10 March 2026

Accelerated Langevin Dynamics Simulation via Neural Network-Driven Importance Sampling

Michael Kim[1], Wei Cai[1]

1. Department of Mechanical Engineering Stanford University

Abstract

Atomistic simulations are powerful tools for understanding the dynamical processes that control material behavior at the molecular scale. Yet the study of many important processes is currently limited by the timescales achievable by direct simulation. This limitation often arises from potential energy landscapes featuring deep energy basins, corresponding to metastable states, in which the system remains trapped for extended periods of time. The long-term evolution of such systems is governed by the transitions between these metastable states, which are rare events that are very inefficient to capture by brute-force simulations. We present an importance sampling framework to accelerate the time scale of Langevin dynamics simulations. Using an importance function to bias the dynamics, this approach significantly enhances the efficiency of sampling the rare transitions, while preserving the relative probabilities between different transition paths. The importance function is parameterized using a neural network for scalability to high-dimensional systems. We also provide a rigorous mathematical formulation to compute the original transition rates between metastable states from the biased dynamics. To control the statistical variance of the estimated transition rates, we introduce a branching random walk (BRW) algorithm to manage the weights of the sampled transition paths. We validate the proposed importance sampling framework on a two-dimensional system and a fourteen-dimensional system to demonstrate its robustness and scalability.

Accelerated Langevin Dynamics Simulation via Neural Network–Driven Importance Sampling

Michael Kim[1] and Wei Cai *[1]

[1]Department of Mechanical Engineering, Stanford University, CA 94305

February 19, 2026

Abstract

Atomistic simulations are powerful tools for understanding the dynamical processes that control material behavior at the molecular scale. Yet the study of many important processes is currently limited by the timescales achievable by direct simulation. This limitation often arises from potential energy landscapes featuring deep energy basins, corresponding to metastable states, in which the system remains trapped for extended periods of time. The long-term evolution of such systems is governed by the transitions between these metastable states, which are rare events that are very inefficient to capture by brute-force simulations. We present an importance sampling framework to accelerate the time scale of Langevin dynamics simulations. Using an importance function to bias the dynamics, this approach significantly enhances the efficiency of sampling the rare transitions, while preserving the relative probabilities between different transition paths. The importance function is parameterized using a neural network for scalability to high-dimensional systems. We also provide a rigorous mathematical formulation to compute the original transition rates between metastable states from the biased dynamics. To control the statistical variance of the estimated transition rates, we introduce a branching random walk (BRW) algorithm to manage the weights of the sampled transition paths. We validate the proposed importance sampling framework on a two-dimensional system and a fourteen-dimensional system to demonstrate its robustness and scalability.

1 Introduction

Atomistic simulations [1–3] are powerful tools to elucidate the fundamental mechanisms of physical processes at the molecular scale. However, the applicability of these methods is fundamentally constrained by the accessible time scales by brute-force simulations. This limitation often stems from the fact that atomistic systems typically remain confined in energy basins, corresponding to metastable states, for long durations. Escaping to neighboring basins (i.e. metastable states) occurs very rarely, but it is precisely these rare events that determines the long term evolution of the system. Therefore, using the brute-force simulations to predict the long-term evolution in such systems can be prohibitively expensive. This phenomenon is known as the rare event problem or the timescale problem [4–6].

To address the rare event problem, various computational methods have been developed. For instance, steered molecular dynamics (SMD) [7], metadynamics (MTD) [8–11], and hyperdynamics (HD) [12, 13]

*Email: caiwei@stanford.edu

introduce external biasing potentials to accelerate transitions between metastable basins. Distinct from these biasing strategies, trajectory-based methods such as forward flux sampling (FFS) [14] and the weighted ensemble (WE) method [15, 16] decompose the configuration space into a sequence of interfaces or bins, and sample short trajectories between them in order to compute the flux across these interfaces. On the other hand, transition path theory [17–26] establishes a mathematical framework to characterize the statistical ensemble of reactive trajectories. Importance sampling [27–30] constitutes another fundamental approach, providing a statistical framework to efficiently sample the reaction pathways and quantify the kinetics of rare events. This paper focuses on applying importance sampling to Langevin dynamics simulations to extend their timescales.

The proposed importance sampling framework addresses two primary objectives. First, it enhances the efficiency of sampling rare transition paths while rigorously preserving the relative probabilities of competing transition channels. Second, it yields an unbiased estimator for the original transition rates by reweighting the ensemble of transition paths generated by the accelerated dynamics. These goals are achieved by modifying the system kinetics via an importance function, which biases the transition probability kernels to preferentially sample successful transition paths. Notably, the normalization condition of this modified probability kernel imposes a unique condition that defines the optimal importance function. Therefore, the rare event problem reduces to optimizing the importance function to bias the system, and constructing an unbiased estimator for the original transition rates.

However, applying this importance sampling framework to high-dimensional systems presents several challenges. First, obtaining the optimal importance function is generally intractable, as it necessitates solving a high-dimensional partial differential equation (PDE) over a complex energy landscape. Second, the accuracy of the transition rate estimator is highly sensitive to the quality of this importance function. Even minor approximation errors can induce large fluctuations in the estimated rates and significantly alter the relative probability of competing transition pathways.

This paper proposes an importance sampling framework that overcomes these challenges in capturing the long term behavior of systems following the Langevin dynamics. To obtain suitable importance function in high-dimensional space, we use a neural network to parameterize the importance function and optimize its parameters via adaptive training. To address the unavoidable approximation error for the optimal importance function, we construct an unbiased estimator for the original transition rate even when the importance function is not strictly optimal, by assigning statistical weights to the sampled transition paths. Furthermore, to increase efficiency and to reduce variance of the estimator, we use a branching random walk (BRW) algorithm [31] to regulate the weights of the transition paths sampled by the biased dynamics.

The rest of the paper is organized as follows. In Section 2, we establish the mathematical framework, formally define the rare event problem, and state the objectives of the importance sampling approach. Section 3 details our importance sampling methodology, describing how an approximate importance function can be used to provide unbiased estimates of the original transition rates and unbiased sampling of transition pathways. The efficacy of the proposed framework is validated in Section 4 through numerical experiments on a 2-dimensional system and a 14-dimensional system. Finally, Section 5 summarizes our key findings and outlines directions for future research.

2 Problem Statement

Consider a system specified by a position vector $\mathbf{x}$ that exists in a domain Ω, on which a potential energy $U(\mathbf{x})$ is defined, as shown in Figure 1. Each vector $\mathbf{x}$ is also called a microscopic state. The potential energy landscape is characterized by two deep basins corresponding to two metastable states, Ω_A and Ω_B. Each metastable state is the collection of all microscopic states that leads to the same local energy minimum when

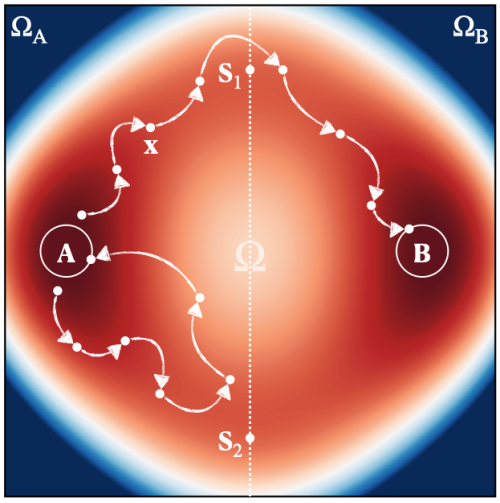


Figure 1: Schematic representation of the configuration space Ω with positions $\mathbf{x}$. The energy landscape $U(\mathbf{x})$ leads to the partition of the entire domain into two basins Ω_A and Ω_B, corresponding to two metastable states. The boundary between Ω_A and Ω_B passes through the saddle points S_1 and S_2. Two small regions A and B (circles), are defined around the local energy minima of the two basins. Arrowed lines depict two representative paths initiated near the boundary of region A: a failure path that returns to A before reaching B and a successful path that reaches B before returning to A.

following a steepest descent trajectory of $U(\mathbf{x})$. The time evolution of the system at temperature T is governed by the overdamped Langevin dynamics [32],

$$\mathrm{d}\mathbf{x}(t) = -\gamma^{-1}\,\nabla U(\mathbf{x}(t))\,\mathrm{d}t + \sqrt{2\gamma^{-1}\,\beta^{-1}}\,\mathrm{d}\boldsymbol{W}(t), \tag{1}$$

where $\beta = 1/(k_B T)$, k_B is the Boltzmann constant, γ is the friction coefficient, and $\boldsymbol{W}(t)$ is the standard Wiener process [33]. Although Fig. 1 depicts a two-dimensional domain, our target is systems with arbitrary dimension d, in which both $\mathbf{x}$ and $\boldsymbol{W}$ are d-dimensional vectors. For simplicity here we assume that domain Ω only contains two metastable states. However, the discussions here can be generalized to systems containing multiple metastable states, in which case Ω_B would represent all mestable states accessible from Ω_A.

The stochastic process described by Eq. (1) has a stationary probability distribution, i.e. at thermal equilibrium, given by the Gibbs–Boltzmann distribution [34],

$$\rho_{\mathrm{eq}}(\mathbf{x}) = Z(\beta)^{-1}\,\mathrm{e}^{-\beta U(\mathbf{x})}, \tag{2}$$

where $Z(\beta)$ is the partition function defined as,

$$Z(\beta) = \int_{\Omega} \mathrm{d}\mathbf{x}\,\mathrm{e}^{-\beta U(\mathbf{x})}. \tag{3}$$

Due to the exponential dependence of the Gibbs-Boltzmann distribution on energy, the stationary probability $\rho_{\mathrm{eq}}(\mathbf{x})$ is heavily concentrated around the local minima of the energy basins (Ω_A and Ω_B). For example, if the system is initially in metastable state Ω_A, it tends to stay within Ω_A for a long time before making a rare transition into Ω_B.

The primary goal of this study is to develop an importance sampling framework to accelerate the simulation of the long-term dynamics of the system. Specifically, our objectives are twofold. First, we modify the system kinetics in order to make the escape from metastable state Ω_A occur faster while maintaining the relative probabilities between different transition mechanisms (e.g. through saddle region near S_1 or S_2). This is

essential for preserving the direction of the long-term evolution of the system, especially for high-dimensional systems with many metastable states. Second, we need to determine the transition rate of the original system (out of Ω_A) through the simulation of the modified system.

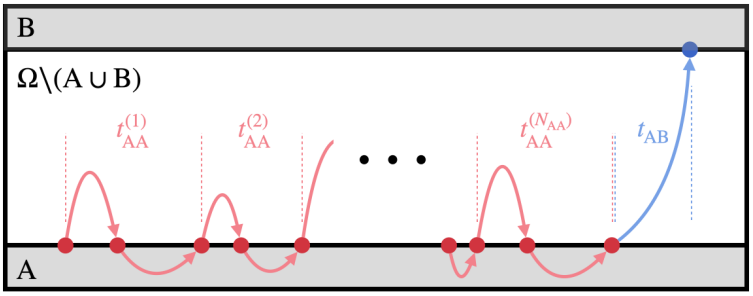


Figure 2: Schematic illustration of the decomposition of a long-time trajectory of the system into short trajectories (paths). The two shaded areas represent regions A and B, respectively. The trajectory is decomposed into two kinds of paths. The "failure" paths (red lines) leave region A and returns to region A before visiting region B. The "success" path (blue line) leaves region A and visits region B before returning A. Under the original dynamics, the rare event problem is equivalent to sampling a very large number of "failure" paths before a single "success" path is observed.

To make progress, we define two small regions A and B around the local energy minima of basins Ω_A and Ω_B, respectively. Assume the system is initialized in metastable state Ω_A. Its trajectory is likely to visit region A repeatedly, for a long time, before it ever visits region B. Therefore, with the introduction of regions A and B, we can break a long-time trajectory of the system into many short trajectories (called paths), as illustrated in Figure 2. In particular, we focus on two kinds of paths. Those leaving region A and returning to region A before visiting region B are called "failure" paths. Those leaving region A and visiting region B before returning A are called "success" paths. When the system follows the original dynamics, the rare event problem is equivalent to the overwhelming probability of generating the "failure" paths and vanishingly small probability of generating the "success" paths. The two goals outlined above can thus be rephrased in terms of the sampling of the paths. First, we aim to enhance the probability of sampling success paths *uniformly*, so that the relative probability of escaping Ω_A via different channels (e.g. near S_1 or S_2) remain unchanged. Second, we need to know exactly by how much we have enhanced the probability of sampling the success paths, so that we can obtain the escape rate of the original system.

In the following, we connect the probability of sampling success paths and the escape rate. The transition rate r_{AB} from metastable state Ω_A to metastable state Ω_B can be obtained from the mean first passage time t_{FPT} from region A to region B,

$$r_{AB} = \langle t_{FPT} \rangle^{-1}. \tag{4}$$

Since the system is most likely to repeatedly visit region A before visiting region B, we expect the mean first passage time to take the following form,

$$\langle t_{FPT} \rangle = \langle N_{AA} \rangle \langle t_{AA} \rangle + \langle t_{AB} \rangle, \tag{5}$$

where $\langle N_{AA} \rangle$ is the average number of failure paths sampled before sampling a successful path, $\langle t_{AA} \rangle$ is the mean time interval between successive escapes from A, and $\langle t_{AB} \rangle$ is the average time of a success path. The average number of failed attempts $\langle N_{AA} \rangle$ can be calculatd from a geometric sum,

$$\langle N_{AA} \rangle = \sum_{n=0}^{\infty} n \, (1 - P_{AB})^n \, P_{AB} = P_{AB}^{-1} - 1. \tag{6}$$

where P_{AB} is the probability of sampling a success path from A. In the following, we shall call P_{AB} the success probability for brevity. When the system exhibit a rare-event problem, $P_{\mathrm{AB}} \ll 1$, and the mean first passage time can be approximated as,

$$\langle t_{\mathrm{FPT}} \rangle \approx P_{\mathrm{AB}}^{-1} \langle t_{\mathrm{AA}} \rangle. \tag{7}$$

This allows us to estimate the transition rate from Ω_{A} to Ω_{B} as,

$$r_{\mathrm{AB}} = \langle t_{\mathrm{FPT}} \rangle^{-1} \approx P_{\mathrm{AB}} \langle t_{\mathrm{AA}} \rangle^{-1} = J_{\mathrm{A}} P_{\mathrm{AB}}. \tag{8}$$

The inverse of $\langle t_{\mathrm{AA}} \rangle$ can be interpreted as the flux J_{A} out of A, known as the Hill relation [35–38]. The mean time interval between escapes $\langle t_{\mathrm{AA}} \rangle$ can be efficiently estimated from unbiased equilibrium simulations near A (see Section 3.6). Since P_{AB} is vanishingly small, computing it by brute-force simulation is very inefficient. Our goal is to modify the system kinetics so that P_{AB} is enhanced in such a way that the relative probability of sampling different success paths (e.g. those going through region near S_1 or S_2) remain unchanged.

3 Methods

3.1 Langevin Dynamics and Path Sampling

The overdamped Langevin dynamics in Eq. (1) is simulated with a timestep size Δt, using the Euler–Maruyama scheme [39], in which the state vector is updated as follows.

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) - \gamma^{-1} \nabla U(\mathbf{x}(t)) \, \Delta t + \sqrt{2 \gamma^{-1} \beta^{-1} \Delta t} \, \boldsymbol{\xi}, \tag{9}$$

where $\boldsymbol{\xi}$ is a vector of independent standard Gaussian random variables. The update consists of a deterministic steepest descent move and a stochastic Gaussian noise displacement. Repeating these updates produces a numerical trajectory representing the time evolution of the system.

We now break an arbitrarily long trajectory of the system generated above into shorter trajectories, called paths. To reduce computational cost, we also coarse-grain time in the representation of paths, as described below. A path is a sequence of microscopic states the system visits along its trajectory,

$$\Gamma = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\} \tag{10}$$

A path is initiated (at $\mathbf{x}_1$) whenever the trajectory leaves region A. In other words, the system is outside region A at the present time step but was inside region A in the previous time step. A path is terminated whenever the trajectory enters region A or region B, i.e. $\mathbf{x}_N \in \mathrm{A} \cup \mathrm{B}$. The paths that terminate by entering region A ($\mathbf{x}_N \in \mathrm{A}$) are called "failure" paths, and are designated by Γ_{AA}. The paths that terminate by entering region B ($\mathbf{x}_N \in \mathrm{B}$) are called "success" paths, and are designated by Γ_{AB}.

To coarse-grain time, successive microscopic states stored in a path Γ, e.g. between $\mathbf{x}_n$ and $\mathbf{x}_{n+1}$ are separated in time by a period of $\tau = k\Delta t$, where k is a positive integer, with the possible exception at $n+1 = N$. The time separation between $\mathbf{x}_{N-1}$ and $\mathbf{x}_N$ is allowed to be less than $k\Delta t$, if the system visits region A or region B before that, causing the path to terminate. Our goal is to enhance the probability of sampling success paths Γ_{AB} uniformly, in such a way that the relative probability of sampling different success paths remain unchanged.

3.2 Importance Sampling of Transition Paths

We start by characterizing the probability density of sampling a successful transition path $\Gamma_{\mathrm{AB}} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, where $\mathbf{x}_1, \cdots \mathbf{x}_{N-1} \notin \mathrm{A} \cup \mathrm{B}$, $\mathbf{x}_N \in \mathrm{B}$. Let $\mathcal{K}_\tau(\mathbf{y} \,|\, \mathbf{x})$ be the probability density kernel of visiting $\mathbf{y}$ as the next

microscopic state on the path, if $\mathbf{x}$ is the current microscopic state on the path. According to Section 3.1, the time separation between visiting these two microscopic states equals to τ or $\tau_{\mathrm{A}\cup\mathrm{B}}$, whichever one is smaller, where $\tau_{\mathrm{A}\cup\mathrm{B}}$ is the first passage time to visit region A or B from $\mathbf{x}$. Therefore,

$$\mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x}) = \mathbb{E}\left[\delta\big(\mathbf{y} - \mathbf{x}(\tau \wedge \tau_{\mathrm{A}\cup\mathrm{B}})\big)\,\big|\,\mathbf{x}(0) = \mathbf{x}\right], \tag{11}$$

where $\delta(\cdot)$ denotes the d-dimensional Dirac delta function, and $\wedge$ means taking the minimum of the two variables on both sides. In terms of this probability density kernel $\mathcal{K}_\tau$, the probability density of sampling path Γ_{AB} can be written as,

$$p(\Gamma_{\mathrm{AB}}) = \rho_{\mathrm{A}}(\mathbf{x}_1)\prod_{i=1}^{N-1}\mathcal{K}_\tau(\mathbf{x}_{i+1}\,|\,\mathbf{x}_i), \tag{12}$$

where $\rho_{\mathrm{A}}(\mathbf{x}_1)$ denotes the probability density of the initial state $\mathbf{x}_1$ sampled at the moment of exiting the boundary of A (see Section 3.6). Let $\{\Gamma_{\mathrm{AB}}\}$ represent the set of all successful paths from A to B and $\{\Gamma_{\mathrm{AA}}\}$ as the set of all failure paths returning to A. Using these definitions, the probability P_{AB} of sampling any success path, as defined in Section 2, can be written as an integral over all success paths,

$$P_{\mathrm{AB}} = \int_{\Gamma\in\{\Gamma_{\mathrm{AB}}\}} p(\Gamma)\,\mathrm{d}\Gamma. \tag{13}$$

For systems exhibiting rare event behavior, the probability density of sampling any success path is very small compared to that of a failure path. This is the underlying reason for the inefficiency of brute-force simulations.

To address this challenge, we employ importance sampling to alter the kinetics of the system. We modify the probability density kernel $\mathcal{K}_\tau$ using an importance function $I(\mathbf{x})$ [28, 29],

$$\mathcal{K}'_\tau(\mathbf{y}\,|\,\mathbf{x}) = \frac{I(\mathbf{y})}{I(\mathbf{x})}\mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x}). \tag{14}$$

However, Eq. (14) is not always possible because the modified probability density kernel $\mathcal{K}'_\tau$ must strictly satisfy the normalization condition,

$$\int_{\mathbf{y}\in\Omega}\mathrm{d}\mathbf{y}\,\mathcal{K}'_\tau(\mathbf{y}\,|\,\mathbf{x}) = 1. \tag{15}$$

We define the importance function that allows the normalization condition to be satisfied everywhere as the optimal importance function $I^{\mathrm{opt}}(\mathbf{x})$.

However, when $I(\mathbf{x})$ is not equal to $I^{\mathrm{opt}}(\mathbf{x})$, the modified kernel must be normalized by introducing a normalization factor $Z_\tau(\mathbf{x})$,

$$\mathcal{K}'_\tau(\mathbf{y}\,|\,\mathbf{x}) = Z_\tau^{-1}(\mathbf{x})\frac{I(\mathbf{y})}{I(\mathbf{x})}\mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x}), \tag{16}$$

$$Z_\tau(\mathbf{x}) = \int_{\mathbf{y}\in\Omega}\mathrm{d}\mathbf{y}\,\frac{I(\mathbf{y})}{I(\mathbf{x})}\mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x}). \tag{17}$$

The algorithms to compute the normalization factor are described in Section 3.7. The new probability density of sampling path Γ_{AB} under $\mathcal{K}'_\tau$ can be expressed as,

$$p'(\Gamma_{\mathrm{AB}}) = \rho_{\mathrm{A}}(\mathbf{x}_1)\prod_{i=1}^{N-1}\mathcal{K}'_\tau(\mathbf{x}_{i+1}\,|\,\mathbf{x}_i). \tag{18}$$

Plugging in Eq. (16) into Eq. (18) yields,

$$p'(\Gamma_{\mathrm{AB}}) = \frac{I(\mathbf{x}_N)}{I(\mathbf{x}_1)}\left[\prod_{i=1}^{N-1}Z_\tau^{-1}(\mathbf{x}_i)\right]p(\Gamma_{\mathrm{AB}}). \tag{19}$$

If the importance function is optimal (i.e., $I(\mathbf{x}) = I^{\mathrm{opt}}(\mathbf{x})$), the normalization factors equal unity everywhere so that the relation in Eq. (19) simplifies into,

$$p'(\Gamma_{\mathrm{AB}}) = \frac{I^{\mathrm{opt}}(\mathbf{x}_N)}{I^{\mathrm{opt}}(\mathbf{x}_1)}\, p(\Gamma_{\mathrm{AB}}). \tag{20}$$

This is a remarkable result because it means that the two primary objectives established in Section 2 have been achieved. First, the probability density of sampling any successful path from $\mathbf{x}_1$ to B is enhanced by a factor of $I^{\mathrm{opt}}(\mathbf{x}_N)/I^{\mathrm{opt}}(\mathbf{x}_1)$, which is independent of the intermediate positions along the path. Second, because this factor is explicitly known, the unbiased path probability density can be rigorously reconstructed from the biased dynamics.

In practice, however, the optimal importance function $I^{\mathrm{opt}}(\mathbf{x})$ is generally unknown and computationally intractable to determine exactly. A standard strategy in importance sampling is to construct a good approximation of $I^{\mathrm{opt}}(\mathbf{x})$. The normalization condition in Eq. (15) provides a target for optimizing the importance function. The optimality condition can be written as,

$$\int_{\mathbf{y}\in\Omega} \mathrm{d}\mathbf{y}\, I(\mathbf{y})\, \mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x}) = I(\mathbf{x}). \tag{21}$$

We set $I(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathrm{A}$ and $I(\mathbf{x}) = 1$ for $\mathbf{x} \in \mathrm{B}$ without loss of generality. Therefore, the optimal importance function quantifies the probability of reaching B before returning to A, which is the same as the committor function [17]. The optimization algorithm of the importance function is described in Section 3.5.

Interestingly, the evolution of the system under $\mathcal{K}'_\tau(\mathbf{y}\,|\,\mathbf{x})$ is equivalent to simulating the following overdamped Langevin dynamics for a time period of $\tau \wedge \tau_{\mathrm{A}\cup\mathrm{B}}$,

$$\mathrm{d}\mathbf{x}(t) = -\gamma^{-1}\,\nabla\Big(U(\mathbf{x}(t)) + U_{\mathrm{bias}}(\mathbf{x}(t))\Big)\,\mathrm{d}t + \sqrt{2\,\gamma^{-1}\,\beta^{-1}}\,\mathrm{d}W(t), \tag{22}$$

where $U_{\mathrm{bias}}(\mathbf{x})$ is an added bias potential that is related to the importance function as,

$$U_{\mathrm{bias}}(\mathbf{x}) = -2\beta^{-1}\,\ln I(\mathbf{x}). \tag{23}$$

This bias potential is often identified as the "optimal controller" [40–42], which is derived from the Doob h-transform [43] or the Girsanov theorem [44–46].

3.3 Success Probability Estimation

Here, we describe the estimation of the probability P_{AB} of sampling a success path from A to B in the original dynamics using importance sampling. We denote $p(\Gamma)$ and $p'(\Gamma)$ as the probability densities of sampling a transition path Γ following $\mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x})$ and $\mathcal{K}'_\tau(\mathbf{y}\,|\,\mathbf{x})$, respectively. We define a reward function for a path Γ,

$$R(\Gamma) = \begin{cases} 1, & \text{if } \Gamma \in \{\Gamma_{\mathrm{AB}}\} \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

Then the success probability in Eq. (12) can be written as,

$$P_{\mathrm{AB}} = \int_\Gamma R(\Gamma)\, p(\Gamma)\, \mathrm{d}\Gamma, \tag{25}$$

where the integral is taken over all paths. This implies that the success probability is equivalent to the statistical expectation of the reward function over all paths sampled using $K_\tau(\mathbf{y}\,|\,\mathbf{x})$,

$$P_{\mathrm{AB}} = \mathbb{E}_\Gamma\left[R(\Gamma)\right]. \tag{26}$$

where $\mathbb{E}_\Gamma[\cdot]$ is the expectation with respect to $p(\Gamma)$. While Eq. (26) offers a straightforward estimator, it becomes computationally intractable for systems exhibiting rare events. In such systems, the vast majority of sampled paths will be failure paths, whose reward is zero.

To overcome this sampling inefficiency, we employ the importance sampling scheme framework in Section 3.2, sampling paths following the biased kernel $\mathcal{K}'_\tau$. The algorithmic details of biased path sampling process is explained in Section 3.8. The success probability can be rewritten using the biased probability density $p'(\Gamma)$ in Eq. (19) as follows.

$$P_{\mathrm{AB}} = \int_\Gamma R(\Gamma)\, \frac{I(\mathbf{x}_1)}{I(\mathbf{x}_N)} \left[\prod_{i=1}^{N-1} Z_\tau(\mathbf{x}_i) \right] p'(\Gamma)\, \mathrm{d}\Gamma. \tag{27}$$

Define the weight of a path Γ as,

$$W(\Gamma) = \prod_{i=1}^{N-1} Z_\tau(\mathbf{x}_i), \tag{28}$$

which is the product of the normalization factors along the path excluding the final point. Then, the success probability can be written as,

$$P_{\mathrm{AB}} = \int_\Gamma R(\Gamma)\, \frac{I(\mathbf{x}_1)}{I(\mathbf{x}_N)}\, W(\Gamma)\, p'(\Gamma)\, \mathrm{d}\Gamma. \tag{29}$$

Therefore, the success probability can be estimated by the statistical expectation over the ensemble of transition paths sampled using $\mathcal{K}'_\tau(\mathbf{y} \mid \mathbf{x})$,

$$P_{\mathrm{AB}} = \mathbb{E}'_\Gamma \left[R(\Gamma)\, W(\Gamma)\, \frac{I(\mathbf{x}_1)}{I(\mathbf{x}_N)} \right], \tag{30}$$

where $\mathbb{E}'_\Gamma[\cdot]$ denotes the expectation with respect to $p'(\Gamma)$. The advantage of this formulation is that the biased kernel $\mathcal{K}'_\tau(\mathbf{y} \mid \mathbf{x})$ preferentially samples successful transition paths with non-zero rewards.

In the ideal limit where the importance function equals the optimal importance function, i.e., $I(\mathbf{x}) = I^{\mathrm{opt}}(\mathbf{x})$, the success probability can be simplified into,

$$P_{\mathrm{AB}} = \int_{\mathbf{x}_1 \in \Omega} \rho_{\mathrm{A}}(\mathbf{x}_1)\, I^{\mathrm{opt}}(\mathbf{x}_1)\, \mathrm{d}\mathbf{x}_1. \tag{31}$$

This result recovers the expression for the "reactive probability" derived in previous studies [38, 47]. However, $I^{\mathrm{opt}}(\mathbf{x}_1)$ is generally unknown. Replacing $I^{\mathrm{opt}}(\mathbf{x}_1)$ by its approximation $I(\mathbf{x}_1)$ in Eq. (31) would lead to uncontrolled errors in the success probability estimate, as we will demonstrate in Section 4. Therefore, one must rely on the full weighted average in Eq. (30) to achieve unbiased estimates.

3.4 Branching Random Walk

Although Eq. (30) provides an unbiased estimate of the success probability P_{AB}, it is necessary to sample a sufficient number of paths to achieve statistical convergence. The number of required paths to ensure the statistical convergence heavily depends on the quality of the importance function $I(\mathbf{x})$. The closer $I(\mathbf{x})$ is to $I^{\mathrm{opt}}(\mathbf{x})$, the smaller variance there is in the path weights $W(\Gamma)$, and the more efficient is the estimation of the success probability. In practice, even a small deviation of $I(\mathbf{x})$ from $I^{\mathrm{opt}}(\mathbf{x})$ may lead to large variations in the path weights and success probability estimates.

We address this problem by combining a branching random walk (BRW) algorithm [31] with the biased path sampling procedure in Section 3.8. The BRW process starts from a single walker with initial state $\mathbf{x}_1$

sampled with respect to $\rho_{\mathrm{A}}(\mathbf{x}_1)$ and its weight set to unity. At every step of the BRW process, the walker evaluates the normalization factor at its current microscopic state defined by Eq. (17). The normalization factor is then multiplied to the walker's weight. Based on the updated weight, the walker may either terminate or branch into multiple walkers following a stochastic branching rule, a technique widely used in Monte Carlo rendering and particle transport simulations [48–50].

The stochastic branching rule is designed to maintain the total weight of the walkers while controlling the variance of individual walker weights. The stochastic branching is executed as follows. Let us denote the weight of a walker after updating with its normalization factor as W. If the weight is within a predefined range, i.e., $W \in [W_{\min}, W_{\max}]$, the walker continues without any branching. Otherwise, the walker branches into $\mathcal{R}(W)$ walkers where $\mathcal{R}(\cdot)$ is a stochastic rounding function defined as,

$$\mathcal{R}(W) = \begin{cases} \lfloor W \rfloor + 1, & \text{with probability } W - \lfloor W \rfloor \\ \lfloor W \rfloor. & \text{otherwise} \end{cases} \tag{32}$$

The weights of each branched walkers are set to unity to preserve the total weight. There exist three major scenarios for the branching process. (1) If $\mathcal{R}(W) = 0$, the walker is terminated. (2) If $\mathcal{R}(W) = 1$, the walker continues without branching but its weight is reset to unity. (3) If $\mathcal{R}(W) > 1$, the walker branches into multiple walkers with their weights set to unity.

Once the BRW process is complete, any walker marked as a "success" receives a reward equal to its weight times $I(\mathbf{x}_1)/I(\mathbf{x}_N)$, while those marked as a "failure" receive zero. The success probability can be estimated by averaging the rewards of all walkers following Eq. (29). The BRW process ensures that the variance of the success probability estimator is controlled by regulating the walker weights to be within a predefined range. Consequently, the BRW process facilitates an accurate estimation of the success probability and transition rate.

3.5 Optimization of Importance Function

In high-dimensional systems, the volume of the configuration space (i.e. domain Ω) increases exponentially with dimensionality. Consequently, standard numerical methods cannot be used to solve for or represent the importance function $I(\mathbf{x})$ across the vast configuration space. To overcome this limitation, we approximate the importance function using a neural network, denoted as $I(\mathbf{x}; \theta)$, where θ represents the trainable parameters [38, 51, 52]. The neural network importance function can be optimized by attempting to satisfy the normalization condition, Eq. (21), as close as possible everywhere. To achieve this, we minimize the following loss function [38, 53],

$$L(\theta) = \int_{\Omega} \mathcal{L}(\mathbf{x}, \theta)\, g(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{33}$$

where,

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \theta) = &\left(\ln \left[\int_{\Omega} \mathrm{d}\mathbf{y}\, I(\mathbf{y}; \theta)\, \mathcal{K}_\tau(\mathbf{y} \,|\, \mathbf{x}) \right] - \ln \left[I(\mathbf{x}; \theta) \right] \right)^2 \\ &+ \left(\ln \left[1 - \int_{\Omega} \mathrm{d}\mathbf{y}\, I(\mathbf{y}; \theta)\, \mathcal{K}_\tau(\mathbf{y} \,|\, \mathbf{x}) \right] - \ln \left[1 - I(\mathbf{x}; \theta) \right] \right)^2, \end{aligned} \tag{34}$$

is a loss density function, and $g(\mathbf{x}) > 0$ is a distribution function that is yet to be specified. It can be easily verified that $L(\theta)$ reaches 0 (the global minimum) when $I(\mathbf{x}; \theta)$ becomes $I^{\mathrm{opt}}(\mathbf{x})$.

In practice, evaluating the loss density function over the entire high-dimensional domain Ω is computationally intractable. Fortunately, this is also unnecessary. Instead, it is important for the importance

Algorithm 1: Adaptive Training of Importance Function

Input: Potential $U(\mathbf{x})$, regions A and B, loss density function $\mathcal{L}(\mathbf{x}, \theta)$

Output: Optimized importance function $I(\mathbf{x}; \theta)$

Initialize neural network parameters θ;

Initialize replay library $\mathcal{M} \leftarrow \emptyset$;

Initialize n walkers near the boundary of A according to $\rho_A(\mathbf{x})$ (defined in Section 3.2);

while *not converged* **do**

`// 1. Sampling Stage`

Propagate walkers using biased kernel $\mathcal{K}_\tau(\mathbf{y} \mid \mathbf{x})$ (Section 3.8) driven by the current $I(\mathbf{x}; \theta)$;

Collect set of newly visited microscopic states, $\mathcal{S}_{\text{new}}$;

if *any walker reaches region B (i.e. "success" path sampled)* **then**

Append states on "success" paths to library $\mathcal{M}$;

`// 2. Training Stage`

Sample newly-visited batch $\mathcal{B}_{\text{new}}$ from $\mathcal{S}_{\text{new}}$;

Sample historical batch $\mathcal{B}_{\text{lib}}$ from $\mathcal{M}$ ($\mathcal{B}_{\text{lib}} = \emptyset$ if $\mathcal{M} = \emptyset$);

Update θ by minimizing loss over training batch $\mathcal{B} \leftarrow \mathcal{B}_{\text{new}} \cup \mathcal{B}_{\text{lib}}$:

$\theta \leftarrow \theta - \eta \nabla_\theta \sum_{\mathbf{x} \in \mathcal{B}} \mathcal{L}(\mathbf{x}, \theta)$;

function to satisfy the normalization condition well only in the relevant regions visited by the transition paths connecting regions A and B. Therefore, we evaluate the loss density function in Eq. (34) using microscopic states $\mathbf{x}$ sampled by the importance sampling procedure itself. This is equivalent to implicitly define the distribution function $g(\mathbf{x})$ such that its samples are the microscopic states visited by paths generated by the importance sampling itself. The result of this procedure is an *adaptive training* scheme. The procedure begins by initializing walkers from microscopic states near the boundary but outside of region A. In the sampling stage, these walkers are propagated according to the biased dynamics described in Section 3.8, using the current neural network approximation $I(\mathbf{x}; \theta)$ with fixed parameters. The states visited during these short trajectories are collected and used in the training stage, where the network parameters θ are updated by minimizing the loss function $\mathcal{L}(\theta)$ evaluated on this current batch of states. This cycle of alternating sampling and incremental training repeats until a complete "success" path connecting region A to region B is sampled. Once a "success" path is identified, its constituent states are stored in a replay library. In subsequent iterations, the training phase makes use of samples drawn from the replay library. While the sampling stage continues to generate new states using the current $I(\mathbf{x}; \theta)$, the optimization step now minimizes the loss over a dataset formed by combining the newly sampled states with a batch of states drawn from the library. This mechanism ensures the network retains knowledge of discovered transition pathways while continuing to explore domain Ω. This cycle of alternating between sampling and optimization is repeated until the neural-network importance function converges. Once convergence is reached, the importance sampling procedure is expected to efficiently generate success paths from region A to region B. Detailed description about the training process is described in Algorithm 1.

3.6 Sampling Initial States of Paths

In Section 2, we have seen that an important step in computing the transition rate is to determine the average time duration t_{AA} between two consecutive escapes from region A. Because region A is a small region surrounding the local energy minima of the basin Ω_A, escaping (and returning to) region A occurs quite

frequently, i.e. it is not a rare event. Hence t_{AA} can be computed straightforwardly using a standard, unbiased Langevin dynamics simulation, with an algorithm described in Section 3.1. The simulation is initialized at a random state within region A and evolved for a brief relaxation period to ensure the system reaches quasi-equilibrium in basin Ω_{A}. The simulation is then continued for a long period during which the system enters and escapes region A repeatedly. Given the rare-event nature of the system, we expect the system to stay inside the basin Ω_{A} during the entire simulation and never escape to basin Ω_{B}. (If escape to basin Ω_{B} does happen, we will need to stop the simulation and restart it from region A again.)

We record the sequence of time at which the system escapes from region A as an array,

$$\{t_1, t_2, \ldots, t_L\}. \tag{35}$$

where L is the total number of times escaping from region A is observed. Then, the mean time interval of successive escapes from region A can be calculated as,

$$\langle t_{\mathrm{AA}} \rangle = \frac{1}{L-1} \sum_{i=1}^{L-1} (t_{i+1} - t_i) = \frac{t_L - t_1}{L-1}\,. \tag{36}$$

3.7 Computing Normalization Factor

A key ingredient of the importance sampling algorithm is to compute the normalization factor $Z_\tau(\mathbf{x})$ defined in Eq. (17). Doing so by brute force requires evaluating an integral over the domain for which $\mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x})$ is non-zero, which is infeasible by numerical quadrature in high dimensions. Furthermore, direct evaluation of the probability density kernel $\mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x})$ requires solving the Fokker-Planck equation [54], which is also computationally intractable in high dimensions. Fortunately, an unbiased statistical estimate of the normalization factor $Z_\tau(\mathbf{x})$ can be constructed in the form of a Monte Carlo integration scheme described below.

First, we generate an ensemble of M independent trajectories initiated at $\mathbf{x}$, each propagated following $\mathcal{K}_\tau(\mathbf{y}\,|\,\mathbf{x})$ for a duration of $\tau \wedge \tau_{\mathrm{A}\cup\mathrm{B}}$ (defined in Section 3.2). Next, we record the terminal position $\{\mathbf{y}_i\}_{i=1}^M$ of each trajectory and evaluate the corresponding importance function $I(\mathbf{y}_i)$. Finally, the normalization factor is estimated as the empirical average,

$$\widehat{Z}_\tau(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^{M} \frac{I(\mathbf{y}_i)}{I(\mathbf{x})}. \tag{37}$$

where $\widehat{Z}_\tau(\mathbf{x})$ denotes the empirical estimate for $Z_\tau(\mathbf{x})$. This estimator $\widehat{Z}_\tau(\mathbf{x})$ converges to the exact value $Z_\tau(\mathbf{x})$ as the number of independent trajectories M increases by the law of large numbers [55]. In practice, we may choose a smaller M during the initial stage of training and increase M as the accuracy of the importance function improves. For simplicity, we choose $M = 100$ in this work.

Conceptually, alternative ways to compute $Z_\tau(\mathbf{x})$ exist [45, 46]. For example, for small enough τ, it is possible to approximate $U(\mathbf{x})$ and $U_{\mathrm{bias}}(\mathbf{x})$ by their Taylor expansions and perform the integral in Eq. (17) as a Gaussian integral [56]. We did not take this approach here because we wish to have a large τ value between consecutive states stored in a path. We also note that the computation of the normalization factor is much easier in discrete systems [57], in which we simply add up the contributions from a finite set of neighboring microscopic states.

3.8 Biased Transition Path Sampling

Here we describe how to sample paths according to the modified probability density kernel $\mathcal{K}'_\tau(\mathbf{y}\,|\,\mathbf{x})$ defined in Eq. (14). For computational efficiency, we make use the same ensemble of M independent trajectories

initiated at $\mathbf{x}$ defined in Section 3.7 for computing the normalization factor, using the sequential importance resampling scheme [58–60]. Each trajectory is propagated following $\mathcal{K}_\tau(\mathbf{y} \,|\, \mathbf{x})$ for a duration of $\tau \wedge \tau_{\mathrm{A}\cup\mathrm{B}}$. Given the terminal position $\{\mathbf{y}_i\}_{i=1}^M$ of each trajectory and the corresponding importance function $I(\mathbf{y}_i)$ already evaluated, we select the next microscopic state $\mathbf{y}$ in the importance-sampled path from the ensemble of $\mathbf{y}_i$'s according to a probability given by,

$$p_i = \frac{I(\mathbf{y}_i)}{\sum_{j=1}^M I(\mathbf{y}_j)}. \tag{38}$$

This approach has the advantage of not needing to compute the gradient of the importance function.

3.9 Reverse Path Sampling

Sometimes it is of interest to simulataneously obtain the transition rate and mechanisms from Ω_{A} to Ω_{B} and those in the reverse direction, i.e. from Ω_{B} to Ω_{A}. Interestingly, for importance sampling of the reverse process, the optimal biased probability kernel is simply,

$$\overline{\mathcal{K}}'_\tau(\mathbf{y} \,|\, \mathbf{x}) = \mathcal{K}_\tau(\mathbf{y} \,|\, \mathbf{x}) \, \frac{1 - I^{\mathrm{opt}}(\mathbf{y})}{1 - I^{\mathrm{opt}}(\mathbf{x})}. \tag{39}$$

This is because if $I^{\mathrm{opt}}(\mathbf{x})$ is the optimal importance function for transitions from A to B, then $1 - I^{\mathrm{opt}}(\mathbf{x})$ is the optimal importance function for the inverse transition, from B to A. Otherwise, the numerical procedure for sampling the paths for the reverse process remains identical to those for the forward process. The only modification lies in the change of importance functions, e.g. from $I(\mathbf{x})$ to $1 - I(\mathbf{x})$, and from $I(\mathbf{y}_i)$ to $1 - I(\mathbf{y}_i)$, in Eqs. (37)-(38).

The adaptive training scheme in Section 3.5 can also be modified accordingly. Trajectories are initiated near the boundaries of regions A and B, simultaneously. Then, the two trajectories are propagated following the biased transition path sampling mechanism described in Section 3.8 using the current neural network approximation $I(\mathbf{x}; \theta)$ with fixed parameters. Once a sufficient number of positions have been accumulated from these paths, the neural network parameters θ are updated by minimizing the loss function evaluated over this set of microscopic states. Subsequently, the algorithm iterates back to the sampling stage, employing the improved neural network importance function to sample microscopic states. This iterative cycle of alternating between sampling and optimization is repeated until the neural network importance function converges. Once the neural network has converged, the sampling process is expected to efficiently generate paths both from A to B and from B to A.

4 Results

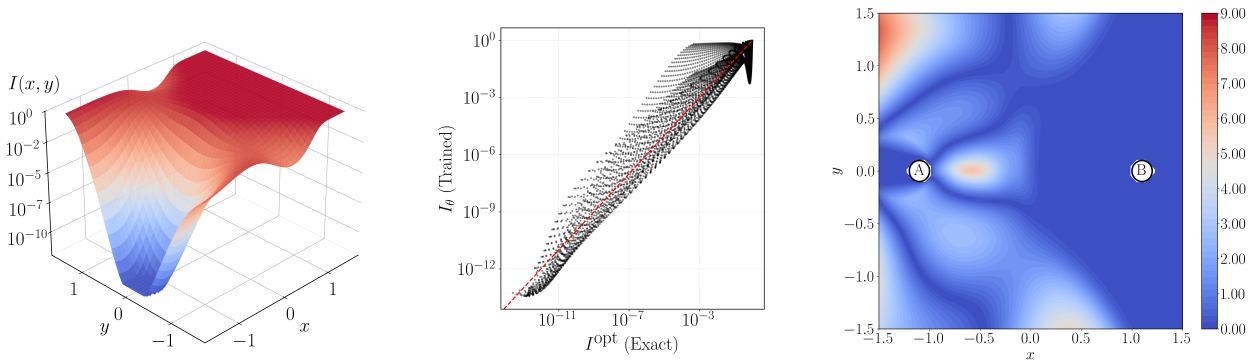
We validate the proposed importance sampling framework on a two-dimensional systems and a fourteen-dimensional system to demonstrate its robustness and scalability.

4.1 Two-Dimensional Potential with Two Reaction Channels

We apply the importance sampling framework to a two-dimensional system characterized by two distinct reaction channels [25, 28, 29, 57]. The potential energy (in units of eV) is defined as,

$$
\begin{aligned}
U_{\mathrm{TC}}(\mathbf{x}) &= \mathbf{c}^{\top}\mathbf{x} - \frac{1}{3} + \frac{1}{6}\sum_{i=1}^{4} w_i \left(\mathbf{x}^{T}\mathbf{Q}_i\,\mathbf{x} - b_i\right)^2, \\
\mathbf{c} &= \begin{bmatrix} 0 & 0.02 \end{bmatrix}^{\top}, \\
\{w_i\} &= \{4,\ \ 2,\ \ 1,\ \ 1\}, \\
\{b_i\} &= \{1,\ \ 2,\ \ 1,\ \ 1\}, \\
\{\mathbf{Q}_i\} &= \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right\},
\end{aligned}
\tag{40}
$$

where $\mathbf{x} = [x_1, x_2]^{\top} = [x, y]^{\top}$ is the position vector (in unit of nm). The system features two energy minima located at $\mathbf{x}_{\mathrm{A}} = [-1.1, 0]^{\top}$ nm and $\mathbf{x}_{\mathrm{B}} = [1.1, 0]^{\top}$ nm and two saddle points at $\mathbf{x}_{\mathrm{S}_1} = [0, 1]^{\top}$ nm and $\mathbf{x}_{\mathrm{S}_2} = [0, -1]^{\top}$ nm. The energy barrier associated with the saddle points are 1.02 eV at S_1 and 0.98 eV at S_2, respectively. The regions A and B are defined as circles of radius 0.1 nm centered at $\mathbf{x}_{\mathrm{A}}$ and $\mathbf{x}_{\mathrm{B}}$, respectively. Transitions from A to B are considered rare events, particularly at low temperatures. The dominant reaction pathways for this process go through the regions near the saddle points S_1 and S_2. The simulations are conducted using the coarse-grained Langevin dynamics in Section 3.1 with a friction coefficient of $\gamma = 1$ eV nm^{-2} fs and a time step size of $\Delta t = 10^{-2}$ fs.



(a) **(b)** **(c)**

Figure 3: (a) Neural network parameterized importance function trained for 2,000 iterations of adaptive training. (b) Parity plot between the neural network importance function and the exact optimal importance function. (c) Absolute error between the log of the neural network importance function and the exact optimal importance function.

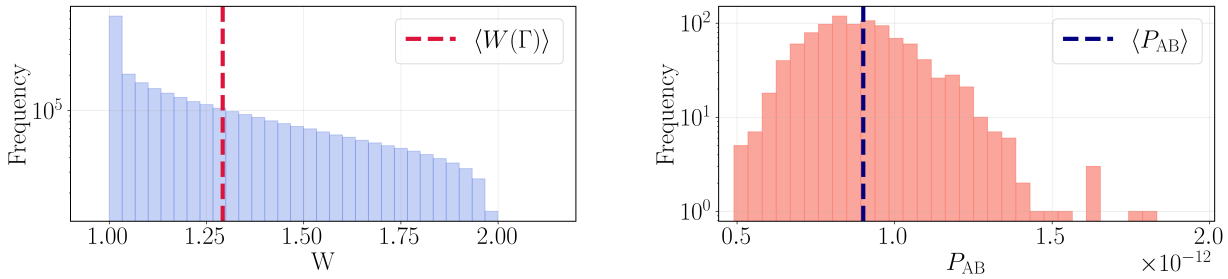
The importance function is parameterized as the logistic sigmoid of a scalar potential $h(\mathbf{x};\,\theta)$,

$$
I(\mathbf{x};\,\theta) = \sigma\big(h(\mathbf{x};\,\theta)\big) = \frac{1}{1 + \mathrm{e}^{-h(\mathbf{x};\,\theta)}}. \tag{41}
$$

The potential $h(\mathbf{x}, \theta)$ is defined as the sum of a fixed Gaussian term and a learnable neural network correction,

$$
h(\mathbf{x};\,\theta) = A\,\exp\left[-\sum_{i=1}^{2} a_i\,(x_i - c_i)^2\right] + \mathrm{MLP}(\mathbf{x};\,\theta). \tag{42}
$$

Here, the Gaussian parameters are set to $A = 3$, $(a_1, a_2) = (1, 1)$, and $(c_1, c_2) = (-1.1, 0)$. The neural network component (MLP) consists of two hidden layers with 50 neurons each, using hyperbolic tangent (`tanh`) activation functions. The neural network importance function is trained using the adaptive algorithm described in Section 3.5. A single walker is initialized at a temperature of $T = 500$ K, and the training proceeds for 1,000 iterations. We use the ADAM optimizer [61] with a learning rate of 5×10^{-3}. Figure 3a shows the trained neural network importance function over the 2-dimensional domain. Figure 3b compares the neural network importance function with the exact optimal importance function obtained using finite element method (FEM) [62]. Figure 3c plots the error between the neural network importance function and the exact optimal importance function over the 2-dimensional domain. Notice that the largest error arises in sparsely sampled regions remote from the dominant transition pathways.



(a) **(b)**

Figure 4: Transition path statistics of two-dimensional system with two reaction channels. (a) Distribution of weights $W(\Gamma)$ for paths sampled from region A to region B. (b) Distribution of the transition probabilities P_{AB} from region A to region B estimated using the sampled transition paths and their corresponding weights.

The trained importance function is then employed to accelerate transition of the system at $T = 500$ K using the importance sampling framework described in Sections 3.2 and 3.8. The transition paths are sampled using the branching random walk (BRW) algorithm in Section 3.4, performing 1,000 independent trials with 1,000 paths sampled per trial. To control the range of the path weights, the branching weight thresholds are constrained to the range $[1.0, 2.0]$. Figure 4a displays the weight distribution of the successful transition paths sampled from A to B. The deviation of the weights from unity is a result of the approximation errors present in the neural network importance function.

The sampled transition paths are used to estimate the success probability P_{AB} following Section 3.3. Figure 4b presents the distributions of the estimates for P_{AB}. The estimated success probability is $P_{AB} = (9.0067 \pm 0.0558) \times 10^{-13}$. In contrast, direct evaluation from Eq. (31), assuming the neural network importance is optimal yields significantly biased value of $P_{AB}^{NS} = 3.2028 \times 10^{-13}$ (no sampling). This discrepancy underscores the need to use the unbiased transition probability estimator in Eq. (29) to correct for approximation errors in the neural network importance function. The flux out of A is estimated as described in Section 3.6. From 10^5 configurations sampled exiting from region A, the flux is estimated to be $J_A = 8.0736$ fs^{-1}. Combining the flux with the estimated success probability yields the transition rate, $r_{AB} = (7.2717 \pm 0.0450) \times 10^{-12}$ fs^{-1}. The estimated rate is in good agreement with the rate obtained from the Kramers rate theory [57] of $r_{AB}^{Kr} = 7.1755 \times 10^{-12}$ fs^{-1}, as shown in Table 1a. In contrast, if the rate is estimated directly from Eq. (31), i.e. without any path sampling, but assuming the neural network trained $I(\mathbf{x})$ is already $I^{opt}(\mathbf{x})$, the result is $r_{AB}^{NS} = 2.5858 \times 10^{-12}$ fs^{-1} (no sampling), which is very different from the correct rate. This comparison illustrates the necessity of performing importance sampling of paths after the training of the importance function has converged, in order to obtain unbiased estimates of the transition rate.

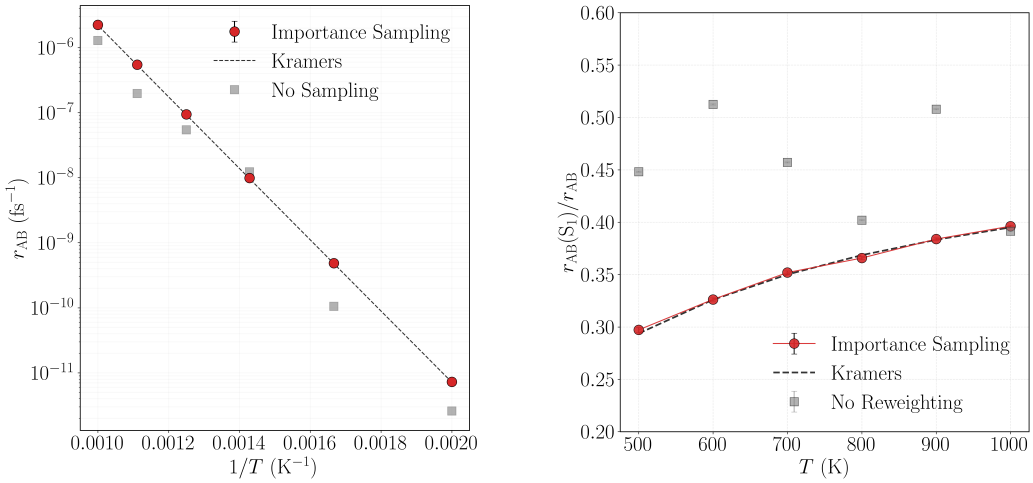
Method	r_{AB}
Importance Sampling	7.2717 ± 0.0450
Kramers	7.1755
No Sampling	2.5858

(a) Transition rates (in 10^{-12} fs^{-1})

Method	$r_{AB}(S_1)/r_{AB}$
Importance Sampling	0.2972 ± 0.0023
Kramers	0.2938
No Reweighting	0.4483

(b) Transition rate fractions

Table 1: Comparison of results for the 2-dimensional system at 500 K. (a) Estimated transition rates obtained from importance sampling, Kramers rate theory, and no sampling (using Eq. (31) assuming $I(\mathbf{x}) \approx I^{\text{opt}}(\mathbf{x})$). (b) Transition rate fraction through saddle point S_1 obtained from importance sampling, Kramers rate theory, and direct counting without reweighting.



(a) **(b)**

Figure 5: (a) Comparison of transition rates estimated at various temperatures using importance sampling (circles), Kramers rate theory (dashed line), and no sampling (squares). (b) Comparison of fraction of rate through the saddle region near S_1 at various temperatures using importance sampling (circles), Kramers rate theory (dashed line), and no reweighting (squares).

The importance sampling framework also enables accurate determination of relative transition probabilities between competing reaction channels. The transition paths are sampled using the branching random walk (BRW) at multiple temperatures for 1,000 trials with 1,000 paths each. The sampled transition paths are classified based on their passage through the saddle regions near S_1 or S_2. Then, the relative transition probabilities are estimated using Eq. (30) considering the weights of the transition paths. At $T = 500$ K, the importance sampling framework yields transition fraction of 0.2972. This estimate aligns closely with the Kramers rate theory prediction of 0.2938, as shown in Table 1b. In contrast, directly counting the number of transition paths without considering their weights produces inaccurate estimates, 0.4483 at 500 K.

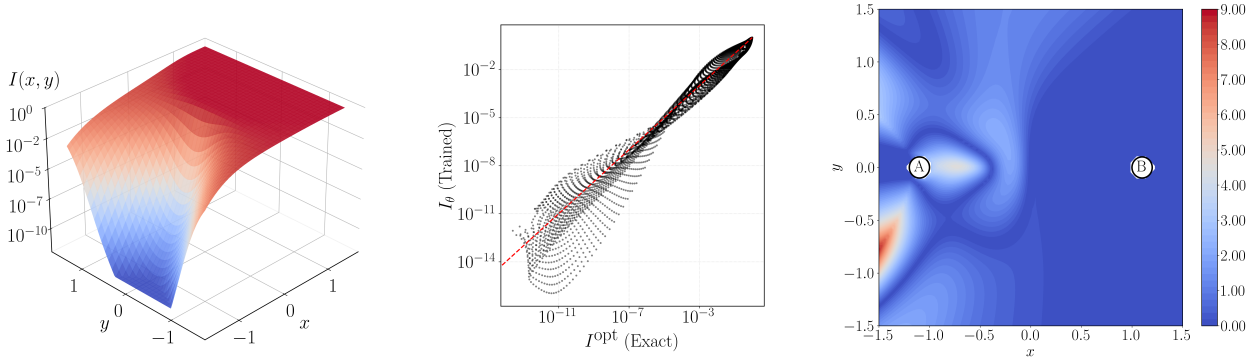
To test the accuracy of the importance sampling framework, the transition rate estimation procedure is repeated at various temperatures ranging from 500 K to 1000 K. Figure 5a shows the Arrhenius plot of the transition rates obtained using direct evaluation, importance sampling estimation, and the Kramers rate theory. Figure 5b shows the fraction of the transitions that proceed near S_1 at various temperatures. The importance sampling results show good agreement with the Kramers rate theory, whereas rate estimates without performing importance sampling show significant errors, especially at low temperatures.

4.2 Fourteen-Dimensional Potential with Two Reaction Channels

We further validate the proposed importance sampling framework on a fourteen-dimensional system characterized by two distinct reaction channels [57]. The potential energy surface is defined as,

$$U_{\mathrm{TC14D}}(\mathbf{x}) = U_{\mathrm{TC}}(\mathbf{x}_{1:2}) + \frac{1}{\sigma^2}\,\mathbf{x}_{3:14}^{\top}\,\mathbf{x}_{3:14}, \tag{43}$$

where $\mathbf{x} \in \mathbb{R}^{14}$ is the position vector and $\mathbf{x}_{i:j}$ denotes the subvector spanning indices i through j (in units of nm). The local minima for the energy basins are located at $\mathbf{x}_{\mathrm{A}} = [-1.1,\, 0,\, 0, \ldots, 0]^{\top}$ nm and $\mathbf{x}_{\mathrm{B}} = [1.1,\, 0,\, 0, \ldots, 0]^{\top}$ nm. The saddle points are located at $\mathbf{x}_{\mathrm{S}_1} = [0,\, 1,\, 0, \ldots, 0]^{\top}$ nm and $\mathbf{x}_{\mathrm{S}_2} = [0,\, -1,\, 0, \ldots, 0]^{\top}$ nm. We define regions A and B as hypercylinders with radius 0.1 nm with center lines passing through $\mathbf{x}_{\mathrm{A}}$ and $\mathbf{x}_{\mathrm{B}}$, respectively, in the (x_1, x_2) subspace. For example, region A consists of all configurations satisfying $\sqrt{(x_1 + 1.1)^2 + x_2^2} < 0.1^2$. By construction, the optimal importance function of this system only depends on the first two dimensions, which enables computation of the optimal importance function via the finite element method (FEM) on the 2D subspace. The parameter for the harmonic confinement width is $\sigma = 2.0\ \mathrm{nm}^{1/2}\ \mathrm{eV}^{-1/2}$. The dynamics are simulated with friction $\gamma = 1\ \mathrm{eV\,nm^{-2}\,fs}$, temperature 500 K, and a time step size of $\Delta t = 10^{-2}$ fs.



(a) **(b)** **(c)**

Figure 6: (a) Neural network parameterized importance function trained for 5,000 iterations of adaptive training. (b) Parity plot between the trained importance function and the exact optimal importance function. (c) Absolute error between the log of the neural network importance function and the exact optimal importance function.

For the 14-dimensional system, the importance function is similarly parameterized as the logistic sigmoid of a scalar potential $h(\mathbf{x};\, \theta)$,

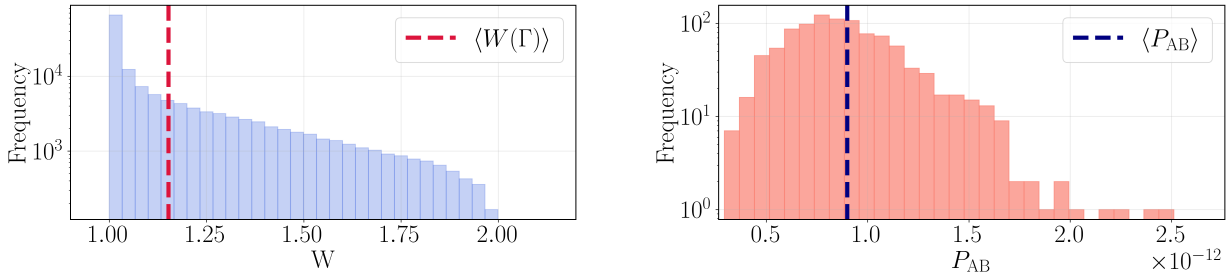
$$I(\mathbf{x};\, \theta) = \sigma\big(h(\mathbf{x};\, \theta)\big) = \frac{1}{1 + \mathrm{e}^{-h(\mathbf{x};\, \theta)}}. \tag{44}$$

The potential $h(\mathbf{x};\, \theta)$ is defined as the superposition of a Gaussian term and a neural network correction,

$$h(\mathbf{x};\, \theta) = A\, \exp\left[-\sum_i a_i\, (x_i - c_i)^2\right] + \mathrm{MLP}(\mathbf{x};\, \theta). \tag{45}$$

Different from the 2-dimensional case, the parameters of the Gaussian term $\{A,\, \{a_i\},\, \{c_i\}\}$ are treated as trainable variables, allowing the guiding potential to adapt dynamically during the optimization. The neural network component (MLP) consists of two hidden layers with 50 neurons each and hyperbolic tangent (`tanh`) activation functions. The total set of parameters $\{A,\, \{a_i\},\, \{c_i\},\, \theta\}$ is optimized using the adaptive training algorithm described in Section 3.5. Training is performed at $T = 500$ K for 5,000 iterations using

the ADAM optimizer [61], with a learning rate of 5×10^{-3}. Figure 6a shows the trained neural network importance function projected to a 2-dimensional domain by setting $\mathbf{x}_i = 0$ for $3 \leq i \leq 14$. Figure 6b compares the trained importance function with the exact optimal importance function obtained using finite element method (FEM) [62]. Figure 6c plots the error between the trained importance function and the exact optimal importance function on the 2-dimensional domain. Notice that the largest error arise in sparsely sampled regions remote from the dominant transition pathways.



(a) **(b)**

Figure 7: Transition path statistics of 14-dimensional system with two reaction channels. (a) Distribution of path weights $W(\Gamma)$ forpaths from region A to region B. (b) Distribution of the success probabilities P_{AB} from region A to region B estimated using the sampled paths and their corresponding weights.

The neural network importance function is then employed to accelerate transition of the system at $T = 500$ K using the importance sampling framework described in Sections 3.2 and 3.8. The transition paths are sampled using the branching random walk (BRW) algorithm in Section 3.4, performing 1,000 independent trials with 1,000 paths sampled per trial. To control the range of the path weights, the branching weight thresholds are set to the range $[1.0, 2.0]$. Figure 7a displays the weight distribution of the successful transition paths sampled from A to B. The deviation of the weights from unity is the result from the approximation errors present in the neural network importance function.

The sampled transition paths are used to estimate the success probability P_{AB} following Section 3.3. Figure 7b present the distribution of the estimated success probabilities for P_{AB}. The estimated success probability is $P_{AB} = (8.9963 \pm 0.0942) \times 10^{-13}$. In contrast, if the success probability is estimated directly from Eq. (31), without importance sampling of paths, but assuming the trained $I(\mathbf{x})$ is already $I^{opt}(\mathbf{x})$, then the result is $P_{AB}^{NS} = 8.4594 \times 10^{-12}$ (no sampling), which is nearly a factor of 10 too high. This discrepancy underscores the use of the unbiased transition probability estimator in Eq. (29) to correct for approximation errors in the neural network importance function. The flux out of A is estimated as described in Section 3.6. From 10^5 sampled configurations exiting region A, the flux is estimated to be $J_A = 8.1180\ s^{-1}$. Combining these fluxes with the estimated success probability yields the transition rate, $r_{AB} = (7.3033 \pm 0.0765) \times 10^{-12}\ fs^{-1}$. The estimated rate is in excellent agreement with the rate obtained from the Kramers rate theory of $r_{AB}^{Kr} = 7.1755 \times 10^{-12}\ fs^{-1}$, as shown in Table 2a. In contrast, if the rate is estimated directly from Eq. (31), i.e. without any path sampling, but assuming the neural network trained $I(\mathbf{x})$ is already $I^{opt}(\mathbf{x})$, the result is $r_{AB}^{NS} = 6.8674 \times 10^{-11}\ fs^{-1}$ (no sampling), which is very different from the correct rate.

The importance sampling framework also enables the determination of relative transition probabilities between competing reaction channels. The transition paths are sampled using the branching random walk (BRW) at $T = 500$ K for 1,000 trials with 1,000 paths each. The sampled transition paths are classified based on their passage through the saddle region near S_1 or S_2. Then, the transition rate fraction through the saddle

Method	r_{AB}
Importance Sampling	7.3033 ± 0.0765
Kramers	7.1755
No Sampling	68.6741

(a) Transition rates (in 10^{-12} fs^{-1})

Method	$r_{AB}(S_1)/r_{AB}$
Importance Sampling	0.3003 ± 0.0059
Kramers	0.2938
No Reweighting	0.4391

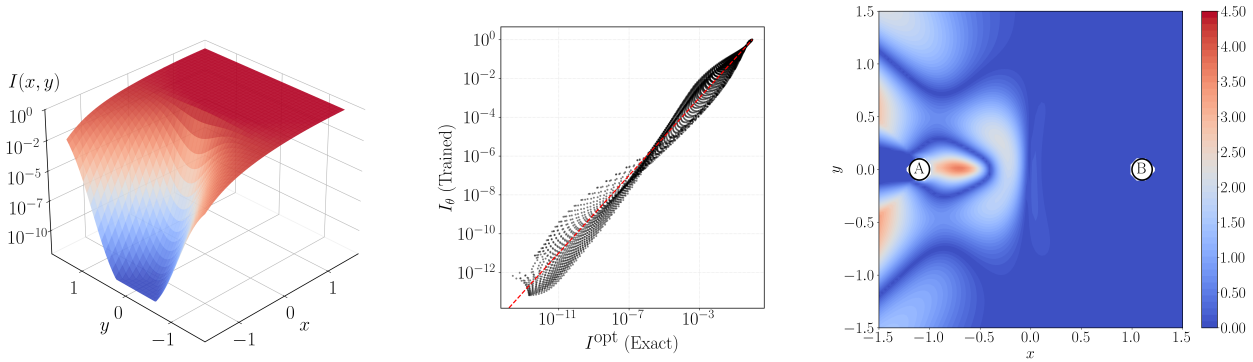
(b) Transition rate fractions

Table 2: Comparison of results for the 14-dimensional system at 500 K. (a) Estimated transition rates obtained from importance sampling, Kramers rate theory, and no sampling (using Eq. (31) assuming $I(\mathbf{x}) \approx I^{\text{opt}}(\mathbf{x})$). (b) Transition rate fraction through saddle point S_1 obtained from importance sampling, Kramers rate theory, and direct counting without reweighting.

point S_1 is estimated using Eq. (30) considering the weights of the transition paths. The importance sampling framework yields transition fraction of 0.3003. This estimate aligns closely with the Kramers rate theory prediction of 0.2938, as shown in Table 2b. In contrast, directly counting the number of transition paths without considering their weights produces an inaccurate estimate, 0.4391.

4.3 Sampling Bidirectional Transitions

Here we demonstrate how to use a single neural network representation of the importance function to boost sample efficiencies of both forward ($\Omega_A \to \Omega_B$) and backward ($\Omega_B \to \Omega_A$) transitions between two metastable states. As described in Section 3.9, the method requires relatively small changes to that used for boosting transitions out of metastable state Ω_A. The demonstrations are performed on the same 2-dimensional and 14-dimensional systems considered in Sections 4.1 and 4.2.



(a) **(b)** **(c)**

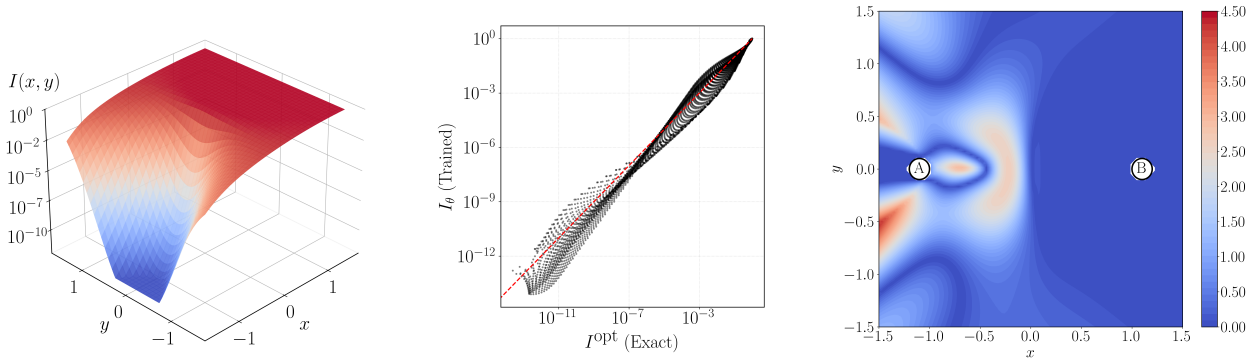
Figure 8: Neural network importance function for 2-dimensional system using bidirectional path sampling. (a) Neural network parameterized importance function trained for 2,000 iterations of adaptive training. (b) Parity plot between the trained importance function and the exact optimal importance function. (c) Absolute error between the log of the neural network importance function and the exact optimal importance function.

We construct the importance function as the sigmoid of a multi-layer perceptron (MLP) output without the superposition of the Gaussian term in the previous sections,

$$I(\mathbf{x};\, \theta) = \sigma\left(\text{MLP}(\mathbf{x};\, \theta)\right). \tag{46}$$

The MLP for the 2-dimensional system consists of two hidden layers with 50 neurons each, utilizing hyperbolic tangent (`tanh`) activation functions. The neural network importance function is trained for 2,000 epochs using the ADAM optimizer [61] with a learning rate of 5×10^{-3}. Figure 8a displays the learned importance

function at 500 K on the 2D surface. Figure 8b compares this result against the exact solution calculated using the finite element method (FEM) [62]. Finally, Figure 8c maps the error between the neural network and the exact solution across the domain. The MLP for the 14-dimensional system consists of two hidden



(a) **(b)** **(c)**

Figure 9: Neural network importance function for 14-dimensional system using bidirectional path sampling. Importance functions are projected onto a 2-dimensional space of $x_i = 0$ for $3 \leq i \leq 14$ for visualization. (a) Neural network parameterized importance function trained for 5,000 iterations of adaptive training. (b) Parity plot between the trained importance function and the exact optimal importance function. (c) Absolute error between the log of the neural network importance function and the exact optimal importance function.

layers with 100 neurons each, utilizing hyperbolic tangent (`tanh`) activation functions. The neural network importance function is trained for 5,000 epochs using the ADAM optimizer [61] with a learning rate of 5×10^{-3}. Figure 9a displays the learned function projected onto 2D by fixing the other dimensions to zero. Figure 9b compares the trained result against the exact solution calculated via FEM [62]. Figure 9c maps the difference between the neural network importance function and the exact solution across the domain. It can be seen that sampling both forward and backward paths during adaptive training leads to higher accuracy of the importance function (compared with the exact solution) after the same number of epochs of adaptive training, compared with only sampling paths escaping Ω_A (as in Sections 4.1 and 4.2). In addition, there is no longer a need to include a Gaussian term (and to select its initial coefficients) in the construction of the importance function, making the method somewhat easier to use.

The transition rates for both the forward ($\Omega_A \to \Omega_B$) and backward ($\Omega_B \to \Omega_A$) directions were estimated using the sampled transition paths. For the 2-dimensional system, we performed 1,000 independent trials for 1,000 paths each using the branching random walk (BRW) with the weight range of $[1.0, 2.0]$. The estimated transition rates r_{AB} and r_{BA} at $T = 500$ K are summarized in Table 3a. The importance sampling results demonstrate excellent agreement with Kramers rate theory across all temperatures. In contrast, the method without importance sampling, which assumes the learned importance function is optimal, yields significantly biased rates. This confirms the necessity of the unbiased estimator to correct for approximation errors in the importance function. Furthermore, the importance sampling method accurately resolves the relative transition probabilities between competing reaction channels. The transition rate fractions through saddle region near the saddle point S_1 are estimated using the path weights. As shown in Table 3b, the importance sampling estimates align closely with theoretical predictions for both forward and backward transitions. Conversely, direct counting of transition paths without reweighting produces inaccurate fractional estimates, failing to correctly capture the preference for the transition mechanism.

We extend this analysis to the 14-dimensional system at $T = 500$ K to demonstrate scalability to higher dimensions. Using the branching random walk (BRW) technique with 1,000 trials of 1,000 paths each and

Method	r_{AB}	r_{BA}
Importance Sampling	7.2093 ±0.0472	7.3635 ±0.0528
Kramers	7.1755	7.1755
No Sampling	21.6420	25.3386

(a) Transition rates (in 10^{-12} fs^{-1})

Method	$r_{AB}(S_1)/r_{AB}$	$r_{BA}(S_1)/r_{BA}$
Importance Sampling	0.2963 ±0.0030	0.2866 ±0.0033
Kramers	0.2938	0.2938
No Reweighting	0.4602	0.4619

(b) Transition rate fractions

Table 3: Comparison of results for the 2-dimensional system at 500 K. (a) Estimated transition rates obtained from importance sampling, Kramers rate theory, and no sampling. (b) Transition rate fraction through saddle point S_1 obtained from importance sampling, Kramers rate theory, and direct counting without reweighting.

a weight range of $[1.0, 2.0]$, we estimate both the transition rates and relative channel probabilities for the forward ($\Omega_A \to \Omega_B$) and backward ($\Omega_B \to \Omega_A$) directions. The results, summarized in Table 4a, show that the importance sampling estimates for the transition rates (r_{AB} and r_{BA}) are in excellent agreement with the Kramers theory predictions. Similarly, the estimated transition fractions through saddle point S_1, $f_{AB}(S_1)$ and $f_{BA}(S_1)$, closely match the theoretical values as shown in Table 4b. In contrast, the direct evaluation without reweighting yields significantly overestimated rates and incorrect transition fractions for both directions. These findings confirm that the importance sampling framework is able to correct for the approximation errors in the importance function when predicting the transition rate and rate fractions in high-dimensional systems.

Method	r_{AB}	r_{BA}
Importance Sampling	7.3067 ±0.0523	7.2655 ±0.0552
Kramers	7.1755	7.1755
No Sampling	12.0419	36.2771

(a) Transition rates (in 10^{-12} fs^{-1})

Method	$r_{AB}(S_1)/r_{AB}$	$r_{BA}(S_1)/r_{BA}$
Importance Sampling	0.2927 ±0.0033	0.2958 ±0.0031
Kramers	0.2938	0.2938
No Reweighting	0.4811	0.4885

(b) Transition rate fractions

Table 4: Comparison of results for the 14-dimensional system at 500 K. (a) Estimated transition rates obtained from importance sampling, Kramers rate theory, and no sampling. (b) Transition rate fraction through saddle point S_1 obtained from importance sampling, Kramers rate theory, and direct counting without reweighting.

5 Conclusions

We developed a rigorous importance sampling framework designed to efficiently sample rare escapes from metastable states in systems governed by Langevin dynamics. This method counteracts the inevitable approximation errors for the optimal importance functions in high dimensions. As a result, our approach achieves two primary objectives: (1) accelerating the sampling of transition paths while preserving the relative probabilities between different transition pathways and (2) providing an unbiased estimator for transition rates. These goals are accomplished by assigning statistical weights to the sampled paths. Furthermore, to ensure computational efficiency and numerical stability, we developed a branching random walk (BRW) algorithm that actively controls the variance of the weights of the paths. We validated the framework on a 2-dimensional and a 14-dimensional system. The results confirm that the method accurately obtains transition rates and relative probabilities of different escape channels, even in the presence of approximation errors in the importance function. Consequently, this work offers a practical solution for the accurate kinetic characterization of complex, high-dimensional systems where determining the exact importance function (a.k.a. committor) is computationally prohibitive and approximations are necessary. Future work will focus on extending this framework to atomistic systems to simulate complex rare events, such as protein conformational changes and crystal defect evolution.

References

[1] David Van Der Spoel et al. "GROMACS: fast, flexible, and free". In: *Journal of computational chemistry* 26.16 (2005), pp. 1701–1718.

[2] Peter Eastman et al. "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics". In: *PLoS computational biology* 13.7 (2017), e1005659.

[3] Aidan P Thompson et al. "LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales". In: *Computer physics communications* 271 (2022), p. 108171.

[4] Arthur F Voter, Francesco Montalenti, and Timothy C Germann. "Extending the time scale in atomistic simulation of materials". In: *Annual review of materials research* 32.1 (2002), pp. 321–346.

[5] Erik Van Der Giessen et al. "Roadmap on multiscale materials modeling". In: *Modelling and Simulation in Materials Science and Engineering* 28.4 (2020), p. 043001.

[6] Lydia Freddolino et al. "Challenges in protein-folding simulations". In: *Nature physics* 6.10 (2010), pp. 751–758.

[7] Sergei Izrailev et al. "Steered molecular dynamics". In: *Computational Molecular Dynamics: Challenges, Methods, Ideas: Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, May 21–24, 1997*. Springer. 1999, pp. 39–65.

[8] Alessandro Laio and Francesco L Gervasio. "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science". In: *Reports on Progress in Physics* 71.12 (2008), p. 126601.

[9] Yong Wang et al. "Frequency adaptive metadynamics for the calculation of rare-event kinetics". In: *The Journal of chemical physics* 149.7 (2018).

[10] Alessandro Laio and Michele Parrinello. "Escaping free-energy minima". In: *Proceedings of the national academy of sciences* 99.20 (2002), pp. 12562–12566.

[11] Gareth A Tribello, Michele Ceriotti, and Michele Parrinello. "A self-learning algorithm for biased molecular dynamics". In: *Proceedings of the National Academy of Sciences* 107.41 (2010), pp. 17509–17514.

[12] Arthur F Voter. "Hyperdynamics: Accelerated molecular dynamics of infrequent events". In: *Physical Review Letters* 78.20 (1997), p. 3908.

[13] LY Chen and NJM Horing. "An exact formulation of hyperdynamics simulations". In: *The Journal of chemical physics* 126.22 (2007).

[14] Rosalind J Allen, Chantal Valeriani, and Pieter Rein Ten Wolde. "Forward flux sampling for rare event simulations". In: *Journal of physics: Condensed matter* 21.46 (2009), p. 463102.

[15] Gary A Huber and Sangtae Kim. "Weighted-ensemble Brownian dynamics simulations for protein association reactions". In: *Biophysical journal* 70.1 (1996), pp. 97–110.

[16] Dhiman Ray, Sharon Emily Stone, and Ioan Andricioaei. "Markovian weighted ensemble milestoning (M-WEM): Long-time kinetics from short trajectories". In: *Journal of Chemical Theory and Computation* 18.1 (2021), pp. 79–95.

[17] Eric Vanden-Eijnden. "Transition path theory". In: *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*. Springer, 2006, pp. 453–493.

[18] Eric Vanden-Eijnden et al. “Transition-path theory and path-finding algorithms for the study of rare events.” In: *Annual review of physical chemistry* 61 (2010), pp. 391–420.

[19] Eric Vanden-Eijnden et al. “Towards a theory of transition paths”. In: *Journal of statistical physics* 123.3 (2006), pp. 503–523.

[20] Christoph Dellago, Peter G Bolhuis, and Phillip L Geissler. “Transition path sampling”. In: *Advances in chemical physics* 123 (2002), pp. 1–78.

[21] E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. “Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes”. In: *Chemical Physics Letters* 413.1-3 (2005), pp. 242–247.

[22] Peter G Bolhuis et al. “Transition path sampling: Throwing ropes over rough mountain passes, in the dark”. In: *Annual review of physical chemistry* 53.1 (2002), pp. 291–318.

[23] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. “Transition path theory for Markov jump processes”. In: *Multiscale Modeling & Simulation* 7.3 (2009), pp. 1192–1219.

[24] Christoph Dellago, Peter G Bolhuis, and David Chandler. “Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements”. In: *The Journal of chemical physics* 108.22 (1998), pp. 9236–9245.

[25] Christoph Dellago et al. “Transition path sampling and the calculation of rate constants”. In: *The Journal of chemical physics* 108.5 (1998), pp. 1964–1977.

[26] David J Wales. “Discrete path sampling”. In: *Molecular physics* 100.20 (2002), pp. 3285–3305.

[27] Peter W Glynn and Donald L Iglehart. “Importance sampling for stochastic simulations”. In: *Management science* 35.11 (1989), pp. 1367–1392.

[28] Wei Cai et al. “Importance sampling of rare transition events in Markov processes”. In: *Physical Review E* 66.4 (2002), p. 046703.

[29] Maurice de Koning et al. “Adaptive importance sampling Monte Carlo simulation of rare transition events”. In: *The Journal of chemical physics* 122.7 (2005).

[30] Pierre L’Ecuyer, Michel Mandjes, and Bruno Tuffin. “Importance sampling in rare event simulation”. In: *Rare event simulation using Monte Carlo methods* (2009), pp. 17–38.

[31] Zhan Shi et al. *Branching random walks*. Vol. 2151. Springer, 2015.

[32] Paul Langevin et al. “Sur la théorie du mouvement brownien”. In: *CR Acad. Sci. Paris* 146.530-533 (1908), p. 530.

[33] Norbert Wiener. “Differential-space”. In: *Journal of Mathematics and Physics* 2.1-4 (1923), pp. 131–174.

[34] Ludwig Boltzmann. “Über die Beziehung eines allgemeinen mechanischen Satzes zum zweiten Hauptsatze der Wärmetheorie”. In: *Kinetische Theorie II: Irreversible Prozesse Einführung und Originaltexte*. Springer, 1970, pp. 240–247.

[35] Terrell Hill. *Free energy transduction in biology: the steady-state kinetic and thermodynamic formalism*. Elsevier, 2012.

[36] Terrell L Hill. *Free energy transduction and biochemical cycle kinetics*. Courier Corporation, 2013.

[37] Manon Baudel, Arnaud Guyader, and Tony Lelièvre. “On the Hill relation and the mean reaction time for metastable processes”. In: *Stochastic Processes and their Applications* 155 (2023), pp. 393–436.

[38] Andrew R Mitchell and Grant M Rotskoff. "Committor guided estimates of molecular transition rates". In: *Journal of Chemical Theory and Computation* 20.21 (2024), pp. 9378–9393.

[39] GN Mil'shtejn. "Approximate integration of stochastic differential equations". In: *Theory of Probability & Its Applications* 19.3 (1975), pp. 557–562.

[40] Carsten Hartmann et al. "Characterization of rare events in molecular dynamics". In: *Entropy* 16.1 (2013), pp. 350–376.

[41] Frédéric Legoll and Tony Lelievre. "Effective dynamics using conditional expectations". In: *Nonlinearity* 23.9 (2010), p. 2131.

[42] Jiaxin Yuan et al. "Optimal control for sampling the transition path process and estimating rates". In: *Communications in Nonlinear Science and Numerical Simulation* 129 (2024), p. 107701.

[43] Joseph L Doob. "Conditional Brownian motion and the boundary limits of harmonic functions". In: *Bulletin de la Société mathématique de France* 85 (1957), pp. 431–458.

[44] Igor Vladimirovich Girsanov. "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures". In: *Theory of Probability & Its Applications* 5.3 (1960), pp. 285–301.

[45] Lorenzo Donati, Carsten Hartmann, and Bettina G Keller. "Girsanov reweighting for path ensembles and Markov state models". In: *The Journal of chemical physics* 146.24 (2017).

[46] Luca Donati and Bettina G Keller. "Girsanov reweighting for metadynamics simulations". In: *The Journal of chemical physics* 149.7 (2018).

[47] Chatipat Lorpaiboon, Jonathan Weare, and Aaron R Dinner. "An exact multiple-time-step variational formulation for the committor and the transition rate". In: *arXiv preprint arXiv:2509.03539* (2025).

[48] Jiří Vorba and Jaroslav Křivánek. "Adjoint-driven Russian roulette and splitting in light transport simulation". In: *ACM Transactions on Graphics (TOG)* 35.4 (2016), pp. 1–11.

[49] James Arvo and David Kirk. "Particle transport and image synthesis". In: *Proceedings of the 17th annual conference on Computer graphics and interactive techniques*. 1990, pp. 63–66.

[50] Alexander Rath et al. "EARS: efficiency-aware russian roulette and splitting". In: *ACM Transactions on Graphics (TOG)* 41.4 (2022), pp. 1–14.

[51] Haoya Li et al. "A semigroup method for high dimensional committor functions based on neural network". In: *Mathematical and Scientific Machine Learning*. PMLR. 2022, pp. 598–618.

[52] John Strahan et al. "Predicting rare events using neural networks and short-trajectory data". In: *Journal of computational physics* 488 (2023), p. 112152.

[53] Enrico Trizio, Peilin Kang, and Michele Parrinello. "Everything everywhere all at once: a probability-based enhanced sampling approach to rare events". In: *Nature Computational Science* (2025), pp. 1–10.

[54] Hannes Risken. "Fokker-planck equation". In: *The Fokker-Planck equation: methods of solution and applications*. Springer, 1989, pp. 63–95.

[55] Paul Erd. "On a new law of large numbers". In: *J. Anal. Muth* 22 (1970), pp. 103–l.

[56] Xinru Hua et al. "Accelerated sampling of rare events using a neural network bias potential". In: *arXiv preprint arXiv:2401.06936* (2024).

[57] Michael Kim and Wei Cai. “Accelerated Markov Chain Monte Carlo Simulation via Neural Network-Driven Importance Sampling”. In: *arXiv preprint arXiv:2602.12294* (2026).

[58] Jun S Liu, Rong Chen, and Tanya Logvinenko. “A theoretical framework for sequential importance sampling with resampling”. In: *Sequential Monte Carlo methods in practice*. Springer, 2001, pp. 225–246.

[59] Justin F Talbot. *Importance resampling for global illumination*. Brigham Young University, 2005.

[60] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. “On adaptive resampling strategies for sequential Monte Carlo methods”. In: (2012).

[61] Diederik P Kingma. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[62] Igor A Baratta et al. “DOLFINx: the next generation FEniCS problem solving environment”. In: (2023).