

TRANSITION PATH SAMPLING WITH IMPROVED OFF-POLICY TRAINING OF DIFFUSION PATH SAMPLERS

Kiyoung Seong¹, Seonghyun Park¹, Seonghwan Kim², Woo Youn Kim², Sungsoo Ahn¹

¹POSTECH, ²KAIST

{kyseong, shpark26, sungsoo.ahn}@postech.ac.kr,

{dmdtka00, wooyoun}@kaist.ac.kr

ABSTRACT

Understanding transition pathways between meta-stable states in molecular systems is crucial to advance material design and drug discovery. However, unbiased molecular dynamics simulations are computationally infeasible due to the high energy barriers separating these states. Although recent machine learning techniques offer potential solutions, they are often limited to simple systems or rely on collective variables (CVs) derived from costly domain expertise. In this paper, we introduce a novel approach that trains diffusion path samplers (DPS) for transition path sampling (TPS) without the need for CVs. We recast the problem as an amortized sampling of the target path measure, minimizing the log-variance divergence between the path measure induced by our DPS and the target path measure. To ensure scalability for high-dimensional tasks, we introduce (1) a new off-policy training objective based on learning control variates with replay buffers and (2) a scale-based equivariant parameterization of the bias forces. We evaluate our approach, coined TPS-DPS, on a synthetic double-well potential and three peptides: Alanine Dipeptide, Polyproline Helix, and Chignolin. Results show that our approach produces more realistic and diverse transition pathways compared to existing baselines. We also provide links to our [project page](#) and [code](#).

1 INTRODUCTION

In material design and drug discovery, it is crucial to understand the mechanisms and kinetics of transitions between meta-stable states of molecular systems, such as protein folding and chemical reactions (Mulholland, 2005; Piana et al., 2012; Ahn et al., 2019; Spotte-Smith et al., 2022). Their comprehensive study requires sampling transition paths (Elber, 2016; Lee et al., 2017), which provides insight into mechanisms and energy landscapes. However, naïvely sampling transition paths by molecular dynamics (MD) simulations is often computationally intractable due to high energy barriers, which cause an exponential decay in probability to make a transition (Pechukas, 1981).

To address this problem, researchers have developed enhanced sampling approaches such as steered MD (SMD; Schlitter et al., 1994; Izrailev et al., 1999), umbrella sampling (Torrie & Valleau, 1977; Kästner, 2011), meta-dynamics (Ensing et al., 2006; Branduardi et al., 2012; Bussi & Branduardi, 2015), on-the-fly probability-enhanced sampling (OPES; Invernizzi & Parrinello, 2020), and adaptive biasing force (ABF; Comer et al., 2015) methods. These methods rely on *bias forces* to facilitate transitions across high energy barriers. They are mainly designed based on collective variables (CVs), which are functions of atomic coordinates that capture the slow modes of the transition. Although effective for some systems, the reliance on expensive domain knowledge limits the applicability of the methods to systems where CVs are less understood.

Recently, machine learning has emerged as a promising paradigm for CV-free transition path sampling (Das et al., 2021; Lelièvre et al., 2023; Holdijk et al., 2024). The key idea is to parameterize the bias force using a neural network and train the corresponding biased MD simulation to sample realistic transition paths. In particular, Lelièvre et al. (2023) considered reinforcement learning to sample paths escaping meta-stable states. Das et al. (2021) and Holdijk et al. (2024) proposed to minimize the Kullback-Leibler (KL) divergence between the path measures induced by the neural network and the target path measure. However, minimizing the KL divergence leads to mode collapse,

which is hazardous for sampling transition paths due to a lack of exploration. Furthermore, Das et al. (2021) and Lelièvre et al. (2023) limited their evaluation to low-dimensional synthetic systems. Designing machine learning algorithms for CV-free transition path sampling of large molecules remains an open problem.

Contribution. In this work, we propose to solve the transition path sampling problem with the diffusion path sampler (TPS-DPS). TPS-DPS trains the bias force to minimize a recently proposed log-variance divergence (Nüsken & Richter, 2021) between the path measure induced by the bias MD and the target path measure. To ensure scalability for high-dimensional tasks, e.g., Chignolin folding, we propose a new off-policy training scheme and a scale-based parameterization of the bias forces.

To be specific, we propose a new training objective to minimize the log-variance divergence, defined between a pair of path measures induced by the biased MD and the unbiased MD conditioned on the target meta-stable state. To minimize the variance of gradient estimators and improve sample efficiency, we propose to use (1) learnable control variates and (2) replay buffers with simulated annealing. Our choice of minimizing the log-variance is motivated by recent works (Richter et al., 2020; Nüsken & Richter, 2021; Richter & Berner, 2024) that demonstrate many desirable properties of the log-variance divergence such as a robust gradient estimator for training stability and avoidance of differentiation through the SDE solver. This allows the use of previously collected samples in a replay buffer, promoting sample efficiency in training and improving the diversity of the generated transition paths by preventing the mode collapse observed in KL divergence.

We also introduce a new $SE(3)$ equivariant scale-based parameterization for the bias force that guides the biased MD to sample meaningful paths more frequently, based on inductive bias. Our key idea is to parameterize the bias force to yield a positive inner product with a direction to decrease the distance between the molecular state and the target meta-stable state, which guarantees the bias force to (locally) decrease the distance. We also use the Kabsch algorithm (Kabsch, 1976) to align the molecule and the bias force to the target meta-stable state, guaranteeing $SE(3)$ equivariance for better generalization across molecular states that are identical up to roto-translations.

We extensively evaluate our method on the synthetic double-well potential and three peptides: Alanine Dipeptide, Polyproline Helix, and Chignolin. We compare TPS-DPS with prior ML approach (PIPS; Holdijk et al., 2024), as well as classical non-ML methods, e.g., steered MD (SMD; Izrailev et al., 1999). Our experiments demonstrate that TPS-DPS consistently generates realistic and diverse transition paths, similar to the ground truth ensemble. In addition, ablation studies of algorithmic components verify the effectiveness of our method.

2 RELATED WORK

Transition path sampling without ML. Metadynamics (Branduardi et al., 2012), on-the-fly probability-enhanced sampling (OPES; Invernizzi & Parrinello, 2020), adaptive biasing force (ABF; Comer et al., 2015), and steered molecular dynamics (SMD; Schlitter et al., 1994; Izrailev et al., 1999) were introduced to explore molecular conformations that are difficult to access by unbiased molecular dynamics (MD) within limited simulation times (Hénin et al., 2022). However, they mostly rely on collective variables (CVs) for high-dimensional problems and are inapplicable to systems with unknown CVs. To sample transition paths without CVs, Dellago et al. (1998) proposed shooting methods that use the Markov chain Monte Carlo (MCMC) procedure on path space.

Data driven ML approaches. Recently, generative models have been trained to sample a new transition path given a dataset of transition paths. Petersen et al. (2023); Triplett & Lu (2023) and Lelièvre et al. (2023) applied diffusion probabilistic models (Ho et al., 2020) and variational auto-encoders (Kingma & Welling, 2013) for transition path sampling, respectively. However, these methods are limited to small systems. Concurrent to our work, Jing et al. (2024) generated molecular trajectories by task-wise conditions given an initial state and viewed transition path sampling as an interpolation task between two molecular states.

We also note some related works that are not directly comparable to our method. Klein et al. (2024) proposed to accelerate MD simulation through coarse-grained path generation that may include multiple meta-stable states. However, this work is incapable of targeting a specific meta-stable state and generating fine-grained transition paths. Duan et al. (2023); Kim et al. (2024) use neural networks to generate transition states of a given chemical reaction, but cannot generate transition paths.

Data free ML approaches. Without a previously collected dataset, Das et al. (2021); Lelièvre et al. (2023); Hua et al. (2024); Holdijk et al. (2024) trained generators to sample transition paths using biased MD. Lelièvre et al. (2023) used reinforcement learning to train the generator but focused on escaping an initial meta-stable state rather than reaching the target meta-stable state. Next, Holdijk et al. (2024) considered transition path sampling as minimizing the reverse KL divergence between path distribution from biased MD and transition path distribution. In this work, we mainly compare our method with Holdijk et al. (2024) and provide quantitative analysis for real-world molecule systems. We also note that, concurrent to our work, Du et al. (2024) also considered variational formulation of the TPS problem using Doob’s h -transform.

3 TRANSITION PATH SAMPLING WITH DIFFUSION PATH SAMPLERS

In this section, we introduce our method, coined transition path sampling with diffusion path sampling (TPS-DPS). Our main idea is to formulate the transition path sampling (TPS) problem as a minimization of log-variance divergence (Nüsken & Richter, 2021) between two path measures: the path measure of transition paths and that of biased MD simulations parameterized by a neural network. Our main methodological contribution is twofold: (1) a new training algorithm that minimizes the log-variance divergence with a learned control variate and replay buffer, and (2) a $SE(3)$ equivariant scale-based parameterization of the biased MD that provides inductive bias for dense training signals in high-dimensional problems.

3.1 PROBLEM SETUP

Our goal is to sample transition paths from one meta-stable state to another meta-stable state given a molecule system. We provide an example of the problem for Alanine Dipeptide in Figure 1. We view this as a task to sample paths from an unbiased molecular dynamics (MD), described by a stochastic differential equation (SDE), conditioned on its starting and ending points. To solve this task, we train the bias force parameterized by a neural network to amortize the sampling procedure.

Molecular dynamics. We consider a MD simulation on time interval $[0, T]$ that describes motion of a molecular state $\mathbf{X}_t = (\mathbf{R}_t, \mathbf{V}_t) \in \mathbb{R}^{6N}$ at time t where N is the number of atoms, $\mathbf{R}_t \in \mathbb{R}^{3N}$ is the atom-wise positions and $\mathbf{V}_t \in \mathbb{R}^{3N}$ is the atom-wise velocities. In particular, we adopt Langevin dynamics (Bussi & Parrinello, 2007) defined as the following SDE:

$$d\mathbf{X}_t = \mathbf{u}(\mathbf{X}_t)dt + \Sigma d\mathbf{W}_t, \quad \mathbf{u}(\mathbf{X}_t) = \left(\mathbf{V}_t, -\frac{\nabla U(\mathbf{R}_t)}{m} - \gamma \mathbf{V}_t \right), \quad \Sigma = \text{diag} \left(\zeta, \sqrt{\frac{2\gamma k_B \lambda}{m}} \right) \quad (1)$$

where U , m , γ , k_B , λ , and \mathbf{W}_t denote the potential energy function, the atom-wise masses, the friction term, the Boltzmann constant, the absolute temperature, and the Brownian motion, respectively, and $\zeta \in \mathbb{R}^{3N}$ is a vector of positive infinitesimal values. We let \mathbb{P}_0 denote the path measure induced by Equation (1) in the space $\mathcal{C}([0, T]; \mathbb{R}^{6N})$ of continuous functions $\mathbf{X} : [0, T] \rightarrow \mathbb{R}^{6N}$.

Transition path sampling. One of the challenges in sampling transition paths through unbiased MD simulations is the meta-stability: a state remains trapped for a long time in the initial meta-stable state $\mathcal{A} \subseteq \mathbb{R}^{3N}$ before transitioning into a distinct meta-stable state $\mathcal{B} \subseteq \mathbb{R}^{3N}$. To capture the rare event where transition from \mathcal{A} to \mathcal{B} occurs, we constrain paths $\mathbf{X} = (\mathbf{X}_t)_{0 \leq t \leq T}$ sampled from unbiased MD to satisfy $\mathbf{R}_0 \in \mathcal{A}$, $\mathbf{R}_T \in \mathcal{B}$ for a fixed time T . Since the meta-stable state \mathcal{A} and \mathcal{B} are not well-specified for many molecular systems, we simplify this task by (1) fixing a local minima $\mathbf{R}_A, \mathbf{R}_B$ of the potential energy function in the meta-stable states \mathcal{A}, \mathcal{B} and (2) sampling a transition path \mathbf{X} that starts from the state $\mathbf{R}_0 = \mathbf{R}_A$ and ends at the vicinity of \mathbf{R}_B .

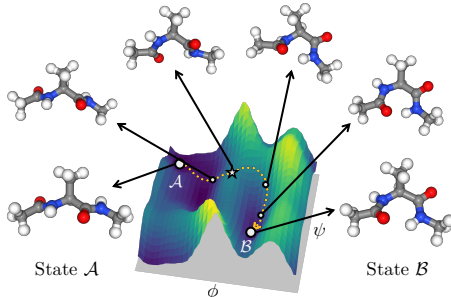


Figure 1: **Problem setup.** Potential energy landscape of Alanine Dipeptide, where the sampled path from state \mathcal{A} to state \mathcal{B} is highlighted in yellow dotted lines. We visualize snapshots in the transition path highlighted in white circles and the transition state highlighted in the white star.

To be specific, we aim to sample from the target path measure \mathbb{Q} which reweights the path measure \mathbb{P}_0 via the Radon–Nikodym derivative, defined as follows:

$$\frac{d\mathbb{Q}}{d\mathbb{P}_0}(\mathbf{X}) = \frac{1_{\mathcal{B}}(\mathbf{X})}{\mathbb{E}_{\mathbb{P}_0}[1_{\mathcal{B}}(\mathbf{X})]}, \quad 1_{\mathcal{B}}(\mathbf{X}) = \begin{cases} 1 & \text{if } \|\mathbf{R}_T - \rho_T \cdot \mathbf{R}_{\mathcal{B}}\| \leq \delta, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where \cdot denotes group action associated with the $SE(3)$ space and $\rho_T \cdot \mathbf{R}_{\mathcal{B}}$ is the aligned position by the optimal roto-translation $\rho_T \in SE(3)$ to minimize its Euclidean distance to \mathbf{R}_T , i.e., $\rho_T = \rho_{\in SE(3)} \|\mathbf{R}_T - \rho \cdot \mathbf{R}_{\mathcal{B}}\|$. Such a transformation can be obtained from the Kabsch algorithm in $O(N)$ complexity (Kabsch, 1976).

Note that one may consider naïve rejection sampling to sample transition paths, based on running unbiased MD to sample a path \mathbf{X} from the path measure \mathbb{P}_0 and accepting if the path \mathbf{X} arrives at the neighborhood of the position $\mathbf{R}_{\mathcal{B}}$ with the radius δ . However, this method does not scale to high-dimensional or low-temperature problems, since the sampled path by unbiased MD rarely reaches the target states due to the high energy barriers, i.e., the rejection ratio is too high.

3.2 LOG-VARIANCE MINIMIZATION

In this section, we propose our algorithm to amortize transition path sampling. Our key idea is to train a neural network to induce a path measure that matches the target path measure \mathbb{Q} , using the log-variance divergence (Nüsken & Richter, 2021) between the path measures. We propose a new training scheme to minimize the log-variance divergence based on learning the control variate of its gradient and a replay buffer to improve sample efficiency and diversity.

Amortizing transition path sampling with log-variance divergence. To match the target path measure \mathbb{Q} , we consider a biased MD defined by a policy v (or bias force \mathbf{b}) as the following SDE:

$$d\mathbf{X}_t = (\mathbf{u}(\mathbf{X}_t) + \Sigma v(\mathbf{X}_t))dt + \Sigma d\mathbf{W}_t, \quad v(\mathbf{X}_t) = \Sigma^{-1} \left(\mathbf{0}, \frac{\mathbf{b}(\mathbf{X}_t)}{m} \right). \quad (3)$$

We also let \mathbb{P}_v denote the path measure induced by the SDE. To amortize transition path sampling, we match the path measure \mathbb{P}_{v_θ} of a parameterized policy v_θ with the target path measure \mathbb{Q} by minimizing the log-variance divergence:

$$D_{\text{LV}}^{\mathbb{P}}(\mathbb{P}_{v_\theta} \|\mathbb{Q}) = \mathbb{V}_{\mathbb{P}} \left[\log \frac{d\mathbb{Q}}{d\mathbb{P}_{v_\theta}} \right] = \mathbb{E}_{\mathbb{P}} \left[\left(\log \frac{d\mathbb{Q}}{d\mathbb{P}_{v_\theta}} - \mathbb{E}_{\mathbb{P}} \left[\log \frac{d\mathbb{Q}}{d\mathbb{P}_{v_\theta}} \right] \right)^2 \right], \quad (4)$$

where \mathbb{P} is an arbitrary reference path measure with $\mathbb{E}_{\mathbb{P}}[\log(d\mathbb{Q}/d\mathbb{P}_{v_\theta})] < \infty$. To express the log-variance divergence in detail, we let $\mathbb{P} = \mathbb{P}_{\tilde{v}}$ for some policy \tilde{v} and apply the Girsanov’s theorem to Equation (4), deriving the following formulation:

$$D_{\text{LV}}^{\mathbb{P}_{\tilde{v}}}(\mathbb{P}_{v_\theta} \|\mathbb{Q}) = \mathbb{E}_{\mathbb{P}_{\tilde{v}}} [(F_{v_\theta, \tilde{v}} - \mathbb{E}_{\mathbb{P}_{\tilde{v}}}[F_{v_\theta, \tilde{v}}])^2], \quad (5)$$

$$F_{v_\theta, \tilde{v}}(\mathbf{X}) = \frac{1}{2} \int_0^T \|\mathbf{v}_\theta(\mathbf{X}_t)\|^2 dt - \int_0^T (\mathbf{v}_\theta \cdot \tilde{\mathbf{v}})(\mathbf{X}_t) dt - \int_0^T \mathbf{v}_\theta(\mathbf{X}_t) \cdot d\mathbf{W}_t + \log 1_{\mathcal{B}}(\mathbf{X}). \quad (6)$$

The first three terms in Equation (6) correspond to the deviation of the biased MD from the unbiased MD integrated over the path sampled from $\mathbb{P}_{\tilde{v}}$. The last term reweights the unbiased MD to the target path measure \mathbb{Q} . As a result, minimizing Equation (5) could be thought as minimizing the variation between \mathbb{P}_{v_θ} and \mathbb{Q} . We provide the full derivation in Appendix A.1. Compared to KL divergence, the log-variance divergence provides a robust gradient estimator and avoids differentiating through the SDE solver. (Richter et al., 2020; Nüsken & Richter, 2021).

Minimizing with learnable control variate. To minimize the log-variance divergence, we consider the following loss function that replaces the estimation of $\mathbb{E}_{\mathbb{P}_{v_\theta}}[F_{v_\theta, v_\theta}]$ by learning a parameter w :

$$\mathcal{L}(\theta, w) = \mathbb{E}_{\mathbb{P}_{v_\theta}} [(F_{v_\theta, v_\theta} - w)^2], \quad (7)$$

where w is a *control variate* that controls the variance of the gradient estimator of $\nabla_{\theta} \mathcal{L}(\theta, w)$ without changing the gradient. Note that we set $\tilde{v} = v_\theta$ in Equation (6), which implies that the gradient of Equation (7) coincides with the KL divergence (Richter et al., 2020; Nüsken & Richter, 2021). When optimized, the control variate w estimates the expectation $\mathbb{E}_{\mathbb{P}_{v_\theta}}[F_{v_\theta, v_\theta}]$ since

$\operatorname{argmin}_w \mathcal{L}(\theta, w) = \mathbb{E}_{\mathbb{P}_{v_\theta}} [F_{v_\theta, v_\theta}]$. Thus, jointly optimizing (θ, w) with the gradient step can be interpreted as jointly minimizing log-variance divergence and estimating $\mathbb{E}_{\mathbb{P}_{v_\theta}} [F_{v_\theta, v_\theta}]$ using w .

Off-policy training with replay buffer. To leverage the degree of freedom in reference path measure for the log-variance divergence, we allow discrepancy between reference path measure and current path measure, called *off-policy training*, which is widely used in discrete-time reinforcement learning (Mnih et al., 2013; Bengio et al., 2021). For the sample efficiency, we reuse the samples with a replay buffer \mathcal{D} which stores path samples from the path measure $\mathbb{P}_{v_{\bar{\theta}}}$ associated with previous policies $v_{\bar{\theta}}$. Our modified loss function $\mathcal{L}^{\mathcal{D}}$ with replay buffer \mathcal{D} is defined as follows:

$$\mathcal{L}^{\mathcal{D}}(\theta, w) = \mathbb{E}_{(v_{\bar{\theta}}, \mathbf{X}) \sim \mathcal{D}} [(F_{v_\theta, v_{\bar{\theta}}}(\mathbf{X}) - w)^2]. \quad (8)$$

Using the replay buffer also prevents mode collapse, using diverse paths from different path measures. Similar to other off-policy training algorithms (Malkin et al., 2022; Kim et al., 2023), we use simulated annealing to collect diverse paths that cross high energy barriers.

Discretization. To implement the algorithm, we discretize Equation (8). Given a discretization step size Δt , we consider the discretized paths $\mathbf{x}_{0:L} = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_L)$ of \mathbf{X} from MD simulations where $L = T/\Delta t$ and $\mathbf{x}_\ell = \mathbf{X}(\ell\Delta t)$. In discrete cases, the discretized paths $\mathbf{x}_{0:L}$ from previous policies $v_{\bar{\theta}}$ and their (gradient-detached) policy values $(v_{\bar{\theta}}(\mathbf{x}_0), \dots, v_{\bar{\theta}}(\mathbf{x}_L))$ are used to approximate the value $F_{v_\theta, v_{\bar{\theta}}}(\mathbf{X})$ in Equation (6) as follows:

$$\hat{F}_{v_\theta, v_{\bar{\theta}}}(\mathbf{x}_{0:L}) = \frac{1}{2} \sum_{\ell=0}^{L-1} \|v_\theta(\mathbf{x}_\ell)\|^2 \Delta t - \sum_{\ell=0}^{L-1} (v_\theta \cdot v_{\bar{\theta}})(\mathbf{x}_\ell) \Delta t - \sum_{\ell=0}^{L-1} v_\theta(\mathbf{x}_\ell) \cdot \ell + \log 1_{\mathcal{B}}(\mathbf{x}_{0:L}), \quad (9)$$

where the noise $\ell = \Sigma^{-1}(\mathbf{x}_{\ell+1} - \mathbf{x}_\ell - (\mathbf{u}(\mathbf{x}_\ell) + \Sigma v_{\bar{\theta}}(\mathbf{x}_\ell))\Delta t)$ is the discretized Brownian motions of the Langevin dynamics with policy $v_{\bar{\theta}}$. For implementation, we further derive a simple discretized loss of Equation (8) from Equation (9) as follows:

$$\mathbb{E}_{\mathbf{x}_{0:L} \sim \hat{\mathcal{D}}} \left[\left(\log \frac{p_0(\mathbf{x}_{0:L}) 1_{\mathcal{B}}(\mathbf{x}_{0:L})}{p_{v_\theta}(\mathbf{x}_{0:L})} - w \right)^2 \right], \quad (10)$$

where the buffer $\hat{\mathcal{D}}$ stores paths $\mathbf{x}_{0:L}$ sampled from the previous policies, and p_0 and p_{v_θ} denote discrete time transition probability induced by Equations (1) and (3), respectively. We provide a formal derivation of the discretized loss in Appendix A.2. Note that the same objective was derived in the name of relative trajectory balance by Venkatraman et al. (2024).

We describe our training algorithm in Appendix A.3. Overall, our off-policy training algorithm iterates through four steps: (1) sampling paths from the biased MD simulation with current policy v_θ at high temperature, (2) storing sampled paths in the replay buffer $\hat{\mathcal{D}}$, (3) sampling a batch of the paths from the replay buffer, and (4) training current policy v_θ by minimizing the loss in Equation (10). After minimization, biased MD simulation can directly sample transition paths by amortizing inference in the target path measure.

3.3 PARAMETERIZATION FOR HIGH DIMENSIONAL TASKS

In this section, we introduce new parameterizations of the bias force and the indicator function for high-dimensional tasks. Our parameterization is designed around alleviating the problem of sparse training signal, where the model is unable to learn anything due to the failure of collecting meaningful paths that end near the meta-stable state. This problem is especially severe in high-dimensional problems such as large molecules.

Bias force parameterization. To frequently sample the meaningful paths, we parameterize the bias force to guarantee (locally) decreasing the distance between the current molecular state and the target meta-stable state. This is done by parameterizing the bias force as positive atom-wise scaling of the force minimizing the distance, which implies that the bias force can be decomposed into two terms: (1) the direction to decrease the Euclidean distance and (2) the direction that leaves the Euclidean distance invariant. The bias force also satisfies roto-translational equivariance to the current state \mathbf{X}_t , aligning with the symmetry of the transition path sampling problem for better generalization.

To be specific, we use a roto-translation $\rho_t \in SE(3)$ to align $\mathbf{R}_{\mathcal{B}}$ with \mathbf{R}_t , as we do for indicator function in Equation (2). To achieve $SE(3)$ equivariant, we parameterize the bias force as follows:

$$\mathbf{b}(\mathbf{X}_t) = \operatorname{diag}(\mathbf{s}_\theta(\rho_t^{-1} \cdot \mathbf{X}_t))(\rho_t \cdot \mathbf{R}_{\mathcal{B}} - \mathbf{R}_t), \quad (11)$$

Table 1: **Benchmark scores on the double-well system and three real-world molecular systems:** Alanine Dipeptide, Polyproline Helix, and Chignolin. All metrics are averaged over 1024 paths for the double-well system, and 64 paths for real-world molecular systems. ETS is computed for paths that hit the target meta-stable state, and the best results are highlighted in bold. TPS-DPS predicting the bias force, potential, and atom-wise scaling are denoted by (F), (P), and (S), respectively. UMD (λ) denotes unbiased MD with temperature λ and SMD (k) denotes steered MD with the force constant k . Unless otherwise specified, paths are generated by MD simulation at 1200K for double-well and 300K for real-world molecules.

Method	RMSD (\downarrow) Å	THP (\uparrow) %	ETS (\downarrow) kJmol ⁻¹	Method	RMSD (\downarrow) Å	THP (\uparrow) %	ETS (\downarrow) kJmol ⁻¹
Double-well				Alanine Dipeptide			
UMD (1200K)	2.21 ± 0.10	0.00	-	UMD (300K)	1.59 ± 0.15	0.00	-
UMD (2400K)	2.11 ± 0.38	3.03	1.69 ± 0.31	UMD (3600K)	1.19 ± 0.32	6.25	812.47 ± 148.80
UMD (3600K)	1.85 ± 0.68	12.60	2.12 ± 0.41	SMD (10)	0.86 ± 0.21	7.81	33.15 ± 6.46
UMD (4800K)	1.54 ± 0.81	21.58	2.77 ± 0.69	SMD (20)	0.56 ± 0.27	54.69	78.40 ± 12.76
SMD (0.5)	0.98 ± 0.90	52.15	1.54 ± 0.21	PIPS (F)	0.66 ± 0.15	43.75	28.17 ± 10.86
SMD (1)	0.14 ± 0.08	99.80	1.85 ± 0.16	PIPS (P)	1.66 ± 0.03	0.00	-
TPS-DPS (F, Ours)	0.01 ± 0.02	99.90	1.38 ± 0.16	TPS-DPS (F, Ours)	0.16 ± 0.06	92.19	19.82 ± 15.88
TPS-DPS (P, Ours)	0.01 ± 0.03	99.71	1.36 ± 0.15	TPS-DPS (P, Ours)	0.16 ± 0.10	87.50	18.37 ± 10.86
TPS-DPS (S, Ours)	0.01 ± 0.03	99.80	1.73 ± 0.20	TPS-DPS (S, Ours)	0.25 ± 0.20	76.00	22.79 ± 13.57
Polyproline Helix				Chignolin			
UMD (300K)	2.22 ± 0.11	0.00	-	UMD (300K)	7.98 ± 0.41	0.00	-
UMD (1200K)	1.38 ± 0.45	10.94	1010.28 ± 38.44	UMD (1200K)	7.23 ± 0.93	1.56	388.17
SMD (5k)	1.68 ± 0.17	54.69	350.58 ± 14.36	SMD (10k)	1.26 ± 0.31	6.25	-527.95 ± 93.58
SMD (10k)	1.26 ± 0.06	100.00	355.62 ± 14.83	SMD (15k)	1.17 ± 0.31	23.44	-237.15 ± 122.29
PIPS (F)	2.64 ± 0.15	0.00	-	PIPS (F)	4.66 ± 0.17	0.00	-
PIPS (P)	1.85 ± 0.13	93.75	574.66 ± 20.49	PIPS (P)	4.67 ± 0.32	0.00	-
TPS-DPS (F, Ours)	1.53 ± 0.12	98.44	418.17 ± 45.54	TPS-DPS (F, Ours)	4.41 ± 0.49	0.00	-
TPS-DPS (P, Ours)	1.35 ± 0.12	100.00	345.00 ± 32.58	TPS-DPS (P, Ours)	3.87 ± 0.42	0.00	-
TPS-DPS (S, Ours)	1.17 ± 0.02	100.00	342.00 ± 20.28	TPS-DPS (S, Ours)	1.17 ± 0.66	59.38	-780.18 ± 216.93

where $s_\theta(\cdot) \in \mathbb{R}_+^{3N}$ is a neural network constrained to have positive output elements. One can observe the positive inner product between the bias force and the direction to minimize the Euclidean distance, i.e., $(\rho_t \cdot \mathbf{R}_B - \mathbf{R}_t)$, as follows:

$$\mathbf{b}(\mathbf{X}_t)^\top (\rho_t \cdot \mathbf{R}_B - \mathbf{R}_t) = (\rho_t \cdot \mathbf{R}_B - \mathbf{R}_t)^\top \text{diag}(s_\theta(\rho_t^{-1} \cdot \mathbf{X}_t)) (\rho_t \cdot \mathbf{R}_B - \mathbf{R}_t) \geq 0, \quad (12)$$

which implies that the bias force can be decomposed into a vector positively aligning with $(\rho_t \cdot \mathbf{R}_B - \mathbf{R}_t)$ and a vector orthogonal to $(\rho_t \cdot \mathbf{R}_B - \mathbf{R}_t)$.

We also note that the bias force is $SE(3)$ equivariant to the current state \mathbf{X}_t since it is the product of an invariant term $\text{diag}(s_\theta(\rho_t^{-1} \cdot \mathbf{X}_t))$ and an equivariant term $(\rho_t \cdot \mathbf{R}_B - \mathbf{R}_t)$. This stems from the fact that the Kabsch algorithm is equivariant to the current state \mathbf{X}_t , i.e., roto-translating position $\mathbf{R}_t \mapsto \tilde{\rho} \cdot \mathbf{R}_t$ leads to roto-translating output of the Kabsch algorithm $\rho_t \mapsto \tilde{\rho} \cdot \rho_t$.

In the experiments, we also consider other equivariant parameterizations that are less constrained: (1) directly predicting the bias force by $\rho_t \cdot \mathbf{b}_\theta(\rho_t^{-1} \cdot \mathbf{X}_t) \in \mathbb{R}^{3N}$ and (2) predicting the force as gradient of the invariant bias potential $\nabla b_\theta(\rho_t^{-1} \cdot \mathbf{X}_t) \in \mathbb{R}$. We observe the parameterizations to be useful for low-dimensional tasks, but suffer from the sparse training signal on high-dimensional tasks.

Indicator function parameterization. We also propose to relax the indicator function 1_B as a radial basis function (RBF) kernel $\tilde{1}_B(\mathbf{X}) = k(\mathbf{R}_T, \rho^{-1} \cdot \mathbf{R}_B; \sigma^2)$ which measures the similarity between two positions and $\sigma > 0$ controls the degree of relaxation. The range of RBF kernel k is bounded by the interval $(0, 1]$ so that $\log \tilde{1}_B(\mathbf{X})$ is well-defined and $\tilde{1}_B(\mathbf{X})$ represents the binary indicator function smoothly. To capture high training signal, we focus on the signal from subtrajectories of sampled paths, taking maximum over RBF kernel values of all intermediate states by $\tilde{1}_B^{\max}(\mathbf{X}) = \max_{t \in [0, T]} k(\mathbf{R}_t, \tilde{\mathbf{R}}_B; \sigma^2)$. To extract the subtrajectory with a high training signal, we can truncate the paths at the time that maximizes RBF kernel values, allowing variable path lengths. Also, the relaxed indicator function is $SE(3)$ invariant to \mathbf{R}_t due to the Kabsch algorithm.

4 EXPERIMENT

In this section, we compare our method, called TPS-DPS, with both classical and ML approaches, assessing the accuracy and diversity of sampled transition paths. We begin with a synthetic double-

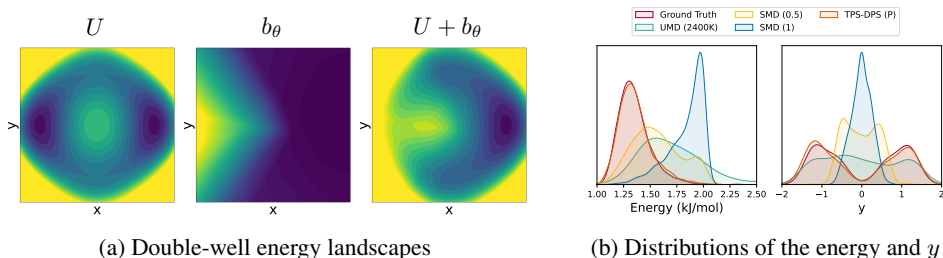


Figure 2: **Visualization of potential energy landscapes and distributions of the double-well system.** (a) Visualization of the neural bias potential b_θ for TPS-DPS (P) in the double-well potential energy landscape. (b) Distributions of the potential energy (left) and y coordinate (right) of transition states from 1024 transition paths sampled by each method.

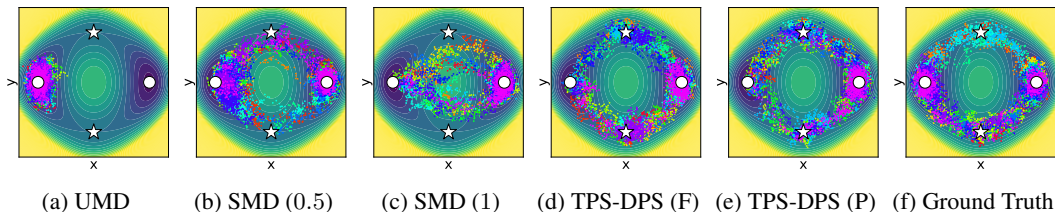


Figure 3: **16 transition path sampled from each method on the potential energy landscape of the double-well system.** White circles indicate the meta-stable states and white stars indicate the saddle points. We sample paths from the left meta-stable state to the right. SMD (k) denotes steered MD with the force constant k . All paths are generated by MD simulation at 300K. SMD (k) denotes steered MD with the force constant k .

well system at 1200K, followed by three real-world molecular systems with various numbers of amino acids: Alanine Dipeptide, Polyproline Helix, and Chignolin. Additionally, we conduct ablation studies to validate the effectiveness of each component in our method. All real-world molecular systems are simulated using the OpenMM library (Eastman et al., 2023). Details on OpenMM simulation and model configurations are provided in Appendices B.1 and B.2, respectively.

Evaluation Metrics. We consider three metrics for evaluation: RMSD, THP, and ETS. The root mean square distance (RMSD) calculates the Euclidean distance between the atoms in the final position of the sampled path and the target position \mathbf{R}_B , with the final position aligned to the target. The target hit percentage (THP) measures the number of paths that successfully arrive at the target meta-stable state B . Finally, the energy of the transition state (ETS) measures the ability to identify probable transition states. For further details, refer to Appendix B.3.

Baselines. We compare TPS-DPS with both non-ML and ML baselines. For non-ML baselines, we consider unbiased MD (UMD) in various temperatures and steered MD (SMD; Izrailev et al., 1999) with various force constants k and collective variables. For ML baselines, we consider a CV-free transition path sampling method, path integral path sampling (PIPS; Holdijk et al., 2024) which also trains a bias force by minimizing the KL divergence between path measures induced by the biased MD and the target path measure. For simplicity, we denote parameterizations for predicting force, potential, and atom-wise scaling as (F), (P) and (S), respectively.

4.1 DOUBLE-WELL SYSTEM

We begin by evaluating our method on a two-dimensional synthetic system, i.e., a double-well potential at 1200K. This system has two global minima (white circles) representing the meta-stable states, and two reaction pathways via saddle points (white stars). We sample transition paths from the left meta-stable state \mathbf{R}_A to the right meta-stable states $B = \{\mathbf{R} \mid \|\mathbf{R} - \mathbf{R}_B\| < 0.5\}$, using over-damped Langevin dynamics. We collect ground truth path ensembles by rejection sampling which proposes paths sampled from the unbiased MD simulations and accepts if the final states are in the target meta-stable states B . We provide more details on the system in Appendix B.4.

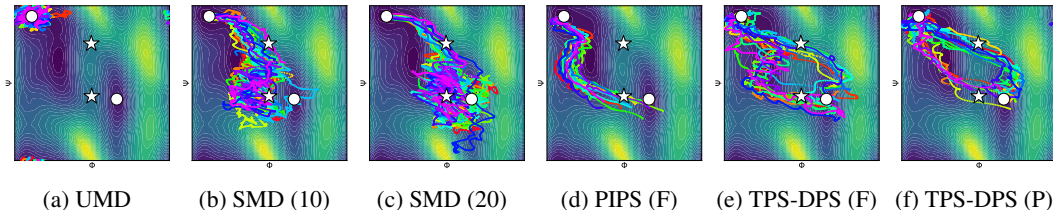


Figure 4: **16 sampled paths from each method on the Ramachandran plot of Alanine Dipeptide.** White circles indicate meta-stable states and white stars indicate saddle points. We sample transition paths from the meta-stable state $C5$ (upper left) to $C7ax$ (lower right). SMD (k) denotes steered MD with the force constant k . Paths are generated by MD simulation at 1200K.

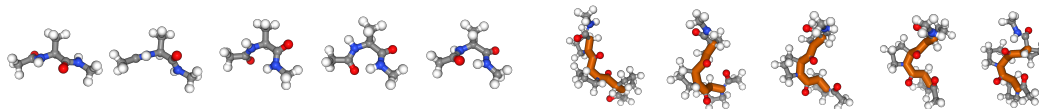


Figure 5: **Snapshots of transition path** for Alanine Dipeptide (left) and Polyproline Helix (right).

In Table 1, TPS-DPS shows superior performance for the double-well potential regardless of the bias force design. In Figure 2, the neural bias potential accelerates the transition by increasing the potential energy near the initial meta-stable state while decreasing the potential energy near the two energy barriers. Moreover, the distribution of energy and y coordinates of the transition states from TPS-DPS is closest to the ground truth compared with other baselines, successfully capturing two reaction channels. In Figure 3, unbiased MD at 1200K fails to escape the initial state while SMD struggles to pass the saddle points. Our approach generates transition paths more similar to the ground truth than other methods.

4.2 ALANINE DIPEPTIDE CONFORMATION CHANGE

We first consider Alanine Dipeptide for a real-world molecule consisting of two alanine residues, sampling transition paths from the $C5$ (upper left) to the $C7ax$ (lower right) as seen in Figure 4. The target meta-stable states are defined as $\mathcal{B} = \{\mathbf{R} \mid \|\xi(\mathbf{R}) - \xi(\mathbf{R}_B)\| < 0.75\}$, where $\xi(\mathbf{R}) = (\phi, \psi)$ is a well-known collective variable which consists of two backbone dihedral angles. Also, Alanine Dipeptide has two reaction channels between the $C5$ and $C7ax$ state passing through the saddle points marked as white stars.

In Table 1 and Figure 4, unbiased MD at 300K fails to escape the initial state, while SMD with the collective variable ξ generates transition paths with less probable transition states. PIPS generates transition paths of only one reaction channel, suffering from mode collapse. Compared to the baselines, our method successfully generates diverse transition paths that pass two reaction channels and show superior performance regardless of the bias force design. For qualitative analysis, we plot the snapshots of the transition path sample from TPS-DPS in Figure 5.

4.3 POLYPROLINE HELIX ISOMERIZATION

Next, we consider a more complex molecule, Polyproline Helix consisting of three proline residues. We sample transition paths from the left-handed state (PPII) in the *cis*-configuration to the right-handed state (PPI) in the *trans*-configuration. For simplicity, we define the target meta-stable state based on handedness H as $\mathcal{B} = \{\mathbf{R} \mid H(\mathbf{R}) > 0\}$. For the formal definition of handedness, refer to Appendix B.4. As seen in Table 1, our models consistently outperform all baselines: unbiased MD, SMD, and PIPS. In Figure 7a, a transition path sampled from TPS-DPS crosses the energy barrier around 0.4ps and changes handedness for 1.5ps. Additionally in Figure 5, we plot the snapshots of the sampled transition path.

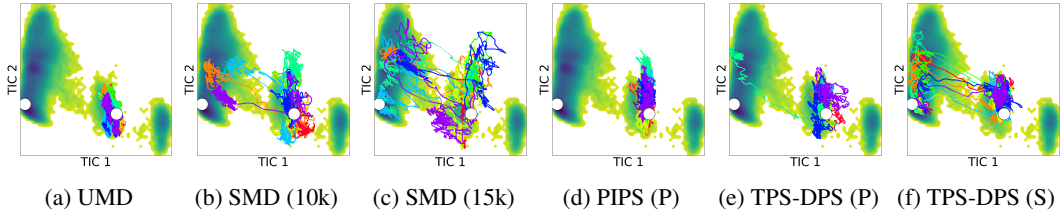


Figure 6: **16 sampled paths from each method projected to the top two TICA components.** White circles indicate meta-stable states. We sample transition paths from the unfolded state (right) to the folded state (left). SMD (k) denotes steered MD with the force constant k . All paths are generated by MD simulation at 300K. (P) and (S) refer to predicting the bias potential and atom-wise scaling.

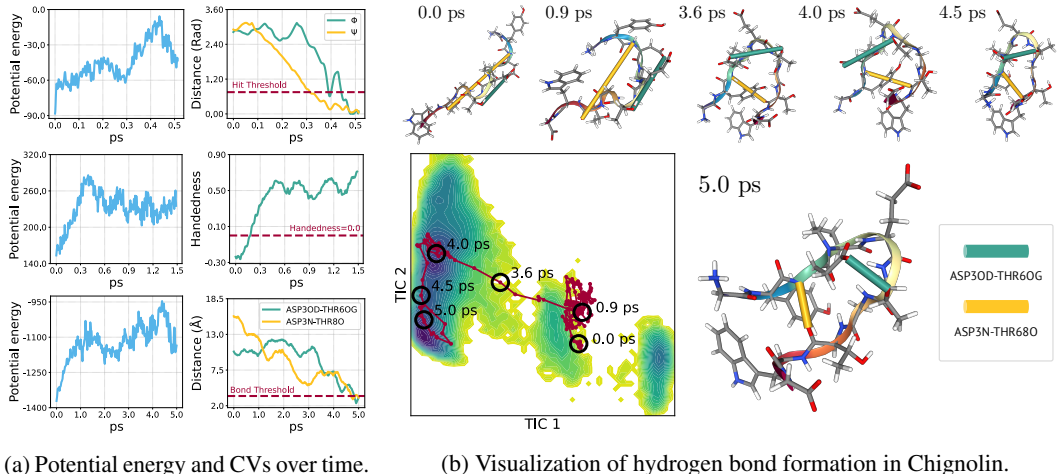


Figure 7: **Qualitative evaluation on transition path sampled from TPS-DPS.** (a) Potential energies and collective variables (CVs) of Alanine dipeptide, Polyporline Helix, and Chignolin from top to bottom. For CVs, we plot the distance of the backbone dihedral angles between the current state and the target state for Alanine Dipeptide, the handedness of the backbone for Polyproline Helix, and the hydrogen bond distance for Chignolin. One can see that the sampled path successfully transitioned to the target meta-stable state. (b) Visualization of hydrogen bond formation for Chignolin in a sampled path. We project a sampled path on the top two TICA components. We visualize the time steps indicated by the black circles, with each hydrogen bond highlighted in green and yellow.

4.4 CHIGNOLIN FOLDING

Finally, we consider a challenging molecule, Chignolin, an artificial protein consisting of 10 amino acids (Honda et al., 2004), which folds into a β -hairpin structure by hydrogen bonds. We sample transition paths from the unfolded state (right white circle) to the folded state (left white circle) in Figure 6. We define the target meta-stable state $\mathcal{B} = \{ \mathbf{R} \mid \|\xi(\mathbf{R}) - \xi(\mathbf{R}_B)\| < 0.75 \}$ where ξ consists of the top two components of time-lagged independent component analysis (TICA; Pérez-Hernández et al., 2013). We further describe TICA in Appendix B.4.

In Table 1 and Figure 6, unbiased MD, PIPS, TPS-DPS (F), and TPS-DPS (P) fail to hit the target meta-stable state. While SMD hits the target meta-stable, its transition paths do not pass probable transition states. Only TPS-DPS (S) successfully samples transition paths that pass probable transition states. In Figure 7, we further validate the sampled paths using the potential energy and donor-acceptance distance of the two key hydrogen bonds. The sampled path forms two key hydrogen bonds, ASP3OD-THR6OG and ASP3N-THR8O, reducing the donor-acceptance distance below the threshold 3.5\AA . Additionally, the transition time of the sampled path by TPS-DPS is 5ps, where the average folding time of Chignolin is known to be $0.6\mu\text{s}$ (Lindorff-Larsen et al., 2011).

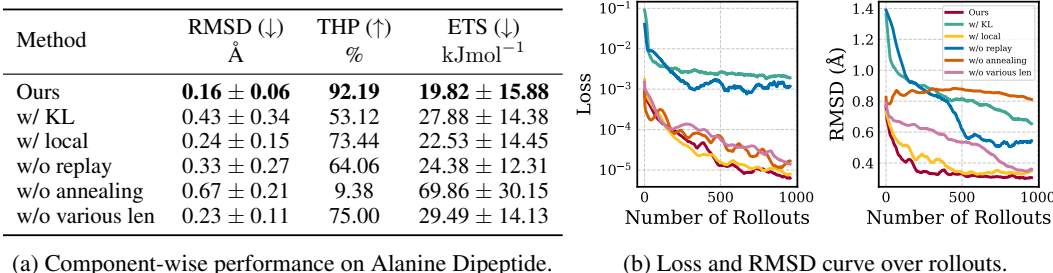


Figure 9: **Ablation studies on the components of TPS-DPS in the Alanine Dipeptide.** (a) Benchmark scores on Alanine Dipeptide. All metrics are averaged over 64 paths, and we highlight the best results in bold. ETS is computed for paths that hit the target meta-stable state. (b) Loss and RMSD curves averaged over 8 different seeds.

4.5 ABLATION STUDY

Effectiveness of algorithmic components. First, we conduct ablation studies to verify the effectiveness of the five proposed components: log-variance loss, learnable control variate, replay buffer, simulated annealing, and maximum over RBF values for various path lengths. To be specific, we (1) replace our loss with the KL divergence, (2) replace the learnable control variate with the local control variate used in Nüsken & Richter (2021), (3) remove the replay buffer and use data only from the current policy, (4) use only one temperature λ , and (5) remove maximum operation over RBF values using only the final state.

As seen in Figure 9, all the proposed components improve performance. Our loss is smaller than the KL divergence by more than two orders of magnitude and significantly improves performance. Learning the control variate slightly improves performance, showing that utilizing data from previous policies is effective. The replay buffer significantly improves training efficiency, and shows that the large performance gap between our loss and KL divergence comes from the replay buffer. Simulated annealing for biased MD simulation is critical to finding transition paths. RMSD does not decrease without simulated annealing while loss decreases significantly. For the relaxed indicator function, maximum operation accelerates convergence and improves performance with frequent training signals from the subtrajectories.

Comparison to steered MD with varying force constant k . The performance of SMD differs by the force constant k . As seen in Figure 8, one can see the trade-off between THP and ETS for various k values. TPS-DPS outperforms SMD in THP and ETS without showing a trade-off between them.

5 CONCLUSION

In this work, we introduced a novel CV-free diffusion path sampler, called TPS-DPS, to amortize transition path sampling, using log-variance divergence with the learnable control variate, and the replay buffer with simulated annealing for off-policy training. We also propose new $SE(3)$ equivariant scale-based parameterization for bias force and relaxed indicator function for frequent training signals. Evaluations on synthetic double-well and real-molecule systems such as Alanine Dipeptide, Polyproline Helix, and Chignolin demonstrate superior accuracy and diversity of our model compared to both classical and ML approaches.

Limitation. While our algorithm successfully samples transition paths for small proteins, we did not evaluate our algorithm for even larger proteins due to the lack of computational budget. In addition, our work does not generalize across unseen pairs of meta-stable states or different molecular systems. These points to an interesting venue for future research, which would be more appealing for practical applications in drug discovery or material design.

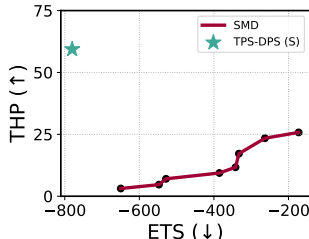


Figure 8: **Trade-off between THP and ETS for various k in SMD.** The further left and higher up (\nearrow), the better.

REFERENCES

- Seihwan Ahn, Mannkyu Hong, Mahesh Sundararajan, Daniel H Ess, and Mu-Hyun Baik. Design and optimization of catalysts based on mechanistic insights derived from quantum chemical reaction modeling. *Chemical reviews*, 119(11):6509–6560, 2019.
- Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. *Advances in Neural Information Processing Systems*, 34:27381–27394, 2021.
- Davide Branduardi, Giovanni Bussi, and Michele Parrinello. Metadynamics with adaptive gaussians. *Journal of chemical theory and computation*, 8(7):2247–2254, 2012.
- Giovanni Bussi and Davide Branduardi. Free-energy calculations with metadynamics: Theory and practice. *Reviews in Computational Chemistry Volume 28*, pp. 1–49, 2015.
- Giovanni Bussi and Michele Parrinello. Accurate sampling using langevin dynamics. *Physical Review E*, 75(5):056707, 2007.
- Jeffrey Comer, James C Gumbart, Jérôme Hénin, Tony Lelièvre, Andrew Pohorille, and Christophe Chipot. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *The Journal of Physical Chemistry B*, 119(3):1129–1151, 2015.
- Avishek Das, Dominic C Rose, Juan P Garrahan, and David T Limmer. Reinforcement learning of rare diffusive dynamics. *The Journal of Chemical Physics*, 155(13), 2021.
- Christoph Dellago, Peter G Bolhuis, and David Chandler. Efficient transition path sampling: Application to lennard-jones cluster rearrangements. *The Journal of chemical physics*, 108(22):9236–9245, 1998.
- Yuanqi Du, Michael Plainer, Rob Brekelmans, Chenru Duan, Frank Noe, Carla P Gomes, Alan Aspuru-Guzik, and Kirill Neklyudov. Doob’s lagrangian: A sample-efficient variational approach to transition path sampling. In *ICML 2024 AI for Science Workshop*, 2024.
- Chenru Duan, Yuanqi Du, Haojun Jia, and Heather J Kulik. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nature Computational Science*, 3(12):1045–1055, 2023.
- Peter Eastman, Raimondas Galvelis, Raúl P Peláez, Charlles RA Abreu, Stephen E Farr, Emilio Gallicchio, Anton Gorenko, Michael M Henry, Frank Hu, Jing Huang, et al. Openmm 8: Molecular dynamics simulation with machine learning potentials. *The Journal of Physical Chemistry B*, 128(1):109–116, 2023.
- Ron Elber. Perspective: Computer simulations of long time dynamics. *The Journal of chemical physics*, 144(6), 2016.
- Bernd Ensing, Marco De Vivo, Zhiwei Liu, Preston Moore, and Michael L Klein. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Accounts of chemical research*, 39(2):73–81, 2006.
- Jérôme Hénin, Tony Lelièvre, Michael R Shirts, Omar Valsson, and Lucie Delemotte. Enhanced sampling methods for molecular dynamics simulations [article v1. 0]. *Living Journal of Computational Molecular Science*, 4(1):1583–1583, 2022.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
- Lars Holdijk, Yuanqi Du, Ferry Hooft, Priyank Jaini, Berend Ensing, and Max Welling. Stochastic optimal control for collective variable free sampling of molecular transition paths. *Advances in Neural Information Processing Systems*, 36, 2024.
- Shinya Honda, Kazuhiko Yamasaki, Yoshito Sawada, and Hisayuki Morii. 10 residue folded peptide designed by segment statistics. *Structure*, 12(8):1507–1518, 2004.

- Xinru Hua, Rasool Ahmad, Jose Blanchet, and Wei Cai. Accelerated sampling of rare events using a neural network bias potential. *arXiv preprint arXiv:2401.06936*, 2024.
- Michele Invernizzi and Michele Parrinello. Rethinking metadynamics: from bias potentials to probability distributions. *The journal of physical chemistry letters*, 11(7):2731–2736, 2020.
- Sergei Izrailev, Sergey Stepaniants, Barry Isralewitz, Dorina Kosztin, Hui Lu, Ferenc Molnar, Willy Wriggers, and Klaus Schulten. 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*, pp. 39–65. Springer, 1999.
- Bowen Jing, Hannes Stark, Tommi Jaakkola, and Bonnie Berger. Generative modeling of molecular dynamics trajectories. In *ICML’24 Workshop ML for Life and Material Science: From Theory to Industry Applications*, 2024.
- Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):922–923, 1976.
- Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.
- Minsu Kim, Joohwan Ko, Dinghuai Zhang, Ling Pan, Taeyoung Yun, Woo Chang Kim, Jinkyoo Park, and Yoshua Bengio. Learning to scale logits for temperature-conditional gflownets. In *NeurIPS 2023 AI for Science Workshop*, 2023.
- Seonghwan Kim, Jeheon Woo, and Woo Youn Kim. Diffusion-based generative ai for exploring transition states from 2d molecular graphs. *Nature Communications*, 15(1):341, 2024.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *The Second International Conference on Learning Representations*, 2013.
- Leon Klein, Andrew Foong, Tor Fjelde, Bruno Mlodozeniec, Marc Brockschmidt, Sebastian Nowozin, Frank Noé, and Ryota Tomioka. Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics. *Advances in Neural Information Processing Systems*, 36, 2024.
- Juyong Lee, In-Ho Lee, InSuk Joung, Jooyoung Lee, and Bernard R Brooks. Finding multiple reaction pathways via global optimization of action. *Nature Communications*, 8(1):15443, 2017.
- Tony Lelièvre, Geneviève Robin, Innas Sekkat, Gabriel Stoltz, and Gabriel Victorino Cardoso. Generative methods for sampling transition paths in molecular dynamics. *ESAIM: Proceedings and Surveys*, 73:238–256, 2023.
- Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins*, 78:1950–1958, 2010.
- Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.
- JA Maier, C Martinez, K Kasavajhala, L Wickstrom, KE Hauser, and C Simmerling ff14SB. Improving the accuracy of protein side chain and backbone parameters from ff99sb., 2015, 11. DOI: <https://doi.org/10.1021/acs.jctc.5b00255>, pp. 3696–3713, 2015.
- Nikolay Malkin, Moksh Jain, Emmanuel Bengio, Chen Sun, and Yoshua Bengio. Trajectory balance: Improved credit assignment in gflownets. *Advances in Neural Information Processing Systems*, 35: 5955–5967, 2022.
- Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning, 2013.
- Mahmoud Moradi, Volodymyr Babin, Christopher Roland, Thomas A Darden, and Celeste Sagui. Conformations and free energy landscapes of polyproline peptides. *Proceedings of the National Academy of Sciences*, 106(49):20746–20751, 2009.

- Adrian J Mulholland. Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug discovery today*, 10(20):1393–1402, 2005.
- Nikolas Nüsken and Lorenz Richter. Solving high-dimensional hamilton–jacobi–bellman pdes using neural networks: perspectives from the theory of controlled diffusions and measures on path space. *Partial differential equations and applications*, 2(4):48, 2021.
- Philip Pechukas. Transition state theory. *Annual Review of Physical Chemistry*, 32(1):159–177, 1981.
- Guillermo Pérez-Hernández, Fabian Paul, Toni Giorgino, Gianni De Fabritiis, and Frank Noé. Identification of slow molecular order parameters for markov model construction. *The Journal of chemical physics*, 139(1), 2013.
- Magnus Petersen, Gemma Roig, and Roberto Covino. Dynamicsdiffusion: Generating and rare event sampling of molecular dynamic trajectories using diffusion models. In *NeurIPS 2023 AI for Science Workshop*, 2023.
- Stefano Piana, Kresten Lindorff-Larsen, and David E Shaw. Protein folding kinetics and thermodynamics from atomistic simulation. *Proceedings of the National Academy of Sciences*, 109(44):17845–17850, 2012.
- Lorenz Richter and Julius Berner. Improved sampling via learned diffusions. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=h4pNROsO06>.
- Lorenz Richter, Ayman Boustati, Nikolas Nüsken, Francisco Ruiz, and Omer Deniz Akyildiz. Vargrad: a low-variance gradient estimator for variational inference. *Advances in Neural Information Processing Systems*, 33:13481–13492, 2020.
- Martin K Scherer, Benjamin Trendelkamp-Schroer, Fabian Paul, Guillermo Pérez-Hernández, Moritz Hoffmann, Nuria Plattner, Christoph Wehmeyer, Jan-Hendrik Prinz, and Frank Noé. Pyemma 2: A software package for estimation, validation, and analysis of markov models. *Journal of chemical theory and computation*, 11(11):5525–5542, 2015.
- Jürgen Schlitter, Michael Engels, and Peter Krüger. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *Journal of molecular graphics*, 12(2):84–89, 1994.
- David A Sivak, John D Chodera, and Gavin E Crooks. Time step rescaling recovers continuous-time dynamical properties for discrete-time langevin integration of nonequilibrium systems. *The Journal of Physical Chemistry B*, 118(24):6466–6474, 2014.
- Evan Walter Clark Spotte-Smith, Ronald L Kam, Daniel Barter, Xiaowei Xie, Tingzheng Hou, Shyam Dwaraknath, Samuel M Blau, and Kristin A Persson. Toward a mechanistic model of solid–electrolyte interphase formation and evolution in lithium-ion batteries. *ACS Energy Letters*, 7(4):1446–1453, 2022.
- Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of computational physics*, 23(2):187–199, 1977.
- Luke Triplett and Jianfeng Lu. Diffusion methods for generating transition paths. *arXiv preprint arXiv:2309.10276*, 2023.
- Siddarth Venkatraman, Moksh Jain, Luca Scimeca, Minsu Kim, Marcin Sendera, Mohsin Hasan, Luke Rowe, Sarthak Mittal, Pablo Lemos, Emmanuel Bengio, et al. Amortizing intractable inference in diffusion models for vision, language, and control. *arXiv preprint arXiv:2405.20971*, 2024.
- Hao Zheng, Zhanlei Yang, Wenju Liu, Jizhong Liang, and Yanpeng Li. Improving deep neural networks using softplus units. In *2015 International joint conference on neural networks (IJCNN)*, pp. 1–4. IEEE, 2015.

A METHOD DETAILS

A.1 LOG VARIANCE FORMULATION

In this section, we derive Equation (5) from Equation (4) to get the explicit expression for log-variance divergence in terms of SDE in Equation (1) and Equation (3). We refer to Nüsken & Richter (2021, Appendix A.1) for the derivation in more general settings.

Our goal is to derive that

$$\mathbb{E}_{\mathbb{P}_{\tilde{\mathbf{v}}}} \left[\left(\log \frac{d\mathbb{Q}}{d\mathbb{P}_{\mathbf{v}_\theta}} - \mathbb{E}_{\mathbb{P}_{\tilde{\mathbf{v}}}} \left[\log \frac{d\mathbb{Q}}{d\mathbb{P}_{\mathbf{v}_\theta}} \right] \right)^2 \right] = \mathbb{E}_{\mathbb{P}_{\tilde{\mathbf{v}}}} [(F_{\mathbf{v}_\theta, \tilde{\mathbf{v}}} - \mathbb{E}_{\mathbb{P}_{\tilde{\mathbf{v}}}}[F_{\mathbf{v}_\theta, \tilde{\mathbf{v}}}]^2], \quad (13)$$

To this end, we focus on calculating $\log \frac{d\mathbb{Q}}{d\mathbb{P}_{\mathbf{v}_\theta}}(\mathbf{X})$ when $\mathbf{X} \sim \mathbb{P}_{\tilde{\mathbf{v}}}$. Following (Nüsken & Richter, 2021, Lemma A.1), we apply Girsanov’s Theorem to calculate the Radon-Nikodym derivative $\frac{d\mathbb{P}_{\mathbf{v}_\theta}}{d\mathbb{P}_0}$ as follows:

$$\frac{d\mathbb{P}_{\mathbf{v}_\theta}}{d\mathbb{P}_0}(\mathbf{X}) = \exp \left(\int_0^T (\mathbf{v}_\theta^T \Sigma^{-1})(\mathbf{X}_t) \cdot d\mathbf{X}_t - \int_0^T (\Sigma^{-1} \mathbf{u} \cdot \mathbf{v}_\theta)(\mathbf{X}_t) dt - \frac{1}{2} \int_0^T \|\mathbf{v}_\theta(\mathbf{X}_t)\|^2 dt \right) \quad (14)$$

since the state \mathbf{X}_t follows the SDE $d\mathbf{X}_t = (\mathbf{u}(\mathbf{X}_t) + \Sigma \tilde{\mathbf{v}}(\mathbf{X}_t))dt + \Sigma d\mathbf{W}_t$. We plug it into Equation (14) and utilize the definition of the target path measure \mathbb{Q} in Equation (2) to compute $\log \frac{d\mathbb{Q}}{d\mathbb{P}_0}$ as follows:

$$\log \frac{d\mathbb{Q}}{d\mathbb{P}_{\mathbf{v}_\theta}}(\mathbf{X}) = \log \frac{d\mathbb{Q}}{d\mathbb{P}_0} \frac{d\mathbb{P}_0}{d\mathbb{P}_{\mathbf{v}_\theta}}(\mathbf{X}) \quad (15)$$

$$= \log 1_{\mathcal{B}}(\mathbf{X}) - \log Z - \int_0^T (\mathbf{v}_\theta^T \Sigma^{-1})(\mathbf{X}_t) \cdot d\mathbf{X}_t \quad (16)$$

$$+ \int_0^T (\Sigma^{-1} \mathbf{u} \cdot \mathbf{v}_\theta)(\mathbf{X}_t) dt + \frac{1}{2} \int_0^T \|\mathbf{v}_\theta(\mathbf{X}_t)\|^2 dt \quad (17)$$

$$= \log 1_{\mathcal{B}}(\mathbf{X}) - \log Z - \int_0^T (\mathbf{v}_\theta \cdot \tilde{\mathbf{v}})(\mathbf{X}_t) dt \quad (18)$$

$$- \int_0^T \mathbf{v}_\theta(\mathbf{X}_t) \cdot d\mathbf{W}_t + \frac{1}{2} \int_0^T \|\mathbf{v}_\theta(\mathbf{X}_t)\|^2 dt \quad (19)$$

$$= F_{\mathbf{v}_\theta, \tilde{\mathbf{v}}}(\mathbf{X}) - \log Z \quad (20)$$

Since $\log Z$ is the constant, it is canceled out in the log-variance divergence as follows:

$$\mathbb{E}_{\mathbb{P}_{\tilde{\mathbf{v}}}} \left[\left(\log \frac{d\mathbb{Q}}{d\mathbb{P}_{\mathbf{v}_\theta}} - \mathbb{E}_{\mathbb{P}_{\tilde{\mathbf{v}}}} \left[\log \frac{d\mathbb{Q}}{d\mathbb{P}_{\mathbf{v}_\theta}} \right] \right)^2 \right] = \mathbb{E}_{\mathbb{P}_{\tilde{\mathbf{v}}}} [(F_{\mathbf{v}_\theta, \tilde{\mathbf{v}}} - \mathbb{E}_{\mathbb{P}_{\tilde{\mathbf{v}}}}[F_{\mathbf{v}_\theta, \tilde{\mathbf{v}}}]^2], \quad (21)$$

A.2 CONNECTION TO EXISTING LOSS FUNCTIONS ON DISCRETE-TIME DOMAIN

In this section, we connect our discretized loss of Equation (8) to the loss function, called relative trajectory balance (Venkatraman et al., 2024, RTB). Like our methods, RTB also amortized inference in target path distribution by training forward distribution on discrete-time domains such as vision, language, and control tasks. When discretized, our loss function is equivalent to the RTB objective.

Our goal is to show that for every paths $\mathbf{x}_{0:L}$ sampled from the path measure $\mathbb{P}_{\mathbf{v}_\theta}$,

$$(\hat{F}_{\mathbf{v}_\theta, \mathbf{v}_\theta}(\mathbf{x}_{0:L}) - w)^2 = \left(\log \frac{p_0(\mathbf{x}_{0:L}) 1_{\mathcal{B}}(\mathbf{x}_{0:L})}{Z_\theta p_{\mathbf{v}_\theta}(\mathbf{x}_{0:L})} \right)^2, \quad (22)$$

where $w = \log Z_\theta$ is a learnable scalar parameter, and path distribution $p_{\mathbf{v}}(\mathbf{x}_{0:L}) = \prod_{\ell=0}^{L-1} p_{\mathbf{v}}(\mathbf{x}_{\ell+1} | \mathbf{x}_\ell)$ is Markovian, and its transition kernel $p_{\mathbf{v}}(\mathbf{x}_{\ell+1} | \mathbf{x}_\ell)$ are derived from Euler-Maruyama discretization of the SDE in Equation (3) as follows:

$$\mathbf{x}_{\ell+1} = \mathbf{x}_\ell + \mathbf{u}(\mathbf{x}_\ell) \Delta t + \Sigma \mathbf{v}(\mathbf{x}_\ell) \Delta t + \Sigma_\ell, \quad (23)$$

where $l \sim \mathcal{N}(\mathbf{0}, \Delta t)$. To this end, we can calculate as follows:

$$\log p_0(\mathbf{x}_{0:L}) - \log p_{\mathbf{v}_\theta}(\mathbf{x}_{0:L}) \quad (24)$$

$$= \sum_{\ell=0}^{L-1} \log p_{\mathbf{v}_\theta}(\mathbf{x}_{\ell+1}|\mathbf{x}_\ell) - \sum_{\ell=0}^{L-1} \log p_0(\mathbf{x}_{\ell+1}|\mathbf{x}_\ell) \quad (25)$$

$$= \frac{1}{2} \sum_{\ell=0}^{L-1} (\Sigma \mathbf{v}_{\bar{\theta}} \Delta t + \Sigma_l - \Sigma \mathbf{v}_\theta \Delta t)^T (\Sigma^T \Sigma \Delta t)^{-1} (\Sigma \mathbf{v}_{\bar{\theta}} \Delta t + \Sigma_l - \Sigma \mathbf{v}_\theta \Delta t) \quad (26)$$

$$- \frac{1}{2} \sum_{\ell=0}^{L-1} (\Sigma \mathbf{v}_{\bar{\theta}} \Delta t + \Sigma_l)^T (\Sigma^T \Sigma \Delta t)^{-1} (\Sigma \mathbf{v}_{\bar{\theta}} \Delta t + \Sigma_l) \quad (27)$$

$$= \frac{1}{2\Delta t} \sum_{\ell=0}^{L-1} (\|\mathbf{v}_{\bar{\theta}} \Delta t + l - \mathbf{v}_\theta \Delta t\|^2 - \|\mathbf{v}_{\bar{\theta}} \Delta t + l\|^2) \quad (28)$$

$$= \frac{1}{2} \sum_{\ell=0}^{L-1} \|\mathbf{v}_\theta(\mathbf{x}_\ell)\|^2 \Delta t - \sum_{\ell=0}^{L-1} (\mathbf{v}_\theta \cdot \mathbf{v}_{\bar{\theta}})(\mathbf{x}_\ell) \Delta t - \sum_{\ell=0}^{L-1} \mathbf{v}_\theta(\mathbf{x}_\ell) \cdot \ell \quad (29)$$

$$= \hat{F}_{\mathbf{v}_\theta, \mathbf{v}_{\bar{\theta}}}(\mathbf{x}_{0:L}) - \log 1_{\mathcal{B}}(\mathbf{x}_{0:L}) \quad (30)$$

which implies

$$\hat{F}_{\mathbf{v}_\theta, \mathbf{v}_{\bar{\theta}}}(\mathbf{x}_{0:L}) = \log \frac{p_0(\mathbf{x}_{0:L}) 1_{\mathcal{B}}(\mathbf{x}_{0:L})}{p_{\mathbf{v}_\theta}(\mathbf{x}_{0:L})}, \quad (31)$$

by subtracting w and squaring both sides, we have

$$(\hat{F}_{\mathbf{v}_\theta, \mathbf{v}_{\bar{\theta}}}(\mathbf{x}_{0:L}) - w)^2 = \left(\log \frac{p_0(\mathbf{x}_{0:L}) 1_{\mathcal{B}}(\mathbf{x}_{0:L})}{Z_\theta p_{\mathbf{v}_\theta}(\mathbf{x}_{0:L})} \right)^2 \quad (32)$$

We can view $p_0(\mathbf{x}_{0:L}) 1_{\mathcal{B}}(\mathbf{x}_{0:L})$ as the unnormalized target distribution discretized from the target path measure \mathbb{Q} , Z_θ as the estimator for normalizing constant $Z = \int p_0(\mathbf{x}_{0:L}) 1_{\mathcal{B}}(\mathbf{x}_{0:L}) d\mathbf{x}_{0:L}$, and $p_{\mathbf{v}_\theta}(\mathbf{x}_{0:L})$ as forward probability distribution to amortize inference in the target distribution. Based on these results, we provide our training algorithm in [Appendix A.3](#).

A.3 TRAINING ALGORITHM

Algorithm 1 Training

- 1: Initialize an empty replay buffer $\hat{\mathcal{D}}$, an policy \mathbf{v}_θ , a scalar parameter w , the number of rollout I and training per rollout J , and an annealing schedule $\lambda_{\text{start}} = \lambda_1 > \dots > \lambda_I = \lambda_{\text{end}}$.
 - 2: **for** $i = 1, \dots, I$ **do**
 - 3: Generate M paths $\{\mathbf{x}_{0:L}^{(m)}\}_{m=1}^M$ from the biased MD simulations with \mathbf{v}_θ at temperature λ_i .
 - 4: Update the replay buffer $\hat{\mathcal{D}} \leftarrow \hat{\mathcal{D}} \cup \{\mathbf{x}_{0:L}^{(m)}\}_{m=1}^M$.
 - 5: **for** $j = 1, \dots, J$ **do**
 - 6: Sample K data $\{\mathbf{x}_{0:L}^{(k)}\}_{k=1}^K$ from $\hat{\mathcal{D}}$.
 - 7: Update θ and w with the gradient of $\frac{1}{K} \sum_{k=1}^K \left(\log \frac{p_0(\mathbf{x}_{0:L}^{(k)}) 1_{\mathcal{B}}(\mathbf{x}_{0:L}^{(k)})}{p_{\mathbf{v}_\theta}(\mathbf{x}_{0:L}^{(k)})} - w \right)^2$
 - 8: **end for**
 - 9: **end for**
-

B EXPERIMENT DETAILS

B.1 OPENMM CONFIGURATIONS

For real-world molecules, Alanine Dipeptide, Polyproline Helix, and Chingolin, we use the VVVR integrator (Sivak et al., 2014) with the step size $\Delta t = 1$ fs and the friction term $\gamma = 1$ ps⁻¹. In the TPS-DPS training algorithm, we start simulations at a temperature $\lambda_{\text{start}} = 600\text{K}$, and end at a temperature $\lambda_{\text{end}} = 300\text{K}$. Other OpenMM configurations are shown in Table 2.

Table 2: **OpenMM configurations for three real-world molecular systems.**

Molecule	Simulation time (T)	Force field	Solvent
Alanine Dipeptide	1000	amber99sbildn (Lindorff-Larsen et al., 2010)	vaccum
Polyproline Helix	10000	amber/protein.ff14SBonlysc (Maier et al., 2015)	implicit/gbn2
Chignolin	5000	amber/protein.ff14SBonlysc	implicit/gbn2

B.2 MODEL CONFIGURATIONS

We use a 3-layer MLP for the double-well system, and a 6-layer MLP for real-world molecules with ReLU activation functions for neural bias force, potential, and scale. To constrain the output of the neural bias scale parameterization to a positive value, we apply Softplus (Zheng et al., 2015) to the MLP output. As an input to the neural network, we concatenate the current position $(\mathbf{R}_t)_i$ of the i -th atom with its distance to the target position $d_i = \|(\tilde{\mathbf{R}}_{\mathcal{B}})_i - (\mathbf{R}_t)_i\|_2$. For real-world molecules, we apply the Kabsch algorithm (Kabsch, 1976) for heavy atoms to align $\mathbf{R}_{\mathcal{B}}$ with \mathbf{R}_t . We update the parameters of the neural network with a learning rate of 0.0001, while the scalar parameter w is updated with a learning rate of 0.001. We clip the gradient norm with 1 to prevent loss from exploding. we train $J = 1000$ times per rollout. We report other model configurations in Table 3. For PIPS, we use the model configurations reported by Holdijk et al. (2024). For CVs of SMD, we use backbone dihedral angles (ϕ, ψ) for Alanine Dipeptide and RMSD for Polyproline Helix and Chignolin.

Table 3: **Model configurations of TPS-DPS.**

System	# of rollouts (I)	# of samples (M)	Batch size (K)	Buffer size	Relaxation (σ)
Double-well	20	512	512	10000	3
Alanine Dipeptide	1000	16	16	1000	0.1
Polyproline Helix	1000	16	4	200	0.2
Chignolin	1000	16	4	200	0.5

B.3 EVALUATION METRICS

Root mean square distance (RMSD). We use the Kabsch algorithm (Kabsch, 1976) for heavy atoms to align the final position with the target position $\mathbf{R}_{\mathcal{B}}$, using the optimal (proper) rotation and translation to superimpose two heavy atom positions. We calculate RMSD between heavy atoms of the final position and the target position $\mathbf{R}_{\mathcal{B}}$.

Target hit percentage (THP). THP measures the success rate of paths arriving at the target meta-stable state \mathcal{B} in a binary manner. Formally, given the final positions $\{R^{(i)}\}_{i=1}^M$ of M paths, THP is defined as follows:

$$\text{THP} = \frac{|\{i : \mathbf{R}^{(i)} \in \mathcal{B}\}|}{M} \quad (33)$$

Energy of transition state (ETS). ETS measures the ability of the method to find probable transition states when crossing the energy barrier. ETS refers to the maximum potential energy among states in a transition path. Formally, given a transition path $\mathbf{x}_{0:L}$ of length L that reaches the target meta-stable state i.e., $\mathbf{R}_L \in \mathcal{B}$, ETS is defined as follows:

$$\text{ETS}(\mathbf{x}_{0:L}) = \max_{\ell \in [0, L]} U(\mathbf{R}_{\ell}) \quad (34)$$

B.4 SYSTEM DETAILS

Double-well potential and dynamics. Double-well system follows the overdamped Langevin dynamics defined as follows:

$$d\mathbf{R}_t = \frac{-\nabla U(\mathbf{R}_t)}{m} dt + \sqrt{\frac{2\gamma k_B \lambda}{m}} d\mathbf{W}_t. \quad (35)$$

For simplicity, we let $\mathbf{R} = (x, y) \in \mathbb{R}^2$, $m = I$, $\gamma = 1$, $\Delta = 0.01$, $T = 10$, and $\lambda = 1200\text{K}$. To evaluate the ability to find diverse transition paths, we consider the following double-well potential (Hua et al., 2024):

$$U(x, y) = \frac{1}{6} (4(1 - x^2 - y^2)^2 + 2(x^2 - 2)^2 + [(x + y)^2 - 1]^2 + [(x - y)^2 - 1]^2 - 2). \quad (36)$$

This potential has global minima and two saddle points, having two meta-stable states and two reaction channels.

Handedness. For four points A, B, C, D , we define a handedness (Moradi et al., 2009) as follows:

$$H_{ABCD} = \frac{\overrightarrow{EF} \cdot (\overrightarrow{CD} \times \overrightarrow{AB})}{|\overrightarrow{EF}| \cdot |\overrightarrow{CD}| \cdot |\overrightarrow{AB}|}. \quad (37)$$

Here, E and F are the midpoints of the vectors \overrightarrow{AB} and \overrightarrow{CD} , respectively. With the backbone atoms X_1, X_2, \dots, X_N of a Polyproline Helix, we can define the handedness for N atoms as $H = \sum_{i=1}^{N-3} H_{X_i X_{i+1} X_{i+2} X_{i+3}}$. In our experiments, we take the alpha carbon X_1 , the carbonyl carbon X_2 of the first proline residue, the alpha carbon X_3 of the second residue and alpha carbon X_4 of third residues to compute the handedness.

Time-lagged independent components (TICA). To extract the collective variable (CV) for Chignolin, we consider components of time-lagged independent component analysis (TICA; Pérez-Hernández et al., 2013). We run $1\mu\text{s}$ unbiased MD simulations and get states per 1ps to collect MD trajectories, using the OpenMM library with the same Chignolin system configuration as in Appendix B.1. For the top two TICA components, we use PyEMMA library (Scherer et al., 2015) with a time lag $\tau = 500\text{ps}$.

Reproducibility. We describe experiment details in Appendix B, including detailed simulation configuration and hyper-parameters. In the anonymous link, we provide the code for TPS-DPS.