajectories (BCTs), which provide insight into the dynamics and properties of various chemical and physical transformations. Unfortunately, a phenomenon known as the separation of timescales occurs, i.e., a long waiting time within a metastable state and a short barrier crossing event [13], so the BCT is rare and computationally expensive to generate.

The foregoing necessitates more efficient sampling algorithms designed to constrain the stochastic system, increasing the rate at which BCTs are generated. Various techniques have been implemented to achieve this, such as umbrella sampling [4,14,15], transition path sampling [16], action-based methods [17], adaptive forcing [18], and string methods [19–21], each coming with their own set of advantages and disadvantages. For many of these, their accuracy is highly dependent on an initial guess; for example, transition state theory highly relies on the choice of a dividing surface in the reaction region [22], and string theories highly depend on the initial guess for the minimum free energy path (MFEP) [21] and only yield a single transition path. For example, consider the simple toy problem illustrated in Fig. 1. Let the energy wells represent three different metastable states of some barrier crossing process. We are interested in only one of the transitions between
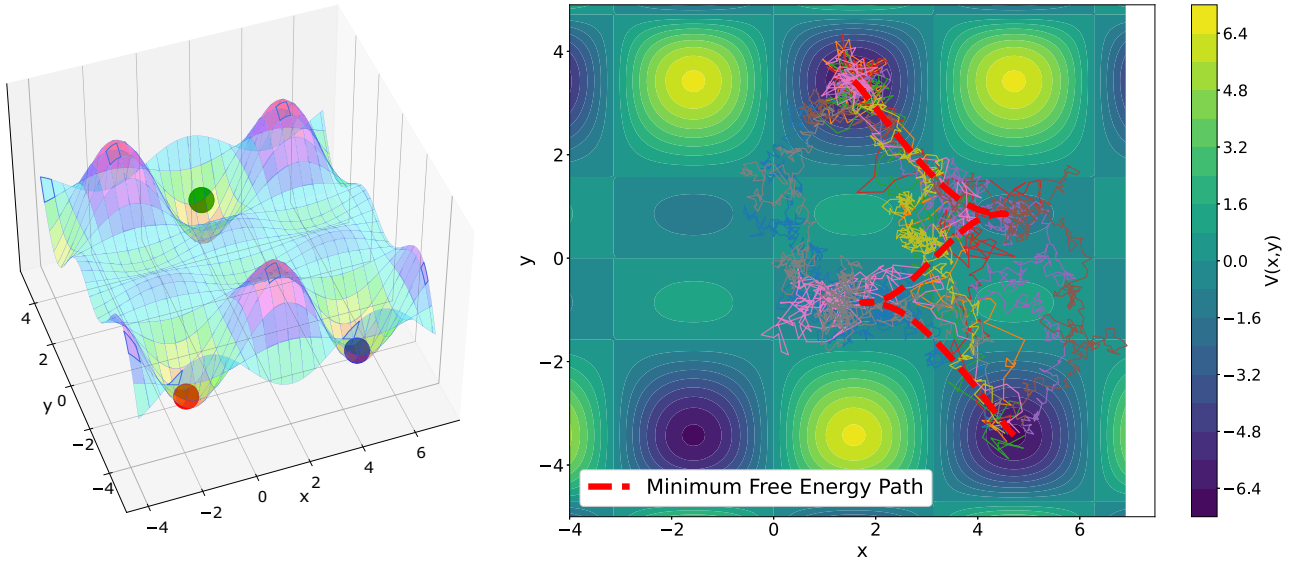
FIG. 1. Example surface for paths which cross from one potential energy minima (upper green dot) to another (rightmost blue dot) before reaching another (leftmost red dot). The minimum free energy path (thick red line) is shown on the contour plot as well as a few sample trajectories. Note that not all trajectories follow the minimum free energy path.

energy wells, i.e., sampling the subset of paths that start at the topmost energy well and reach the bottom right well before the bottom left well.

The string method shows the MFEP for these trajectories, but does not capture all such paths that the system can take to reach that state. So, rate predictions made using this method can ignore some of the possible pathways, resulting in an underestimation of the rates, and may not illuminate possible alternative mechanisms and intermediate states dividing these two surfaces. One way to avoid such a disadvantage is to use a conditioned random walk, i.e., a Brownian bridge (BB). A BB is a type of Doob h-transform [23], which generates only trajectories that satisfy the constraint, which in this case is $\{\mathbf{x}(t) \in \Omega_1, \mathbf{x}(s) \notin \Omega_2 \mid \forall \ s \in [0, t]\}$. This bridge process is illustrated in Fig. 2. Out of the entire path ensemble, Ξ, the bridge process selects only those paths that satisfy the constraint that the system reaches a state within the shaded region. The bridge process is very general and can be applied to many types of stochastic systems, such as extremely high or low temperature (noise) systems and those far from equilibrium, and it may also be used to study a wide variety of rare events depending on the constraints applied to the system. The bridge process has far-reaching applicability to the study of rare event processes; however, its main drawback is that it requires solving the backward Fokker-Planck equation [10], which for complex potential energy surfaces or in high-dimensional spaces becomes cumbersome to numerically calculate. Therefore, the main idea of this paper is the development of efficient and accurate ways to calculate its solution in order to construct these bridges and sample rare events.

II. BROWNIAN BRIDGE CONSTRUCTION

A Brownian bridge modifies the underlying stochastic process (2) to focus on a specific subset of paths. Consider

a stochastic process $\mathbf{X}_t$ defined within a domain Ω and let Ξ represent the set of all possible trajectories or path ensemble. Our interest lies only in the set of paths Γ that reach a region Ω_1 before Ω_2. Since $\Gamma \subseteq \Omega$, the bridge process is associated with an entropic penalty. To understand this phenomenon, let P describe the probability measure of Ξ [24]; then define its entropy $S(P)$ using the Kullback-Leiber (KL) divergence. A property of KL divergence is that conditioning reduces the entropy, so $S(P|\Gamma) \leqslant S(P)$ [25]. Consequently, there is an accompanying entropic force caused by the set partition $\Xi = \{\Gamma, \ \Xi \setminus \Gamma\}$ [26]. This entropic force is accounted for in the bridge construction as an additional biasing force added to the drift term of (2).
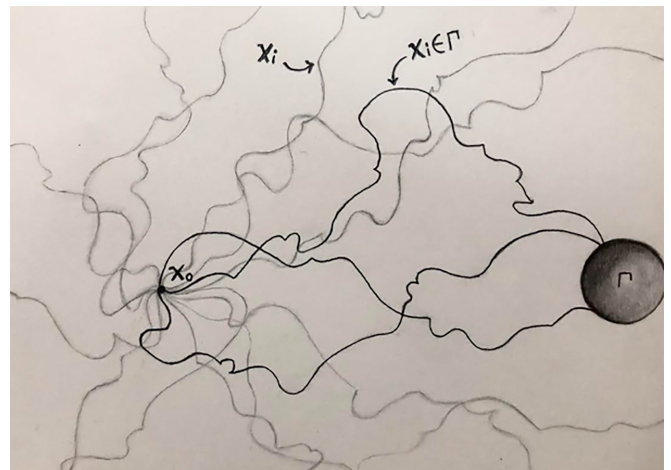

FIG. 2. Path ensemble, $\{\mathbf{X}_i\} \in \Xi$, for a stochastic system beginning at a state $\mathbf{x}_0$. We are interested in the subset of paths, $\Gamma \subset \Xi$, that reach a specified region of the phase space, represented by the shaded circle. The bridge process generates only such paths so that each sample $\mathbf{X}_i(t) \in \Gamma$.
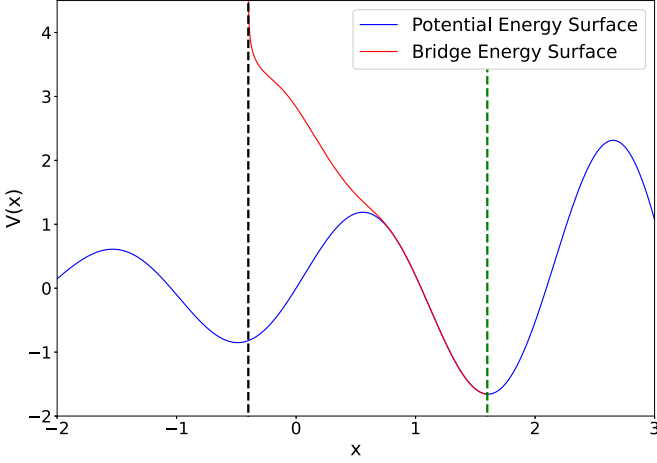
FIG. 3. Comparison of a potential energy surface (blue) and bridge surface (red). The bridge surface is the equivalent energy surface for the conditional sampling of paths that reach $x_h = 1.6$ (green line) before reaching $x_a = -0.4$ (black line), for any initial point $x_0 \in (-0.4, 1.6)$.

Thus, the modified SDE describing the bridge processes satisfies [3,4,10]

$$d\mathbf{x}^{Br} = [\mathbf{A} + \nabla_{\mathbf{x}} \cdot \mathbf{D} + (\mathbf{B}\mathbf{B}^T)\nabla_{\mathbf{x}} \ln q(\mathbf{x}, t)]dt + \mathbf{B}d\mathbf{W}, \quad (3)$$

where $q(\mathbf{x}, t)$ is the hitting probability. We define that probability as

$$q(\mathbf{x}, t) = \mathbb{P}[T_1 < T_2 | \mathbf{x}(t)], \quad (4)$$

where T_1 and T_2 are the first passage times to regions Ω_1 and Ω_2, respectively, i.e., the probability that the remaining path for the original SDE (2) reaches Ω_1 before Ω_2, given that it starts at $\mathbf{x}(t)$. The hitting probability satisfies the backward Fokker-Planck (BFP) equation, and since the hitting time is not specified, we use the time homogeneous equation, that is, $\frac{\partial q}{\partial t} = 0$. The BFP equation is given by [27]

$$(\mathbf{A} + \nabla_{\mathbf{x}} \cdot \mathbf{D}) \cdot \nabla_{\mathbf{x}} q + \frac{1}{2}(\mathbf{B}\mathbf{B}^T) : \nabla_{\mathbf{x}}\nabla_{\mathbf{x}} q = 0,$$
$$q(\mathbf{x} \in \partial\Omega_1) = 1; \quad q(\mathbf{x} \in \partial\Omega_2) = 0, \quad (5)$$

where $\partial\Omega_1$ and $\partial\Omega_2$ are the boundaries for regions 1 and 2.

We note that the bridge SDE (3) takes the same form as the original SDE (2) with a modified drift, $\mathbf{A}^* = \mathbf{A} + (\mathbf{B}\mathbf{B}^T)\nabla_{\mathbf{x}} \ln q$. Take an overdamped Langevin equation where $\mathbf{A} = -\frac{1}{2}(\mathbf{B}\mathbf{B}^T)\nabla_{\mathbf{x}}V(\mathbf{x}) = -\mathbf{D}\nabla_{\mathbf{x}}V(\mathbf{x})$, where $V(\mathbf{x})$ is a PES; then the bridge is equivalent to sampling on a modified PES, $V^* = V - 2\ln q$. Thus, sampling on the modified PES V^* is equivalent to conditional sampling on the original surface V, for paths that reach region Ω_1 before Ω_2. In Fig. 3, we show what the modified PES would be for one-dimensional (1D) diffusion ($B = 0.5$) on $V(x) = 2\sin(3x)e^{\frac{x}{\pi}}$, for the subset of paths that reach $x_h = 1.6$ (bottom of right well) before they reach $x_a = -0.4$ (bottom of left well). Note that as we approach the avoidance region, $x \to x_a$, we have $V^*(x) \to \infty$. The singularity implies that it would take infinite energy for the system to reach the region we wish to avoid; therefore, all trajectories sampled from this bridge must first reach the hitting location, x_h.

A strength of calculating these modified potential energy surfaces is that it retains a great deal of flexibility. For a given hitting and avoiding region, the starting location can be placed anywhere, $\mathbf{x}(0) \in \Omega$, as long as $\mathbf{x}(0) \notin \Omega_1 \cup \Omega_2$, without altering the modified surface.

III. DETERMINATION OF HITTING PROBABILITY

To implement the bridge construction, we require an efficient and accurate solution technique to the BFP equation. In this section, we develop an approximation method to quickly and accurately determine the hitting probability. We derive an analytical approximation for the 1D time-independent BFP equation through asymptotic expansions, which are valid in the case of either larger potential energy barriers or in the low-noise limit. We call this approximation technique the reflection method since it results in simply reflecting the potential energy surface in specific locations.

A. Asymptotic analysis: Reflection method

Let $\mathbf{X}$ represent a trajectory of a 1D stochastic process and $x(t)$ represent the state of the process at time t. The system is governed by

$$dx = -\frac{dV}{dx}dt + BdW, \quad (6)$$

where $V(x)$ is a PES and, for simplicity, we assume B is constant; however, B can be made to depend on x with very little modification to this theory. We are interested in the subset of trajectories where $\mathbf{X} \in \{x(t) = x_h | x(0) = y, x(s) \neq x_a \; \forall \; s < t\}$, where y is the initial position anywhere in the interval $y \in (x_a, x_h)$. The associated BFP equation is then

$$-\frac{dV}{dx}\frac{dq}{dx} + \frac{1}{2}B^2\frac{d^2q}{dx^2} = 0,$$
$$q(x_a) = 0; \quad q(x_h) = 1. \quad (7)$$

After integrating twice, we find that the solution can be written as

$$q(x) = \mathcal{N}\int_{x_a}^{x} e^{\beta V(x')}dx', \quad (8)$$

where $\beta = \frac{2}{B^2}$ and $\mathcal{N}$ is a normalization constant. The SDE for the conditioned process is then

$$dx^{Br} = \left(-\frac{dV}{dx} + B^2\frac{d\ln q}{dx}\right)dt + BdW, \quad (9)$$

so the modified drift for capturing the conditional statistics is $A^* = -\frac{dV}{dx} + B^2\frac{d}{dx}\ln q$. The analytical solution to (8) can only be found for simple PES and so we wish to find an analytical representation of $q(x)$ that approximates the exact solution. Assume that $\sup_{y,z \in (x_a, x_h)} \beta|V(y) - V(z)| \gg 1$, that is, the system is either in the low-noise limit or the existence of a large potential energy barrier. Then examine the Laplace-type integral,

$$I(x; \beta) = \int_{x_a}^{x} e^{\beta V(x')}dx', \quad (10)$$

which for $\beta \gg 1$ can be accurately approximated by the Laplace method [28,29]. Then define the following sets:

$$M = \left\{ x_1, x_2, \ldots, x_N \mid x_i \in (x_a, x_h), \frac{dV(x_i)}{dx} = 0, \right.$$

$$\left. \times \frac{d^2 V(x_i)}{dx^2} < 0 \right\},$$

$$m = \{\eta_i \in (x_i, x_{i+1}) \mid V(\eta_i) = V(x_i)\}, \quad (11)$$

where M is the set of all locations corresponding to local maxima of $V(x)$ and the set m is the corresponding points on the PES between the two local maxima and has a value equal to the previous maximum value. Note that since $\beta V(x_1) \gg 1$, $I(x)$ is sharply peaked at $x = x_1$. Here we split into two cases: *Case 1*: $x < x_1$, before reaching the potential energy maxima; and *Case 2*: $x_1 < x < \eta_1$, after reaching the potential energy maxima but before growing larger than that maxima.

Case 1: $x < x_1$. At $x < x_1$ before reaching the first maxima, $V'(x) > 0 \, \forall \, x \in (x_a, x_1)$, and therefore $I(x)$ is sharply peaked at x. Thus, we can asymptotically expand the exponent around x,

$$I(x) \approx \int_{x_0}^{x} e^{\beta[V(x)+V'(x)(y-x)]} dy \approx e^{\beta V(x)} \left(\frac{e^{\beta V'(x)(x-x_0)} - 1}{\beta V'(x)} \right)$$

$$\approx e^{\beta V(x)}, \quad (12)$$

so $I(x) \approx e^{\beta V(x)}$ for all $x \in [0, x_1]$.

Case 2: $x_1 < x < \eta_1$. For $x_1 < x < \eta_1$, $V(x_1) > V(y) \, \forall \, y \in (x_1, \eta_1)$ so $I(x)$ is sharply peaked at x_1. Thus, we asymptotically expand the exponent around x_1, $V(y) \approx V(x_1) + \frac{V''(x_1)}{2}(y - x)^2$, and substitute this into $I(x)$ to yield

$$I(x) \approx \int_{x_1}^{x} e^{\beta[V(x_1) + \frac{1}{2} V''(x_1)(y-x_1)]^2} dy,$$

$$= \frac{\sqrt{\pi}}{2} \{ \text{erf}[\xi(x_1 - x_0)] + \text{erf}[\xi(x - x_1)] \} e^{\beta V(x_1)},$$

where $\xi = \sqrt{\frac{-\beta V''(x_1)}{2}}$. Then, for $\xi(x - x_1) \gg 1$, we have

$$I(x) \approx \sqrt{\pi} e^{\beta V(x_1)} = \text{const}, \quad (13)$$

so $I(x)$ is constant for $x \in (x_1, \eta_1)$. We repeat this same analysis for each pair (x_i, η_i) so

$$I(x) \approx \begin{cases} e^{\beta V(x)}; & \forall x \in (x_i, \eta_i), \ i = 1, 2, \ldots, N \\ \sqrt{\pi} e^{\beta V(x_i)}; & \forall x \in (\eta_i, x_{i+1}), \ i = 1, 2, \ldots, N-1, \end{cases} \quad (14)$$

which results in the approximation

$$A^* = A + B^2 \ln I(x)$$

$$= \begin{cases} -A; & \forall x \in (x_i, \eta_i), \ i = 1, 2, \ldots, N \\ A; & \forall x \in (\eta_i, x_{i+1}), \ i = 1, 2, \ldots, N-1. \end{cases} \quad (15)$$

This formulation can be extended to nonconstant diffusion problems as well. The simple change is that the sets m, M will no longer depend on the maxima of $\beta V(x)$, but instead depend on the maxima of $\int \frac{2V'(x)}{B^2(x)} dx$, and will still result in the same
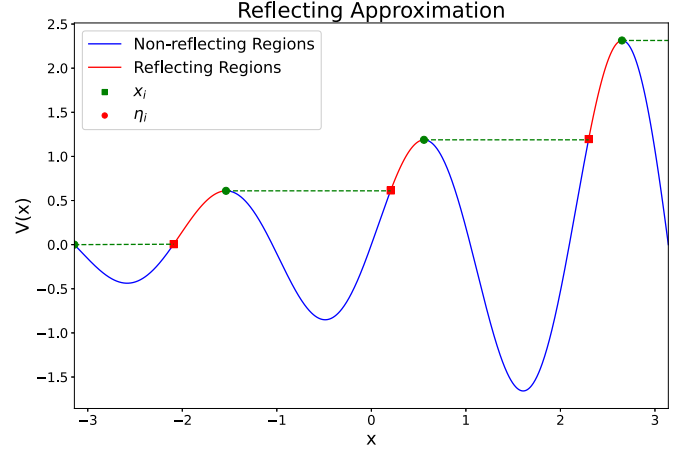

FIG. 4. Visual representation of asymptotic regions on a potential energy surface for trajectories which reach $x = \pi$ before they reach $x = -\pi$. Case 1 holds in the red (reflecting) regions and case 2 holds in the blue (nonreflecting) regions, respectively. Green markers are placed at each $x_i \in M$ and red markers are placed at each $\eta_i \in m$.

reflection behavior. To illustrate and better explain the reason for defining sets M and m, in Fig. 4 we show where each are to be placed and the regions where the potential surface should be reflected.

B. Reflected potential

From the previous result in (14), we can begin to construct the approximate bridges. Plugging the approximation into the SDE in (3) yields the modified drift velocity as

$$A^* \approx \begin{cases} -\frac{dV}{dx} & \text{if } x \in (x_i, \eta_i) \\ +\frac{dV}{dx} & \text{if } x \in (\eta_i, x_{i+1}), \end{cases} \quad (16)$$

which clearly implies that the modified potential energy surface is simply reflected about the potential energy maxima up to a constant,

$$V^* = \begin{cases} V(x) & \text{if } x \in (x_i, \eta_i) \\ -V(x) + c_i & \text{if } x \in (\eta_i, x_{i+1}), \end{cases} \quad (17)$$

where c_i is a set of constants that are used to enforce the continuity of $V^*(x)$.

In Fig. 5, we examine a 1D diffusion ($B = 1$) over a double-well potential, $V(x) = U_0[4(x^4 - x^2) + 1]$. We plot the modified potential energy surface (bridge surface) when one examines the subset of paths that reach the rightmost well ($x_h = \frac{1}{\sqrt{2}}$) before the leftmost well ($x_a = \frac{-1}{\sqrt{2}}$), as well as the potential energy surface from the approximation in (17) that comes from "reflecting" the potential energy surface in specific regions.

Note that as the barrier height increases, the reflected bridge surfaces converge to the exact bridge surface. The sole difference between the reflected and exact bridge for very large potential barriers is that the reflected bridge does not exhibit the singular behavior near the avoidance region, which forfeits the impossibility of a trajectory reaching that location. Those trajectories become extremely rare nonetheless.
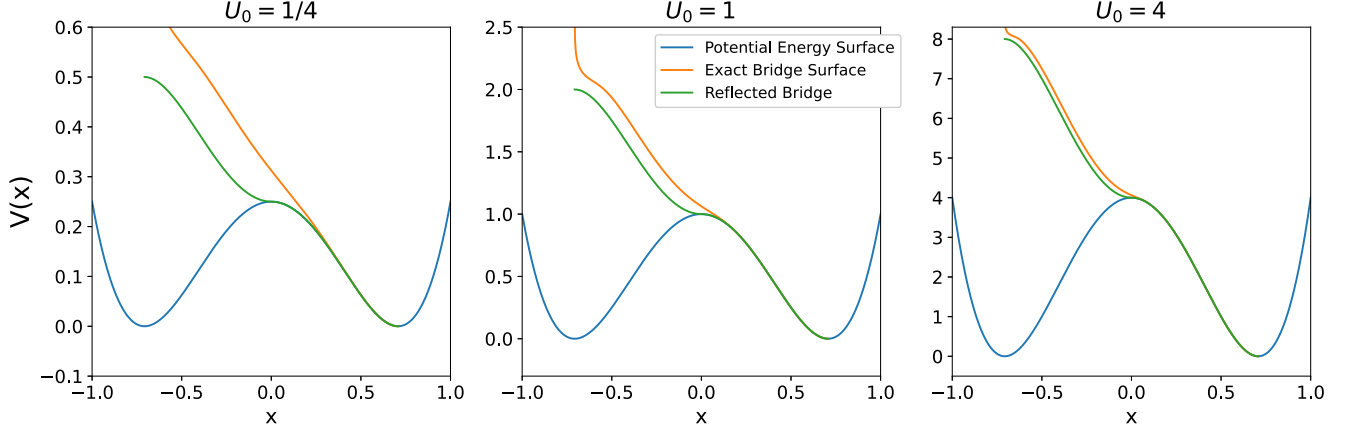
FIG. 5. Simple barrier crossing over 1D potential, $V(x) = U_0[4(x^4 - x^2) + 1]$ [blue (bottom) line], with diffusion, $B = 1$. Comparison of the exact bridge [orange (upper) line] with the reflection approximation [green (middle) line], over several different barrier heights, for the subset of paths which reach $x_h = \frac{1}{\sqrt{2}}$ before reaching $x_a = \frac{-1}{\sqrt{2}}$. As the barrier height U_0 increases, the approximation converges to the integral solution.

As the barrier height decreases, this reflection method becomes less viable. However, in our previous work [10], we found an importance sampling method to correct for errors incurred by the approximation. Let q^* be an approximation of the hitting probability q, and define the functional

$$G[\ln q^*] = \frac{\partial \ln q^*}{\partial t} + \nabla_{\mathbf{x}} V \cdot \nabla_{\mathbf{x}} \ln q^* + \frac{1}{2}(\mathbf{B}\mathbf{B}^T) :$$
$$\times [\nabla_{\mathbf{x}} \nabla_{\mathbf{x}} \ln q^* + (\nabla_{\mathbf{x}} \ln q^*) \cdot (\nabla_{\mathbf{x}} \ln q^*)^T], \quad (18)$$

where, as $q^* \to q$, then $G[\ln q^*] \to 0$. We then assign weights w_i to each trajectory $\mathbf{X}_t$ through path integration,

$$w_i = \exp\left\{\int G[\ln q^*][x_i(t)]dt\right\}. \quad (19)$$

So for any observable f of the trajectories, the reweighted expectation is

$$\mathbb{E}[f] = \frac{\sum_i f_i w_i}{\sum_i w_i}. \quad (20)$$

Thus, we can use the approximate hitting probability to sample the bridge, and any errors in the conditional statistics can be alleviated using this weighted average. A drawback of this importance sampling scheme is that a bad approximation leads to a wide variation of the weights w_i, and therefore to incorrect conditional statistics. Using the reflection method sidesteps this issue, producing a valid approximation in all but the most extreme cases.

We note that the problem formulation differs from that in Wang *et al.* [10] as they derived a method to produce accurate conditional statistics for an SDE that had a fixed time horizon. In fixed time horizon problems, one needs to approximate the solution to a time-dependent BFP equation in the importance sampling procedure, typically using a singular perturbation expansion near the endpoint. In the problem in this paper, we are examining an SDE where the time horizon is not fixed, i.e., the first hitting time occurs randomly. The BFP equation (5) in this case is time independent, and the appropriate approximation to use in the importance sampling procedure involves an asymptotic expansion in an interior region of the domain.

IV. RESULTS AND ANALYSIS

In this section, we analyze barrier crossing trajectories for a few different potential energy surfaces and compare brute-force sampling, bridge sampling, and sampling on the approximate bridge, i.e., using the reflection method. For clarity we define these methods (for constant matrix $\mathbf{B}$) as

$$d\mathbf{x} = \mathbf{A}(\mathbf{x}, t)dt + \mathbf{B}d\mathbf{W},$$
$$d\mathbf{x}^{Br} = [\mathbf{A}(\mathbf{x}, t) + (\mathbf{B}\mathbf{B}^T) \cdot \nabla \ln q]dt + \mathbf{B}d\mathbf{W},$$
$$d\mathbf{x}^{Br}_{\text{approx}} = \pm\mathbf{A}(\mathbf{x}, t)dt + \mathbf{B}d\mathbf{W}. \quad (21)$$

The sign on the latter is selected based on the results from the reflection method.

A. Single barrier

We begin by looking at a normal double-well potential, $V(x) = k(x^4 - x^2)$, as shown in Fig. 5, where the barrier height is U_0. We wish to sample the subset of trajectories that reach the rightmost minima $x_h = \frac{1}{\sqrt{2}}$ before reaching the leftmost one, $x_a = \frac{-1}{\sqrt{2}}$. To accomplish this, we compare the stopping times of brute-force sampling, bridge sampling, and sampling using the reflection method. The latter is also subjected to the reweighting procedure outlined in Ref. [10].

In Fig. 6, we show the results of sampling such paths as we increase the barrier height. Note that brute-force sampling is inefficient, as the ratio of samples that reach the hitting region first compared to the total number of samples is low and decreases exponentially with increasing barrier height. Alternatively, sampling on the bridge surface boasts perfect sampling efficiency. We also see that the reflection method has a much greater efficiency than that of brute-force sampling. As the barrier height increases, the brute-force sampling efficiency falls, whereas the reflection method improves as it converges to the real bridge. Additionally, the stopping times are also in good agreement. Notably, the reflection method also yields very good results after reweighting, matching the brute-force samples nearly perfectly. Thus, this approximation
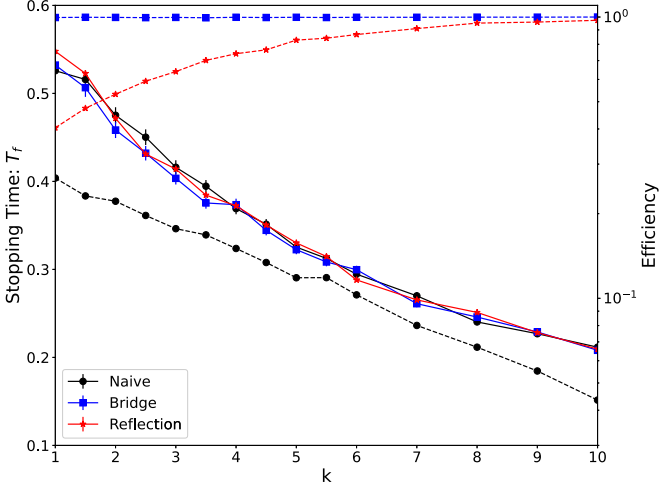
FIG. 6. Statistics of conditional path sampling on a potential energy surface $V(x) = k(x^4 - x^2)$, where k represents the energy barrier. Conditional sampling was performed for paths that reach $x_h = \frac{1}{\sqrt{2}}$ before reaching $x_a = \frac{-1}{\sqrt{2}}$ and start at $x(0) = -0.3$. The stopping time T_f (solid lines) is defined as the time at which the paths reach x_f, and the efficiency (dotted lines) is defined as the ratio of the number of successful paths to the total number of samples. Sampling is performed by brute force (naïve), bridge (numerically integrated), and the reflection approximation that was reweighted in Eqs. (18)–(20). The error bars indicate the standard error in the samples for successful paths $N = 1000$.

method can be used even in cases where the large barrier assumption does not hold, as long as the reweighting procedure is carried out. It should be noted that if the reweighting method is not carried out, these statistics do not match well with the brute-force results unless the barrier height is sufficiently large.

B. 1D multiple barriers

In Fig. 7, we choose a PES with several potential energy barriers (green), $V(x) = 15x \cos(x) \sin(3x)$, with diffusion
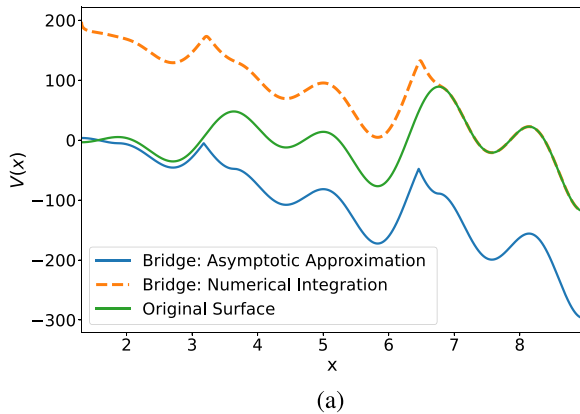
constant $D = \frac{1}{2}B^2 = 1$. We construct a bridge that conditions the trajectories to reach $x_a = 8.97$ before they reach $x_a = 1.31$, by direct numerical integration (dotted orange) and the reflection method (blue). Note that the derivative of $V(x)$ is the actual drift force when sampling trajectories; therefore, we can arbitrarily add any constant to $V(x)$ without altering the system's dynamics. In Fig. 7(b), the original PES is removed and the bridge surfaces are shifted to overlap. In this case, where the barriers are very large $U_0 \sim O(10\text{--}100)$, the reflection approximation lies almost entirely along the exact bridge surface, confirming that the multibarrier approximation is accurate. Trajectories sampled from these surfaces will yield only such paths that reach x_h before x_a. However, there is considerable variability in such trajectories. This is caused by the presence of large internal potential energy barriers. As such, there can be extended waiting times within the interior wells before escaping to adjacent wells. This variability contributes to a different sort of sampling inefficiency, which the bridge surfaces generated via the time-homogeneous BFP are not equipped to handle.

C. Schögl reaction example

In the last example, we show that Brownian bridges (BBs) can act as a valuable tool to simulate chemical reaction systems. Consider the Schögl reaction system [30] defined by

$$A + 2X \underset{k_1}{\overset{k_2}{\rightleftharpoons}} 3X, \quad B \underset{k_3}{\overset{k_4}{\rightleftharpoons}} X, \tag{22}$$
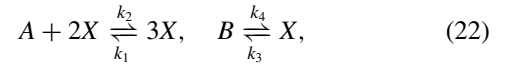
where A, B, and X are chemical species and the reaction rates are $k_1 - k_4$. From the master equation that satisfies the above reaction system, we can determine a chemical Langevin equation. Let $p_n(t) = \mathbb{P}\{n_X(t) = n\}$ be the probability of having n molecules of species X at time t; then the master equation governing the system is

$$\frac{dp_0}{dt} = \mu_1 p_1 - \lambda_0 p_0, \tag{23}$$

$$\frac{dp_n}{dt} = \lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1} - (\lambda_n + \mu_n) p_n,$$
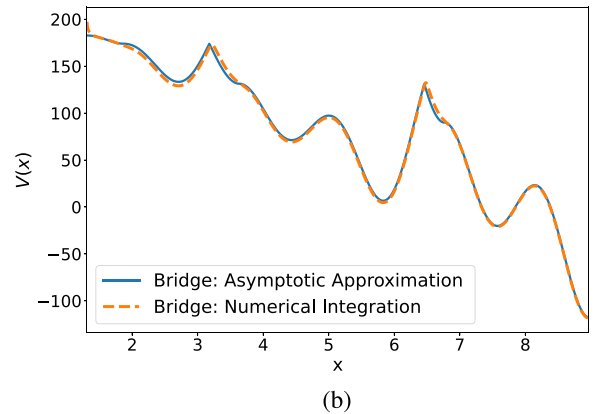

(a)

(b)

FIG. 7. Potential energy surface (green) and equivalent bridge surfaces for paths that reach $x_h = 8.97$ before reaching $x_a = 1.31$ (two minima of the surface). The exact bridge (orange) is calculated via numerical integration and the reflection approximation (blue) is calculated using the reflection method outlined above. (a) Potential energy surface and the equivalent potential energy surfaces (exact and approximate) for paths reaching one region before another. (b) The same bridge surfaces as in (a), shifted to overlap. Note how well the reflection method approximates the bridge surface.
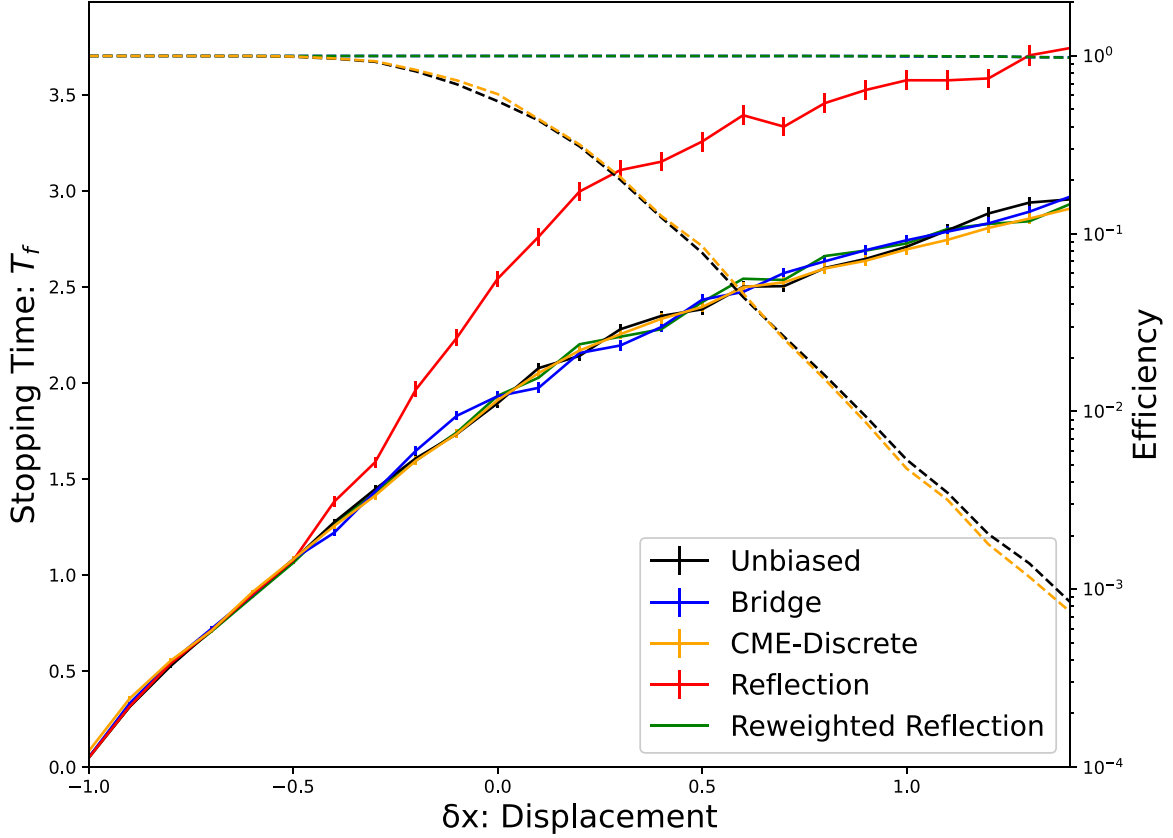
FIG. 8. Stopping time and efficiency comparison of the Schögl reaction system (22). The unbiased, bridge, and reflection cases are simulated under the continuum approximation using the Langevin equation (27), whereas the CME-discrete case is modeled by the chemical master equation (23) and (24) using the Gillespie algorithm [32]. The bridge and reflection cases are simulated using the Langevin equation with an additional drift force (3), the former using the exact hitting probability and the latter using an approximate hitting probability via the reflection method. The samples are initiated at varying concentrations, $x_0 = x_c + \delta x$, about the central unstable state, $x_c = 1.204$, conditional on hitting $x_h = 0.094$ before $x_a = 3.702$. Solid lines indicate the mean stopping time and error bars indicate the standard error at each initial concentration after simulating $N = 1000$ successful samples. The dotted lines indicate the sampling efficiency, i.e., the ratio of successful samples to the total number of simulated samples.

$$n = 1, 2, \ldots, \infty, \tag{24}$$

where $\lambda_n = \hat{k}_3 n_B + \hat{k}_1 n_A n(n-1)$, $\mu_n = \hat{k}_4 n + \hat{k}_2 n(n-1)(n-2)$, and $\hat{k}_i = k_1/V^{m-1}$, where (n_A, n_B) are the number of molecules of chemical species (A, B), respectively, and m is the number of reactant molecules. In the continuum limit $V \to \infty$, the master equation can be approximated by the chemical Langevin equation. The drift and diffusion coefficients are determined via

$$\frac{d\langle X \rangle}{dt} = k_1 a x^2 - k_2 x^3 + k_3 x - k_4 b, \tag{25}$$

$$\frac{d\sigma^2(X)}{dt} = \frac{d\langle X^2 \rangle}{dt} - 2\langle X \rangle \frac{d\langle X \rangle}{dt}$$
$$= k_1 a x^2 + k_2 x^3 + k_3 x + k_4 b, \tag{26}$$

which is calculated from the master equation [31]. Therefore, the reaction system is approximated by

$$dx = (k_1 a x^2 - k_2 x^3 + k_3 x - k_4 b)dt$$
$$+ \sqrt{\frac{k_1 a x^2 + k_2 x^3 + k_3 x + k_4 b}{V}}dW, \tag{27}$$

where $x = X/V$, $a = A/V$, $b = B/V$ are concentrations. The choice of parameters as $k_1 = 3$, $k_1 = 0.6$, $k_3 = 0.25$, $k_4 = 2.95$, $a = b = 1$, and $V = 40$ results in a chemical system that contains three steady states, including an unstable state at $x_c = 1.204$ flanked by two stable states at $x_h = 0.094$ and $x_a = 3.702$. The central unstable state acts as a potential energy barrier between the two stable states. We look at the paths that reach the stable low-concentration steady state before the stable high-concentration steady state, $\tau_{x_h} < \tau_{x_a}$.

In Fig. 8, the initial concentration is $x_0 = x_c + \delta x$, where $x_c = 1.204$ is the central unstable steady state and δx is the displacement from the unstable state. We plot the average stopping time T_f, conditioned for the case that paths hit the leftmost stable state first, $\tau_{x_h} < \tau_{x_a}$, for (a) the discrete unbiased case, given by Eqs. (23) and (24), solved via the Gillespie algorithm [32], (b) the continuous unbiased case (27), and (c)–(d) bridge cases (3), i.e., biasing using the exact bridge solution and the approximate reflection scheme (modified to take into account nonconstant diffusion). The standard reflection method overestimates the stopping time; however, the importance sampling procedure again produces accurate conditional statistics. Furthermore, we see that while the unbiased (discrete and continuous) sampling efficiency decreases

exponentially as x_0 increases past x_c (i.e., $\delta x > 0$), both the bridge and reflection sampling methods retain near perfect sampling efficiency. This validates the use of the bridge to generate these conditional statistics for this chemical reaction problem and, in fact, can easily be modified to follow more general reaction schemes. Further, it shows that the simple reflection approximation can be used even within a nonconstant diffusion problem.

V. SUMMARY

Paths that reach one region before another often arise in chemistry and physics since the interesting dynamics of a system often occurs during some sort of barrier crossing event. The study of these types of paths is hindered by the fact that they usually are rare events. There are many different methods commonly used in the literature that circumvent this issue; however, they are generally very sensitive to their user-defined initial conditions. In this paper, we demonstrate that Brownian bridges (BBs) can be used as an effective tool to generate these paths and do not require any guesswork. Additionally, often times other methods, such as string techniques, can overlook valid barrier crossing pathways and reveal only the minimum free energy path. In some cases, this may not be a huge concern as the MFEP will dominate a majority of these trajectories. When that is not the case, these methods can completely overlook a lot of dynamics and result in inaccurate property calculations. The bridge does not suffer from this because it samples from the other pathways with the same probability as the unbiased trajectories. In so doing, it generates a path ensemble that is more faithful to the real unbiased ensemble.

The main drawback of this methodology is that the construction of the bridge requires the hitting probability, that is, a solution to the backward Fokker-Planck (BFP) equation. The BFP equation can be difficult to solve, especially on high-dimensional and/or complex potential energy surfaces. In order to ameliorate that difficulty, we derived two solution methods. The first, which we call the reflection method, is an approximation method that is extremely simple to apply and yields an analytical representation of the solution. This method is best for trajectories that cross large energy barriers and we show that it converges to the hitting probability as the barrier height becomes very large. Although this method does not yield the perfect sampling efficiency that the exact bridge does, it does show a drastic improvement in efficiency compared to brute-force sampling, as well as an improvement in its efficiency with increasing barrier height. We show that the reflection method may also be used in conjunction with importance sampling and still yield accurate conditional statistics even for smaller barriers. This method is currently restricted to 1D stochastic processes, but we intend to use the insights the 1D method provides to attempt to generalize this approximation technique to higher dimensions.

We have shown that these methods can accurately construct Brownian bridges and can capture rare events, e.g., barrier crossing events, much more effectively than traditional sampling. They are also quite versatile, as they can be applied to a wide variety of systems and accurately capture many types of rare events. In the future, we would like to use these methods to examine a wider range of chemical reaction problems.

ACKNOWLEDGMENT

[1] R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena*, revised 2nd ed. (Wiley, New York, 2007).

[2] J. Z. Chen, Theory of wormlike polymer chains in confinement, Prog. Polym. Sci. **54-55**, 3 (2016).

[3] S. Krishnaswami, D. Ramkrishna, and J. M. Caruthers, Statistical-mechanically exact simulation of polymer conformation in an external field, J. Chem. Phys. **107**, 5929 (1997).

[4] S. Wang, D. Ramkrishna, and V. Narsimhan, Exact sampling of polymer conformations using Brownian bridges, J. Chem. Phys. **153**, 034901 (2020).

[5] V. Shneidman and P. Hänggi, Continuous approximation of a random walk, Phys. Rev. E **49**, 894 (1994).

[6] T. Aquino and M. Dentz, Chemical continuous time random walks, Phys. Rev. Lett. **119**, 230601 (2017).

[7] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* of *Springer Series in Synergetics*, 3rd ed. (Springer-Verlag, Berlin, 2004), Vol. 13.

[8] J.-P. Bouchaud and A. Georges, Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications, Phys. Rep. **195**, 127 (1990).

[9] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, Berlin, 1996).

[10] S. Wang, A. Venkatesh, D. Ramkrishna, and V. Narsimhan, Brownian bridges for stochastic chemical processes—An approximation method based on the asymptotic behavior of the backward Fokker–Planck equation, J. Chem. Phys. **156**, 184108 (2022).

[11] R. D. Skeel, *Integration Schemes for Molecular Dynamics and Related Applications* (Springer, Berlin, 1999), pp. 119–176.

[12] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations* (Cambridge University Press, Cambridge, 2019).

[13] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: Fifty years after Kramers, Rev. Mod. Phys. **62**, 251 (1990).

[14] R. W. W. Hooft, B. P. van Eijck, and J. Kroon, An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol, J. Chem. Phys **97**, 6690 (1992).

[15] G. Ciccotti and M. Ferrario, Blue moon approach to rare events, Mol. Simul. **30**, 787 (2004).

[16] C. Dellago, P. G. Bolhuis, and D. Chandler, On the calculation of reaction rate constants in the transition path ensemble, J. Chem. Phys. **110**, 6617 (1999).

[17] H. Fujisaki, M. Shiga, K. Moritsugu, and A. Kidera, Multiscale enhanced path sampling based on the Onsager-Machlup action:

Application to a model polymer, J. Chem. Phys. **139**, 054117 (2013).

[18] 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**, 1451 (2007).

[19] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Transition path sampling: Throwing ropes over rough mountain passes, in the dark, Annu. Rev. Phys. Chem. **53**, 291 (2002).

[20] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, String method in collective variables: Minimum free energy paths and isocommittor surfaces, J. Chem. Phys. **125**, 024106 (2006).

[21] Weinan E and E. Vanden-Eijnden, Transition-path theory and path-finding algorithms for the study of rare events, Annu. Rev. Phys. Chem. **61**, 391 (2010).

[22] Weinan E and E. Vanden-Eijnden, Towards a theory of transition paths, J. Stat. Phys. **123**, 503 (2006).

[23] J. L. Doob, Conditional Brownian motion and the boundary limits of harmonic functions, Air Force Office of Scientific Research Technical Note No. AFOSR TN 57-625 (United States Air Force, Office of Scientific Research, Washington, DC, 1957).

[24] B. Øksendal, *Stochastic Differential Equations*, 6th ed. (Springer, Berlin, 2003).

[25] R. M. Gray, *Entropy and Information Theory* (Springer, New York, 1990).

[26] A. D. Wissner-Gross and C. E. Freer, Causal entropic forces, Phys. Rev. Lett. **110**, 168702 (2013).

[27] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1992), Vol. 1.

[28] N. M. Temme, *Asymptotic Methods for Integrals* (World Scientific, Singapore, 2014).

[29] R. B. Paris, *Hadamard Expansions and Hyperasymptotic Evaluation: An Extension of the Method of Steepest Descents*, Encyclopedia of Mathematics and its Applications (Cambridge University Press, Cambridge, 2011).

[30] M. Vellela and H. Qian, Stochastic dynamics and nonequilibrium thermodynamics of a bistable chemical system: the Schlogl model revisited, J. R. Soc. Interface **6**, 925 (2009).

[31] Y. Cao and J. Liang, Nonlinear Langevin model with product stochasticity for biological networks: The case of the Schnakenberg model, J. Syst. Sci. Complex **23**, 896 (2010).

[32] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. **81**, 2340 (1977).