JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

Committor Guided Estimates of Molecular Transition Rates

Published as part of Journal of Chemical Theory and Computation special issue "Machine Learning and Statistical Mechanics: Shared Synergies for Next Generation of Chemical Theory and Computation".

Andrew R. Mitchell and Grant M. Rotskoff*

Cite This: J. Chem. Theory Comput. 2024, 20, 9378−9393

Read Online

ACCESS | Metrics & More | Article Recommendations

Downloaded via UNIV OF CHICAGO on May 13, 2025 at 08:53:57 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ABSTRACT: The probability that a configuration of a physical system reacts, or transitions from one metastable state to another, is quantified by the committor function. This function contains richly detailed mechanistic information about transition pathways, but a full parametrization of the committor requires the construction of a high-dimensional function, a generically challenging task. Recent efforts to leverage neural networks as a means to solve high-dimensional partial differential equations, often called "physics-informed" machine learning, have brought the committor into computational reach. Here, we build on the semigroup approach to learning the committor and assess its utility for predicting dynamical quantities such as transition rates. We show that a careful reframing of the objective function and improved adaptive sampling strategies provide highly accurate representations of the committor. Furthermore, by directly applying the Hill relation, we show that these committors provide accurate transition rates for molecular systems.

1. INTRODUCTION

A statistically representative collection of transition pathways in a molecular system can yield profound physical insight about a complex reaction.[1] Experimentally characterizing reaction pathways with atomistic resolution in biophysical systems, however, is exceedingly difficult due to the disparate time scales and collective molecular fluctuations at play. Molecular dynamics simulation provides a toolkit that, in principle, provides a window into these dynamical events, but the temporal and spatial scales accessible with atomic resolution molecular dynamics lead to a complementary set of challenges. The kinetics of many molecular processes of interest are dominated by rare events inaccessible to current computational methods, and dimensionality reduction and approximations must thus be invoked.[2]

Contemporary enhanced sampling methodology is sophisticated and diverse, but most modern methods still rely on low dimensional reaction coordinates[3] which can limit the accuracy of resulting free energy surfaces. Alternative approaches based on estimating high-dimensional committor[4−12] functions have recently gained traction due to the efficacy of neural networks as representations of high-dimensional functions. In large part, efforts to produce a representation of the committor have used data collected from enhanced sampling simulations, e.g., metadynamics[13] that incorporate a low-dimensional reaction coordinate, or have leveraged data from transition path sampling to optimize a representation of the committor,[11] which is itself nontrivial to collect.

Efficient data acquisition that is unbiased by a choice of reaction coordinate remains a significant bottleneck for these calculations, and our previous work[14] incorporates adaptive importance sampling, which guides sampling using the online approximation of the committor function. The variational objective function in that work and others,[4,5,14] however, requires diffusive, overdamped Langevin dynamics, which renders it inapplicable to most condensed phase molecular dynamics simulations, as underdamped dynamics is used to improve numerical stability. An alternative that avoids this

Received: July 31, 2024
Revised: October 1, 2024
Accepted: October 2, 2024
Published: October 18, 2024

ACS Publications © 2024 American Chemical Society

constraint appears in Li et al.,[7] but still requires the Dirichlet form and Boltzmann sampling. Similarly, Chen et al.[15] uses a semigroup formulation for the committor, representing the solution as the minimizer of a Boltzmann weighted mean-squared loss. These works formulate the committor using the Markov semigroup for the dynamics after some lag time τ, which ensures Markovianity and yields a stable objective. Using a stopped transition operator,[16] Strahan et al.[8] showed that the committor can be formulated as the solution of a Feynman-Kac equation which uses a mean-squared error but, crucially, does not require Boltzmann reweighting, tremendously simplifying evaluation of the objective. Still, the utility of representations trained with these approaches has not yet been thoroughly examined, especially in the setting of molecular simulations.

Here, we characterize the use of self-consistent committor optimization with active importance sampling and assess whether the representation can be made sufficiently accurate to compute transition rates. We characterize the performance in 2-dimensional toy potentials and in two minimal molecular systems to analyze the effect of the assumptions we make, while also having a clear ground truth comparison. We obtain highly accurate rate estimates, typically an exceedingly challenging task for molecular simulations.[17,18] Because our approach does not postulate a low dimensional reaction coordinate, it has the ability to simultaneously represent multiple reaction channels.

We use the Hill relation[19] to estimate transition rates, consistent with standard practice in algorithms like adaptive multilevel splitting. However, we find that rate estimates are extremely sensitive to imprecision in the committor near each metastable basin, because the inverse of the committor appears in the estimator. To overcome this, we develop and employ a biased estimator for the committor; interestingly, as we demonstrate both analytically and numerically, the bias in this estimator can be controlled with modest sample sizes. While the Hill relation applies most simply to the setting of single barrier crossing kinetics, we also test our approach on an example that features significant intermediate metastability. Rate estimates are more computationally demanding in this setting but still accurate, and the additional information provided by the committor produces useful insight into the reaction pathways.

2. THEORY

We consider a system with position and momenta coordinates $(\boldsymbol{x}, \boldsymbol{p}) \in \Omega \subset \mathbb{R}^d \times \mathbb{R}^d$ evolving under a stochastic dynamics in which the stationary probability measure of the configurational degrees of freedom is given by the Gibbs-Boltzmann probability measure,

$$\mu(\mathrm{d}\boldsymbol{x}) = Z^{-1}(\beta)e^{-\beta U(\boldsymbol{x})}\mathrm{d}\boldsymbol{x} \tag{1}$$

where $\beta = 1/(k_\mathrm{B}T)$ and $U\colon \mathbb{R}^d \to \mathbb{R}$ is the potential energy and $Z = \int_\Omega e^{-\beta(U(\boldsymbol{x})+\frac{1}{2}\boldsymbol{p}^T\mathbf{M}^{-1}\boldsymbol{p})}\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{p}$ is the partition function and $\mathbf{M}$ is the mass matrix. For molecular systems, we typically work with underdamped Langevin dynamics,

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

where the initial positions are sampled from some initial distribution $\boldsymbol{x}_0 \sim \rho_0$ and the momenta is selected from a Maxwell−Boltzmann distribution $\boldsymbol{p}_0 \sim \rho_\mathrm{MB}(\beta)$.

We now develop the variational formulation of the partial differential equation defining the probability of a reactive event. Throughout, we focus on a transition from one metastable state $A \subset \Omega$ to a distinct state $B \subset \Omega/A$. We assume A and B are open, bounded subsets of Ω. The transition rate between A and B can be expressed in terms of hitting times

$$\tau_S(\boldsymbol{x}_0) = \inf_t\{\boldsymbol{x}_t^{\boldsymbol{x}_0} \in S\} \tag{3}$$

where $S \subset \Omega$ and $\boldsymbol{x}_t^{\boldsymbol{x}_0}$ is the solution to initial condition $\boldsymbol{x}_0$. For a given configuration, the probability of reacting $A \to B$ is quantified through the forward committor function,[20]

$$q^+(\boldsymbol{x}) = \mathbb{P}[\tau_B(\boldsymbol{x}) < \tau_A(\boldsymbol{x})] \tag{4}$$

where the path measure $\mathbb{P}$ is determined by the choice of stochastic dynamics. The backward committor function is defined analogously; when the dynamics is reversible, $q^+ = 1 - q^-$ and henceforth we denote $q := q^+$ because we only consider equilibrium transitions in this work. The phase space committor, $\overline{q}$, solves a high-dimensional Dirichlet problem,

$$\begin{cases} (\mathcal{L}\overline{q})(\boldsymbol{x}, \boldsymbol{p}) = 0\boldsymbol{x} \in \Omega\backslash(A \cup B), \\ \overline{q}(\boldsymbol{x}, \boldsymbol{p}) = 0\boldsymbol{x} \in A, \\ \overline{q}(\boldsymbol{x}, \boldsymbol{p}) = 1\boldsymbol{x} \in B, \end{cases} \tag{5}$$

with the infinitesimal generator

$$\mathcal{L} = \boldsymbol{p}\cdot\nabla_{\boldsymbol{q}} - \nabla_{\boldsymbol{x}}\cdot U(\boldsymbol{x})\nabla_{\boldsymbol{p}} + \gamma(-\boldsymbol{p}\nabla_{\boldsymbol{p}} + \beta^{-1}\Delta_{\boldsymbol{p}}) \tag{6}$$

The generator (eq 6) is Markovian and hence enables a representation for the backward Kolmogorov eq 5.[21] We have shown that even high-dimensional committor functions can be determined using active importance sampling and a neural network ansatz[6] using a PINN loss,[22] but only in the case of overdamped dynamics. For molecular systems, this equation could be solved directly with the infinitesimal generator for underdamped Langevin by extending the committor to the full phase space coordinates.

Generally, we want to analyze the commitment probabilities in state space, rather than the full phase space, which motivates a distinct approach that we outline now. If we introduce a sufficiently large lag time τ, we can define an associated Markovian propagator in the state space

$$\mathcal{P}_\tau f(\boldsymbol{x}_0) = \mathbb{E}[f(\boldsymbol{x}_\tau^{\boldsymbol{x}_0})] \tag{7}$$

where $\mathcal{P}_\tau$ is the Markov semigroup operator associated with the dynamics and $f\colon \Omega \to \mathbb{R}$ is some observable. In principle, a posteriori checks to ensure that the propagator is Markovian on the time scale τ could be carried out if there is ambiguity. In the results shown below, we evaluate the accuracy of the committor without directly checking the Markovianity of the propagator on the time scale τ.

Using the Markovian semigroup[7] provides an alternative to (eq 5) and allows a distinct expression for the committor,

$$q(\boldsymbol{x}_0) = \mathbb{E}[q(\boldsymbol{x}_t^{\boldsymbol{x}_0})] \equiv \mathcal{T}_\tau^{\Omega^c}q(\boldsymbol{x}_0),\ x \in \Omega\backslash(A \cup B) \tag{8}$$

where the expectation is computed over dynamical trajectories of duration τ and $\mathcal{T}_\tau^{\Omega^c}$ denotes the stopped transition operator,[16] which terminates when hitting $A \cup B$; formally, it is defined as $\mathbb{E}f(\boldsymbol{X}_t^{\boldsymbol{x}})$ where $t = \min[\tau, \tau_{A\cup B}(\boldsymbol{x})]$. This equation quantifies the intuitive relation that the average value of the committor for a collection of trajectories $\{\boldsymbol{x}_\tau^{\boldsymbol{x}_0}\}$ evolving independently from the

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 1. (a) Equilibrium simulations (black lines) are initiated in each basin, and configurations that exit the boundaries χ_A and χ_b (red dots) are collected, along with the mean time between exits. (b) From each exit configuration, swarms of trajectories are launched and their end points (black borders) are used to train the committor. The sampling points from which the next swarm will be generated (red borders) are chosen from previous swarm end points based on their estimated committor values. Once an unbiased transition path is sampled, the algorithm repeats starting with a new exit configuration. (c) After many iterations, this algorithm collects many samples in the transition state region (blue borders) and an accurate global committor is learned; the committor is then evaluated at the exit configurations from (a) (red borders) and combined with the mean exit time to compute a rate estimate via (eq 9).

initial condition $\boldsymbol{x}_0$ must equal the committor of $\boldsymbol{x}_0$. The relation is clearest in the limit $\tau \gg \tau_B(\boldsymbol{x}_0)$, because in this case the right-hand side of (eq 8) is directly an empirical estimate of the fraction of trajectories that first hit B. The expression 8 is a particular instantiation of a Feynman-Kac equation, and the general strategy for solving equations of this type with a neural network ansatz was first developed in ref 8. When the dynamics is described by a time-homogeneous Markov process, the transition semigroup can be written in terms of an infinitesimal generator, $\mathcal{L}$ where $\mathcal{P}_\tau = e^{\tau\mathcal{L}}$.

While the committor is often heralded as the "ideal reaction coordinate", it also enables efficient estimates of transition rates.[23] To estimate rates, we employ the Hill relation,[19,24,25] which decomposes the transition rate into an outward flux j_{A^-} from the boundary of basin A, denoted χ_A, and reactive probabilities at the boundary,

$$k_{AB} = j_{A^-} \int_{\chi_A} q(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{9}$$

While one could hypothetically use any boundary (e.g., the isosurface $q = 0.5$) to calculate the rate using the Hill relation, the use of small committor values near the state boundary minimizes the length of the equilibrium simulation needed to calculate the outward flux j_{A^-}, avoiding the necessity of an equilibrium simulation of length comparable to the mean first passage time itself. This formalism is closely related to path-based methods such as the widely used forward flux sampling algorithm,[26] transition interface sampling,[2,27] and adaptive multilevel splitting.[28] In each of these approaches, the integral on the right-hand side of (eq 9) is estimated by reweighting transitions between a sequence of interfaces. With a direct parametrization of the committor, however, this estimate is considerably simpler. The general computational approach is illustrated in Figure 1 and discussed in detail in Section 4.

In this work, we optimize a variational objective to solve (eq 8) using a neural network ansatz for the committor function q in the configuration space for biomolecular systems, allowing rate estimates using (eq 9). To do so, we must create an objective function that, when minimized, satisfies (eq 8). In particular, we represent the committor as a parametric function, q_θ: $\Omega \to [0, 1]$ and solve

$$q_* = \min_\theta L(\theta) \tag{10}$$

Producing an objective function for (eq 8) is straightforward in this case; previous works[7,8,15] have used a mean-squared error. Our approach differs in an important way; we instead use a mean-squared discrepancy on the log-committor so that

$$L(\theta) = \frac{1}{2} \int_\Omega (\log(q_\theta(\boldsymbol{x})) - \log(\mathcal{T}_\tau^{\Omega^c} q_\theta(\boldsymbol{x})))^2 \mathrm{d}\nu(\boldsymbol{x}) \tag{11}$$

where the expectation is computed over initial conditions drawn from some sampling distribution ν which is not necessarily the equilibrium measure μ. While (eq 11) carries a statistical bias, we show in Section 3, that this bias is controlled and, in fact, as detailed below, the use of this biased loss improves our rate estimates and aids in sampling (see Section 7).

3. LOGARITHMIC COMMITTOR LOSS

Accurately estimating transition rates is the central goal of our approach, which in turn requires accurate committor probabilities near the reactant state A. When the time scale for a typical reaction to occur is much longer than the molecular relaxation time, typical committor values near the metastable state A will concentrate near zero. Therefore, the logarithmic difference between $q_\theta(\boldsymbol{x})$ and $\mathcal{P}_t q_\theta(\boldsymbol{x})$ in the objective scales these small probabilities to prioritize this region of configuration space.

The success of our algorithm at predicting the committor in regions close to the basins relies on the logarithmic bias in (eq 11), and we analyze this bias to show that it is controlled. Using (eq 8), the action of the semigroup can be estimated by the unbiased sample mean, giving

$$q(\boldsymbol{x}) \approx \frac{1}{n} \sum_{i=1}^{n} q(\boldsymbol{x}_\tau^{\boldsymbol{x},i}) := q\hat{q}(\boldsymbol{x}) \tag{12}$$

for a collection of i end points $\boldsymbol{x}_\tau^{\boldsymbol{x},i}$ of trajectories of duration τ initialized from $\boldsymbol{x}$. Taking the logarithm of both sides yields our estimator for the logarithm of the committor

$$\log q(\boldsymbol{x}) \approx \log\left(\frac{1}{n} \sum_{i=1}^{n} q(\boldsymbol{x}_t^{\boldsymbol{x},i})\right) = \log \hat{q}(\boldsymbol{x}) \tag{13}$$

However, this estimator is biased, as is revealed by Jensen's inequality,

$$\mathbb{E}[\log \hat{q}(\boldsymbol{x})] \leq \log(\mathbb{E}[\hat{q}(\boldsymbol{x})]) = \log q(\boldsymbol{x}) \tag{14}$$

We aim to show that this bound is tight for the committor, or, equivalently, that the "Jensen gap",

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 2. (a) Relative error in the second-order Taylor approximation (eq 18) for the sample mean estimator $\hat{q}$ of exponential random variables (chosen to maximize the coefficient of variation at small q) at various means q and sample sizes n. (b) Coefficient of variation for the sample mean estimator $\hat{q}$ for the same set of exponential variables. (c) Jensen gap for the same set of exponential variables, with upper bounds (eq 20) and (eq 21). The variables shown overshoot the low-q bound due to small underestimates in the Taylor approximation; dividing the bound by the relative error in the approximation corrects this. (d) Coefficient of variation for sets of trajectory end points generated by our algorithm in a two-channel potential. The coefficient clearly respects a large-q bound.

$$\frac{\mathbb{E}[\log \hat{q}(\boldsymbol{x})]}{\log(\mathbb{E}[\hat{q}(\boldsymbol{x})])} \approx 1 \tag{15}$$

To show this, we assume that for n large enough, the Central Limit Theorem (CLT) assures that $\hat{q}(\boldsymbol{x})$ is normally distributed with mean $q(\boldsymbol{x})$ and variance $\frac{\sigma^2}{n}$, where $\sigma^2(\boldsymbol{x})$ is the variance of the committor evaluated at the end of a trajectory of fixed length τ initialized at $\boldsymbol{x}$. In the case of the committor, which takes on bounded non-negative values, the coefficient of variation for the collection of trajectory end points is bounded above by the sample size,[29]

$$\frac{\sigma(\boldsymbol{x})}{q(\boldsymbol{x})} \leq \sqrt{n} \tag{16}$$

and, accordingly, the bound on the coefficient of variation for the estimator $\hat{q}$ is

$$\frac{\sigma(\boldsymbol{x})}{\sqrt{n}\, q(\boldsymbol{x})} \leq 1 \tag{17}$$

As a result, we can expand $\log \hat{q}(\boldsymbol{x})$ around its mean $q(\boldsymbol{x})$ to obtain,

$$\begin{aligned}\mathbb{E}[\log \hat{q}(\boldsymbol{x})] &= \mathbb{E}[q(\boldsymbol{x})] + \mathbb{E}\left[\frac{\hat{q}(\boldsymbol{x}) - q(\boldsymbol{x})]}{q(\boldsymbol{x})}\right] - \mathbb{E}\left[\frac{(\hat{q}(\boldsymbol{x}) - q(\boldsymbol{x}))^2}{2q(\boldsymbol{x})^2}\right] + \ldots \\ &\approx \log q(\boldsymbol{x}) - \frac{\sigma^2(\boldsymbol{x})}{2nq(\boldsymbol{x})^2}\end{aligned} \tag{18}$$

The Jensen gap then reduces to

$$\begin{aligned}\frac{\mathbb{E}[\log \hat{q}(\boldsymbol{x})]}{\log(\mathbb{E}[\hat{q}(\boldsymbol{x})])} &\approx \frac{\log(q(\boldsymbol{x})) - \frac{\sigma^2(\boldsymbol{x})}{2nq(\boldsymbol{x})^2}}{\log(q(\boldsymbol{x}))} \\ &= 1 - \frac{\sigma^2(\boldsymbol{x})}{2nq(\boldsymbol{x})^2 \log(q(\boldsymbol{x}))}\end{aligned} \tag{19}$$

and due to the bound on the coefficient of variation (eq 17), the upper bound as a function of $q(\boldsymbol{x})$ is

$$\frac{\mathbb{E}[\log(\hat{q}(\boldsymbol{x}))]}{\log(\mathbb{E}[\hat{q}(\boldsymbol{x})])} \leq 1 - \frac{1}{2\log(q(\boldsymbol{x}))} \tag{20}$$

For points near the basin, where $q(\boldsymbol{x}) \ll 1$, the Jensen gap ensures that the bias is well controlled. It may appear from this bound that the use of the logarithmic objective is inappropriate for large values of $q(\boldsymbol{x})$; however, as $q(\boldsymbol{x})$ becomes larger, stricter bounds on the coefficient of variation (eq 16) become applicable. Specifically, the bounded nature of the committor restricts the end point variance $\sigma^2(\boldsymbol{x})$ to no greater than $q(\boldsymbol{x})(1 - q(\boldsymbol{x}))$; this bounds the coefficient of variation by $\sqrt{\frac{1 - q(\boldsymbol{x})}{q(\boldsymbol{x})}}$ leading to another upper bound on the Jensen gap

$$\frac{\mathbb{E}[\log(\hat{q}(\boldsymbol{x}))]}{\log(\mathbb{E}[\hat{q}(\boldsymbol{x})])} \leq 1 - \frac{1 - q(\boldsymbol{x})}{2nq(\boldsymbol{x})\log(q(\boldsymbol{x}))} \tag{21}$$

which is loose at small $q(\boldsymbol{x})$ but is tighter than (eq 20) at larger values. For example, at the transition state, where $q(\boldsymbol{x}) = 1 - q(\boldsymbol{x}) = 0.5$, the upper bound becomes

$$\frac{\mathbb{E}[\log(\hat{q}(\boldsymbol{x}))]}{\log(\mathbb{E}[\hat{q}(\boldsymbol{x})])} \leq 1 + \frac{1}{2n\log(2)} \tag{22}$$

which is negligible for large n.

By identifying the intersection of the bounds (eqs 20 and 21), an absolute upper bound independent of $q(\boldsymbol{x})$ can be calculated.

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

The intersection occurs when the two upper bounds on the coefficients of variation coincide

$$\sqrt{n} = \sqrt{\frac{1 - q(\boldsymbol{x})}{q(\boldsymbol{x})}} \rightarrow q(\boldsymbol{x}) = \frac{1}{n+1} \tag{23}$$

leading to the absolute bound

$$\frac{\mathbb{E}[\log(\hat{q}(\boldsymbol{x}))]}{\log(\mathbb{E}[\hat{q}(\boldsymbol{x})])} \leq 1 + \frac{1}{2\log(n+1)} \tag{24}$$

Surprisingly, the bias is at a maximum at this intermediate value of $q(\boldsymbol{x})$, and ameliorates slowly as $q(\boldsymbol{x}) \rightarrow 0$ and sharply as $q(\boldsymbol{x}) \rightarrow 1$, as shown in Figure 2.

Although the bias is bounded, there is a systematic underestimate of $q(\boldsymbol{x})$ present on the order of a few percent. Because we symmetrize our loss function (25) in practice, this ultimately suggests that our learned committor $q_\theta(\boldsymbol{x})$ may be slightly more peaked around the transition state than the true committor, leading to an underestimation of rates in both directions.

4. COMPUTATIONAL DETAILS

We optimize the committor and estimate the rate "on-the-fly". The algorithm we employ combines local sampling within the reactant and product basins with importance sampling to obtain data for the committor away from the basins. This section systematically outlines the computational approach and details how we integrate the committor update with the rate estimates. The algorithm we use requires many independent simulations of short, "swarm" trajectories,[30] which are trivially parallelizable.

First, an equilibrium simulation is run in each basin until each trajectory has exited the basin N_{exit} times. Each time the trajectory exits the basin, its configuration and the time at which the exit took place are stored. This information constitutes an ensemble of points sampled from the basin boundaries χ_A and χ_B and exit times that can be used to calculate the fluxes j_{A^-} and j_{B^-}. The uncertainty in these estimates, and subsequently the uncertainty in the estimated transition rate, depends on the number of exit events observed.

The importance sampling scheme we employ selects points from one boundary ensemble and attempts to sample points along a reactive trajectory, eventually sampling points near the opposite basin. First, a boundary configuration $\boldsymbol{x}_i$ from one basin is chosen uniformly at random, from which a swarm of k independent trajectories $\boldsymbol{x}_\tau^{\boldsymbol{x}_i,j}$ is initiated and evolved. The sample $\boldsymbol{x}_i$ and the end points of each trajectory in its associated swarm $\boldsymbol{x}_\tau^{\boldsymbol{x}_i,j}$ are saved and used to train the committor q_θ via n_{iter} iterations of stochastic gradient descent on the symmetrized objective

$$L_{n,k,\tau}(\theta) = \frac{1}{2n}\sum_{i=1}^{n}\left(\log(q_\theta(\boldsymbol{x}_i)) - \log\left(\frac{1}{k}\sum_{j=1}^{k} q_\theta(\boldsymbol{x}_\tau^{\boldsymbol{x}_i,j})\right)\right)^2 + \frac{1}{2n}\sum_{i=1}^{n}\left(\log(1 - q_\theta(\boldsymbol{x}_i)) - \log\left(\frac{1}{k}\sum_{j=1}^{k}[1 - q_\theta(\boldsymbol{x}_\tau^{\boldsymbol{x}_i,j})]\right)\right)^2 \tag{25}$$

where the conditions $q_\theta(\boldsymbol{x} \in A) = 0$ and $q_\theta(\boldsymbol{x} \in B) = 1$ are enforced explicitly by setting the committor value of any end points that enter A or B to 0 or 1, respectively. This estimator requires sampling k trajectories from the same initial condition $\boldsymbol{x}_i$, which is straightforward in molecular simulations. Nevertheless, this requirement can be relaxed, as pointed out in 10. It is also important to note that in our implementation, we detach the "target" committor value $\frac{1}{k}\sum_{j=1}^{k} q_\theta(\boldsymbol{x}_\tau^{\boldsymbol{x}_i,j}))$ from the computational graph and treat it as a nonparametric value. Thus, gradient updates are based only on the "estimate" $q_\theta(\boldsymbol{x}_i)$. While in theory it is possible to recompute $\frac{1}{k}\sum_{j=1}^{k} q_\theta(\boldsymbol{x}_\tau^{\boldsymbol{x}_i,j}))$ periodically during this training process as the committor evolves, we choose to compute it only once at the outset of each training step to ensure stability.

Our algorithm chooses the next sampling point $\boldsymbol{x}_{i+1}$ as the swarm end point $\boldsymbol{x}_\tau^{\boldsymbol{x}_i,j}$ with the highest committor value under q_θ. This algorithm resembles forward propagation strategies that underlie forward flux sampling and steered transition path sampling.[2,27,28,31] Another swarm is generated from $\boldsymbol{x}_{i+1}$, and the committor is updated again based on a data set now consisting of $i + 1$ sampled points and their associated swarms. This process continues iteratively, with the next configuration chosen from the set of all previously sampled trajectory end points, the configuration with highest committor value being chosen as the next sample. Allowing the algorithm to sample a configuration from any previous swarm affords flexibility in the construction of a transition pathway as the committor q_θ evolves and estimates of transition probabilities change.

As the committor evolves and we sample configurations with larger estimated committor values, members of a swarm should eventually terminate in the basin opposite the basin from which the original sample $\boldsymbol{x}_i$ was drawn. When we reach this point, the algorithm has sampled configurations in an intermediate region of state space connecting basins A and B. We refer to this collection of samples connecting the basins as one sampling "chain". Next, to continue to facilitate sampling in intermediate regions of state space, a new boundary sample is randomly drawn and a new chain of samples is generated. After we acquire new samples, the committor q_θ is updated using stochastic gradient descent, evaluating the loss function (eq 25) on all previously sampled configurations. Crucially, when choosing the next configuration in a given chain, our algorithm only considers swarm end points generated from the current chain, and not previous chains. This prevents sampling from concentrating near the opposite basin and instead produces a new chain that is only correlated to the existing chains via their influence on the learning of the committor. As a result, this new chain could, in principle, explore alternative reaction pathways.

In practice, we find that it is helpful to simultaneously sample points originating from both basins A and B. Thus, in one iteration of our algorithm, we collect two sample points, each associated with a chain originating from each basin. This approach allows the committor to learn the vital boundary conditions $q_\theta(\boldsymbol{x} \in A) = 0$ and $q_\theta(\boldsymbol{x} \in B) = 1$ at the outset of sampling. We do not enforce these boundary conditions in the objective by training on a separate set of samples from within the basins, as in other approaches.[4,7,14]

At any point in this process, rates k_{AB} and k_{BA} can be estimated in an online manner using (eq 9); the flux is estimated from the initial equilibrium simulation, and the mean committor value on the boundary of each basin can be estimated by evaluating q_θ on the associated boundary ensemble. The convergence of the rate estimates for both transitions is a useful proxy for the convergence of q_θ itself, and we demonstrate below that rate estimates generated by this algorithm converge stably in applications to biomolecular systems.

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 3. Results on the Müller-Brown potential. Panels (a−d) utilize MSE loss, while panels (e−h) utilize MSLE loss. (a, e) Learned committor after 500 iterations of the algorithm. (b, f) Online rate estimates against total simulation time. (c, g) Error in the committor compared to the finite-element solution. (d, h) Error in the log-committor compared to the finite-element solution.

5. RESULTS AND DISCUSSION

We applied our methodology to two benchmark two-dimensional potentials to assess the assumptions and convergence properties of the algorithm. We chose these relatively simple examples because ground truth rates from unbiased simulations are straightforward to estimate. We also illustrate the scalability of the algorithm by studying transitions in molecular dynamics simulations of the miniprotein Chignolin and the synthetic peptide Aib9. Throughout, we represent the committor with a minimalistic neural network ansatz because we want to focus attention on the sampling procedure. We also embed the coordinates using a relatively naïve approach—a flattened vector of pairwise distances for all atoms in each molecule. For chignolin, this amounts on an input dimension $d = 13695$; for Aib9, we used all atoms alongside its 18 dihedral angles leading to an input dimension $d = 8274$. We used a multilayer perceptron with three hidden layers and 100 neurons per layer using a leaky ReLU nonlinear activation function to represent the committor q. Because the output represents a probability, we pass the final output through a sigmoid function to ensure that the range is (0, 1).

5.1. Illustration on Two-Dimensional Potentials. We first examine our algorithm's performance on the Müller-Brown potential, defined as

$$V(x) = \sum_{i=1}^{4} A_i e^{(x-\mu_i)^T \Sigma_i (x-\mu_i)}$$

$$A = [-20 \quad -10 \quad -17 \quad 1.5]$$

$$\mu = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \\ -0.5 & 1.5 \\ -1 & 1 \end{bmatrix}$$

$$\Sigma = \left[\begin{bmatrix} -1 & 0 \\ 0 & -10 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -10 \end{bmatrix} \begin{bmatrix} -6.5 & 5.5 \\ 5.5 & -6.5 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \right] \tag{26}$$

We simulate using overdamped dynamics with $\gamma = \beta = 1$ and step size $\Delta t = 10^{-3}$, shooting 100 swarms of maximum length $\tau = 5\ \Delta t$ from each sampling point, stopping early if a member of the swarm enters a basin. State A is defined as a circle of radius 0.1 around $[-0.5, 1.5]$ and state B is defined similarly around $[0.5, 0]$. In order to examine the effect of using the logarithmic loss function (eq 25), we ran 500 iterations of our algorithm using as a loss function either a standard mean-squared error (MSE) or the mean-squared logarithmic error (MSLE) given by (eq 25).

The results are illustrated in Figure 3. After 500 iterations of the algorithm, the MSE approach has not converged, and both the representation of the committor and the rate estimate deviate significantly from the analytical solution. However, the approach utilizing the logarithmic objective (eq 25) converges after about 400 iterations and agrees well with both the analytical rates and the analytical committor solution, even on a

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 4. Results on a two-channel potential. Panels (a−d) utilize MSE loss, while panels (e−h) utilize MSLE loss. (a, e) Learned committor after 300 iterations of the algorithm. (b, f) Online rate estimates against total simulation time. (c, g) Error in the committor compared to the finite-element solution. (d, h) Error in the log-committor compared to the finite-element solution.

logarithmic scale. An analysis of the sampling dynamics in this system, including the failure of the MSE approach to sample a single transition in the $B \to A$ direction, can be found in Section 7. This failure mode is not observed in the logarithmic case.

To illustrate our algorithm's performance on a potential will multiple reaction pathways, we also examined the two-channel system defined by

$$V(x) = \sum_{i=1}^{4} A_i e^{-\mathbf{1}^T (x-\mu_i)^2} + 0.2(\mathbf{1}^T (x - \mu_0))^4$$

$$A = [-3 \quad -3 \quad -5 \quad 5]$$

$$\mu = \begin{bmatrix} 0 & 1/3 \\ 0 & 5/3 \\ -1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$\mu_0 = [0 \quad 1/3] \tag{27}$$

We simulate overdamped dynamics with $\gamma = 1$, $\beta = 6.67$ and step size $\Delta t = 10^{-2}$, shooting 1000 swarms of maximum length $\tau = 5$ Δt from each sampling point, utilizing the same early stopping criterion as the Müller-Brown system. State A is defined as a circle of radius 0.2 around $[-1, 0]$ and state B is defined similarly around $[1, 0]$ Again, we compare our logarithmic loss approach with a MSE approach. The results are illustrated in Figure 4, and an analysis of the sampling dynamics can be found in Section 7.

5.2. Estimating Unfolding Rates for Chignolin. We next examined a condensed molecular system—the miniprotein Chignolin in solvent. Chignolin is a 10-residue, 166-atom peptide that transitions between a folded β-hairpin motif and a disordered unfolded conformation in the condensed phase. We use the sum of the distance between four pairs of atoms (Tyr1 N—Tyr10 O, Tyr1 O—Tyr10 H, Asp3 N—Thr8 O, Asp3 O—Thr8 N) involved in the hydrogen bonds that characterize the β-hairpin as definitions of states A (<1.75 nm) and B (<3 nm) respectively. We simulate this peptide in solvent using the Amber-ff99sb force field at 340 K with $\tau = 2$ ps. The results are illustrated in Figure 5.

Our algorithm is able to produce over 200 unfolding paths within the roughly 1 μs sampling duration of the simulation. The unfolding transition is sampled at a significantly higher frequency than the folding transition, which was only explored by about 30 paths during the entire simulation. We posit that this is due to the fact that the folded state represents a specific set of conformations in a well-defined metastable basin whereas the unfolded state represents a wider ensemble of conformations and exists in a more diffusive regime. Accordingly, the rate of the unfolding transition agrees well with rate estimates generated from a long equilibrium trajectory[18] where the unfolding rate has not converged at this point in the simulation. This example illustrates the potential for our algorithm to both generate transition-relevant conformations and calculate rates on time scales much faster than those of the unbiased transitions, but also highlights a potential shortfall in simulating diffusive dynamics.

5.3. Estimating Transition Rates with Metastable Intermediates: Aib9 Isomerization. We also examined the

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 5. Results on solvated Chignolin at 340 K. (a) Committor isosurfaces on the pontential of mean force in the space of two H-bond distances (between residues ASP3 and THR8) that characterize the folded state. (b) Running rate estimates for the folding and unfolding processes, along with a smoothed running average and standard deviation based on the previous 300 iterations. The rate estimates from a long equilibrium trajectory[18] are shown as dotted lines. (c) Representative unfolding path initiated from the folded boundary, which passes through a transition state with $q(\boldsymbol{x}) \approx 0.5$ to the unfolded boundary. (d) Final estimates of the mean first passage time compared with estimates from a long equilibrium trajectory.

performance of our rate estimation algorithm on a relatively adversarial example that features challenges more characteristic of complex biophysical dynamics. While the system is still small, the synthetic peptide Aib9 features intermediate metastability,[32] which complicates the computation of a rate and challenges the sampling ability of our algorithm. Intermediate metastability results in transit times that constitute an appreciable fraction of the mean first passage time itself. Because each sample can only progress along the current reaction path by a maximum length τ, and multiple τ-length segments are simulated from each sample, the simulation time required to observe a transition event in our algorithm can approach the unbiased mean first passage time in these systems.

We simulated this system using the Amber15-IPQ force field[33] in vacuum at 500 K. In Figure 6(a,c) we show the estimated high-dimensional committor projected onto the backbone dihedral angles of the fifth amino acid in the nine amino acid chain. The potential of mean force for this coordinate is shown as filled contours for reference, which we obtained from a long unbiased simulation. Because of the metastable intermediates, the total sampling time needed to estimate the transition rate exceeds the mean first passage time by roughly 2 orders of magnitude, and separate realizations of the algorithm display appreciable variance. While a naive implementation of our algorithm is unlikely to be computationally efficient in this case, the rate estimates we make, shown in Figure 6(b,d), are within error of the empirical estimate.

Furthermore, we obtain independently useful information about the transition state through knowledge of the committor. For example, we can sample the transition state, as shown in Figure 7, which illustrates the nonobvious characteristic of the reaction pathway that the synthetic peptide collapses during isomerization. This analysis illustrates a powerful use of the committor to design and test intuitive reaction coordinates with no additional simulation requirements.

6. CONCLUSIONS

The formulation of efficient and stable objective functions for the committor have led to numerous works seeking to learn this ideal reaction coordinate from molecular dynamics simulation data. These algorithms build upon general purpose machine learning strategies for solving high-dimensional PDEs with a neural network ansatz.[22] We have shown that combining a variational objective for the Feynman−Kac formulation of the committor with an iterative importance sampling scheme enables accurate parametrization of this high-dimensional function. This algorithm is agnostic to the underlying sampling

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 6. Results on the AIB9 system. (a) Example committor isosurfaces on the pontential of mean force in the $[\phi, \psi]$—dihedral space of the middle residue. (b) Running rate estimates for both directions of helix isomerization, represented as a smoothed average of 3 different realizations of the algorithm. The rate estimate from a long equilibrium simulation is shown for comparison. (c) Example relative transition probability $q(1-q)$ on a logarithmic scale in the $[\phi, \psi]$—dihedral space of the middle residue. (d) Final estimates of the mean first passage time across 3 realizations of the algorithm compared with estimates from a long equilibrium trajectory.

distribution,[7] meaning that reweighted adaptive importance sampling can be avoided so long as the transition interfaces are well sampled.

Access to a representation of the high-dimensional committor also enables rate estimates using the Hill relation. We show that for single barrier isomerization reactions, these estimators are computationally efficient and provide very accurate rate estimates. Our estimator performs well in part due to the optimization of the committor using a logarithmic loss function, which we show has a statistical bias but that this bias is easily controlled. For a more complex reaction featuring multiple metastable intermediates, the computational cost of rate estimates remains high due to a breakdown of the assumption of Markovianity. Furthermore, diffusive transitions with nonuniform transition paths also incur a high computational cost as the committor must acquire enough data to generalize between all possible transitions.

Because our focus in the present work is the overall sampling scheme, we intentionally used simple neural network architectures and naïve embeddings of molecular structure. Of course, we believe there will be advantages in exploring the rapidly growing literature on neural network architectures appropriate for chemical systems,[34−36] and it will be beneficial to assess representations that are more tailored to proteins in the future. Additionally, there are natural opportunities to combine the algorithms that we have developed with reaction coordinate learning[3,37] and other enhanced sampling strategies.[38,39] Furthermore, combining the approach we have taken here with algorithms more naturally suited to metastable intermediates, such as Markov state models,[40] presents a natural opportunity.

7. CONNECTION TO PATH SAMPLING METHODS

The algorithm we present here attempts to generate conformations that interpolate two states A and B. In the case of overdamped dynamics, these chains of sampled configurations represent points along a transition path connecting A and B, and in the underdamped case, while not technically transition paths due to velocity randomization, these sampling chains still cross transition states and can thus serve as approximations to true underdamped transition paths. While our algorithm is not explicitly concerned with the construction of path ensembles or the exhaustive exploration of potential transition pathways, it does implicitly depend on sampling along

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 7. With only information characterizing the two metastable helical configurations of AIB9, our algorithm samples a transition state ensemble. The end-to-end distance of the transition state ensemble is shown to be much smaller than the helical configuration, suggesting that the formation of contacts between the terminal residues of AIB9 is an important step in the isomerization process.

low-energy transition pathways to calculate accurate rate estimates and committor functions. It is therefore natural to compare the sampling procedure we employ in this algorithm to existing path sampling methods.

Traditional path sampling approaches include Transition Path Sampling (TPS),[41] Transition Interface Sampling (TIS),[27] and Adaptive Multilevel Splitting (AMS).[28] TPS presupposes access to an initial transition path and makes Monte Carlo moves in path space to generate a transition path ensemble representative of unbiased system dynamics, from which a variety of observables can be calculated. TIS and AMS both presuppose access to an order parameter that interpolates states A and B; TIS generates paths from A to B incrementally between fixed order parameter isosurfaces, while AMS adaptively places isosurfaces to standardize incremental transition probabilities.

In generating sampling chains between states A and B, we note distinct advantages that our algorithm possesses over these existing approaches. Firstly, our algorithm can quickly generate initial transition chains between states using unbiased system dynamics, whereas the initial transition path in the aforementioned approaches is either a prerequisite or the end product of the entire algorithm. Secondly, our approach uses the committor as an order parameter, which changes in response to sampled data, rather than a static parameter chosen a priori that may bias calculations. Finally, our algorithm generates transition chains that are only loosely correlated to previous paths via the learned committor, rather than paths that overlap significantly with each other. In the TPS case, the sampling of an alternative reaction channel would require an unlikely series of Monte Carlo acceptances, whereas TIS and AMS are more capable of exploring alternative transition pathways but generate overlapping trajectory ensembles within each channel.

7.1. q-Greedy Algorithm and Connection to Doob Conditioning. The importance of learning an accurate representation of $\log q(\boldsymbol{x})$ near each state boundary not only allows for rate estimation using the Hill relation 9, but also aids in generating transition chains between A and B. To illustrate this, consider a overdamped Langevin dynamics initiated at a point $\boldsymbol{x}$ outside of state A that has been conditioned to transition into state B after time t. This conditioning was originally characterized by Doob[42] and, in the special case of the hitting point distribution for committor[43] obeys the dynamics

$$\gamma \mathrm{d}\boldsymbol{x}_t = [-\nabla V(\boldsymbol{x}_t) + 2\beta^{-1}\nabla \log h(t, \boldsymbol{x}_t)]\mathrm{d}t + \sqrt{2\gamma\beta^{-1}}\,\mathrm{d}\boldsymbol{W}_t \tag{28}$$

where $h(t, \boldsymbol{x})$ is the time-dependent probability that $\boldsymbol{x}$ will be in B at time t. If we treat A and B as absorbing states, then we recover the committor as

$$\lim_{t\to\infty} h(t, \boldsymbol{x}) = q(\boldsymbol{x}) \tag{29}$$

which leads to a conditioned dynamics of

$$\gamma \mathrm{d}\boldsymbol{x}_t = [-\nabla V(\boldsymbol{x}_t) + 2\beta^{-1}\nabla \log q(\boldsymbol{x}_t)]\mathrm{d}t + \sqrt{2\gamma\beta^{-1}}\,\mathrm{d}\boldsymbol{W}_t \tag{30}$$

This can be interpreted as adding a biasing potential that sharply increases around A, preventing entry and ensuring an eventual transition into B. This potential induces a new probability distribution on the state space

$$P(\boldsymbol{x}) \propto e^{-\beta V(\boldsymbol{x})} q(\boldsymbol{x})^2 \tag{31}$$

In order to generate transition paths given a committor $q(\boldsymbol{x})$, one could sample biased dynamics using (eq 30) or sample unbiased dynamics and use (eq 31) as a Metropolis-Hastings acceptance criterion. Given our need to sample unbiased dynamics to estimate the committor, the latter approach is appropriate in our case; however, during testing on 2D benchmark systems, we observed that using (eq 31) as an acceptance criterion led to a dramatic slowdown of transition sampling. This was largely due to the fact that both of the 2D systems we studied exhibit some metastability near intermediate values of $q(\boldsymbol{x})$, where $\nabla q(\boldsymbol{x})$ is relatively flat. Thus, sampling chains became stuck in these intermediate metastable regions and convergence suffered.

Instead, we implemented a q-greedy approach that samples the largest committor value available in the sampling chain. A theoretical treatment of this approach is left to future work, but generally speaking this can be thought of as an attempt to accelerate transitions toward state B, rather than simply preventing them from entering state A as suggested by the Doob-transformed dynamics.

An attempt to understand the high-level behavior of this sampling approach can be made by considering the case in which swarms are generated by a single step of time-discretized overdamped dynamics and the committor is locally flat. The probability that any individual step of the dynamics makes positive progress along the committor is then given by the probability that the overdamped step has a positive projection onto the committor gradient

$$\begin{aligned} P_{q+}(\boldsymbol{x}) &\approx P\left((-\gamma^{-1}\nabla V(\boldsymbol{x})\Delta t + \gamma^{-1}\sqrt{2\beta^{-1}}\,\Delta W_t)\cdot\frac{\nabla q(\boldsymbol{x})}{\|\nabla q(\boldsymbol{x})\|} > 0\right) \\ &= P\left(-\gamma^{-1}\nabla V(\boldsymbol{x})\Delta t\cdot\frac{\nabla q(\boldsymbol{x})}{\|\nabla q(\boldsymbol{x})\|} > -\sqrt{2\gamma^{-1}\beta^{-1}}\,\Delta W_t\cdot\frac{\nabla q(\boldsymbol{x})}{\|\nabla q(\boldsymbol{x})\|}\right) \end{aligned} \tag{32}$$

This probability is minimized when the gradient of the potential aligns with the gradient of the committor; thus

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 8. Transition paths sampled by our algorithm on 2D potentials. (a) MSE approach on the Müeller-Brown potential. The sole errant $B \to A$ transition chain is highlighted in red. (b) Logarithmic approach on the Müeller-Brown potential. (c) MSE approach on a two-channel potential. (d) Logarithmic approach on a two-channel potential.

$$P_{q^+}(\boldsymbol{x}) \geq P\left(\left\|\gamma^{-1}\nabla V(\boldsymbol{x})\right\|\Delta t < \sqrt{2\gamma^{-1}\beta^{-1}}\,\Delta W_t \cdot \frac{\nabla q(\boldsymbol{x})}{\|\nabla q(\boldsymbol{x})\|}\right) = \Phi\left(-\sqrt{\frac{\beta\Delta t}{2\gamma}}\left\|\nabla V(\boldsymbol{x})\right\|\right) \tag{33}$$

where Φ is the cumulative distribution function of the standard normal distribution. Finally, given n trajectories in a swarm, the probability that at least one sample makes positive progress along the committor is

$$P_{q_n^+}(\boldsymbol{x}) \geq 1 - \left(\Phi\left(\sqrt{\frac{\beta\Delta t}{2\gamma}}\left\|\nabla V(\boldsymbol{x})\right\|\right)\right)^n \tag{34}$$

We see that the probability of continuing along a putative transition path still depends on the gradient of the potential, but it is also attenuated by large swarms, smaller timesteps, and higher temperatures. Though this picture is complicated somewhat by underdamped dynamics, longer swarms, and an adaptive committor, preliminary experiments seem to indicate that these general trends continue to hold in those cases.

In conclusion, the above theoretical treatment suggests several reasons for our algorithm's success in path sampling. First, the learning of $\log q(\boldsymbol{x})$ rather than $q(\boldsymbol{x})$ itself resembles the conditioned transition dynamics (eq 30) and quickly pushes samples away from their basin of origin, especially as sampling begins near the boundary of each basin. Second, sampling chains are able to climb relatively steep potentials as indicated by (eq 34), but will still encounter resistance if the potential becomes too extreme; in this case, the committor should tend to self-correct based on the self-consistent criterion. Finally, local committor minima and maxima outside of states A and B are also discouraged by application of the self-consistent criterion. Taken together, these factors suggest that the learned log-committor is able to efficiently guide sampling between the states. Though we make no theoretical guarantees regarding the speed at which an initial transition or subsequent transitions will occur, our experiments in molecular systems provide strong evidence that this algorithm is successful in generating transition paths orders of magnitude more quickly than unbiased dynamics.

7.2. Transition Paths in 2D Systems. The above theoretical treatment also suggests a reason for the superiority of the logarithmic approach in sampling transition paths, as opposed to a MSE approach that learns the value of $q(\boldsymbol{x})$ directly. Figure 8 illustrates the transition paths sampled by each algorithm on both 2D potentials studied in this work. We see that in both cases, the MSE approach favors relatively direct transition paths but few others. In the Müeller-Brown system, we additionally observe the first sampling chain initiated from B shoot past state

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 9. Results on overdamped alanine dipeptide. (a) Example committor isosurfaces on the pontential of mean force in $[\phi, \psi]$—dihedral space. (b) Running rate estimates for both directions of the $C_7^{ax} \leftrightarrow C_7^{eq}$, represented as a smoothed average over 5 different realizations of the algorithm. (c) Example relative transition probability $q(1-q)$ on a logarithmic scale in $[\phi, \psi]$—dihedral space. (d) Final estimates of mean first passage times across 5 different realization of the algorithm compared with estimates from a long equilibrium trajectory.

A and remain in a high-energy state for the duration of the simulation. In this region, because the committor is close to 0, there is little incentive for the committor to adapt via gradient descent; thus, the committor gradient continues to point away from state A at the location of the current sample. The MSE approach performs better in the two-channel potential, but still undersamples the dominant top channel relative to the shorter but less likely bottom channel.

The logarithmic approach appears to strike a useful balance between exploration and exploitation of transition paths in both systems. Initially, in the Müeller-Brown system, paths with very low probability are sampled; quickly, however, the minimum energy path is identified and sampling begins to explore this path repeatedly. In the two-channel system, the logarithmic approach switches between the top and bottom reaction channels depending on the location of the initial sample on the state boundary. This ability to sample two separate transition channels independently is a longstanding difficulty of traditional path sampling methods.

One possibile explanation for the superiority of the logarithmic approach, as discussed previously, is that points sampled near state boundaries are preferentially generated radially away from the state due to the large value of $\log q(\boldsymbol{x})$ in that region, rather than immediately toward the location of the transition state isosurface as is likely the case in the MSE regime. This could encourage the sampling of alternative reaction channels rather than biasing all initial samples in the same direction. Especially in the two-channel case, the path distribution sampled by the logarithmic approach also appear to more closely follow the flow lines of the reactive flux as depicted in 44, indicating a more precise characterization of the committor near the basins.

We recognize the primitive nature of the theoretical understanding of this sampling approach. Further investigation into the sampling dynamics of our algorithm is a promising direction for future work.

8. CASE STUDY: OVERDAMPED VERSUS UNDERDAMPED DYNAMICS FOR ALANINE DIPEPTIDE

In transitioning from overdamped dynamics on toy 2D systems to molecular simulations, which typically employ underdamped dynamics, we sought to understand the effect of employing underdamped dynamics using the semigroup objective (eq 8). To do this, we directly compare rate calculations for the peptide fragment alanine dipeptide in the underdamped and over-

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Figure 10. Results on underdamped alanine dipeptide. (a) Example committor isosurfaces on the pontential of mean force in $[\phi, \psi]$—dihedral space. (b) Running rate estimates for both directions of the $C_7^{ax} \leftrightarrow C_7^{eq}$, represented as a smoothed average of 4 different realizations of the algorithm. (c) Example relative transition probability $q(1-q)$ on a logarithmic scale in $[\phi, \psi]$—dihedral space. (d) Final estimates of mean first passage times across 4 different realizations of the algorithm compared with estimates from a long equilibrium trajectory.

damped case. Alanine dipeptide is not a particularly challenging benchmark, but it offers a useful comparison between both overdamped and underdamped dynamics because its sampling is tractable even when using the 0.01 fs time step required for numerical stability with an overdamped integrator. We simulated alanine dipeptide with the Charmm27 force field[45] in vacuum at 300 K.

In Figure 9(a), we show the high-dimensional committor projected onto the backbone dihedrals. The potential of mean force is shown as filled contours, which we obtained using umbrella sampling. The transition isosurface where $q = \frac{1}{2}$ sits at the transition states for the $C_7^{ax} \leftrightarrow C_7^{eq}$ isomerization, consistent with expectations. We also plot the quantity $q(1-q)$ in Figure 9 (b), which obtains its maximal value when $q = \frac{1}{2}$ and symmetrizes the reactive probability flow lines. By taking the negative logarithm of this quantity, we see that the variation in probability near basins is well resolved.

Using the online estimate of the committor together with the reactive flux out of the basins A and B, we estimate transition rates. Figure 9 (c) shows a rolling average of the mean rate estimate of 5 separate realizations of the system as a function of total sampling time. At early times, the representation of the committor function is imprecise, causing substantial error in the rate. As the representation of the committor improves, the rate stabilizes within the correct order of magnitude. Note that the mean first passage time for the transition is 38 ns and with less than 3 ns of total sampling time, we obtain rate estimates accurate to within an order of magnitude.

The underdamped case, for which it is difficult to guarantee the assumption of Markovianity a priori, shows remarkably similar performance. In the case of underdamped dynamics, we initialize each trajectory in a swarm with a velocity randomly sampled from the Maxwell-Botlzmann distribution, thus effectively integrating out the velocity portion of phase space, and we continue to condition our estimate of the committor on the configuration of the molecule only. As shown in Figure 10(a,b), we obtain committor isosurfaces that sit near the saddles of the free energy surface plotted in dihedral coordinates. The moving average of the online rate estimates again starts with a significant overestimate, but as the committor representation is trained, the predicted rate approaches a value close to the empirically measured rate for both the $A \to B$ and $B \to A$ transition. As before, the total sampling time required to predict both transition rates accurately is a small fraction of the mean

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

first passage time. Of course, the simulations run for our rate estimator are also trivially parallelizable.

In both the overdamped and underdamped case, we observe a slight but systematic underestimate of the $A \to B$ transition rate. The cause of this bias is not immediately clear, but it may result from the bias discussed in Section 3, the broad nature of the energy minimum around state A, or the undersampling of alternative reaction channels. We hope to further understand the nature of such biases in future work.

9. COMPUTATIONAL SPECIFICATIONS AND PARAMETER CHOICE

All of our simulations were run with OpenMM 8.1.0.[46] A GitHub repository containing the code we used to run our experiments can be found at https://github.com/rotskoff-group/sc_committor_rates.

9.1. Choice of Lag Time. In principle, for the importance sampling phase, τ should be chosen to be on a similar scale to the molecular relaxation time. This prevents our committor-greedy sampling scheme from inefficiently generating correlated samples and from exploring energetically unreasonable regions of state space (see Section 7) while still allowing for fluctuations in the direction of reaction (which would relax away with τ too large). In practice, we have found it advantageous to avoid the small-τ regime by setting a time interval t_{stride} and evolve each swarm in multiples of this interval before checking for a stopping criterion. In other words, only every t_{stride} do we check the basin membership of each trajectory in the swarm, and we stop evolving the swarm if one or more trajectories has entered a basin. Thus, t_{stride} can be thought of as both an effective minimum value for τ and a discretization scale for the dynamics of the system. In the case of a t_{stride} equal to a single integration, we found that the stopping criterion led to swarms with τ on the order of only a few integration steps near a basin, leading to issues associated with the small-τ regime. While choosing a larger t_{stride} could theoretically bias committor estimation near the basins, our results demonstrate that such a bias has minimal effect on rate estimation. The nature of the effect of t_{stride} on committor estimation is left to future work.

9.2. Overdamped Alanine Dipeptide. We simulated the overdamped alanine dipeptide system using OpenMM's Brownian integrator, with a time step of 0.01 fs, a friction coefficient of 1/ps, and hydrogen-bond restraints. Basins A and B were defined by a circle of radius 10° in $[\phi, \psi]$-dihedral space around the coordinates $[-150°, 170°]$ and $[90°, -50°]$, respectively. We ran initial equilibrium simulations in each basin and collected a boundary ensemble of 1000 configurations from each. For the sampling process, we chose swarms of size $k = 100$ and maximum length $\tau = 10$ fs, with $t_{\text{stride}} = 1$ fs, stopping early if a swarm trajectory re-entered a basin. We ran each simulation for a total of 48 h on a single GPU, during which each algorithm sampled roughly 8000 configurations and 100 transitions in each direction; this represents 8×10^5 swarm end points used in the process of committor training, and about 2.5 ns of total simulation time, equivalent to 2.5×10^8 total integration steps. For reference, the mean first passage time for the $A \to B$ transition in this system is on the order of 10^9 steps.

The neural network used to compute the committor was implemented in PyTorch using the architecture described in Section 5. After each sampling step, the committor was trained using full-batch stochastic gradient descent using the ADAM optimizer[47] with a learning rate of 10^{-4} for 100 iterations. Alternate approaches that involve batching, alternative network architectures, or periodic recomputations of the target committor value $\frac{1}{k}\sum_{j=1}^{k} q_\theta(\boldsymbol{x}_\tau^{\boldsymbol{x}_{i,j}})$) are possible, but we leave these to future work.

9.3. Underdamped Alanine Dipeptide. We simulated underdamped alanine dipeptide system using the Langevin BAOAB integrator[48] from the OpenMMTools package, with a time step of 1 fs, a friction coefficient of 1/ps, and hydrogen-bond restraints. Basins A and B were defined identically to the overdamped case, and we collected the same number of configurations from each boundary ensemble. The swarm size was again chosen to be $k = 100$; however, we chose $\tau = 100$ fs and chose the t_{stride} to be an identical value; in the case of underdamped dynamics, the small τ regime is especially dangerous as the assumption of Markovianity in configuration space breaks down. Again, we terminated swarm evolution early if a swarm trajectory re-entered a basin. We also each of these simulations for a total of 48 h on a single GPU, sampling roughly 5800 configurations and 100 transitions in each direction over 40 ns each. The training regime was identical to the overdamped case.

9.4. Underdamped Chignolin in Solvent. We simulated Chignolin in solvent using the Langevin BAOAB integrator from the OpenMMTools package, with a time step of 2 fs, a friction coefficient of 1/ps, and hydrogen-bond restraints.

Basins A and B were defined by the average distance between 4 hydrogen-bond pairs that characterize the β-folded state. We ran initial equilibrium simulations in each basin and collected a boundary ensemble of 100 configurations from each. We chose swarms of size $k = 100$ and $t_{\text{stride}} = 2$ ps; τ was set as a single stride. We ran the simulation for a total of 144 h on a single GPU, during which each algorithm sampled roughly 2200 configurations, 224 unfolding transitions, and 36 folding transitions, amounting to roughly 900 ns of total simulation time. The neural network architecture used for this system involved pairwise distances between all atoms in the miniprotein (d = 13 695).

9.5. Underdamped AIB9. We simulated the AIB9 system using the Langevin BAOAB integrator from the OpenMMTools package, with a time step of 1 fs, a friction coefficient of 1/ps, and hydrogen-bond restraints. Basins A and B were defined by a circle of radius 10 in $[\phi, \psi]$-dihedral space around the coordinates $[-60, -40]$ and $[60, 40]$, respectively. We ran initial equilibrium simulations in each basin and collected a boundary ensemble of 1000 configurations from each. We chose swarms of size $k = 100$ and $t_{\text{stride}} = 1$ ps with τ again equal to a single stride. We ran each simulation for a total of 96 h on a single GPU, during which each algorithm sampled roughly 8400 configurations; however, only a few dozen transitions were observed in each case. A total of roughly 2.75 μs was required to sample these configurations and their swarms. The neural network architecture used for this system mirrors that of alanine dipeptide.

■ AUTHOR INFORMATION

Corresponding Author

Grant M. Rotskoff − *Department of Chemistry, Stanford University, Stanford, California 94305, United States;* orcid.org/0000-0002-7772-5179; Email: rotskoff@stanford.edu

https://doi.org/10.1021/acs.jctc.4c00997
J. Chem. Theory Comput. 2024, 20, 9378−9393

Author

Andrew R. Mitchell − *Department of Chemistry, Stanford University, Stanford, California 94305, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c00997

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Bolhuis, P. G.; Dellago, C.; Geissler, P. L.; Chandler, D. Transition path sampling: throwing ropes over mountains in the dark. *J. Phys.: Condens. Matter* **2000**, *12* (8A), A147−A152. ISSN 0953−8984, 1361−648X. URL https://iopscience.iop.org/article/10.1088/0953-8984/12/8A/316.

(2) Hall, S. W.; Díaz Leines, G.; Sarupria, S.; Rogal, J. Practical guide to replica exchange transition interface sampling and forward flux sampling. *J. Chem. Phys.* **2022**, *156* (20), No. 200901. ISSN 0021−9606. URL 10.1063/5.0080053.

(3) Mehdi, S.; Smith, Z.; Herron, L.; Zou, Z.; Tiwary, P. Enhanced Sampling with Machine Learning. *Annu. Rev. Phys. Chem.* **2024**, *75*, 347−370. ISSN 0066−426X, 1545−1593

(4) Li, Q.; Lin, B.; Ren, W. Computing committor functions for the study of rare events using deep learning. *J. Chem. Phys.* **2019**, *151* (5), No. 054112. ISSN 0021−9606, 1089−7690. URL http://aip.scitation.org/doi/10.1063/1.5110439.

(5) Khoo, Y.; Lu, J.; Ying, L. Solving for high-dimensional committor functions using artificial neural networks. *Res. Math. Sci.* **2019**, *6* (1), 1. ISSN 2522−0144, 2197−9847. URL http://link.springer.com/10.1007/s40687-018-0160-2.

(6) Rotskoff, G. M.; Mitchell, A. R.; Vanden-Eijnden, E. Active Importance Sampling for Variational Objectives Dominated by Rare Events: Consequences for Optimization and Generalization. In *Mathematical and Scientific Machine Learning*; PMLR, 2022; pp 757−780.

(7) Li, H.; Khoo, Y.; Ren, Y.; Ying, L. A semigroup method for high dimensional committor functions based on neural network. In. In *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*; Bruna, J.; Hesthaven, J.; Zdeborova, L., Eds.; PMLR, 2022. URL; p 598618. https://proceedings.mlr.press/v145/li22a.html

(8) Strahan, J.; Finkel, J.; Dinner, A. R.; Weare, J. Predicting rare events using neural networks and short-trajectory data. *J. Comput. Phys.* **2023**, *488*, No. 112152. ISSN 0021−9991. URL https://www.sciencedirect.com/science/article/pii/S0021999123002474.

(9) Liang, S.; Singh, A. N.; Zhu, Y.; Limmer, D. T.; Yang, C. Probing reaction channels via reinforcement learning. *Mach. Learn.: Sci. Technol.* **2023**, *4* (4), No. 045003. ISSN 2632−2153

(10) Strahan, J.; Guo, S. C.; Lorpaiboon, C.; Dinner, A. R.; Weare, J. Inexact iterative numerical linear algebra for neural network-based spectral estimation and rare-event prediction. *J. Chem. Phys.* **2023**, *159* (1), No. 014110. ISSN 0021−9606. URL 10.1063/5.0151309.

(11) Jung, H.; Covino, R.; Arjun, A.; Leitold, C.; Dellago, C.; Bolhuis, P. G.; Hummer, G. Machine-guided path sampling to discover mechanisms of molecular self-organization. *Nat. Comput. Sci.* **2023**, *3* (4), 334−345. ISSN 2662−8457. URL https://www.nature.com/articles/s43588-023-00428-z. Number: 4 Publisher: Nature Publishing Group.

(12) Evans, L.; Cameron, M. K.; Tiwary, P. Computing committors in collective variables via Mahalanobis diffusion maps. *Appl. Comput. Harmonic Anal.* **2023**, *64*, 62−101. ISSN 1063−5203

(13) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1* (5), 826−843. ISSN 1759−0884. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.31. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.31.

(14) Rotskoff, G. M.; Mitchell, A. R.; Vanden-Eijnden, E.Active Importance Sampling for Variational Objectives Dominated by Rare Events: Consequences for Optimization and Generalization. In *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*; PMLR, 2022; pp 757−780. PMLR, April URL https://proceedings.mlr.press/v145/rotskoff22a.html.

(15) Chen, H.; Roux, B.; Chipot, C. Discovering Reaction Pathways, Slow Variables, and Committor Probabilities with Machine Learning. *J. Chem. Theory Comput.* **2023**, 4414−4426, DOI: 10.1021/acs.jctc.3c00028.

(16) Strahan, J.; Antoszewski, A.; Lorpaiboon, C.; Vani, B. P.; Weare, J.; Dinner, A. R. Long-Time-Scale Predictions from Short-Trajectory Data: A Benchmark Analysis of the Trp-Cage Miniprotein. *J. Chem. Theory Comput.* **2021**, *17* (5), 2948−2963. ISSN 1549−9618

(17) Palacio-Rodriguez, K.; Vroylandt, H.; Stelzl, L. S.; Pietrucci, F.; Hummer, G.; Cossio, P. Transition Rates and Efficiency of Collective Variables from Time-Dependent Biased Simulations. *J. Phys. Chem. Lett.* **2022**, *13* (32), 7490−7496. ISSN 1948−7185, 1948−7185. URL https://pubs.acs.org/doi/10.1021/acs.jpclett.2c01807. tex.ids= palacio-rodriguez_transition_2022−1.

(18) Lazzeri, G.; Jung, H.; Bolhuis, P. G.; Covino, R. Molecular Free Energies, Rates, and Mechanisms from Data-Efficient Path Sampling Simulations. *J. Chem. Theory Comput.* **2023**, *19* (24), 9060−9076. ISSN 1549−9618. URL 10.1021/acs.jctc.3c00821. Publisher: American Chemical Society.

(19) Baudel, M.; Guyader, A.; Lelièvre, T. On the Hill relation and the mean reaction time for metastable processes. *Stochastic Processes Appl.* **2022**, *155*, 393−436, DOI: 10.1016/j.spa.2022.10.014.

(20) Vanden-Eijnden, E.; Tal, F. A. Transition state theory: Variational formulation, dynamical corrections, and error estimates. *J. Chem. Phys.* **2005**, *123* (18), No. 184103. ISSN 0021−9606. URL http://aip.scitation.org/doi/full/10.1063/1.2102898. Publisher: American Institute of Physics.

(21) Pavliotis, G. A. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Texts in Applied Mathematics; Springer: New York, 2014. http://link.springer.com/10.1007/978-1-4939-1323-7

(22) Karniadakis, G. E.; Karniadakis, I. G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3* (6), 422−440. ISSN 2522−5820. URL http://www.nature.com/articles/s42254-021-00314-5.

(23) Lopes, L. J. S.; Lelièvre, T. Analysis of the adaptive multilevel splitting method on the isomerization of alanine dipeptide. *J. Comput. Chem.* **2019**, *40* (11), 1198−1208. ISSN 1096−987X. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.25778. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.25778.

(24) Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7* (4), 284−304. ISSN 0031−8914

(25) Zuckerman, D. M.; Chong, L. Weighted ensemble simulation: Review of methodology, applications, and software. *Ann. Rev. Biophys.* **2017**, *46*, 43−57.

(26) Allen, R. J.; Valeriani, C.; Rein ten Wolde, P. Forward flux sampling for rare event simulations. *J. Phys.: Condens. Matter* **2009**, *21* (46), No. 463102. ISSN 0953−8984. URL 10.1088/0953-8984/21/46/463102. Publisher: IOP Publishing.

(27) van Erp, T. S.; Bolhuis, P. G. Elaborating transition interface sampling methods. *J. Comput. Phys.* **2005**, *205* (1), 157−181. ISSN 0021−9991

(28) Cérou, F.; Guyader, A. Adaptive Multilevel Splitting for Rare Event Analysis. *Stochastic Anal. Appl.* **2007**, *25* (2), 417−443. ISSN 0736−2994

(29) Katsnelson, J.; Kotz, S. On the upper limits of some measures of variability. *Arch. Meteorol., Geophys. Bioklimatol., Ser. B* **1957**, *8* (1), 103−107. ISSN 1434−4483. URL 10.1007/BF02260299.

(30) Pan, A. C.; Sezer, D.; Roux, B. Finding Transition Pathways Using the String Method with Swarms of Trajectories. *J. Phys. Chem. B* **2008**, *112* (11), 3432−3440. ISSN 1520−6106

(31) Guttenberg, N.; Dinner, A. R.; Weare, J. Steered transition path sampling. *J. Chem. Phys.* **2012**, *136* (23), No. 234103. ISSN 0021−9606

(32) Lorpaiboon, C.; Guo, S. C.; Strahan, J.; Weare, J.; Dinner, A. R. Accurate estimates of dynamical statistics using memory. *J. Chem. Phys.* **2024**, *160* (8), No. 084108. ISSN 0021−9606. URL 10.1063/5.0187145.

(33) Bogetti, A. T.; Piston, H. E.; Leung, J. M. G.; Cabalteja, C. C.; Yang, D. T.; DeGrave, A. J.; Debiec, K. T.; Cerutti, D. S.; Case, D. A.; Seth Horne, W.; Chong, L. T. A twist in the road less traveled: The AMBER ff15ipq-m force field for protein mimetics. *J. Chem. Phys.* **2020**, *153* (6), No. 064101. ISSN 0021−9606

(34) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet − A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148* (24), No. 241722. ISSN 0021−9606. URL https://aip.scitation.org/doi/full/10.1063/1.5019779. Publisher: American Institute of Physics.

(35) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. In *Advances in Neural Information Processing Systems*; Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; Oh, A., Eds.; Curran Associates, Inc., 2022; pp 11423−11436. https://proceedings.neurips.cc/paper_files/paper/2022/file/4a36c3c51af11ed9f34615b81edb5bbc-Paper-Conference.pdf

(36) Gasteiger, J.; Becker, F.; Günnemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. In *Advances in Neural Information Processing Systems* 2022; pp 6790−6802. http://arxiv.org/abs/2106.08903

(37) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9* (1), No. 5. ISSN 2041−1723. URL http://www.nature.com/articles/s41467-017-02388-1. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Machine learning;Molecular modelling;Theoretical chemistry Subject_term_id: machine-learning;-molecular-modelling;theoretical-chemistry.

(38) Giardina, C.; Kurchan, J.; Lecomte, V.; Tailleur, J. Simulating Rare Events in Dynamical Processes. *J. Stat. Phys.* **2011**, *145* (4), 787−811. ISSN 0022−4715, 1572−9613. URL http://link.springer.com/10.1007/s10955-011-0350-4.

(39) Heymann, M.; Vanden-Eijnden, E. The geometric minimum action method: A least action principle on the space of curves. *Commun. Pure Appl. Math.* **2008**, *61* (8), 1052−1117. ISSN 00103640, 10970312. URL https://onlinelibrary.wiley.com/doi/10.1002/cpa.20238.

(40) Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135−144. ISSN 0959−440X

(41) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53* (1), 291−318. URL 10.1146/annurev.physchem.53.082301.113146. _eprint: 10.1146/annurev.physchem.53.082301.113146tex.ids=bolhuis_t_2002.

(42) Doob, J. L.; Doob, J. I. *Classical Potential Theory and Its Probabilistic Counterpart*; Springer: New York, 1984.

(43) Hartmann, C.; Banisch, R.; Sarich, M.; Badowski, T.; Schütte, C. Characterization of Rare Events in Molecular Dynamics. *Entropy* **2014**, *16* (1), 350−376. ISSN 1099−4300. URL https://www.mdpi.com/1099-4300/16/1/350. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

(44) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.* **2006**, *125* (8), No. 084110. ISSN 0021−9606, 1089−7690. URL http://aip.scitation.org/doi/10.1063/1.2335447.

(45) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31* (4), 671−690. ISSN 1096−987X. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21367. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21367.

(46) Eastman, P.; Galvelis, R.; Peláez, R. P.; Abreu, C. R. A.; Farr, S. E.; Gallicchio, E.; Gorenko, A.; Henry, M. M.; Hu, F.; Huang, J.; Krämer, A.; Michel, J.; Mitchell, J. A.; Pande, V. S.; Rodrigues, J.; Rodriguez-Guerra, J.; Simmonett, A. C.; Singh, S.; Swails, J.; Turner, P.; Wang, Y.; Zhang, I.; Chodera, J. D.; De Fabritiis, G.; Markland, T. E. Openmm 8: Molecular dynamics simulation with machine learning potentials. *J. Phys. Chem. B* **2024**, *128* (1), 109−116. URL 10.1021/acs.jpcb.3c06662. PMID: 38154096.

(47) Kingma, D. P. Adam: A Method for Stochastic Optimization *arXiv preprint arXiv:1412.6980* 2017.

(48) Leimkuhler, B.; Matthews, C. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. *Applied Mathematics Research eXpress* **2013**, *1*, 34−56. ISSN 1687−1200, 1687−1197. URL https://academic.oup.com/amrx/article-lookup/doi/10.1093/amrx/abs010.