Entangled Schrödinger Bridge Matching

Sophia Tang[1], **Yinuo Zhang**[2], **Pranam Chatterjee**[1,2]

[1]Department of Computer and Information Science, University of Pennsylvania
[2]Department of Bioengineering, University of Pennsylvania

Correspondence to: sophtang@engineering.upenn.edu, pranam@upenn.edu

Abstract

Simulating trajectories of multi-particle systems on complex energy landscapes is a central task in molecular dynamics (MD) and drug discovery, but remains challenging at scale due to computationally expensive and long simulations. Previous approaches leverage techniques such as flow or Schrödinger bridge matching to implicitly learn joint trajectories from data snapshots. However, many systems, including biomolecular systems and heterogeneous cell populations, undergo *dynamic* interactions that evolve over their trajectory and cannot be captured through static snapshots. To close this gap, we introduce **Entangled Schrödinger Bridge Matching (EntangledSBM)**, a framework that learns the first- and second-order stochastic dynamics of interacting, multi-particle systems where the direction and magnitude of each particle's path depend dynamically on the paths of the other particles. We define the Entangled Schrödinger Bridge (EntangledSB) problem as solving a coupled system of bias forces that *entangle* particle velocities. We show that our framework accurately simulates heterogeneous cell populations under perturbations and rare transitions in high-dimensional biomolecular systems.

1 Introduction

Schrödinger bridge matching (SBM) has enabled progress in problems ranging from molecular dynamics simulation [Holdijk et al., 2023] to cell state modeling [Tong et al., 2024] by learning dynamic trajectories between initial and target distributions, with extensions for state-cost minimization [Liu et al., 2024] and branching dynamics [Tang et al., 2026]. Most of these frameworks treat particles as independent or capture interactions only implicitly through static snapshots, and existing approaches for interacting particle systems rely on a mean-field assumption in which particles are exchangeable and interact only through averaged effects [Liu et al., 2022, Zhang et al., 2025a]. This assumption fails for heterogeneous dynamics such as those of conformationally dynamic proteins [Vincoff et al., 2025, Chen et al., 2023], heterogeneous cell populations [Smela et al., 2023], or interacting token sequences [Geshkovski et al., 2023], where each particle's motion depends on the evolving configuration of the others.

Modeling such dependencies requires a coupled distribution over the joint n-particle phase space $\boldsymbol{X}_t = (\boldsymbol{R}_t, \boldsymbol{V}_t) \in \mathcal{X}$, transporting $\pi_{\mathcal{A}}(\boldsymbol{X}_0)$ to $\pi_{\mathcal{B}}(\boldsymbol{X}_T)$, and learning second-order dynamics in which interactions between velocity fields evolve over time – a regime that remains largely unexplored. We introduce **Entangled Schrödinger Bridge Matching (EntangledSBM)**, a framework that learns interacting second-order dynamics of n-particle systems, capturing dependencies on both positions and velocities at each step.

Our **main contributions** can be summarized in the following points:

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

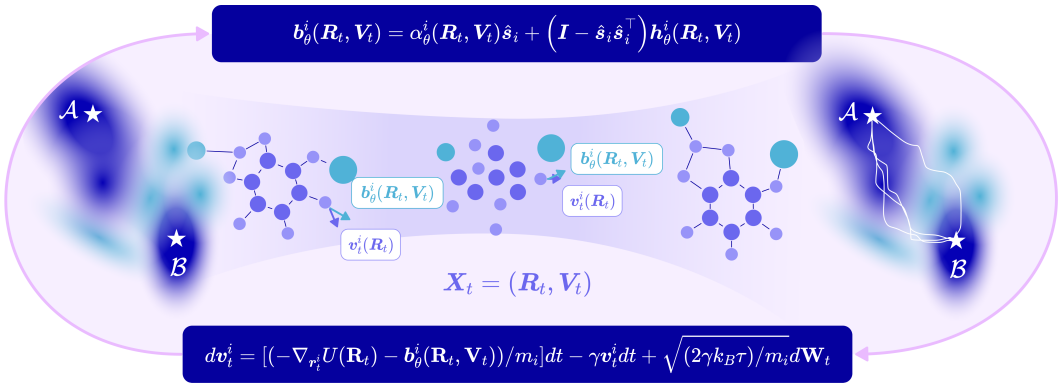

Figure 1: **Entangled Schrödinger Bridge Matching.** We consider the problem of simulating interacting multi-particle systems, where each particle's velocity depends dynamically on the velocities of the other particles in the system, and introduce **EntangledSBM**, a framework that parameterizes an entangled bias force capturing the dynamic interactions between particles.

1. **Problem Formulation:** We formulate the Entangled Schrödinger Bridge (EntangledSB) problem over the joint phase space of positions and velocities under second-order Langevin dynamics, and reduce it to a stochastic optimal control (SOC) problem with a non-factorizable control that assumes no exchangeable, mean-field interactions (Sec 3).
2. **Entangled Schrödinger Bridge Matching:** We propose an importance-weighted cross-entropy objective that is convex in the controlled path measure, fully off-policy, and defined in discrete time, so that a replay buffer can reuse each simulated rollout over many gradient steps (Sec 4.2).
3. **Target-Conditioned Parameterization:** We introduce a target-conditioned bias force with a non-negative component along the target direction and an unconstrained orthogonal correction. This makes sparse-reward transition path sampling trainable and allows conditioning on unseen targets at inference time (Sec 4.1).

We evaluate EntangledSBM on drug-perturbed cell dynamics (Sec 6.1), pancreatic β-cell differentiation (Sec 6.2), and all-atom transition path sampling (Sec 6.3), demonstrating stronger performance across baselines while modeling strictly more expressive dynamics.

2 Preliminaries

Langevin Dynamics A time-evolving n-particle molecular system can be represented as $\boldsymbol{X}_t = (\boldsymbol{R}_t, \boldsymbol{V}_t)$, where $\boldsymbol{R}_t = \{\boldsymbol{r}^i_t \in \mathbb{R}^d\}_{i=1}^n$ denotes the set of coordinates and $\boldsymbol{V}_t = \{\boldsymbol{v}^i_t \in \mathbb{R}^d\}_{i=1}^n$ denotes the set of velocities of each particle i. The evolution of the positions and velocities of the system, given a potential energy function $U : \mathcal{X} \to \mathbb{R}$, can be modeled with *Langevin dynamics*, which effectively captures the motion of particles under **conservative forces** between particles and **stochastic collisions** with the surrounding environment [Bussi and Parrinello, 2007, Yang et al., 2006] using a pair of stochastic differential equations (SDEs) defined as

$$\begin{cases} d\boldsymbol{r}^i_t &= \boldsymbol{v}^i_t dt \\ d\boldsymbol{v}^i_t &= \frac{-\nabla_{\boldsymbol{r}^i_t} U(\boldsymbol{R}_t)}{m_i} dt - \gamma \boldsymbol{v}^i_t dt + \sqrt{\frac{2\gamma k_B \tau}{m_i}} d\boldsymbol{W}^i_t \end{cases} \tag{1}$$

where m_i is the mass of each particle, γ is the friction coefficient, k_B is the Boltzmann constant, τ is the temperature, and $d\boldsymbol{W}_t$ is standard Brownian motion. In molecular dynamics (MD) simulations of biomolecules, the particles undergo *underdamped* Langevin dynamics with small γ, where *inertia* is not negligible.

Many biological systems, including cell clusters for cell-state trajectory simulation, can be modeled with *overdamped* Langevin dynamics, where inertia is negligible but the system still undergoes external forces that define its motion. This can be represented with the first-order SDE given by

$$d\boldsymbol{r}^i_t = \frac{-\nabla_{\boldsymbol{r}^i_t} U(\boldsymbol{R}_t)}{\gamma} dt + \sqrt{\frac{2k_B\tau}{\gamma}} d\boldsymbol{W}^i_t \tag{2}$$

Schrödinger Bridge Matching Tasks such as simulating cell state trajectories and transition path sampling aim to simulate the Langevin dynamics from an initial distribution to a desired target state or distribution. Given an initial distribution $\pi_{\mathcal{A}}$ and a target distribution $\pi_{\mathcal{B}}$, we define the distribution of paths $\boldsymbol{X}_{0:T} := (\boldsymbol{X}_t)_{t\in[0,T]}$ satisfying the endpoint constraints $\boldsymbol{R}_0 \sim \pi_{\mathcal{A}}$ and $\boldsymbol{R}_T \sim \pi_{\mathcal{B}}$ as the *optimal bridge distribution* $\mathbb{P}^\star(\boldsymbol{X}_{0:T})$ defined as

$$\mathbb{P}^\star(\boldsymbol{X}_{0:T}) = \frac{1}{Z(\boldsymbol{X}_0)}\mathbb{P}^0(\boldsymbol{X}_{0:T})\pi_{\mathcal{B}}(\boldsymbol{R}_T), \quad Z(\boldsymbol{X}_0) = \mathbb{E}_{\mathbb{P}^0}\left[\pi_{\mathcal{B}}(\boldsymbol{R}_T) \mid \boldsymbol{X}_0\right] \tag{3}$$

where $\mathbb{P}^0(\boldsymbol{X}_{0:T})$ is the base path distribution generated from the SDEs (1) or (2) and Z is the normalizing constant (see App D.3 for how this soft-terminal bridge relates to the standard two-marginal SB problem). Schrödinger Bridge Matching (SBM) aims to parameterize a **control** or **bias force** $\boldsymbol{b}_\theta$ that tilts the path distribution so that $\mathbb{P}^{b_\theta}(\boldsymbol{X}_{0:T})$ minimizes the KL-divergence from the bridge path distribution $\mathbb{P}^\star$ given by

$$\boldsymbol{b}^\star = \arg\min_{\boldsymbol{b}_\theta} D_{\mathrm{KL}}\left(\mathbb{P}^{b_\theta} \| \mathbb{P}^\star\right) \quad \text{s.t.} \quad \begin{cases} d\boldsymbol{r}_t^i = \boldsymbol{v}_t^i dt \\ d\boldsymbol{v}_t^i = \frac{-\nabla_{\boldsymbol{r}_t^i} U(\boldsymbol{R}_t) + \boldsymbol{b}_\theta^i(\boldsymbol{R}_t)}{m_i} dt - \gamma \boldsymbol{v}_t^i dt + \sqrt{\frac{2\gamma k_B \tau}{m_i}} d\boldsymbol{W}_t^i \end{cases} \tag{4}$$

In the case of transition path sampling (TPS) where there is a single initial state $\boldsymbol{R}_{\mathcal{A}}$ and target state $\boldsymbol{R}_{\mathcal{B}}$, we define the target distribution as the relaxed indicator function $\pi_{\mathcal{B}}(\boldsymbol{R}_T) = \mathbf{1}_{\mathcal{B}}(\boldsymbol{R}_T)$ centered around $\boldsymbol{R}_{\mathcal{B}}$.

3 Learning Interacting Multi-Particle Dynamics

The **key challenge** with simulating the dynamics of multi-particle systems lies in the question: *how can we simulate dynamic trajectories from static snapshot data?* The emergence of flow and Schrödinger bridge matching frameworks has effectively approached this problem by defining a **parameterized velocity field** that learns feasible trajectories from data snapshots. However, current strategies remain limited in their ability to model *interacting* multi-particle systems where the velocities carry inherent dependencies that cannot be captured via static snapshots of the system. Prior attempts to capture interactions between particles rely on the *mean-field assumption* [Liu et al., 2024, 2022, Zhang et al., 2025a], where each particle interacts with the others only through their averaged effect, which **does not hold for heterogeneous biomolecular systems** [Illarionov et al., 2023].

To address this challenge, we propose a framework that introduces an additional degree of freedom through a **bias force** that implicitly captures the dependencies between *both* the static positions and dynamic velocities of each particle in the system to control the joint trajectories. We leverage a Transformer architecture and treat each particle as an individual token with features that attend to the features of all other tokens. Crucially, our approach requires no handcrafted features, making it scalable to high-complexity systems.

In Sec 3.1, we formalize this problem as the **Entangled Schrödinger Bridge (EntangledSB)** problem, which aims to determine the optimal set of trajectories to a target state while capturing the dynamic interactions between particles. We then provide a tractable approach to solve the EntangledSB problem with stochastic optimal control (SOC) theory in Sec 3.2. We illustrate the high-level framework of learning **entangled** bias forces in Alg 1 and detail our specific implementation and parameterization in Sec 4.

3.1 Entangled Schrödinger Bridge Problem

Here, we formalize the **Entangled Schrödinger Bridge (EntangledSB) problem**, which aims to find a set of optimal bias forces for each particle in the system that depends *dynamically* on the positions and velocities across the entire multi-particle system to guide the Schrödinger bridge trajectory to a target distribution. Specifically, we consider an n-particle system $\boldsymbol{X}_t = (\boldsymbol{R}_t, \boldsymbol{V}_t)$ with $\boldsymbol{R}_t = \{\boldsymbol{r}_t^i \in \mathbb{R}^d\}_{i=1}^n$ and $\boldsymbol{V}_t = \{\boldsymbol{v}_t^i \in \mathbb{R}^d\}_{i=1}^n$, where each particle evolves over the time horizon $t \in [0, T]$ via the **bias-controlled** SDE given by

$$d\boldsymbol{r}_t^i = \frac{1}{\gamma}\left(-\nabla_{\boldsymbol{r}_t^i} U(\boldsymbol{R}_t) + \boldsymbol{b}^i(\boldsymbol{R}_t, \boldsymbol{V}_t)\right) dt + \sqrt{\frac{2k_B\tau}{\gamma}} d\boldsymbol{W}_t^i \tag{5}$$

Algorithm 1 Framework for **Entangled Schrödinger Bridge Matching**

Input: Parameterized bias force network $\boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) : \mathbb{R}^{n\times d} \times \mathbb{R}^{n\times d} \to \mathbb{R}^{n\times d}$, initial and target joint distributions of the system $\pi_{\mathcal{A}}, \pi_{\mathcal{B}}$, energy function $U(\boldsymbol{R}_t) : \mathbb{R}^{n\times d} \to \mathbb{R}$, objective function in path space $\mathcal{L}(\mathbb{P}^\star, \mathbb{P}^{b_\theta})$ where $\mathbb{P}^\star = \arg\min_{\mathbb{P}^{b_\theta}} \mathcal{L}(\mathbb{P}^\star, \mathbb{P}^{b_\theta})$
while not converged **do**
  `Sample` initial positions $\boldsymbol{R}_0 \sim \pi_{\mathcal{A}}$
  for $t = 0$ to $T - 1$ **do**
    `Evaluate` unconditional velocities $\boldsymbol{V}_t(\boldsymbol{R}_t)$
    `Compute` entangled bias force $\boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)$
    `Simulate` step over a discrete time step Δt to get the updated state $(\boldsymbol{R}_{t+1}, \boldsymbol{V}_{t+1})$
  end for
  `Compute` reward on terminal states $r(\boldsymbol{X}_T)$
  `Evaluate` path objective $\mathcal{L}(\mathbb{P}^\star, \mathbb{P}^{b_\theta})$ on simulated paths
  `Update` θ using the gradient $\nabla_\theta \mathcal{L}(\mathbb{P}^\star, \mathbb{P}^{b_\theta})$
end while

The velocities follow the potential energy landscape defined over the joint coordinates of the system $U(\boldsymbol{R}_t)$ with an *entangled* bias force $\boldsymbol{b}^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$ that depends dynamically on the position and velocity of *all* particles in the system. We can further extend this framework to **second-order** Langevin dynamics via the pair of SDEs

$$d\boldsymbol{r}_t^i = \boldsymbol{v}_t^i dt, \quad d\boldsymbol{v}_t^i = \frac{-\nabla_{\boldsymbol{r}_t^i} U(\boldsymbol{R}_t) + \boldsymbol{b}^i(\boldsymbol{R}_t, \boldsymbol{V}_t)}{m_i} dt - \gamma \boldsymbol{v}_t^i dt + \sqrt{\frac{2\gamma k_B \tau}{m_i}} d\boldsymbol{W}_t^i \tag{6}$$

Formally, we seek the optimal $\boldsymbol{b}^\star(\boldsymbol{R}_t, \boldsymbol{V}_t) := \{\boldsymbol{b}^i(\boldsymbol{R}_t, \boldsymbol{V}_t)\}_{i=1}^n$ that solves the EntangledSB problem defined below.

Definition 3.1 (Entangled Schrödinger Bridge Problem)**.** *Given an initial distribution $\pi_{\mathcal{A}}$, a target distribution $\pi_{\mathcal{B}}$, the EntangledSB problem aims to determine the set of optimal bias forces $\boldsymbol{b}^\star := \{\boldsymbol{b}^i(\boldsymbol{R}_t, \boldsymbol{V}_t)\}_{i=1}^n$ for each particle i in the system that solves the optimization problem*

$$\boldsymbol{b}^\star = \arg\min_{\boldsymbol{b}_\theta} D_{KL}\left(\mathbb{P}^{b_\theta} \| \mathbb{P}^\star\right) \text{ s.t. } \begin{cases} \mathbb{P}^\star = \frac{1}{Z}\mathbb{P}^0 \pi_{\mathcal{B}}(\boldsymbol{R}_T) \\ \mathbb{P}_0^\star = \pi_{\mathcal{A}}(\boldsymbol{X}_0) \end{cases}$$

where $\mathbb{P}^0$ is the base path measure with joint dynamics that follow the SDEs in (6).

3.2 Solving EntangledSB with Stochastic Optimal Control

To solve the EntangledSB problem, we leverage stochastic optimal control (SOC) theory (see App D.3 for the formal connection between Schrödinger bridges and SOC), where we aim to find the set of optimal control drifts $\boldsymbol{b}^\star(\boldsymbol{R}_t, \boldsymbol{V}_t)$ given a target distribution $\pi_{\mathcal{B}}$.

Proposition 3.1 (Solving EntangledSB with Stochastic Optimal Control)**.** *We can solve the EntangledSB problem with the stochastic optimal control (SOC) objective given by*

$$\boldsymbol{b}^\star = \arg\min_{\boldsymbol{b}_\theta} \mathbb{E}_{\boldsymbol{X}_{0:T}\sim\mathbb{P}^{b_\theta}} \left[\int_0^T \frac{1}{2}\|\boldsymbol{b}_\theta\|^2 dt - r(\boldsymbol{X}_T)\right]$$

such that (6) where $\boldsymbol{b}_\theta \equiv \boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)$ and $r(\boldsymbol{X}_T) := \log \pi_{\mathcal{B}}(\boldsymbol{R}_T)$ is the terminal reward that measures the log-probability under the target distribution.

The proof is provided in App E.1. The SOC problem is optimized over *unconstrained* trajectories. The terminal reward acts as a soft penalty rather than a hard constraint, and the terminal distribution of $\mathbb{P}^{b_\theta}$ matches that of $\mathbb{P}^\star$ only at the optimum. In Sec 4.2, we introduce an importance-weighted cross-entropy objective that shares this optimum and can be optimized off-policy.

4 Entangled Schrödinger Bridge Matching

In this section, we introduce **Entangled Schrödinger Bridge Matching** (EntangledSBM), a novel framework for learning to simulate trajectories of n-particle systems with a **bias force that dynamically depends on the positions and velocities of the other particles** in the system and can generalize to *unseen* target distributions without further training. Our parameterization ensures that a bias-only update does not increase the distance to some target state for small enough Δt, and its orthogonal component remains unconstrained (Sec 4.1). We introduce a **weighted cross-entropy** objective that enables efficient *off-policy learning* with a replay buffer of simulated trajectories (Sec 4.2).

4.1 Parameterizing the Entangled Bias Force

Bias Force Parameterization To ensure that the bias force does not increase the distance from the target distribution $\pi_{\mathcal{B}}$, we ensure that the projection of the predicted bias force onto the direction of the gradient of the target distribution is non-negative for all particles i in the system

$$\langle \boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t), \nabla_{\boldsymbol{r}_t^i} \log \pi_{\mathcal{B}} \rangle \geq 0 \tag{7}$$

Let $\boldsymbol{s}_i = \nabla_{\boldsymbol{r}_t^i} \log \pi_{\mathcal{B}} \in \mathbb{R}^d$ denote the direction towards the target state. To ensure that the bias force is within the cone surrounding $\boldsymbol{s}_i$, we use the following parameterization

$$\boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t) = \underbrace{\alpha_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)\hat{\boldsymbol{s}}_i}_{\text{parallel component}} + \underbrace{\left(\boldsymbol{I} - \hat{\boldsymbol{s}}_i \hat{\boldsymbol{s}}_i^\top\right) \boldsymbol{h}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)}_{\text{orthogonal component}} \tag{8}$$

where $\hat{\boldsymbol{s}}_i = \boldsymbol{s}_i / \|\boldsymbol{s}_i\|$ and $\alpha_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t) := \texttt{softplus}(a_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)) \geq 0$, with raw network output a_θ^i, is a scaling factor for the unit vector $\hat{\boldsymbol{s}}_i$ and $\boldsymbol{h}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t) \in \mathbb{R}^d$ is the per-atom correction vector that is projected onto the plane orthogonal to $\hat{\boldsymbol{s}}_i$, both parameterized with neural networks θ. Since the orthogonal component does not affect the dot product with $\hat{\boldsymbol{s}}_i$, the non-negativity constraint $\langle \boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t), \nabla_{\boldsymbol{r}_t^i} \log \pi_{\mathcal{B}} \rangle \geq 0$ in (7) is guaranteed by the first term. Intuitively, the orthogonal component enables the bias force to have greater flexibility in moving sideways (e.g., to avoid infeasible regions or add rotational/collective effects) while the distance from some target state stays non-increasing under the bias-only update in Proposition 4.1.

Proposition 4.1 (Non-Increasing Distance from Target Distribution)**.** *For small enough Δt, the distance from some target state $\boldsymbol{R}_{\mathcal{B}} \in \text{supp}(\pi_{\mathcal{B}})$ after an update with only the bias force $\boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$ defined in (8) is non-increasing, such that*

$$\exists \boldsymbol{R}_{\mathcal{B}} \in \text{supp}(\pi_{\mathcal{B}}) \;\; \textit{s.t.} \;\; \|\boldsymbol{R}_{t+\Delta t} - \boldsymbol{R}_{\mathcal{B}}\| \leq \|\boldsymbol{R}_t - \boldsymbol{R}_{\mathcal{B}}\|$$

where $\boldsymbol{R}_{t+\Delta t}$ collects the per-particle updates $\boldsymbol{r}_{t+\Delta t}^i = \boldsymbol{r}_t^i + (\boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)/m_i)\Delta t$, which omit the potential, friction, and noise terms.

The proof is provided in App E.2. In contrast to previous works that constrain the bias force to point strictly in the direction of a fixed target position, our approach allows greater flexibility in the direction of the orthogonal component.

Model Architecture To integrate dependencies on the positions and velocities across n particles, we leverage a Transformer-based architecture where each particle has input features $\texttt{Cat}(\boldsymbol{r}_t^i, \boldsymbol{v}_t^i) \in \mathbb{R}^{2d}$ and the full system features are $\texttt{Cat}(\boldsymbol{R}_t, \boldsymbol{V}_t) \in \mathbb{R}^{n \times 2d}$, where $\texttt{Cat}$ denotes concatenation. Each particle additionally receives its target-relative offset $\boldsymbol{r}_{\mathcal{B}}^i - \boldsymbol{r}_t^i$ and distance $\|\boldsymbol{r}_{\mathcal{B}}^i - \boldsymbol{r}_t^i\|$. Attention couples these features across all n particles, so the bias force is conditioned on the full target-relative configuration of the system, acting as a goal-conditioned correction rather than a fixed directional field. This is what allows conditioning on unseen targets at inference. The Transformer encoder enables efficient and expressive propagation of feature information across all pairs of particles in the system to generate context-aware embeddings for bias force parameterization. For MD systems, we ensure invariance of the coordinate frame using the Kabsch algorithm [Kabsch, 1976], which aligns the position $\boldsymbol{R}_t$ with the target position $\boldsymbol{R}_{\mathcal{B}}$ before input into the model. Further details are provided in App F and H.

4.2 Off-Policy Learning with Weighted Cross-Entropy

Log-Variance Objective To train the bias force to match the optimal $\boldsymbol{b}^\star$, we can adapt the *log-variance* (LV) divergence [Seong et al., 2025, Nüsken and Richter, 2021] defined as

$$\mathcal{L}_{\mathrm{LV}}(\mathbb{P}^\star, \mathbb{P}^{b_\theta}) = \mathrm{Var}_{\mathbb{P}^v}\left[\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}\right] = \mathbb{E}_{\mathbb{P}^v}\left[\left(\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}} - \mathbb{E}_{\mathbb{P}^v}\left[\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}\right]\right)^2\right] \tag{9}$$

where $\boldsymbol{v}$ is an arbitrary control that enables *off-policy learning* from trajectories that need not be generated by the current bias force $\boldsymbol{b}_\theta$. While the optimal solution to the LV objective is exactly the optimal bias force $\boldsymbol{b}^\star$, the LV objective is non-convex with respect to the path measure $\mathbb{P}^{b_\theta}$, unlike the cross-entropy objective below. We provide more details in App D.

Cross-Entropy Objective To achieve a more theoretically favorable optimization problem, we propose a cross-entropy objective that is *convex* with respect to the biased path measure $\mathbb{P}^{b_\theta}$ given by

$$\mathcal{L}_{\mathrm{CE}}(\mathbb{P}^\star, \mathbb{P}^{b_\theta}) := D_{\mathrm{KL}}(\mathbb{P}^\star \| \mathbb{P}^{b_\theta}) = \mathbb{E}_{\mathbb{P}^\star}\left[\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}\right] = \mathbb{E}_{\mathbb{P}^v}\left[\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^v} \log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}\right] \tag{10}$$

which equals the path cross-entropy $\mathbb{E}_{\mathbb{P}^\star}[-\log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0}]$ up to a constant independent of θ, where $\mathbb{P}^\star$ is the target bridge measure and $\mathbb{P}^v$ is the path measure generated with an arbitrary control $\boldsymbol{v}$. To avoid taking the expectation with respect to the unknown path measure $\mathbb{P}^\star$, we define an importance weight $w^\star(\boldsymbol{X}_{0:T}) := \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^v}(\boldsymbol{X}_{0:T})$ independent of $\boldsymbol{b}_\theta$ that enables optimization using trajectories generated from an arbitrary control $\boldsymbol{v}$. We note that similar CE objectives have been adopted in earlier work for different applications [Kappen and Ruiz, 2016, Zhu et al., 2025, Tang et al., 2025], which we discuss in App B.

Proposition 4.2 (Convexity and Uniqueness of Cross-Entropy Objective)**.** *The cross-entropy objective $\mathcal{L}_{CE}$ is convex in $\mathbb{P}^{b_\theta}$, and there exists a unique minimizer $\boldsymbol{b}^\star$ that is the solution to the EntangledSOC problem in Proposition 3.1.*

The proof is given in App E.3. To amortize the cost of simulation and reinforce high-reward trajectories, we define the arbitrary control as the biased path measure from the previous iteration $\boldsymbol{v} := \bar{\boldsymbol{b}} = \texttt{stopgrad}(\boldsymbol{b}_\theta)$, which allows us to reuse the trajectories over multiple training steps by maintaining a replay buffer $\mathcal{R}$ that contains the trajectories $\boldsymbol{X}_{0:T}$ and their importance weights.

Proposition 4.3 (Equivalence of Variational and Path Integral Objectives)**.** *The cross-entropy objective can be expressed in path-integral form as*

$$\mathcal{L}_{CE}(\theta) = \mathbb{E}_{\mathbb{P}^v}\left[w^\star(\boldsymbol{X}_{0:T})\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}(\boldsymbol{X}_{0:T})\right] \quad s.t. \quad \begin{cases} w^\star(\boldsymbol{X}_{0:T}) = \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^v}(\boldsymbol{X}_{0:T}) = \frac{e^{r(\boldsymbol{X}_T)}}{Z}\frac{\mathrm{d}\mathbb{P}^0}{\mathrm{d}\mathbb{P}^v}(\boldsymbol{X}_{0:T}) \\ \mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}(\boldsymbol{X}_{0:T}) = \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)\|^2 dt \\ -\int_0^T (\boldsymbol{b}_\theta^\top \boldsymbol{v})(\boldsymbol{R}_t, \boldsymbol{V}_t)dt - \int_0^T \boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)^\top d\boldsymbol{W}_t \end{cases}$$

where the off-policy control drift $\boldsymbol{v} = \bar{\boldsymbol{b}} := \texttt{stopgrad}(\boldsymbol{b}_\theta)$ comes from the previous iteration, and we drop the θ-independent term $w^\star(\boldsymbol{X}_{0:T})(r(\boldsymbol{X}_T) - \log Z)$.

The proof is given in App E.4. Since Z is intractable, we use the self-normalized estimate $w^\star(\boldsymbol{X}^{(j)}_{0:T}) \approx \texttt{softmax}_j\big(r(\boldsymbol{X}^{(j)}_T) + \log p^0(\boldsymbol{X}^{(j)}_{0:T}) - \log p^{\bar{b}}(\boldsymbol{X}^{(j)}_{0:T})\big)$ over the N trajectories in a batch. This estimate is consistent but biased at order $\mathcal{O}(1/N)$. The replay buffer $\mathcal{R}$ retains the highest-reward trajectories. When a batch contains a single trajectory (i.e., for large systems), the self-normalized weight is trivially 1. We therefore draw trajectories from $\mathcal{R}$ in the TPS experiments with probability proportional to $\exp(r(\boldsymbol{X}_T))$, so that the reweighting by the terminal reward happens at the sampling stage. Setting $\boldsymbol{v} = \bar{\boldsymbol{b}}$ keeps the proposal close to the current model.

Discrete Time Objective Since we want to train on off-policy trajectories from previous iterations to dynamically update the learned bias force given the velocities of the remaining particles, we require storing the simulated trajectories in a discretized form $\boldsymbol{X}_{0:K} := (\boldsymbol{X}_k)_{k\in\{0,\dots,K\}}$ with step size Δt.

To compute the term $\mathcal{F}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}$ in the CE loss, we discretize it as

$$\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:K}) = \frac{1}{2}\sum_{k=0}^{K-1}\|\boldsymbol{b}_\theta(\boldsymbol{X}_k)\|^2\Delta t - \sum_{k=0}^{K-1}(\boldsymbol{b}_\theta\cdot\bar{\boldsymbol{b}})(\boldsymbol{X}_k)\Delta t - \sum_{k=0}^{K-1}\boldsymbol{b}_\theta(\boldsymbol{X}_k)\cdot\Delta\boldsymbol{W}_k \tag{11}$$

where $\Delta\boldsymbol{W}_k = \boldsymbol{\Sigma}^{-1}[\boldsymbol{X}_{k+1} - \boldsymbol{X}_k - (\boldsymbol{f}(\boldsymbol{X}_k) + \boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_k))\Delta t]$. Now, we can establish a simplified version of the cross-entropy objective in Prop 4.3.

Proposition 4.4 (Discretized Cross-Entropy)**.** *Given the discretized $\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:K})$ in (11), we can derive a simplified loss function as*

$$\hat{\mathcal{L}}_{CE}(\theta) = \mathbb{E}_{\boldsymbol{X}_{0:K}\sim\mathbb{P}^{\bar{b}}}\left[\underbrace{\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{\bar{b}}}(\boldsymbol{X}_{0:K})}_{w^\star(\boldsymbol{X}_{0:K})}\log\left(\frac{p^0(\boldsymbol{X}_{0:K})\exp(r(\boldsymbol{X}_K))}{p^{b_\theta}(\boldsymbol{X}_{0:K})}\right)\right] \tag{12}$$

where $w^\star(\boldsymbol{X}_{0:K})$ is the importance weight of the discrete time trajectory $\boldsymbol{X}_{0:K}$.

The proof is provided in App E.5. This allows us to train by sampling trajectories, tracking their log-probabilities under the biased and base path measures, computing the terminal reward $r(\boldsymbol{X}_K)$, storing the values in the replay buffer $\mathcal{R}$, and reusing the trajectories over N_{epochs} training iterations.

5 Related Works

Schrödinger Bridges Bridge-matching methods such as DSBM [Shi et al., 2023] learn first-order dynamics by repeatedly sampling the reference process pinned at both endpoints. Closed-form pinned bridges exist for linear references, including second-order phase-space dynamics [Chen and Georgiou, 2016, Chen et al., 2024], but not for underdamped Langevin dynamics under a nonlinear force field $-\nabla U$. We therefore train on simulated trajectories with an off-policy SOC objective. Interacting-particle extensions use mean-field couplings among exchangeable particles [Liu et al., 2022, Zhang et al., 2025a], whereas we make no exchangeability assumption.

Transition Path Sampling ML-based TPS methods include PIPS [Holdijk et al., 2023] and TPS-DPS [Seong et al., 2025], which learn bias forces with SOC objectives, and Doob's Lagrangian [Du et al., 2024]. Steered MD and PIPS bias along a fixed direction, and TPS-DPS uses a rank-one bias parallel to the target direction. Our parameterization contains the latter as a special case and adds $d-1$ orthogonal degrees of freedom (Sec 4.1).

An extended discussion of related works is in App B.

6 Experiments

We evaluate **EntangledSBM** on trajectory simulation tasks involving n-particle systems with interacting dynamics: drug-perturbed cell cluster dynamics (Sec 6.1), pancreatic β-cell differentiation (Sec 6.2), and transition path sampling (TPS) for all-atom protein dynamics (Sec 6.3).

6.1 Simulating Interacting Cell Dynamics Under Perturbation

Cell populations undergo dynamic signaling interactions that shift their state under drug perturbation, a process central to drug discovery. We use **EntangledSBM** to parameterize an *entangled* bias force that learns **dynamic interactions between cells**, guiding a batch of cells toward the perturbed state. Beyond reconstructing in-distribution targets along the data manifold, EntangledSBM generates paths to *unseen* target distributions, demonstrating its potential to **scale to diverse perturbations and cell types under sparse data**.

Setup and Baselines We use two perturbations from Tahoe-100M [Zhang et al., 2026], Clonidine and Trametinib (both at $5\mu M$), reduced to the top 2000 HVGs, followed by PCA. To probe generalization to *unseen* targets, we partition the perturbed cells into disjoint clusters (two for Clonidine,

Table 1: **Results for simulating cell cluster dynamics under Clonidine and Trametinib perturbation with EntangledSBM.** Best values for each PC dimension are **bolded**. We report RBF-MMD for all PCs and $\mathcal{W}_1$ and $\mathcal{W}_2$ distances of the top 2 PCs between ground truth and reconstructed clusters after 100 simulation steps and cluster size set to $n = 16$. The mean and standard deviation of metrics from 5 independent simulations are reported. We simulate PCs $d = \{50, 100, 150\}$ for Clonidine and $d = 50$ for Trametinib to cells sampled from the training target distribution and unseen target distributions, and compare against the learned bias force with no velocity conditioning.

Clonidine Perturbation	**Trained Target Distribution**			**Unseen Target Distribution**		
Model	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)
Base Dynamics (50 PCs)	$0.677_{\pm 0.001}$	$5.947_{\pm 0.005}$	$6.015_{\pm 0.005}$	$0.784_{\pm 0.001}$	$8.217_{\pm 0.005}$	$8.384_{\pm 0.005}$
EntangledSBM w/o Velocity Conditioning						
50 PCs	$0.440_{\pm 0.000}$	$1.741_{\pm 0.003}$	$1.857_{\pm 0.004}$	$0.478_{\pm 0.000}$	$2.907_{\pm 0.006}$	$3.022_{\pm 0.006}$
100 PCs	$0.494_{\pm 0.000}$	$2.315_{\pm 0.004}$	$2.423_{\pm 0.004}$	$0.539_{\pm 0.000}$	$4.110_{\pm 0.004}$	$4.249_{\pm 0.003}$
150 PCs	$0.510_{\pm 0.000}$	$2.497_{\pm 0.006}$	$2.620_{\pm 0.006}$	$0.560_{\pm 0.000}$	$4.573_{\pm 0.006}$	$4.716_{\pm 0.007}$
EntangledSBM w/ CE						
50 PCs	$\mathbf{0.401}_{\pm 0.000}$	$\mathbf{0.342}_{\pm 0.002}$	$\mathbf{0.400}_{\pm 0.001}$	$\mathbf{0.419}_{\pm 0.000}$	$\mathbf{0.538}_{\pm 0.013}$	$\mathbf{0.705}_{\pm 0.030}$
100 PCs	$\mathbf{0.455}_{\pm 0.000}$	$\mathbf{0.953}_{\pm 0.025}$	$\mathbf{1.015}_{\pm 0.025}$	$\mathbf{0.500}_{\pm 0.001}$	$\mathbf{0.899}_{\pm 0.006}$	$\mathbf{1.055}_{\pm 0.008}$
150 PCs	$\mathbf{0.478}_{\pm 0.000}$	$\mathbf{0.753}_{\pm 0.008}$	$\mathbf{0.826}_{\pm 0.007}$	$\mathbf{0.506}_{\pm 0.000}$	$\mathbf{0.700}_{\pm 0.009}$	$\mathbf{0.811}_{\pm 0.011}$

Trametinib Perturbation	**Trained Target Distribution**			**Unseen Target Distribution 1**			**Unseen Target Distribution 2**		
Method	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)
Base Dynamics	$0.938_{\pm 0.001}$	$7.637_{\pm 0.005}$	$7.653_{\pm 0.006}$	$0.900_{\pm 0.000}$	$7.766_{\pm 0.009}$	$7.877_{\pm 0.009}$	$0.754_{\pm 0.001}$	$1.201_{\pm 0.009}$	$1.455_{\pm 0.012}$
EntangledSBM w/o Velocity Conditioning	$0.449_{\pm 0.000}$	$1.506_{\pm 0.005}$	$1.544_{\pm 0.005}$	$0.476_{\pm 0.000}$	$2.116_{\pm 0.005}$	$2.197_{\pm 0.005}$	$0.480_{\pm 0.000}$	$0.505_{\pm 0.004}$	$0.627_{\pm 0.005}$
EntangledSBM w/ CE	$\mathbf{0.428}_{\pm 0.000}$	$\mathbf{0.392}_{\pm 0.005}$	$\mathbf{0.434}_{\pm 0.006}$	$\mathbf{0.409}_{\pm 0.000}$	$\mathbf{0.453}_{\pm 0.008}$	$\mathbf{0.561}_{\pm 0.009}$	$\mathbf{0.451}_{\pm 0.000}$	$\mathbf{0.394}_{\pm 0.003}$	$\mathbf{0.469}_{\pm 0.004}$

three for Trametinib; Fig. 3), train on one, and hold out the rest. The held-out clusters are spatially disjoint from the training cluster but lie on the same gene-expression manifold. This setting therefore tests in-manifold generalization, not zero-shot generalization to unseen perturbations. The bias force is trained on batches of $n = 16$ cells with positions and velocities $\boldsymbol{X}_t = (\boldsymbol{R}_t, \boldsymbol{V}_t)$ over d PCs, where the energy landscape places low energy at high data density (App F.1). After 100 simulation steps, we report RBF-MMD across all PCs and Wasserstein distances ($\mathcal{W}_1$, $\mathcal{W}_2$) over the top two PCs. We ablate velocity conditioning by training on positions alone, and compare the cross-entropy objective $\mathcal{L}_{\text{CE}}$ against the log-variance divergence $\mathcal{L}_{\text{LV}}$ (App I, F).

Clonidine EntangledSBM accurately steers base dynamics – which otherwise stay in the initial high-density region – to the perturbed target across $d \in \{50, 100, 150\}$ (Fig. 5; Table 1), and **generalizes to unseen perturbed clusters** that diverge from the training distribution by conditioning on the target. Masking the velocity input while keeping the architecture fixed degrades reconstruction on every metric. The gap is largest on unseen targets ($\mathcal{W}_1$ $0.538 \rightarrow 2.907$), where a position-only model must extrapolate (Table 1). The two objectives also behave differently. $\mathcal{L}_{\text{LV}}$ attains lower terminal error but produces near-straight, abrupt paths, whereas $\mathcal{L}_{\text{CE}}$ traces smooth trajectories along the data manifold (Table 8; Fig. 5; App I).

Trametinib Trametinib induces three distinct perturbed distributions (Fig. 3B). Trained on only one, EntangledSBM with $\mathcal{L}_{\text{CE}}$ accurately reconstructs the remaining two *without* retraining and substantially outperforms the velocity-ablated variant (Table 1; Fig. 6). $\mathcal{L}_{\text{LV}}$ again fails to capture intermediate dynamics (Table 9; Fig. 6).

6.2 Pancreatic β-cell Differentiation

Reconstructing differentiation trajectories from sparse population snapshots tests whether a model captures the structure of cell-fate dynamics or merely interpolates observed marginals. We show that EntangledSBM, supervised only on the initial and terminal distributions, improves on mean-field and unbalanced Schrödinger bridge baselines at late time points, even though both baselines are trained on every intermediate snapshot.

Setup and Baselines We compare against CytoBridge [Zhang et al., 2025a], a mean-field Schrödinger bridge model of cell-cell interactions, and DeepRUOT [Zhang et al., 2025b], which learns unbalanced path measures through regularized unbalanced optimal transport. Both baselines are trained on every intermediate snapshot $t \in \{1, \ldots, 6\}$ following their intended protocol, and we report their published numbers. EntangledSBM receives only $t = 0$ and $t = 7$ as endpoint

Table 2: **Results for pancreatic β-cell differentiation modeling.** 1-Wasserstein distances at each observed time point (lower is better). Baselines are trained on all intermediate snapshots; EntangledSBM is trained only on $t = 0$ and $t = 7$. † indicates values taken from Zhang et al. [2025a]. **Bold** indicates the best result per column.

Model	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$
DeepRUOT†	$\mathbf{8.0447}_{\pm 0.0005}$	$8.0773_{\pm 0.0021}$	$7.6301_{\pm 0.0032}$	$\mathbf{8.0064}_{\pm 0.0042}$	$\mathbf{7.9018}_{\pm 0.0117}$	$8.3977_{\pm 0.0102}$	$7.8346_{\pm 0.0109}$
CytoBridge†	$8.0448_{\pm 0.0005}$	$\mathbf{8.0771}_{\pm 0.0021}$	$\mathbf{7.6299}_{\pm 0.0032}$	$8.0066_{\pm 0.0043}$	$\mathbf{7.9018}_{\pm 0.0117}$	$8.3974_{\pm 0.0102}$	$7.8343_{\pm 0.0109}$
EntangledSBM w/ LV	$17.9857_{\pm 0.0015}$	$16.0905_{\pm 0.0014}$	$12.7845_{\pm 0.0019}$	$11.1558_{\pm 0.0015}$	$9.0821_{\pm 0.0017}$	$8.3907_{\pm 0.0007}$	$4.3539_{\pm 0.0006}$
EntangledSBM w/ CE	$13.2059_{\pm 0.0083}$	$14.2824_{\pm 0.0140}$	$12.8030_{\pm 0.0097}$	$10.6162_{\pm 0.0130}$	$8.8382_{\pm 0.0138}$	$\mathbf{8.1769}_{\pm 0.0070}$	$\mathbf{4.2109}_{\pm 0.0052}$

distributions. Intermediate structure enters only through the base potential U, the RBF data-manifold metric of App F.1, fit on cells pooled across all time points without time labels. Since $\mathcal{L}_{\text{CE}}$ penalizes control cost relative to these base dynamics, paths through low-density regions are costly, so the learned bridge follows high-density intermediate regions without being trained on the intermediate marginals (App G).

Results EntangledSBM is less accurate than the baselines at early time points ($t \leq 5$), but it is the most accurate at $t = 6$ and $t = 7$, reducing terminal-distribution error by 46% relative to CytoBridge (4.21 vs. 7.83 in Table 2). The terminal state corresponds to mature β-cell commitment, where fine-grained cell-cell dependencies are expected to matter most. This suggests that pairwise entangled dynamics resolve late-stage interactions that mean-field couplings average out. $\mathcal{L}_{\text{CE}}$ also outperforms $\mathcal{L}_{\text{LV}}$ on 5 of the 6 intermediate marginals, consistent with $\mathcal{L}_{\text{CE}}$ learning path measures with meaningful intermediate dynamics.

6.3 Transition Path Sampling of Protein Folding Dynamics

Simulating rare transitions across high-energy barriers in molecular dynamics (MD) is challenging due to the high dimensionality of biomolecules and long timescales involved. We show that EntangledSBM generates feasible *all-atom* transition paths with higher target accuracy than baselines.

Setup and Baselines Following Seong et al. [2025], we simulate positions and velocities with OpenMM [Eastman et al., 2017] under an annealed temperature schedule and report the Kabsch-aligned RMSD over 64 paths (**RMSD**; Å) [Kabsch, 1976], target hit percentage (**THP**; %), and highest-energy transition state along successful paths (**ETS**; kJ/mol). Baselines include unbiased MD (**UMD**) with sampling temperatures at 3600K for Alanine Dipeptide and 1200K for the fast-folding proteins, steered MD (**SMD** [Schlitter et al., 1994, Izrailev et al., 1999]) at 10K and 20K, path integral path sampling (**PIPS** [Holdijk et al., 2023]), and TPS with diffusion path samplers (**TPS-DPS** [Seong et al., 2025], scale parameterization); see App H for temperatures and hyperparameters. The terminal reward is computed over all Kabsch-aligned heavy atoms, and collective variables are used only to define hits for THP.

Alanine Dipeptide For alanine dipeptide (22 atoms), we target $\pi_{\mathcal{B}} = \{\boldsymbol{R} \mid \|\xi(\boldsymbol{R}) - \xi(\boldsymbol{R}_{\mathcal{B}})\| < 0.75\}$ defined by backbone dihedrals $\xi(\boldsymbol{R}) = (\phi, \psi)$. EntangledSBM generates feasible trajectories through *both* saddle points, recovering the two reaction channels (Fig. 2A), and achieves the highest THP and lowest RMSD among all baselines (Table 3).

Fast-Folding Proteins We scale to all-atom transition paths for three fast folders: **Chignolin** (10 aa, 138 atoms), **Trpcage** (20 aa, 284 atoms), and **BBA** (28 aa, 504 atoms). A hit reaches $\pi_{\mathcal{B}}$ defined over the top two TICA components [Pérez-Hernández et al., 2013]. EntangledSBM produces *diverse* transition paths over high-energy barriers (Fig. 2), achieving the best THP across all proteins and competitive or lower RMSD (Table 3). The higher average ETS reflects the larger volume and diversity of successful paths. We verify in App J that this corresponds to mechanistically distinct folding routes rather than energy-landscape mismatch, and in App K that the resulting trajectories remain physically feasible. Given that base dynamics $\mathbb{P}^0$ rarely escape the initial metastable basin, these results show EntangledSBM effectively learns the target bridge dynamics $\mathbb{P}^\star$ even across substantial energy barriers.

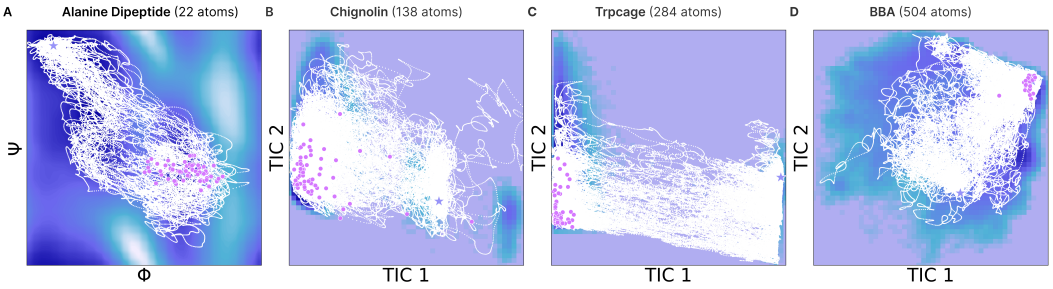

Figure 2: **Transition paths generated with EntangledSBM.** Energy landscapes (dark navy: low; light purple: high) are plotted over backbone dihedrals (ϕ, ψ) for Alanine Dipeptide and the top two TICA components for the fast-folding proteins. Purple stars mark the start $\boldsymbol{R}_0$ and target $\boldsymbol{R}_{\mathcal{B}}$ states; pink circles mark the simulated endpoints $\boldsymbol{R}_T$.

Table 3: **Transition path sampling benchmarks with EntangledSBM.** Best values are **bolded**. All metrics are averaged over 64 paths. Unless specified in brackets, paths are generated at 300K for Alanine Dipeptide and Chignolin and 400K for the others. † denotes values taken from Seong et al. [2025].

Protein	**Alanine Dipeptide**			**Chignolin**			
Method	RMSD (↓)	THP (↑)	ETS (↓)	**Method**	RMSD (↓)	THP (↑)	ETS (↓)
UMD †	$1.19_{\pm 0.32}$	6.25	$812.47_{\pm 148.80}$	UMD †	$7.23_{\pm 0.93}$	1.56	388.17
SMD (10) †	$0.86_{\pm 0.21}$	7.81	$33.15_{\pm 6.46}$	SMD (10k) †	$1.26_{\pm 0.31}$	6.25	$-527.95_{\pm 93.58}$
SMD (20) †	$0.56_{\pm 0.27}$	54.69	$78.40_{\pm 12.76}$	SMD (20k) †	$\mathbf{0.85}_{\pm 0.24}$	34.38	$179.52_{\pm 138.87}$
PIPS (Force) †	$0.66_{\pm 0.15}$	43.75	$28.17_{\pm 10.86}$	PIPS (Force) †	$4.66_{\pm 0.17}$	0.00	–
TPS-DPS (Scale Param.) †	$0.25_{\pm 0.20}$	76.00	$\mathbf{22.79}_{\pm 13.57}$	TPS-DPS (Scale Param.) †	$1.17_{\pm 0.66}$	59.38	$\mathbf{-780.18}_{\pm 216.93}$
EntangledSBM (Ours)	$\mathbf{0.18}_{\pm 0.07}$	**92.19**	$47.91_{\pm 22.76}$	**EntangledSBM (Ours)**	$0.92_{\pm 0.13}$	**64.06**	$2825.61_{\pm 318.94}$

Protein	**Trpcage**			**BBA**			
Method	RMSD (↓)	THP (↑)	ETS (↓)	**Method**	RMSD (↓)	THP (↑)	ETS (↓)
UMD †	$8.27_{\pm 1.13}$	0.00	–	UMD †	$10.81_{\pm 1.05}$	0.00	-
SMD (10k) †	$1.68_{\pm 0.23}$	3.12	$-312.54_{\pm 20.67}$	SMD (10k) †	$2.89_{\pm 0.32}$	0.00	-
SMD (20k) †	$1.20_{\pm 0.20}$	42.19	$-226.40_{\pm 85.59}$	SMD (20k) †	$1.66_{\pm 0.30}$	26.56	$-3104.95_{\pm 97.57}$
PIPS (Force) †	$7.47_{\pm 0.19}$	0.00	-	PIPS (Force) †	$9.84_{\pm 0.18}$	0.00	-
TPS-DPS (Scale Param.) †	$\mathbf{0.76}_{\pm 0.12}$	81.25	$\mathbf{-317.61}_{\pm 140.89}$	TPS-DPS (Scale Param.) †	$1.21_{\pm 0.09}$	84.38	$\mathbf{-3801.68}_{\pm 139.38}$
EntangledSBM (Ours)	$1.04_{\pm 0.22}$	**82.81**	$765.74_{\pm 155.28}$	**EntangledSBM** (Ours)	$\mathbf{0.84}_{\pm 0.08}$	**96.88**	$1453.80_{\pm 367.84}$

7 Conclusion

In this work, we present **Entangled Schrödinger Bridge Matching (EntangledSBM)**, a principled framework for learning the second-order dynamics of interacting multi-particle systems through entangled bias forces and an unconstrained cross-entropy objective. EntangledSBM captures dependencies between particle positions and velocities, enabling the modeling of complex dynamics across biological scales. For perturbation modeling, EntangledSBM reconstructs perturbed cell states while generalizing to divergent target states not seen during training, and for molecular dynamics (MD), it generates physically plausible transition paths for fast-folding proteins at an all-atom resolution.

Limitations and Future Work To conclude, we highlight some limitations and future directions of our work with an extended discussion in App C. First, our cell-state results show in-manifold generalization to held-out clusters of perturbations seen during training, and we have not yet tested new perturbations. Conditioning the bias force (e.g. on perturbation identity) is a promising route toward generalizing predictions to unseen systems (e.g. unseen drugs and cell types). Second, our protein validation is structural and trajectory-based, so we do not recover rates, committors, or free-energy surfaces. Since EntangledSBM tracks path likelihoods during simulation, future work could reweight generated paths to estimate these kinetic and thermodynamic quantities. Finally, all-particle attention costs $\mathcal{O}(n^2)$, which limits scaling to larger systems. Sparse or local attention could extend EntangledSBM to model multi-domain complexes, functional interactions between proteins and ligands, and other interacting multi-particle systems.

Declarations

Acknowledgments We thank Mark III Systems for providing database and hardware support that has contributed to the research reported within this manuscript.

Author Contributions S.T. conceived the study. S.T. devised and developed model architectures and theoretical formulations, and trained and benchmarked models. Y.Z. advised on model design and theoretical framework. S.T. drafted the manuscript and designed the figures. P.C. supervised and directed the study, and reviewed and finalized the manuscript.

Data and Materials Availability The codebase is freely accessible to the academic community at https://github.com/sophtang/EntangledSBM and https://huggingface.co/ChatterjeeLab/EntangledSBM.

Funding Statement This research was supported by NIH grant R35GM155282 to the lab of P.C.

Competing Interests P.C. is a co-founder of Gameto, Inc., UbiquiTx, Inc., AtomBioworks, Inc., and Kosha Sciences, Inc. and advises companies involved in therapeutics development. P.C.'s interests are reviewed and managed by the University of Pennsylvania in accordance with their conflict-of-interest policies. S.T. and Y.Z. have no conflicts of interest to declare.

References

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*, 36:79540–79556, 2023.

Alexander Y. Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching. *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, 238:1279–1287, 2024.

Guan-Horng Liu, Yaron Lipman, Maximilian Nickel, Brian Karrer, Evangelos A. Theodorou, and Ricky T. Q. Chen. Generalized schrödinger bridge matching. *International Conference on Learning Representations*, 2024.

Sophia Tang, Yinuo Zhang, Alexander Tong, and Pranam Chatterjee. Branched schrödinger bridge matching. In *International Conference on Learning Representations*, 2026.

Guan-Horng Liu, Tianrong Chen, Oswin So, and Evangelos A. Theodorou. Deep generalized schrödinger bridge. *Advances in Neural Information Processing Systems*, 2022.

Zhenyi Zhang, Zihan Wang, Yuhao Sun, Tiejun Li, and Peijie Zhou. Modeling cell dynamics and interactions with unbalanced mean field schrödinger bridge. In *Advances in Neural Information Processing Systems*, volume 38, 2025a.

Sophia Vincoff, Shrey Goel, Kseniia Kholina, Rishab Pulugurta, Pranay Vure, and Pranam Chatterjee. Fuson-plm: a fusion oncoprotein-specific language model via adjusted rate masking. *Nature Communications*, 16(1):1436, 2025.

Tianlai Chen, Lauren Hong, Vivian Yudistyra, Sophia Vincoff, and Pranam Chatterjee. Generative design of therapeutics that bind and modulate protein states. *Current Opinion in Biomedical Engineering*, 28:100496, 2023.

Merrick D Pierson Smela, Christian C Kramme, Patrick RJ Fortuna, Jessica L Adams, Rui Su, Edward Dong, Mutsumi Kobayashi, Garyk Brixi, Venkata Srikar Kavirayuni, Emma Tysinger, et al. Directed differentiation of human ipscs to functional ovarian granulosa-like cells via transcription factor overexpression. *Elife*, 12:e83291, 2023.

Borjan Geshkovski, Cyril Letrouit, Yury Polyanskiy, and Philippe Rigollet. The emergence of clusters in self-attention dynamics. *Advances in Neural Information Processing Systems*, 36:57026–57037, 2023.

Giovanni Bussi and Michele Parrinello. Accurate sampling using langevin dynamics. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 75(5):056707, 2007.

Sichun Yang, José N Onuchic, and Herbert Levine. Effective stochastic dynamics on a protein folding energy landscape. *The Journal of chemical physics*, 125(5), 2006.

Alexey Illarionov, Serzhan Sakipov, Leonid Pereyaslavets, Igor V Kurnikov, Ganesh Kamath, Oleg Butin, Ekaterina Voronina, Ilya Ivahnenko, Igor Leontyev, Grzegorz Nawrocki, et al. Combining force fields and neural networks for an accurate representation of chemically diverse molecular interactions. *Journal of the American Chemical Society*, 145(43):23620–23629, 2023.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Foundations of Crystallography*, 32(5):922–923, 1976.

Kiyoung Seong, Seonghyun Park, Seonghwan Kim, Woo Youn Kim, and Sungsoo Ahn. Transition path sampling with improved off-policy training of diffusion path samplers. In *The Thirteenth International Conference on Learning Representations*, 2025.

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.

Hilbert Johan Kappen and Hans Christian Ruiz. Adaptive importance sampling for control and inference. *Journal of Statistical Physics*, 162(5):1244–1266, 2016.

Yuchen Zhu, Wei Guo, Jaemoo Choi, Guan-Horng Liu, Yongxin Chen, and Molei Tao. MDNS: Masked diffusion neural sampler via stochastic optimal control. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Sophia Tang, Yuchen Zhu, Molei Tao, and Pranam Chatterjee. Tr2-d2: Tree search guided trajectory-aware fine-tuning for discrete diffusion. *arXiv preprint arXiv:2509.25171*, 2025.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. *Advances in Neural Information Processing Systems*, 2023.

Yongxin Chen and Tryphon Georgiou. Stochastic bridges of linear systems. *IEEE Transactions on Automatic Control*, 61(2):526–531, 2016. doi: 10.1109/TAC.2015.2440567.

Tianrong Chen, Jiatao Gu, Laurent Dinh, Evangelos Theodorou, Joshua M. Susskind, and Shuangfei Zhai. Generative modeling with phase stochastic bridges. In *International Conference on Learning Representations*, 2024.

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. *Advances in Neural Information Processing Systems*, 37:65791–65822, 2024.

Jesse Zhang, Airol A Ubas, Valentine Svensson, Nicole Thomas, et al. Tahoe-100m: Mapping drug-induced molecular phenotypes at single-cell resolution. *Cell*, 189(19):5945–5961.e8, 2026. doi: 10.1016/j.cell.2026.08.035.

Zhenyi Zhang, Tiejun Li, and Peijie Zhou. Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport. *International Conference on Learning Representations*, 2025b.

Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7):e1005659, 2017.

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.

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*, pages 39–65. Springer, 1999.

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.

Adrian Veres, Aubrey L Faust, Henry L Bushnell, Elise N Engquist, Jennifer Hyoje-Ryu Kenty, George Harb, Yeh-Chuin Poh, Elad Sintov, Mads Gürtler, Felicia W Pagliuca, et al. Charting cellular identity during human in vitro β-cell differentiation. *Nature*, 569(7756):368–373, 2019.

Peter G Bolhuis, David Chandler, Christoph Dellago, and Phillip L Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual review of physical chemistry*, 53 (1):291–318, 2002.

Christoph Dellago, Peter G Bolhuis, Félix S Csajka, and David Chandler. Transition path sampling and the calculation of rate constants. *The Journal of chemical physics*, 108(5):1964–1977, 1998.

Eric Vanden-Eijnden et al. Transition-path theory and path-finding algorithms for the study of rare events. *Annual review of physical chemistry*, 61:391–420, 2010.

Ferry Hooft, Alberto Perez de Alba Ortiz, and Bernd Ensing. Discovering collective variables of molecular transitions via genetic algorithms and neural networks. *Journal of chemical theory and computation*, 17(4):2294–2306, 2021.

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.

Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the national academy of sciences*, 99(20):12562–12566, 2002.

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.

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*, pages 1–49, 2015.

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.

Michele Invernizzi and Michele Parrinello. Rethinking metadynamics: from bias potentials to probability distributions. *The journal of physical chemistry letters*, 11(7):2731–2736, 2020.

Mohammad M Sultan and Vijay S Pande. Automated design of collective variables using supervised machine learning. *The Journal of chemical physics*, 149(9), 2018.

Jutta Rogal, Elia Schneider, and Mark E Tuckerman. Neural-network-based path collective variables for enhanced sampling of phase transformations. *Physical Review Letters*, 123(24):245701, 2019.

Wei Chen and Andrew L Ferguson. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *Journal of computational chemistry*, 39(25):2079–2102, 2018.

Lixin Sun, Jonathan Vandermause, Simon Batzner, Yu Xie, David Clark, Wei Chen, and Boris Kozinsky. Multitask machine learning of collective variables for enhanced sampling of rare events. *Journal of Chemical Theory and Computation*, 18(4):2341–2353, 2022.

Martin Sipka, Andreas Erlebach, and Lukas Grajciar. Constructing collective variables using invariant learned representations. *Journal of Chemical Theory and Computation*, 19(3):887–901, 2023a.

Aditya N Singh and David T Limmer. Variational deep learning of equilibrium transition path ensembles. *The Journal of Chemical Physics*, 159(2), 2023.

Jiawei Yan, Hugo Touchette, and Grant M Rotskoff. Learning nonequilibrium control forces to characterize dynamical phase transitions. *Physical Review E*, 105(2):024115, 2022.

Martin Sipka, Johannes CB Dietschreit, Lukáš Grajciar, and Rafael Gómez-Bombarelli. Differentiable simulations for enhanced sampling of rare events. In *International Conference on Machine Learning*, pages 31990–32007. PMLR, 2023b.

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.

Sanjeev Raja, Martin Šípka, Michael Psenka, Tobias Kreiman, Michal Pavelka, and Aditi S Krishnapriyan. Action-minimization meets generative modeling: Efficient transition path sampling with the onsager-machlup functional. In *International Conference on Machine Learning*, volume 267, pages 50972–51008, 2025.

Denis Blessing, Julius Berner, Lorenz Richter, Carles Domingo-Enrich, Yuanqi Du, Arash Vahdat, and Gerhard Neumann. Trust region constrained measure transport in path space for stochastic optimal control and inference. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Ophir Shalem, Neville E Sanjana, Ella Hartenian, Xi Shi, David A Scott, Tarjei S Mikkelsen, Dirk Heckl, Benjamin L Ebert, David E Root, John G Doench, et al. Genome-scale crispr-cas9 knockout screening in human cells. *Science*, 343(6166):84–87, 2014.

Christian Kramme, Alexandru M. Plesa, Helen H. Wang, Bennett Wolf, Merrick Pierson Smela, Xiaoge Guo, Richie E. Kohman, Pranam Chatterjee, and George M. Church. An integrated pipeline for mammalian genetic screening. *Cell Reports Methods*, 1(6):100082, October 2021.

Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P Fulco, Livnat Jerby-Arnon, Nemanja D Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. *cell*, 167(7): 1853–1866, 2016.

George I Gavriilidis, Vasileios Vasileiou, Aspasia Orfanou, Naveed Ishaque, and Fotis Psomopoulos. A mini-review on perturbation modelling across single-cell omic modalities. *Computational and Structural Biotechnology Journal*, 23:1886, 2024.

Mutsumi Kobayashi, Misato Kobayashi, Junko Odajima, Keiko Shioda, Young Sun Hwang, Kotaro Sasaki, Pranam Chatterjee, Christian Kramme, Richie E Kohman, George M Church, et al. Expanding homogeneous culture of human primordial germ cell-like cells maintaining germline features without serum or feeder layers. *Stem Cell Reports*, 17(3):507–521, 2022.

Merrick Pierson Smela, Christian C Kramme, Patrick RJ Fortuna, Bennett Wolf, Shrey Goel, Jessica Adams, Carl Ma, Sergiy Velychko, Ursula Widocki, Venkata Srikar Kavirayuni, et al. Rapid human oogonia-like cell specification via transcription factor-directed differentiation. *EMBO Reports*, pages 1–30, 2025.

Grace Hui Ting Yeo, Sachit D Saksena, and David K Gifford. Generative modeling of single-cell time series with prescient enables prediction of cell trajectories with interventions. *Nature communications*, 12(1):3222, 2021.

Yuhui Zhang, Yuchang Su, Chenyu Wang, Tianhong Li, Zoe Wefers, Jeffrey Nirschl, James Burgess, Daisy Ding, Alejandro Lozano, Emma Lundberg, et al. Cellflow: Simulating cellular morphology changes via flow matching. *arXiv preprint arXiv:2502.09775*, 2025c.

Martin Rohbeck, Charlotte Bunne, Edward De Brouwer, Jan-Christian Huetter, Anne Biton, Kelvin Y. Chen, Aviv Regev, and Romain Lopez. Modeling complex system dynamics with flow matching across time and conditions. *The Thirteenth International Conference on Learning Representations*, 2025.

Lazar Atanackovic, Xi Zhang, Brandon Amos, Mathieu Blanchette, Leo J Lee, Yoshua Bengio, Alexander Tong, and Kirill Neklyudov. Meta flow matching: Integrating vector fields on the wasserstein manifold. In *International Conference on Learning Representations*, 2025.

Dongyi Wang, Yuanwei Jiang, Zhenyi Zhang, Xiang Gu, Peijie Zhou, and Jian Sun. Joint velocity-growth flow matching for single-cell dynamics modeling. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Sayali Anil Alatkar and Daifeng Wang. Artemis integrates autoencoders and schrodinger bridges to predict continuous dynamics of gene expression, cell population and perturbation from time-series single-cell data. *bioRxiv*, pages 2025–01, 2025.

Kacper Kapuśniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. *Advances in Neural Information Processing Systems*, 2024.

Guan-Horng Liu, Jaemoo Choi, Yongxin Chen, Benjamin Kurt Miller, and Ricky TQ Chen. Adjoint schrödinger bridge sampler. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky TQ Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *International Conference on Learning Representations*, 2025.

Jhanvi Garg, Xianyang Zhang, and Quan Zhou. Soft-constrained schrödinger bridge: a stochastic control approach. In *International Conference on Artificial Intelligence and Statistics*, pages 4429–4437. PMLR, 2024.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

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.

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: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.

Hai Nguyen, Daniel R Roe, and Carlos Simmerling. Improved generalized born solvent model parameters for protein simulations. *Journal of chemical theory and computation*, 9(4):2020–2034, 2013.

Appendix

A Broader Impact **17**

B Related Works **17**

C Limitations and Future Directions **18**

D Extended Theoretical Preliminaries **19**

D.1 Notation 19

D.2 Stochastic Optimal Control 19

D.3 Connection Between the Schrödinger Bridge Objective and Stochastic Optimal Control 22

D.4 Justification for Entangled Bias Forces 23

E Theoretical Proofs **23**

E.1 Proof of Proposition 3.1 24

E.2 Proof of Proposition 4.1 25

E.3 Proof of Proposition 4.2 26

E.4 Proof of Proposition 4.3 26

E.5 Proof of Proposition 4.4 28

F Cell Perturbation Experiment Details **29**

F.1 Experiment Setup 29

F.2 Evaluation Metrics 31

G Pancreatic β-Cell Experiment Details **32**

G.1 Experiment Setup 32

H Transition Path Sampling Experiment Details **32**

H.1 Experiment Setup 32

H.2 Evaluation Metrics 34

H.3 Compute Time 34

I Comparison to Log-Variance Divergence **35**

J Trajectory Diversity Analysis **36**

J.1 Metrics 36

J.2 Results 37

K Physical Feasibility of Generated Trajectories **38**

K.1 Metrics 38

K.2 Results 38

L Algorithms **39**

A Broader Impact

This work introduces a computational framework for modeling interacting multi-particle dynamics with applications to protein folding and cell state perturbations. The cell perturbation experiments (Section 6.1) use the publicly available Tahoe-100M [Zhang et al., 2026] and pancreatic β-cell [Veres et al., 2019] datasets, cited with attribution and used under their original licenses. All biomolecular simulations (Section 6.3) were performed on standard test proteins (Alanine Dipeptide, Chignolin, Trpcage, BBA) using publicly available force fields and simulation protocols, with no human or animal subjects involved. No proprietary or sensitive data were employed. The work does not provide experimental protocols or actionable therapeutic targets, and is restricted to methodological contributions in generative modeling. While such frameworks may be broadly useful for molecular or cellular design, the immediate scope of EntangledSBM is fundamental methodology, and the dual-use risk is minimal.

B Related Works

Off-Policy Learning Importance-weighted cross-entropy (CE) objectives have previously been used to sample the target distribution $g^\star \propto h f_u$ by aligning a proposal distribution f_v to the intractable target with $\mathrm{KL}(g^\star \| f_v)$ [Kappen and Ruiz, 2016]. The main difference is that our approach aims to match a **controlled path measure** $\mathbb{P}^{b_\theta}$ with an energy-minimizing path measure $\mathbb{P}^\star$ rather than a static target distribution. Furthermore, we adapt our path CE objective to apply to discrete-time paths, similar to the approach taken by Seong et al. [2025], which discretizes the log-variance (LV) divergence [Nüsken and Richter, 2021]. While the LV objective has theoretical optimality guarantees, it is non-convex in $\mathbb{P}^{b_\theta}$. Since the CE objective is convex in $\mathbb{P}^{b_\theta}$, it yields an ideal optimization landscape for θ and demonstrates improved performance as shown in Sec 6.1. We also highlight that our optimization scheme is a form of *off-policy* learning, as we minimize the objective with respect to the path distribution generated from an auxiliary control $\boldsymbol{v}$, which we set to the non-gradient-tracking bias force $\boldsymbol{v} := \bar{\boldsymbol{b}} = \texttt{stopgrad}(\boldsymbol{b}_\theta)$. This objective is inspired by recent work that leverages off-policy reinforcement learning in the discrete state space for discrete diffusion sampling [Zhu et al., 2025] and fine-tuning [Tang et al., 2025].

Transition Path Sampling To overcome the challenge of generating feasible transition paths across high energy barriers for transition path sampling, several non-ML and ML-based approaches have been introduced [Bolhuis et al., 2002, Dellago et al., 1998, Vanden-Eijnden et al., 2010]. Non-ML approaches often rely on collective variables (CVs), which reduce the dimensionality of the molecular conformation to a function of coordinates that are known to be involved in transition states [Hooft et al., 2021]. These include steered MD [Schlitter et al., 1994, Izrailev et al., 1999], umbrella sampling [Torrie and Valleau, 1977, Kästner, 2011], meta-dynamics [Laio and Parrinello, 2002, Ensing et al., 2006, Branduardi et al., 2012, Bussi and Branduardi, 2015], adaptive biasing force [Comer et al., 2015], and on-the-fly probability-enhanced sampling [Invernizzi and Parrinello, 2020]. To overcome the lack of well-understood CVs for large biomolecular systems, ML-based methods have been explored for determining or constructing CVs [Sultan and Pande, 2018, Rogal et al., 2019, Chen and Ferguson, 2018, Sun et al., 2022, Sipka et al., 2023a], but modeling transition paths at all-atom resolution is still desirable due to systems of increased complexity, and CVs are uncertain.

Several ML-based approaches for simulating transition paths have been developed [Singh and Limmer, 2023, Yan et al., 2022, Sipka et al., 2023b, Das et al., 2021]. Notably, Path Integral Path Sampling (PIPS; Holdijk et al. [2023]) learns a bias force with stochastic optimal control theory, Doob's Lagrangian [Du et al., 2024] defines an optimal transition path with a Lagrangian objective that solves the Doob's h-transform, Transition Path Sampling with Diffusion Path Samplers (TPS-DPS; Seong et al. [2025]) trains a bias force with an off-policy log-variance objective, and Action Minimization with Onsager-Machlup (OM) Functional [Raja et al., 2025] generates a discrete interpolation between endpoints by minimizing the OM functional. Furthermore, Blessing et al. [2025] introduces a trust-region-constrained SOC optimization algorithm, which is applied to TPS. While these approaches demonstrate the promise of ML-based TPS, their applicability remains limited to small-scale biomolecular systems (Alanine Dipeptide, tetrapeptides, etc.) for all-atom simulation [Holdijk et al., 2023, Du et al., 2024] or require coarse-grained representations when scaling up to larger fast-folding proteins [Raja et al., 2025]. We directly compare against PIPS [Holdijk et al., 2023] and TPS-DPS [Seong et al., 2025] and achieve the highest target hit percentage

on all-atom simulations of larger fast-folding proteins. Furthermore, we highlight that our framework extends beyond molecular dynamics (MD) simulations, with applications to single-cell simulation and generalization to unseen target distributions.

Modeling Cell Dynamics Under Perturbation Predicting the dynamics of heterogeneous cell populations under perturbations such as treatment with a drug candidate, gene editing, and knockouts, or protein expression, has critical applications in drug discovery and predictive medicine [Shalem et al., 2014, Kramme et al., 2021, Dixit et al., 2016, Gavriilidis et al., 2024, Zhang et al., 2026, Kobayashi et al., 2022, Smela et al., 2023, Pierson Smela et al., 2025, Yeo et al., 2021]. Generative modeling frameworks, including flow matching [Zhang et al., 2025c, Rohbeck et al., 2025, Atanackovic et al., 2025, Tong et al., 2024, Wang et al., 2025], Schrödinger Bridge Matching [Alatkar and Wang, 2025, Tong et al., 2024, Kapuśniak et al., 2024, Tang et al., 2026], and stochastic optimal control [Zhang et al., 2025b,a], have enabled significant advances in parameterizing velocities that evolve dynamically over time on non-linear energy manifolds by learning on static snapshots across the trajectory. However, these models often require *explicit training on a specific target distribution* and fail to generalize to unseen target distributions at inference, especially if they diverge from those seen in the training data, which limits their scalability to cell types and perturbations that are not seen during training. Furthermore, many of these approaches simulate cells as populations of particles that evolve independently from each other, which fails to capture the diverse intercellular signaling and interactions that occur naturally and under perturbations. Although there have been some early developments in learning these interactions from data [Zhang et al., 2025a], they operate under the mean-field assumption, where each cell takes the average effect of all surrounding cells rather than simulating individual pairwise interactions. This assumption does not account for the complex interactions between cells and limits its applicability to heterogeneous cell populations.

C Limitations and Future Directions

While our results demonstrate the effectiveness of EntangledSBM across both cellular and molecular regimes, several limitations should be noted to properly contextualize the scope of our claims.

First, the cell-state results reported here characterize in-manifold generalization: the model is evaluated on held-out cells drawn from the same perturbation distributions seen during training, rather than on entirely unseen perturbations. The bridge formulation can therefore be expected to interpolate trajectories within a perturbation manifold it has already observed, but our experiments do not establish that the model extrapolates to novel chemical or genetic perturbations whose endpoint distributions lie outside the training support. Extending EntangledSBM to a true zero-shot perturbation setting would require conditioning on perturbation identity in a way that supports compositional or out-of-distribution generalization, which we leave to future work.

Second, our protein-system validation is primarily structural and trajectory-based. We assess the physical plausibility of generated transitions (e.g., energy of transition state, smoothness, etc.) but do not conduct full kinetic or thermodynamic validation. We do not recover transition rates, committor probabilities, or free-energy surfaces at quantitative accuracy comparable to enhanced-sampling molecular dynamics, both because such validation lies outside the scope of the bridge formulation and because it would require substantially greater computational resources. The trajectories should therefore be interpreted as physically plausible interpolants rather than calibrated samples from the underlying Boltzmann transition density.

Third, the all-particle attention mechanism that underlies the entanglement step incurs $O(n^2)$ time and memory cost in the number of particles. We demonstrate that this remains tractable for systems of $500+$ atoms, which covers the protein and cellular regimes considered here, but the quadratic scaling presents a clear barrier to substantially larger systems such as full multi-domain complexes, membrane assemblies, or dense tissue-scale simulations. Scaling further will likely require sparse, hierarchical, or locality-aware attention variants – for example, neighborhood-restricted attention, low-rank approximations, or message-passing schemes that exploit the geometric locality of physical interactions – and integrating such architectures into the bridge framework without disrupting its probabilistic guarantees is left for future work.

D Extended Theoretical Preliminaries

In this section, we provide relevant theoretical background in stochastic optimal control (SOC) theory for path measures. We use the terms *bias force* and *optimal control* interchangeably, denoted by $\boldsymbol{b}$, which is less common in general SOC literature where $\boldsymbol{u}$ is often used. For a more comprehensive analysis of SOC theory, we refer the reader to Nüsken and Richter [2021].

D.1 Notation

In this work, we consider an n-particle system $\boldsymbol{X}_t = (\boldsymbol{R}_t, \boldsymbol{V}_t)$ where $\boldsymbol{R}_t \in \mathbb{R}^{n\times d}$ and $\boldsymbol{V}_t \in \mathbb{R}^{n\times d}$ denote the positions and velocities of the full system that evolves over the time horizon $t \in [0, T]$. The notation $\boldsymbol{X}$ indicates a random variable and $\boldsymbol{x}$ denotes a deterministic realization. Each particle i is defined by its position $\boldsymbol{r}_t^i \in \mathbb{R}^d$ and velocity $\boldsymbol{v}_t^i \in \mathbb{R}^d$ which lie in d-dimensional Euclidean space. The base dynamics evolve according to a potential energy function denoted $U(\boldsymbol{R}_t) : \mathbb{R}^{n\times d} \to \mathbb{R}$ and the biased dynamics evolve with an additional parameterized bias force $\boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) : \mathbb{R}^{n\times d} \times \mathbb{R}^{n\times d} \to \mathbb{R}^{n\times d}$ that captures the implicit dependencies within the system and controls the base dynamics towards a target distribution $\pi_{\mathcal{B}}(\boldsymbol{R}_T)$ from a state in the initial distribution $\boldsymbol{R}_{\mathcal{A}} \in \pi_{\mathcal{A}}(\boldsymbol{R}_0)$. The corresponding path measures over trajectories $\boldsymbol{X}_{0:T} := (\boldsymbol{X}_t)_{t\in[0,T]}$ are denoted $\mathbb{P}^0$ for the base process and $\mathbb{P}^{b_\theta}$ for the biased process. The optimal path measure that solves the EntangledSB problem is denoted $\mathbb{P}^\star \equiv \mathbb{P}^{b^\star}$ with bias force $\boldsymbol{b}^\star(\boldsymbol{R}_t, \boldsymbol{V}_t)$.

In the SOC derivations, we write the controlled SDE as in (13), where the control enters the drift as $\boldsymbol{\Sigma b}$, so that the running cost is $\frac{1}{2}\|\boldsymbol{b}\|^2$. The physical bias force maps to this form after rescaling by $\boldsymbol{\Sigma}_V^{-1}\boldsymbol{M}^{-1}$ in the second-order dynamics (6), with $\boldsymbol{\Sigma}_V = \operatorname{diag}\big(\sqrt{2\gamma k_B\tau/m_i}\big)$, and by $(\gamma\boldsymbol{\Sigma})^{-1}$ in the first-order dynamics, with $\boldsymbol{\Sigma} = \sqrt{2k_B\tau/\gamma}\,\boldsymbol{I}$.

D.2 Stochastic Optimal Control

Controlled Path Measures First, we consider the controlled path measure $\mathbb{P}^b$ defined by the SDE:

$$d\boldsymbol{X}_t = [\boldsymbol{f}(\boldsymbol{X}_t, t) + \boldsymbol{\Sigma b}(\boldsymbol{X}_t, t)]dt + \boldsymbol{\Sigma} d\boldsymbol{W}_t \tag{13}$$

where $\boldsymbol{b} : [0, T] \times \mathcal{X} \to \mathcal{X} \in \mathcal{U}$ is known as the **control**[1] that tilts the path measure from the unconditional dynamics $\mathbb{P}^0$. Now, we will define the **Radon-Nikodym derivative (RND)** of the path measures corresponding to the controlled and unconditional SDEs, which is crucial for defining our objective.

Lemma D.1 (Radon-Nikodym Derivative)**.** *Given a pair of SDEs with and without a control drift $\boldsymbol{b}(\boldsymbol{X}_t, t)$ on $t \in [0, T]$ with the same diffusion $\boldsymbol{\Sigma}$ defined as*

$$\mathbb{P}^0 : d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{\Sigma} d\boldsymbol{W}_t, \qquad \boldsymbol{X}_0 = \boldsymbol{x}_0 \tag{14}$$

$$\mathbb{P}^b : d\boldsymbol{X}_t^b = [\boldsymbol{f}(\boldsymbol{X}_t, t) + \boldsymbol{\Sigma b}(\boldsymbol{X}_t, t)]\, dt + \boldsymbol{\Sigma} d\boldsymbol{W}_t, \qquad \boldsymbol{X}_0^b = \boldsymbol{x}_0 \tag{15}$$

Then, the Radon-Nikodym derivative satisfies

$$\log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = \int_0^T (\boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1})(\boldsymbol{X}_t, t) d\boldsymbol{X}_t - \int_0^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{f} \cdot \boldsymbol{b})(\boldsymbol{X}_t, t)dt - \frac{1}{2}\int_0^T \|\boldsymbol{b}(\boldsymbol{X}_t, t)\|^2 dt$$

or equivalently:

$$\log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = \int_0^T \boldsymbol{b}(\boldsymbol{X}_t, t)^\top d\boldsymbol{W}_t - \frac{1}{2}\int_0^T \|\boldsymbol{b}(\boldsymbol{X}_t, t)\|^2 dt \tag{16}$$

Proof. First, we use Girsanov's theorem to get the Radon-Nikodym derivative with respect to the zero-drift reference path measure $\mathbb{Q}$ defined by the SDE $d\boldsymbol{X}_t = \boldsymbol{\Sigma}(\boldsymbol{X}_t, t)d\boldsymbol{W}_t$, we have

$$\log \frac{\mathrm{d}\mathbb{P}^0}{\mathrm{d}\mathbb{Q}} = \int_0^T \left(\boldsymbol{f}(\boldsymbol{X}_t, t) \cdot \boldsymbol{\Sigma}^{-2}(\boldsymbol{X}_t, t)\right) d\boldsymbol{X}_t - \frac{1}{2}\int_0^T \left(\boldsymbol{f} \cdot \boldsymbol{\Sigma}^{-2}\boldsymbol{f}\right)(\boldsymbol{X}_t, t)dt \tag{17}$$

[1] $\mathcal{U}$ is the set of admissible controls continuously differentiable in $(\boldsymbol{x}, t)$ with bounded linear growth in $\boldsymbol{x}$

and for the controlled path measure $\mathbb{P}^b$, we have

$$\log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{Q}}(\boldsymbol{X}_{0:T}) = \int_0^T (\boldsymbol{f} + \boldsymbol{\Sigma}\boldsymbol{b})(\boldsymbol{X}_t, t) \cdot \boldsymbol{\Sigma}^{-2}(\boldsymbol{X}_t, t) d\boldsymbol{X}_t - \frac{1}{2}\int_0^T \left((\boldsymbol{f} + \boldsymbol{\Sigma}\boldsymbol{b}) \cdot \boldsymbol{\Sigma}^{-2}(\boldsymbol{f} + \boldsymbol{\Sigma}\boldsymbol{b})\right)(\boldsymbol{X}_t, t) dt \tag{18}$$

Then, using the identity

$$\log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{P}^0} = \log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{Q}} \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}^0} = \log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{Q}} + \log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}^0} = \log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{Q}} - \log \frac{\mathrm{d}\mathbb{P}^0}{\mathrm{d}\mathbb{Q}} \tag{19}$$

Substituting in (17) and (18) and canceling terms, we get

$$\log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = \int_0^T (\boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1})(\boldsymbol{X}_t, t) d\boldsymbol{X}_t - \int_0^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{f} \cdot \boldsymbol{b})(\boldsymbol{X}_t, t) dt - \frac{1}{2}\int_0^T \|\boldsymbol{b}(\boldsymbol{X}_t, t)\|^2 dt \tag{20}$$

Equivalently, since we can write $d\boldsymbol{W}_t = \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_t, t)(d\boldsymbol{X}_t - \boldsymbol{f}(\boldsymbol{X}_t, t)dt)$ (the Brownian motion under $\mathbb{P}^0$), we can write

$$\log \frac{\mathrm{d}\mathbb{P}^b}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = \int_0^T \boldsymbol{b}(\boldsymbol{X}_t, t)^\top d\boldsymbol{W}_t - \frac{1}{2}\int_0^T \|\boldsymbol{b}(\boldsymbol{X}_t, t)\|^2 dt \tag{21}$$

which concludes the proof. □

Remark D.1. *In the under-damped Langevin model, the diffusion acts only on the velocity coordinates, so the full diffusion matrix on $\boldsymbol{X}_t = (\boldsymbol{R}_t, \boldsymbol{V}_t)$ is degenerate, i.e., $\boldsymbol{\Sigma} = \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{\Sigma}_V \end{bmatrix}$. We therefore apply Girsanov's theorem to the velocity SDE only, where the diffusion $\boldsymbol{\Sigma}_V$ is invertible, and treat positions as deterministic integrals of velocity (i.e., $\boldsymbol{R}_t = \boldsymbol{R}_0 + \int_0^t \boldsymbol{V}_s ds$). Therefore, Girsanov's theorem is first defined on velocity paths as*

$$\log \frac{\mathrm{d}\mathbb{P}_V^b}{\mathrm{d}\mathbb{P}_V^0}(\boldsymbol{X}_{0:T}) = \int_0^T \boldsymbol{b}(\boldsymbol{X}_t, t)^\top d\boldsymbol{W}_t - \frac{1}{2}\int_0^T \|\boldsymbol{b}(\boldsymbol{X}_t, t)\|^2 dt \tag{22}$$

and pushed forward to the full path measure via $(\boldsymbol{R}_t, \boldsymbol{V}_t) = (\boldsymbol{R}_0 + \int_0^t \boldsymbol{V}_s ds, \boldsymbol{V}_t)$. For simplicity in notation, we define $\boldsymbol{\Sigma}^{-1} \equiv \boldsymbol{\Sigma}_V^{-1}$ in the theoretical proofs.

Stochastic Optimal Control The reward-guided **stochastic optimal control (SOC)** problem aims to determine the **optimal control** $\boldsymbol{b}^\star = \arg\min_{\boldsymbol{b}\in\mathcal{U}} J(\boldsymbol{b}; \boldsymbol{x}, t)$ that minimizes the **cost functional** $J(\boldsymbol{b}_\theta; \boldsymbol{x}, t)$ with reward $r : \mathcal{X} \to \mathbb{R}$ given by

$$J(\boldsymbol{b}; \boldsymbol{x}, t) = \mathbb{E}_{p_t}\left[\int_t^T \left(\boldsymbol{f}(\boldsymbol{X}_s, s) + \frac{1}{2}\|\boldsymbol{b}(\boldsymbol{X}_s, s)\|^2\right) ds - r(\boldsymbol{X}_T) \middle| \boldsymbol{X}_t = \boldsymbol{x}\right] \tag{23}$$

which measures the *cost-to-go* from the current state $\boldsymbol{X}_t = \boldsymbol{x}$ at time t under the controlled dynamics defined in (13) to the terminal state $\boldsymbol{X}_T$. The infimum of the cost functional is defined as

$$J(\boldsymbol{b}^\star; \boldsymbol{x}, t) = \inf_{\boldsymbol{b}\in\mathcal{U}} J(\boldsymbol{b}; \boldsymbol{x}, t) =: V_t(\boldsymbol{x}) \tag{24}$$

which is also referred to as the *value function* $V_t(\boldsymbol{x})$. To derive the path integral representation of the optimal control $\boldsymbol{b}^\star$, we introduce the *work functional* which computes the running cost subtracted from the terminal reward of a state path $\boldsymbol{X}_{0:T}$,

$$\mathcal{W}(\boldsymbol{X}_{t:T}, t) := \int_t^T \boldsymbol{f}(\boldsymbol{X}_s, s) ds - r(\boldsymbol{X}_T) \tag{25}$$

We can now define a series of statements that connect the optimal drift $\boldsymbol{b}^\star$ and the value function V_t.

Theorem D.1. *Let V_t be the value function and $\boldsymbol{b}^\star$ be the optimal control. Then, the following are true:*

(a) The optimal control satisfies $\boldsymbol{b}^\star(\boldsymbol{x}) = -\boldsymbol{\Sigma}^\top \nabla_{\boldsymbol{x}} V_t(\boldsymbol{x})$

(b) The Radon-Nikodym derivative of the optimal path measure $\mathbb{P}^\star$ can be defined by taking state paths from the unconditional path measure $\mathbb{P}^0$ and reweighting them with the output of the work functional $\mathcal{W}(\boldsymbol{X}_{0:T}, 0)$.

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = \frac{1}{Z} e^{-\mathcal{W}(\boldsymbol{X}_{0:T},0)}, \quad Z = \mathbb{E}_{\boldsymbol{X}_{0:T}\sim\mathbb{P}^0}\left[e^{-\mathcal{W}(\boldsymbol{X}_{0:T},0)}\right] \tag{26}$$

(c) For any $(\boldsymbol{x}, t) \in \mathcal{X} \times [0, T]$, the value function can be written in path integral form by taking the expectation over the uncontrolled path measure $\mathbb{P}^0$ as

$$V_t(\boldsymbol{x}) = -\log \mathbb{E}\left[e^{-\mathcal{W}(\boldsymbol{X}_{t:T},t)} \middle| \boldsymbol{X}_t = \boldsymbol{x}\right] \tag{27}$$

(d) Combining (a) and (c), we can write the optimal control as

$$\boldsymbol{b}^\star(\boldsymbol{x}) = \boldsymbol{\Sigma}^\top \nabla_{\boldsymbol{x}} \log \mathbb{E}_{\boldsymbol{X}_{t:T}\sim\mathbb{P}^0}\left[e^{-\mathcal{W}(\boldsymbol{X}_{t:T},t)} \middle| \boldsymbol{X}_t = \boldsymbol{x}\right] \tag{28}$$

While Theorem D.1 provides a closed-form solution to the optimal control $\boldsymbol{b}^\star$, it remains impractical to compute as $\mathbb{P}^0$ contains an infinite number of paths from any point $\boldsymbol{x}$. This problem motivates parameterizing $\boldsymbol{b}^\star$ with a neural network $\boldsymbol{b}_\theta$. The natural objective for obtaining an accurate approximation of $\boldsymbol{b}^\star$ that induces the path measure $\mathbb{P}^\star$ is the KL-divergence $\mathcal{L}(\mathbb{P}^\star, \mathbb{P}^{b_\theta}) := D_{\mathrm{KL}}(\mathbb{P}^{b_\theta} \| \mathbb{P}^\star)$ as its minimizer is exactly $\mathbb{P}^\star$.

However, taking the gradient $\nabla_\theta D_{\mathrm{KL}}(\mathbb{P}^{b_\theta} \| \mathbb{P}^\star)$ requires differentiating through the full stochastic trajectories $\boldsymbol{X}_{0:T} \sim \mathbb{P}^{b_\theta}$ generated by the Euler-Maruyama SDE solver due to the expectation over $\mathbb{P}^{b_\theta}$, resulting in a significant computational bottleneck. To overcome this bottleneck, previous work [Nüsken and Richter, 2021, Seong et al., 2025] leverages alternative path-measure objectives that do not involve expectations over $\mathbb{P}^{b_\theta}$ while preserving optimality guarantees. We focus on two objectives: the **log-variance divergence** used in Seong et al. [2025] and the **cross-entropy objective**, which we formalize in the present work for solving the EntangledSB problem.

Log-Variance Objective Here, we describe the **log-variance (LV) divergence**, which is introduced in Nüsken and Richter [2021] and applied to transition path sampling (TPS) in TPS-DPS [Seong et al., 2025]. Given the path measure $\mathbb{P}^{b_\theta}$ generated with the parameterized bias force $\boldsymbol{b}_\theta$ and the target path measure $\mathbb{P}^\star$, the log-variance divergence $\mathcal{L}_{\mathrm{LV}}$ is defined as

$$\mathcal{L}_{\mathrm{LV}}(\mathbb{P}^\star, \mathbb{P}^{b_\theta}) := \mathrm{Var}_{\mathbb{P}^v}\left[\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}\right] = \mathbb{E}_{\mathbb{P}^v}\left[\left(\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}} - \mathbb{E}_{\mathbb{P}^v}\left[\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}\right]\right)^2\right] \tag{29}$$

where $\mathbb{P}^v$ can be defined as any arbitrary path measure generated from a control $\boldsymbol{v}$. To allow simulated trajectories to be used across multiple training iterations, it is common to define $\boldsymbol{v}$ as the bias force with frozen parameters $\boldsymbol{v} = \bar{\boldsymbol{b}} := \texttt{stopgrad}(\boldsymbol{b}_\theta)$. Using Lemma D.1, we define $\mathcal{L}_{\mathrm{LV}}(\mathbb{P}^\star, \mathbb{P}^{b_\theta}) = \mathbb{E}_{\mathbb{P}^v}[(\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}} - \mathbb{E}_{\mathbb{P}^v}[\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}])^2]$, where $\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}$ is defined as

$$\begin{aligned}\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}} &:= \log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}} = \log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^0}\frac{\mathrm{d}\mathbb{P}^0}{\mathrm{d}\mathbb{P}^{b_\theta}} \\ &= r(\boldsymbol{X}_T) - \log Z + \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt - \int_0^T (\boldsymbol{b}_\theta^\top \boldsymbol{v})(\boldsymbol{X}_t) dt - \int_0^T \boldsymbol{b}_\theta^\top d\boldsymbol{W}_t\end{aligned} \tag{30}$$

using similar justification as in App E.4. Since the expectation $\mathbb{E}_{\mathbb{P}^v}[\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}]$ is computationally intractable, Seong et al. [2025] introduces a scalar parameter w that is jointly optimized with θ such that $\arg\min_w \mathcal{L}_{\mathrm{LV}}(\theta, w) = \mathbb{E}_{\mathbb{P}^v}[\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}]$. Substituting in w for $\mathbb{E}_{\mathbb{P}^v}[\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}]$ and taking the discretization over K steps, the LV objective becomes

$$\mathcal{L}_{\mathrm{LV}}(\theta, w) = \mathbb{E}_{\boldsymbol{X}_{0:K}\sim\mathbb{P}^v}\left[(\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}(\boldsymbol{X}_{0:K}) - w)^2\right] = \mathbb{E}_{\boldsymbol{x}_{0:K}\sim\mathbb{P}^v}\left[\left(\log \frac{p^0(\boldsymbol{x}_{0:K})\exp(r(\boldsymbol{X}_K))}{p^{b_\theta}(\boldsymbol{x}_{0:K})} - w\right)^2\right]$$

where $\boldsymbol{v} = \bar{\boldsymbol{b}} := \texttt{stopgrad}(\boldsymbol{b}_\theta)$.

Regression Form and Why We Do Not Use It From Prop. 4.3, $\mathcal{L}_{\mathrm{CE}}$ differs only by a θ-independent term from the regression objective

$$\mathcal{L}_{\mathrm{L2}}(\theta) = \mathbb{E}_{\mathbb{P}^v}\left[w^\star(\boldsymbol{X}_{0:T})\frac{1}{2}\int_0^T \left\|\boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) - \boldsymbol{b}^\star(\boldsymbol{R}_t, \boldsymbol{V}_t)\right\|^2 dt\right] \tag{31}$$

which typically has lower variance. It is unusable here because $\boldsymbol{b}^\star = \boldsymbol{\Sigma}^\top \nabla \log h(\boldsymbol{x}, t)$, where h solves the backward Kolmogorov equation of the underdamped reference with terminal condition $\pi_{\mathcal{B}}$. Under a nonlinear force field, computing h is as hard as the original problem. In first-order settings with Brownian or Ornstein-Uhlenbeck references, the analogous target is a closed-form Gaussian bridge drift, which is what makes the L2 form attractive there. $\mathcal{L}_{\mathrm{CE}}$ needs no pointwise regression target and evaluates r only at the endpoint. Our variance-reduced baseline is instead the log-variance divergence (App I).

D.3 Connection Between the Schrödinger Bridge Objective and Stochastic Optimal Control

The Schrödinger Bridge (SB) problem stated in the main text is not written in its most standard form, but is an equivalent reformulation that is better suited to our setting. Similar formulations have appeared in prior work [Holdijk et al., 2023]. In this section, we make the connection between the standard SB problem, our reformulation, and the stochastic optimal control (SOC) objective explicit.

Standard SB Problem to a Tilted Path Measure The standard SB problem is defined as

$$\min_{\mathbb{P}} D_{\mathrm{KL}}\left(\mathbb{P} \| \mathbb{P}^0\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \pi_{\mathcal{A}}, \quad \mathbb{P}_T = \pi_{\mathcal{B}} \tag{32}$$

which seeks the path measure closest in KL to the reference $\mathbb{P}^0$ subject to *both* marginal constraints. Its solution has the form $\frac{\mathrm{d}\mathbb{P}^{\mathrm{SB}}}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = f(\boldsymbol{X}_0)\, g(\boldsymbol{X}_T)$, where the Schrödinger potentials (f, g) solve a coupled system that is intractable for nonlinear references. Following Holdijk et al. [2023] and the SOC view of Liu et al. [2025], we instead fix $\boldsymbol{X}_0 \sim \pi_{\mathcal{A}}$ and replace the hard terminal constraint by a terminal reward $r = \log \pi_{\mathcal{B}}$. The resulting optimal path measure is the conditional tilt

$$\mathbb{P}^\star(\boldsymbol{X}_{0:T}) = \pi_{\mathcal{A}}(\boldsymbol{X}_0)\, \mathbb{P}^0(\boldsymbol{X}_{0:T} \mid \boldsymbol{X}_0)\, \frac{\pi_{\mathcal{B}}(\boldsymbol{R}_T)}{Z(\boldsymbol{X}_0)}, \qquad Z(\boldsymbol{X}_0) = \mathbb{E}_{\mathbb{P}^0}\left[\pi_{\mathcal{B}}(\boldsymbol{R}_T) \mid \boldsymbol{X}_0\right] \tag{33}$$

which satisfies $\mathbb{P}_0^\star = \pi_{\mathcal{A}}$ exactly. Its terminal marginal equals $\pi_{\mathcal{B}}$ reweighted by the reference transition density, not $\pi_{\mathcal{B}}$ itself. The two coincide in special cases, e.g. for memoryless reference processes with a corrected terminal cost [Domingo-Enrich et al., 2025, Liu et al., 2025], and they are close when $\pi_{\mathcal{B}}$ is sharply concentrated relative to the reference transition density, as in TPS where $\pi_{\mathcal{B}}$ is a narrow Gaussian around $\boldsymbol{R}_{\mathcal{B}}$. In general, (33) is a soft-constrained Schrödinger bridge [Garg et al., 2024], and this is the problem EntangledSBM solves.

Reformulation as an Unconstrained KL Minimization Let $\mathbb{P}^{b_\theta}$ denote the path measure induced by the controlled Langevin SDEs in (6) of the main text. Substituting (33) into the KL divergence between $\mathbb{P}^{b_\theta}$ and the optimal path measure $\mathbb{P}^\star$ yields:

$$D_{\mathrm{KL}}\left(\mathbb{P}^{b_\theta} \| \mathbb{P}^\star\right) = D_{\mathrm{KL}}\left(\mathbb{P}^{b_\theta} \| \mathbb{P}^0\right) - \mathbb{E}_{\mathbb{P}^{b_\theta}}\left[\log \pi_{\mathcal{B}}(\boldsymbol{R}_T)\right] + \mathbb{E}_{\pi_{\mathcal{A}}}\left[\log Z(\boldsymbol{X}_0)\right] \tag{34}$$

Because the base path measure (induced by the uncontrolled SDE in (1)) and the controlled path measure $\mathbb{P}^{b_\theta}$ (6) differ only by the learned control $\boldsymbol{b}_\theta$, we may apply Girsanov's theorem to express $D_{\mathrm{KL}}(\mathbb{P}^{b_\theta} \| \mathbb{P}^0)$ as a quadratic running cost in $\boldsymbol{b}_\theta$, recovering the form stated in Proposition 3.1.

Connection to Stochastic Optimal Control (SOC) (34) has a direct interpretation as a SOC problem. Defining the terminal reward as:

$$r(\boldsymbol{X}_T) := \log \pi_{\mathcal{B}}(\boldsymbol{R}_T) \tag{35}$$

with the log-density of the target distribution, the objective decomposes into a running control cost and a terminal reward, and its optimal solution coincides exactly with the tilted SB measure: $\mathbb{P}^{b^\star} = \mathbb{P}^\star$.

Equivalently, the optimal control $\boldsymbol{b}^\star$ steers the base process toward high-probability regions of $\pi_{\mathcal{B}}$ while incurring minimal deviation, in the KL sense, from the reference dynamics.

Although (34) is not the textbook statement of the SB problem, this SOC-style formulation has been used in prior work [see, e.g., Holdijk et al., 2023] and recasts the constrained problem in (32) as an *unconstrained* optimization over per-trajectory path KL and terminal log-probability. This unconstrained form is the foundation of our practical algorithm: it admits a tractable trajectory-level Monte Carlo estimator and avoids the need to enforce the terminal marginal constraint explicitly during training.

Requirements on $\pi_{\mathcal{B}}$ EntangledSBM never samples the terminal marginal of the bridge. It only needs to evaluate $\log \pi_{\mathcal{B}}$ and its gradient pointwise, up to an additive constant. The constant cancels both in the score $\nabla_{\boldsymbol{r}_t^i} \log \pi_{\mathcal{B}}$ (Sec 4.1) and in the self-normalized importance weights (Sec 4.2). This covers **(i)** transition path sampling, where $\pi_{\mathcal{B}}$ is a Gaussian around a target structure (App H), **(ii)** bridges between data distributions, where $\pi_{\mathcal{B}}$ is a Gaussian smoothing of target samples (App F), and **(iii)** Boltzmann or reward-tilted targets $\pi_{\mathcal{B}} \propto e^{-U}$ or $\pi_{\mathcal{B}} \propto e^{r}$. When the target is known only through samples, as in our cell experiments, we construct $\pi_{\mathcal{B}}$ by Gaussian smoothing of the samples. The learned bridge then matches this smoothed density rather than the true target distribution, and the approximation degrades in high dimensions or with few samples.

D.4 Justification for Entangled Bias Forces

Proposition D.1 (Monotone Optimality of Entangled Bias Forces)**.** *Let $\mathcal{F}_{ind}$ be the hypothesis class of bias forces that depend only on particle positions $\boldsymbol{b}_\theta^i := \boldsymbol{b}_\theta^i(\boldsymbol{R}_t)$ and $\mathcal{F}_{ent}$ be the hypothesis class of entangled bias forces $\boldsymbol{b}_\theta^i := \boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$. Then, under the cross-entropy objective,*

$$\inf_{\boldsymbol{b}_\theta \in \mathcal{F}_{ent}} \mathcal{L}_{CE}(\boldsymbol{b}_\theta) \leq \inf_{\boldsymbol{b}_\theta \in \mathcal{F}_{ind}} \mathcal{L}_{CE}(\boldsymbol{b}_\theta) \tag{36}$$

with strict improvement when no position-only bias force can realize the optimal path measure.

Proof. Let $\mathbb{P}^\star$ denote the optimal path measure that solves the EntangledSB problem from Definition 3.1 and $\mathbb{P}^{b_\theta}$ be the path measure induced by the bias force $\boldsymbol{b}_\theta$. Let $\mathcal{F}_{\text{ind}}$ be the hypothesis class of bias forces independent of the full system velocities $\boldsymbol{b}_\theta^i := \boldsymbol{b}_\theta^i(\boldsymbol{R}_t)$ and $\mathcal{F}_{\text{ent}}$ be the hypothesis class of entangled bias forces $\boldsymbol{b}_\theta^i := \boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$. Clearly, we have $\mathcal{F}_{\text{ind}} \subseteq \mathcal{F}_{\text{ent}}$. It follows that

$$\{\mathbb{P}^{b_\theta} : \boldsymbol{b}_\theta \in \mathcal{F}_{\text{ind}}\} \subseteq \{\mathbb{P}^{b_\theta} : \boldsymbol{b}_\theta \in \mathcal{F}_{\text{ent}}\} \tag{37}$$

Taking the infimum over the KL-divergence functional $D_{\text{KL}}(\mathbb{P}^\star \| \cdot) \geq 0$ over a larger set of functions cannot increase the value.

$$\inf_{\boldsymbol{b}_\theta \in \mathcal{F}_{\text{ent}}} D_{\text{KL}}(\mathbb{P}^\star \| \mathbb{P}^{b_\theta}) \leq \inf_{\boldsymbol{b}_\theta \in \mathcal{F}_{\text{ind}}} D_{\text{KL}}(\mathbb{P}^\star \| \mathbb{P}^{b_\theta}) \tag{38}$$

Furthermore, if $\mathbb{P}^\star \notin \overline{\{\mathbb{P}^{b_\theta} : \boldsymbol{b}_\theta \in \mathcal{F}_{\text{ind}}\}}$ which denotes the closure of the set of path measures induced by $\mathcal{F}_{\text{ind}}$ and there exists $\boldsymbol{b}^\star \in \overline{\mathcal{F}_{\text{ent}}}$ where $\mathbb{P}^{\boldsymbol{b}^\star} = \mathbb{P}^\star$, then we have

$$\inf_{\boldsymbol{b}_\theta \in \mathcal{F}_{\text{ent}}} D_{\text{KL}}(\mathbb{P}^\star \| \mathbb{P}^{b_\theta}) = 0 \quad \text{while} \quad \inf_{\boldsymbol{b}_\theta \in \mathcal{F}_{\text{ind}}} D_{\text{KL}}(\mathbb{P}^\star \| \mathbb{P}^{b_\theta}) > 0 \tag{39}$$

yielding strict improvement. By Proposition 4.2, we have that $\mathbb{P}^{b^\star} = \mathbb{P}^\star$ is the unique minimizer of $D_{\text{KL}}(\mathbb{P}^\star \| \cdot)$. □

E Theoretical Proofs

For notational simplicity, we will drop the explicit dependence on $(\boldsymbol{R}_t, \boldsymbol{V}_t)$ and simply denote the stochastic path of the system as $(\boldsymbol{X}_t)_{t \in [0,T]}$.

E.1 Proof of Proposition 3.1

Lemma E.1. *The bias force $\boldsymbol{b}^\star$ that minimizes the SOC objective $J(\boldsymbol{b}_\theta; \boldsymbol{x}, t)$ generates the path measure $\mathbb{P}^b$ that minimizes the KL-divergence with the optimal tilted path measure $\mathbb{P}^\star$ that satisfies*

$$\mathbb{P}^\star = \frac{1}{Z}\mathbb{P}^0 e^{r(\boldsymbol{X}_T)}, \quad Z = \mathbb{E}_{\mathbb{P}^0_T}\left[e^{r(\boldsymbol{X}_T)}\right] \tag{40}$$

where $\mathbb{P}^0$ is the base path measure.

Proof. Let the dynamics of $\boldsymbol{X}_t = (\boldsymbol{R}_t, \boldsymbol{V}_t)$ under the base path measure $\mathbb{P}^0$ be defined as

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t)dt + \boldsymbol{\Sigma} d\boldsymbol{W}_t \tag{41}$$

and the dynamics under the biased path measure $\mathbb{P}^{b_\theta}$ be defined as

$$d\boldsymbol{X}_t = \left[\boldsymbol{f}(\boldsymbol{X}_t) + \boldsymbol{\Sigma}\boldsymbol{b}_\theta(\boldsymbol{X}_t)\right]dt + \boldsymbol{\Sigma} d\boldsymbol{W}_t \tag{42}$$

By Lemma D.1, we have that the logarithm of the Radon-Nikodym derivative of the biased measure $\mathbb{P}^{b_\theta}$ with respect to the base measure $\mathbb{P}^0$ is given by

$$\log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = \int_0^T \boldsymbol{b}_\theta(\boldsymbol{X}_t)^\top d\boldsymbol{W}_t - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt \tag{43}$$

Now, taking the expectation with respect to $\mathbb{P}^{b_\theta}$, we derive the expression for the KL-divergence

$$\begin{aligned}
D_{\mathrm{KL}}(\mathbb{P}^{b_\theta}\|\mathbb{P}^0) &= \mathbb{E}_{\mathbb{P}^{b_\theta}}\left[\log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0}\right] \\
&= \mathbb{E}_{\mathbb{P}^{b_\theta}}\left[\int_0^T \boldsymbol{b}_\theta(\boldsymbol{X}_t)^\top d\boldsymbol{W}_t - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt\right] \\
&= \mathbb{E}_{\mathbb{P}^{b_\theta}}\left[\int_0^T \boldsymbol{b}_\theta(\boldsymbol{X}_t)^\top (d\boldsymbol{W}^b_t + \boldsymbol{b}_\theta(\boldsymbol{X}_t)dt) - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt\right] \\
&= \underbrace{\mathbb{E}_{\mathbb{P}^{b_\theta}}\left[\int_0^T \boldsymbol{b}_\theta(\boldsymbol{X}_t)^\top d\boldsymbol{W}^b_t\right]}_{=0} + \mathbb{E}_{\mathbb{P}^{b_\theta}}\left[\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt\right] \\
&= \frac{1}{2}\mathbb{E}_{\mathbb{P}^{b_\theta}}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt
\end{aligned} \tag{44}$$

Substituting this into the SOC objective in Proposition 3.1, we get

$$\begin{aligned}
J(\boldsymbol{b}_\theta) &= \mathbb{E}_{\mathbb{P}^{b_\theta}}\left[\int_0^T \frac{1}{2}\|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt - r(\boldsymbol{X}_T)\right] \\
&= D_{\mathrm{KL}}(\mathbb{P}^{b_\theta}\|\mathbb{P}^0) - \mathbb{E}_{\mathbb{P}^{b_\theta}}[r(\boldsymbol{X}_T)]
\end{aligned} \tag{45}$$

By adding $\log Z$ to both sides, we do not change the minimizer of $J(\boldsymbol{b}_\theta)$, and the SOC objective becomes the KL-divergence between the controlled path measure $\mathbb{P}^{b_\theta}$ and the optimal path measure $\mathbb{P}^\star$:

$$\begin{aligned}
J(\boldsymbol{b}_\theta) + \log Z &= D_{\mathrm{KL}}(\mathbb{P}^{b_\theta}\|\mathbb{P}^0) - \mathbb{E}_{\mathbb{P}^{b_\theta}}[r(\boldsymbol{X}_T)] + \log Z \\
&= \underbrace{\mathbb{E}_{\mathbb{P}^{b_\theta}}\left[\log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0} - r(\boldsymbol{X}_T) + \log Z\right]}_{D_{\mathrm{KL}}(\mathbb{P}^{b_\theta}\|\mathbb{P}^\star)}
\end{aligned} \tag{46}$$

This shows that the minimizer $\boldsymbol{b}^\star$ to the SOC objective also generates the path measure $\mathbb{P}^{b^\star}$ that is closest in KL-divergence to the optimal path measure $\mathbb{P}^\star$. Therefore, the optimal control solution satisfies

$$\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = r(\boldsymbol{X}_T) - \log Z \iff \mathbb{P}^\star = \frac{1}{Z}\mathbb{P}^0 e^{r(\boldsymbol{X}_T)} \tag{47}$$

which concludes our proof. □

Proposition 3.1 (Solving EntangledSB with Stochastic Optimal Control)**.** *We can solve the EntangledSB problem with the stochastic optimal control (SOC) objective given by*

$$\boldsymbol{b}^\star = \arg\min_{\boldsymbol{b}_\theta} \mathbb{E}_{\boldsymbol{X}_{0:T}\sim\mathbb{P}^{b_\theta}} \left[\int_0^T \frac{1}{2}\|\boldsymbol{b}_\theta\|^2 dt - r(\boldsymbol{X}_T) \right]$$

such that (6) where $\boldsymbol{b}_\theta \equiv \boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)$ and $r(\boldsymbol{X}_T) := \log \pi_{\mathcal{B}}(\boldsymbol{R}_T)$ is the terminal reward that measures the log-probability under the target distribution.

We aim to recover the optimal path measure $\mathbb{P}^\star$ that minimizes the KL divergence from the base dynamics $\mathbb{P}^0$ while matching the terminal distribution $\pi_{\mathcal{B}}$ defined as

$$\mathbb{P}^\star(\boldsymbol{X}_{0:T}) = \frac{1}{Z}\mathbb{P}^0 \pi_{\mathcal{B}}(\boldsymbol{R}_T) \tag{48}$$

From Lemma E.1, we have that the bias force $\boldsymbol{b}_\theta$ that minimizes the SOC objective defined as

$$\boldsymbol{b}^\star = \arg\min_{\boldsymbol{b}_\theta} \mathbb{E}_{\boldsymbol{X}_{0:T}\sim\mathbb{P}^{b_\theta}} \left[\int_0^T \frac{1}{2}\|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt - r(\boldsymbol{X}_T) \right] \tag{49}$$

also generates the path measure $\mathbb{P}^{b^\star}$ that minimizes the KL-divergence $D_{\mathrm{KL}}(\mathbb{P}^{b_\theta}\|\mathbb{P}^\star)$ with the optimal path measure $\mathbb{P}^\star$ defined as

$$\mathbb{P}^\star = \frac{1}{Z}\mathbb{P}^0 e^{r(\boldsymbol{X}_T)} \tag{50}$$

Therefore, setting $r(\boldsymbol{X}_T) := \log \pi_{\mathcal{B}}(\boldsymbol{R}_T)$ recovers the solution to the EntangledSB problem. □

E.2 Proof of Proposition 4.1

Proposition 4.1 (Non-Increasing Distance from Target Distribution)**.** *For small enough Δt, the distance from some target state $\boldsymbol{R}_{\mathcal{B}} \in \mathrm{supp}(\pi_{\mathcal{B}})$ after an update with only the bias force $\boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$ defined in (8) is non-increasing, such that*

$$\exists \boldsymbol{R}_{\mathcal{B}} \in \mathrm{supp}(\pi_{\mathcal{B}}) \;\; \textit{s.t.} \;\; \|\boldsymbol{R}_{t+\Delta t} - \boldsymbol{R}_{\mathcal{B}}\| \le \|\boldsymbol{R}_t - \boldsymbol{R}_{\mathcal{B}}\|$$

where $\boldsymbol{R}_{t+\Delta t}$ collects the per-particle updates $\boldsymbol{r}_{t+\Delta t}^i = \boldsymbol{r}_t^i + (\boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)/m_i)\Delta t$, which omit the potential, friction, and noise terms.

Proof. We aim to show that our parameterization of the bias force in (8) ensures that the distance between the current position $\boldsymbol{R}_t$ and some target state $\boldsymbol{R}_T \sim \pi_{\mathcal{B}}$ in the target distribution is *non-increasing*. For each atom $i \in \{1, \ldots, n\}$, we define the bias force as

$$\boldsymbol{b}_\theta^i := \alpha_\theta^i \hat{\boldsymbol{s}}_i + \left(\boldsymbol{I} - \hat{\boldsymbol{s}}_i \hat{\boldsymbol{s}}_i^\top\right) \boldsymbol{h}_\theta^i, \qquad \alpha_\theta^i \ge 0, \;\; \hat{\boldsymbol{s}}_i = \frac{\nabla_{\boldsymbol{r}_t^i} \log \pi_{\mathcal{B}}}{\|\nabla_{\boldsymbol{r}_t^i} \log \pi_{\mathcal{B}}\|} \tag{51}$$

After the update in the direction of the potential energy $-\nabla_{\boldsymbol{r}_t^i} U(\boldsymbol{R}_t)$, the Euler update to the position with time step Δt using the bias force is

$$\boldsymbol{r}_{t+\Delta t}^i = \boldsymbol{r}_t^i + \frac{\Delta t}{m_i}\left(\alpha_\theta^i \hat{\boldsymbol{s}}_i + \left(\boldsymbol{I} - \hat{\boldsymbol{s}}_i \hat{\boldsymbol{s}}_i^\top\right) \boldsymbol{h}_\theta^i\right) = \boldsymbol{r}_t^i + \frac{\Delta t}{m_i}\boldsymbol{b}_\theta^i \tag{52}$$

Let $\boldsymbol{r}_{\mathcal{B}}^i \in \mathrm{supp}(\pi_{\mathcal{B}})$ be a target point chosen on the ascent ray of $\hat{\boldsymbol{s}}_i$ given by

$$\boldsymbol{r}_{\mathcal{B}}^i \in \left\{ \boldsymbol{r}_t^i + \lambda\, \hat{\boldsymbol{s}}_i : \;\; \lambda \ge 0 \right\} \cap \mathrm{supp}(\pi_{\mathcal{B}}) \tag{53}$$

We denote displacement to this target as $\boldsymbol{d}_i := \boldsymbol{r}_{\mathcal{B}}^i - \boldsymbol{r}_t^i = \rho_i\, \hat{\boldsymbol{d}}_i$ with $\hat{\boldsymbol{d}}_i := \frac{\boldsymbol{d}_i}{\|\boldsymbol{d}_i\|}$ and $\rho_i := \|\boldsymbol{d}_i\| \ge 0$. By construction of the ray, $\hat{\boldsymbol{d}}_i = \hat{\boldsymbol{s}}_i$ (i.e., $\boldsymbol{d}_i$ is parallel with $\hat{\boldsymbol{s}}_i$). After a single step, the squared

distance is given by

$$\begin{aligned}\|\boldsymbol{r}_{\mathcal{B}}^i - \boldsymbol{r}_{t+\Delta t}^i\|^2 &= \left\|\boldsymbol{r}_{\mathcal{B}}^i - \left(\boldsymbol{r}_t^i + \frac{\Delta t}{m_i}\boldsymbol{b}_\theta^i\right)\right\|^2 \\ &= \left\|\boldsymbol{d}_i - \frac{\Delta t}{m_i}\boldsymbol{b}_\theta^i\right\|^2 \\ &= \|\boldsymbol{d}_i\|^2 - \frac{2\Delta t}{m_i}\langle \boldsymbol{d}_i, \boldsymbol{b}_\theta^i\rangle + \frac{\Delta t^2}{m_i^2}\|\boldsymbol{b}_\theta^i\|^2 \end{aligned} \tag{54}$$

Using the decomposition of $\boldsymbol{b}_\theta^i$ and the fact that $\boldsymbol{d}_i \parallel \hat{\boldsymbol{s}}_i$ and $(\boldsymbol{I} - \hat{\boldsymbol{s}}_i\hat{\boldsymbol{s}}_i^\top)\boldsymbol{h}_\theta^i \perp \hat{\boldsymbol{s}}_i$, we obtain

$$\langle \boldsymbol{d}_i, \boldsymbol{b}_\theta^i\rangle = \alpha_\theta^i \langle \boldsymbol{d}_i, \hat{\boldsymbol{s}}_i\rangle + \langle \boldsymbol{d}_i, \underbrace{(\boldsymbol{I} - \hat{\boldsymbol{s}}_i\hat{\boldsymbol{s}}_i^\top)\boldsymbol{h}_\theta^i}_{\perp \hat{\boldsymbol{s}}_i}\rangle = \alpha_\theta^i \langle \rho_i \hat{\boldsymbol{s}}_i, \hat{\boldsymbol{s}}_i\rangle = \alpha_\theta^i \rho_i \geq 0 \tag{55}$$

Thus, the first-order term in (54) always decreases (or preserves) the distance, and the orthogonal correction cannot increase it at first order because it is orthogonal to $\boldsymbol{d}_i$. When $\Delta t \to 0$, the second-order term becomes negligible, and the first-order term dominates. We determine the exact threshold for Δt that guarantees a non-increasing distance to be

$$\frac{2\Delta t}{m_i}\alpha_\theta^i \rho_i \geq \frac{\Delta t^2}{m_i^2}\|\boldsymbol{b}_\theta^i\|^2 \implies \Delta t \leq \frac{2m_i \alpha_\theta^i \rho_i}{\|\boldsymbol{b}_\theta^i\|^2} \tag{56}$$

under which the decrease due to the first-order term dominates the quadratic increase, yielding

$$\|\boldsymbol{r}_{t+\Delta t}^i - \boldsymbol{r}_{\mathcal{B}}^i\| \leq \|\boldsymbol{r}_t^i - \boldsymbol{r}_{\mathcal{B}}^i\| \tag{57}$$

In the continuous-time limit as $\Delta t \to 0$, the non-increasing guarantee follows immediately from (55):

$$\frac{\partial}{\partial t}\|\boldsymbol{r}_{\mathcal{B}}^i - \boldsymbol{r}_t^i\|^2 = \frac{\partial}{\partial t}\|\boldsymbol{d}_i\|^2 = -\frac{2}{m_i}\langle \boldsymbol{d}_i, \boldsymbol{b}_\theta^i\rangle = -\frac{2}{m_i}\alpha_\theta^i \rho_i \leq 0 \tag{58}$$

Summing over $i = 1, \ldots, n$ (or equivalently using the concatenated coordinates), we conclude that there exists a target configuration $\boldsymbol{R}_{\mathcal{B}} \in \operatorname{supp}(\pi_{\mathcal{B}})$ such that the distance to $\boldsymbol{R}_{\mathcal{B}}$ is *non-increasing* under the bias force update, which proves the claim. □

E.3 Proof of Proposition 4.2

Proposition 4.2 (Convexity and Uniqueness of Cross-Entropy Objective)**.** *The cross-entropy objective $\mathcal{L}_{CE}$ is convex in $\mathbb{P}^{b_\theta}$, and there exists a unique minimizer $\boldsymbol{b}^\star$ that is the solution to the EntangledSOC problem in Proposition 3.1.*

Proof. First, we show that the cross-entropy objective is convex in the path measure $\mathbb{P}^{b_\theta}$. The functional $\mathbb{Q} \mapsto D_{\mathrm{KL}}(\mathbb{P}^\star \| \mathbb{Q})$ is convex on the space of probability measures, since $x \mapsto -\log x$ is convex. The set of controlled path measures $\{\mathbb{P}^b\}$ need not be convex, so this statement concerns the objective as a functional of the measure. Since $D_{\mathrm{KL}}(\mathbb{P}^\star \| \mathbb{Q}) \geq 0$, with equality if and only if $\mathbb{Q} = \mathbb{P}^\star$, the minimizer of $\mathcal{L}_{\mathrm{CE}}$ is unique and equals $\mathbb{P}^{b^\star} \equiv \mathbb{P}^\star$. From Proposition 3.1, we have that the unique minimizer of $\mathcal{L}_{\mathrm{CE}}$ is exactly the solution to the EntangledSB problem. While this does not necessarily imply convexity in the neural network parameters θ, it yields a more favorable optimization objective. □

E.4 Proof of Proposition 4.3

Proposition 4.3 (Equivalence of Variational and Path Integral Objectives)**.** *The cross-entropy objective can be expressed in path-integral form as*

$$\mathcal{L}_{CE}(\theta) = \mathbb{E}_{\mathbb{P}^{\boldsymbol{v}}}\left[w^\star(\boldsymbol{X}_{0:T})\mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}(\boldsymbol{X}_{0:T})\right] \quad s.t. \quad \begin{cases} w^\star(\boldsymbol{X}_{0:T}) = \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{\boldsymbol{v}}}(\boldsymbol{X}_{0:T}) = \frac{e^{r(\boldsymbol{X}_T)}}{Z}\frac{\mathrm{d}\mathbb{P}^0}{\mathrm{d}\mathbb{P}^{\boldsymbol{v}}}(\boldsymbol{X}_{0:T}) \\ \mathcal{F}_{\boldsymbol{b}_\theta,\boldsymbol{v}}(\boldsymbol{X}_{0:T}) = \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)\|^2 dt \\ -\int_0^T (\boldsymbol{b}_\theta^\top \boldsymbol{v})(\boldsymbol{R}_t, \boldsymbol{V}_t)dt - \int_0^T \boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)^\top d\boldsymbol{W}_t \end{cases}$$

where the off-policy control drift $\boldsymbol{v} = \bar{\boldsymbol{b}} := \texttt{stopgrad}(\boldsymbol{b}_\theta)$ comes from the previous iteration, and we drop the θ-independent term $w^\star(\boldsymbol{X}_{0:T})(r(\boldsymbol{X}_T) - \log Z)$.

From the cross-entropy objective, we have

$$\mathcal{L}_{\text{CE}}(\mathbb{P}^\star, \mathbb{P}^{b_\theta}) := D_{\text{KL}}(\mathbb{P}^\star \| \mathbb{P}^{b_\theta}) = \mathbb{E}_{\mathbb{P}^\star}\left[\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}\right] = \mathbb{E}_{\mathbb{P}^{\bar{b}}}\left[\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{\bar{b}}} \log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}\right] \tag{59}$$

Setting $w^\star := \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{\bar{b}}}$, we can derive

$$w^\star(\boldsymbol{X}_{0:T}) = \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{\bar{b}}}(\boldsymbol{X}_{0:T}) = \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^0}\frac{\mathrm{d}\mathbb{P}^0}{\mathrm{d}\mathbb{P}^{\bar{b}}}(\boldsymbol{X}_{0:T}) \tag{60}$$

From Lemma E.1, we have

$$w^\star(\boldsymbol{X}_{0:T}) = \frac{e^{r(\boldsymbol{X}_T)}}{Z}\frac{\mathrm{d}\mathbb{P}^0}{\mathrm{d}\mathbb{P}^{\bar{b}}}(\boldsymbol{X}_{0:T}) \tag{61}$$

which is our definition for $w^\star$. Now, expanding $\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}$, we have

$$\begin{aligned}
\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}} &= \log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^0} + \log \frac{\mathrm{d}\mathbb{P}^0}{\mathrm{d}\mathbb{P}^{b_\theta}} \\
&= \log \frac{e^{r(\boldsymbol{X}_T)}}{Z} - \log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0} \\
&= r(\boldsymbol{X}_T) - \log Z - \log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0}
\end{aligned} \tag{62}$$

Applying Girsanov's theorem from Lemma D.1, we expand the second term $\log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0}$ as

$$\log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) = \int_0^T (\boldsymbol{b}_\theta(\boldsymbol{X}_t)^\top \boldsymbol{\Sigma}^{-1}) d\boldsymbol{X}_t - \int_0^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{f}(\boldsymbol{X}_t))^\top \boldsymbol{b}_\theta(\boldsymbol{X}_t) dt - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt \tag{63}$$

Given that $d\boldsymbol{X}_t$ evolves via the SDE

$$d\boldsymbol{X}_t = (\boldsymbol{f}(\boldsymbol{X}_t) + \boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_t))dt + \boldsymbol{\Sigma} d\boldsymbol{W}_t \tag{64}$$

Applying $(\boldsymbol{b}_\theta^\top \boldsymbol{\Sigma}^{-1})$ to both sides of the equation, we have

$$\begin{aligned}
(\boldsymbol{b}_\theta^\top \boldsymbol{\Sigma}^{-1}) d\boldsymbol{X}_t &= (\boldsymbol{b}_\theta^\top \boldsymbol{\Sigma}^{-1})(\boldsymbol{f}(\boldsymbol{X}_t) + \boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_t))dt + (\boldsymbol{b}_\theta^\top \boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma} d\boldsymbol{W}_t \\
&= \boldsymbol{b}_\theta^\top(\boldsymbol{\Sigma}^{-1}\boldsymbol{f})(\boldsymbol{X}_t)dt + (\boldsymbol{b}_\theta^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\bar{\boldsymbol{b}})(\boldsymbol{X}_t)dt + (\boldsymbol{b}_\theta^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}) d\boldsymbol{W}_t \\
&= (\boldsymbol{\Sigma}^{-1}\boldsymbol{f}(\boldsymbol{X}_t))^\top \boldsymbol{b}_\theta(\boldsymbol{X}_t)dt + (\boldsymbol{b}_\theta^\top \bar{\boldsymbol{b}})(\boldsymbol{X}_t)dt + \boldsymbol{b}_\theta^\top d\boldsymbol{W}_t
\end{aligned} \tag{65}$$

Substituting this into (63), we get

$$\begin{aligned}
\log \frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T}) &= \int_0^T (\boldsymbol{b}_\theta(\boldsymbol{X}_t)^\top \boldsymbol{\Sigma}^{-1}) d\boldsymbol{X}_t - \int_0^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{f}(\boldsymbol{X}_t))^\top \boldsymbol{b}_\theta(\boldsymbol{X}_t) dt - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt \\
&= \int_0^T \left((\boldsymbol{\Sigma}^{-1}\boldsymbol{f}(\boldsymbol{X}_t))^\top \boldsymbol{b}_\theta(\boldsymbol{X}_t)dt + (\boldsymbol{b}_\theta^\top \bar{\boldsymbol{b}})(\boldsymbol{X}_t)dt + \boldsymbol{b}_\theta^\top d\boldsymbol{W}_t\right) \\
&\quad - \int_0^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{f}(\boldsymbol{X}_t))^\top \boldsymbol{b}_\theta(\boldsymbol{X}_t) dt - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt \\
&= \int_0^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{f}(\boldsymbol{X}_t))^\top \boldsymbol{b}_\theta(\boldsymbol{X}_t) dt + \int_0^T (\boldsymbol{b}_\theta^\top \bar{\boldsymbol{b}})(\boldsymbol{X}_t)dt + \int_0^T \boldsymbol{b}_\theta^\top d\boldsymbol{W}_t \\
&\quad - \int_0^T (\boldsymbol{\Sigma}^{-1}\boldsymbol{f}(\boldsymbol{X}_t))^\top \boldsymbol{b}_\theta(\boldsymbol{X}_t) dt - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt \\
&= \int_0^T (\boldsymbol{b}_\theta^\top \bar{\boldsymbol{b}})(\boldsymbol{X}_t)dt + \int_0^T \boldsymbol{b}_\theta^\top d\boldsymbol{W}_t - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt
\end{aligned} \tag{66}$$

Putting everything together into the expression for the cross-entropy objective, we get

$$\begin{aligned}
\mathcal{L}_{\text{CE}}(\boldsymbol{b}_\theta) &= \mathbb{E}_{\mathbb{P}^v}\left[w^\star(\boldsymbol{X}_{0:T})\left(r(\boldsymbol{X}_T) - \log Z - \log\frac{\mathrm{d}\mathbb{P}^{b_\theta}}{\mathrm{d}\mathbb{P}^0}(\boldsymbol{X}_{0:T})\right)\right] \\
&= \mathbb{E}_{\mathbb{P}^v}\left[w^\star(\boldsymbol{X}_{0:T})\left(r(\boldsymbol{X}_T) - \log Z - \left[\int_0^T (\boldsymbol{b}_\theta^\top \bar{\boldsymbol{b}})(\boldsymbol{X}_t)dt + \int_0^T \boldsymbol{b}_\theta^\top d\boldsymbol{W}_t - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt\right]\right)\right] \\
&= \mathbb{E}_{\mathbb{P}^v}\left[w^\star(\boldsymbol{X}_{0:T})\left(r(\boldsymbol{X}_T) - \log Z\right) - w^\star(\boldsymbol{X}_{0:T})\left(\int_0^T (\boldsymbol{b}_\theta^\top \bar{\boldsymbol{b}})(\boldsymbol{X}_t)dt + \int_0^T \boldsymbol{b}_\theta^\top d\boldsymbol{W}_t - \frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt\right)\right] \\
&= \mathbb{E}_{\mathbb{P}^v}\Bigg[\underbrace{w^\star(\boldsymbol{X}_{0:T})\left(r(\boldsymbol{X}_T) - \log Z\right)}_{\text{independent of } \theta} + w^\star(\boldsymbol{X}_{0:T})\underbrace{\left(\frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{X}_t)\|^2 dt - \int_0^T (\boldsymbol{b}_\theta^\top \bar{\boldsymbol{b}})(\boldsymbol{X}_t)dt - \int_0^T \boldsymbol{b}_\theta^\top d\boldsymbol{W}_t\right)}_{:=\mathcal{F}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:T})}\Bigg]
\end{aligned}$$

Since $w^\star(\boldsymbol{X}_{0:T})\left(r(\boldsymbol{X}_T) - \log Z\right)$ is independent of θ, we can drop it and write the cross-entropy objective as

$$\mathcal{L}_{\text{CE}}(\theta) = \mathbb{E}_{\mathbb{P}^{\bar{b}}}\left[w^\star(\boldsymbol{X}_{0:T})\mathcal{F}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:T})\right] \tag{67}$$

where we define $\bar{\boldsymbol{b}} := \texttt{stopgrad}(\boldsymbol{b}_\theta)$. □

E.5 Proof of Proposition 4.4

Proposition 4.4 (Discretized Cross-Entropy). *Given the discretized $\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:K})$ in (11), we can derive a simplified loss function as*

$$\hat{\mathcal{L}}_{CE}(\theta) = \mathbb{E}_{\boldsymbol{X}_{0:K}\sim\mathbb{P}^{\bar{b}}}\Bigg[\underbrace{\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{\bar{b}}}(\boldsymbol{X}_{0:K})}_{w^\star(\boldsymbol{X}_{0:K})}\log\left(\frac{p^0(\boldsymbol{X}_{0:K})\exp(r(\boldsymbol{X}_K))}{p^{b_\theta}(\boldsymbol{X}_{0:K})}\right)\Bigg] \tag{12}$$

where $w^\star(\boldsymbol{X}_{0:K})$ is the importance weight of the discrete time trajectory $\boldsymbol{X}_{0:K}$.

We will show that $\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:K})$, together with the terminal reward, can be written as

$$\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:K}) + r(\boldsymbol{X}_K) = \log\left(\frac{p^0(\boldsymbol{X}_{0:K})\exp(r(\boldsymbol{X}_K))}{p^{b_\theta}(\boldsymbol{X}_{0:K})}\right) \tag{68}$$

First, we expand the expression as

$$\begin{aligned}
\log\left(\frac{p^0(\boldsymbol{X}_{0:K})\exp(r(\boldsymbol{X}_K))}{p^{b_\theta}(\boldsymbol{X}_{0:K})}\right) &= \log p^0(\boldsymbol{X}_{0:K}) + r(\boldsymbol{X}_K) - \log p^{b_\theta}(\boldsymbol{X}_{0:K}) \\
&= \sum_{k=0}^{K-1}\left[\log p^0(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k) - \log p^{b_\theta}(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k)\right] + r(\boldsymbol{X}_K)
\end{aligned} \tag{69}$$

Throughout this proof, $\boldsymbol{X}_{k+1} - \boldsymbol{X}_k$ denotes the increment after subtracting the base drift $\boldsymbol{f}(\boldsymbol{X}_k)\Delta t$. This drift is common to both densities and cancels in their log-ratio. The Gaussian normalizing constants cancel for the same reason, and $\boldsymbol{\Sigma}$ is diagonal, so $\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$.

For second-order dynamics, these densities apply to the velocity increments, as in the remark after Lemma D.1. Under the uncontrolled dynamics, $\boldsymbol{X}_{k+1}$ is Gaussian with mean $\boldsymbol{X}_k$ and covariance $\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\Delta t$. The log-density becomes

$$\begin{aligned}
\log p^0(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k) &= \log\mathcal{N}\left(\boldsymbol{X}_{k+1}\middle|\boldsymbol{X}_k, \boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\Delta t\right) \\
&= -\frac{1}{2}(\boldsymbol{X}_{k+1} - \boldsymbol{X}_k)^\top(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\Delta t)^{-1}(\boldsymbol{X}_{k+1} - \boldsymbol{X}_k)
\end{aligned} \tag{70}$$

Similarly, the log-density under the bias force $\boldsymbol{b}_\theta$ is given by

$$\begin{aligned}
\log p^{b_\theta}(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k) &= \log\mathcal{N}\left(\boldsymbol{X}_{k+1}\middle|\boldsymbol{X}_k + \boldsymbol{\Sigma}\boldsymbol{b}_\theta(\boldsymbol{X}_k)\Delta t, \boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\Delta t\right) \\
&= -\frac{1}{2}(\boldsymbol{X}_{k+1} - \boldsymbol{X}_k - \boldsymbol{\Sigma}\boldsymbol{b}_\theta(\boldsymbol{X}_k)\Delta t)^\top(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\Delta t)^{-1}(\boldsymbol{X}_{k+1} - \boldsymbol{X}_k - \boldsymbol{\Sigma}\boldsymbol{b}_\theta(\boldsymbol{X}_k)\Delta t)
\end{aligned} \tag{71}$$

Now, we write the increment from time k to $k+1$ as

$$\begin{aligned} \boldsymbol{X}_{k+1} &= \boldsymbol{X}_k + \boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_k)\Delta t + \boldsymbol{\Sigma}\Delta \boldsymbol{W}_k \\ \boldsymbol{X}_{k+1} - \boldsymbol{X}_k &= \boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_k)\Delta t + \boldsymbol{\Sigma}\Delta \boldsymbol{W}_k \end{aligned} \tag{72}$$

and substitute into (70) and (71) to get

$$\begin{aligned} &\log p^0(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k) - \log p^{b_\theta}(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k) \\ &= -\frac{1}{2}\big(\boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_k)\Delta t + \boldsymbol{\Sigma}\Delta \boldsymbol{W}_k\big)^\top \big(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma}\Delta t\big)^{-1}\big(\boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_k)\Delta t + \boldsymbol{\Sigma}\Delta \boldsymbol{W}_k\big) \\ &+ \frac{1}{2}\big(\boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_k)\Delta t + \boldsymbol{\Sigma}\Delta \boldsymbol{W}_k - \boldsymbol{\Sigma}\boldsymbol{b}_\theta(\boldsymbol{X}_k)\Delta t\big)^\top \big(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma}\Delta t\big)^{-1}\big(\boldsymbol{\Sigma}\bar{\boldsymbol{b}}(\boldsymbol{X}_k)\Delta t + \boldsymbol{\Sigma}\Delta \boldsymbol{W}_k - \boldsymbol{\Sigma}\boldsymbol{b}_\theta(\boldsymbol{X}_k)\Delta t\big) \end{aligned}$$

Given that $\big(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma}\Delta t\big)^{-1} = \frac{1}{\Delta t}(\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma})^{-1} = \frac{1}{\Delta t}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-\top})$, and rewriting $\boldsymbol{a} = \bar{\boldsymbol{b}}\Delta t + \Delta \boldsymbol{W}_k - \boldsymbol{b}_\theta \Delta t$ and $\bar{\boldsymbol{a}} = \bar{\boldsymbol{b}}\Delta t + \Delta \boldsymbol{W}_k$, we have

$$\begin{aligned} &\log p^0(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k) - \log p^{b_\theta}(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k) \\ &= \frac{1}{2\Delta t}\Big[-(\boldsymbol{\Sigma}\bar{\boldsymbol{a}})^\top(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-\top})(\boldsymbol{\Sigma}\bar{\boldsymbol{a}}) + (\boldsymbol{\Sigma}\boldsymbol{a})^\top(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-\top})(\boldsymbol{\Sigma}\boldsymbol{a})\Big] \\ &= \frac{1}{2\Delta t}\Big[-\bar{\boldsymbol{a}}^\top\left(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-\top}\boldsymbol{\Sigma}\right)\bar{\boldsymbol{a}} + \boldsymbol{a}^\top\left(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-\top}\boldsymbol{\Sigma}\right)\boldsymbol{a}\Big] \\ &= \frac{1}{2\Delta t}\big[-\|\bar{\boldsymbol{a}}\|^2 + \|\boldsymbol{a}\|^2\big] \\ &= \frac{1}{2\Delta t}\big[-\|\bar{\boldsymbol{b}}\Delta t + \Delta \boldsymbol{W}_k\|^2 + \|\bar{\boldsymbol{b}}\Delta t + \Delta \boldsymbol{W}_k - \boldsymbol{b}_\theta\Delta t\|^2\big] \\ &= \frac{1}{2\Delta t}\big[-\|\bar{\boldsymbol{b}}\Delta t + \Delta \boldsymbol{W}_k\|^2 + \|\bar{\boldsymbol{b}}\Delta t + \Delta \boldsymbol{W}_k\|^2 - 2(\boldsymbol{b}_\theta\Delta t)\left(\bar{\boldsymbol{b}}\Delta t + \Delta \boldsymbol{W}_k\right) + \|\boldsymbol{b}_\theta\Delta t\|^2\big] \\ &= \frac{1}{2\Delta t}\big[-2(\boldsymbol{b}_\theta\cdot\bar{\boldsymbol{b}})(\Delta t)^2 - 2(\boldsymbol{b}_\theta\cdot\Delta \boldsymbol{W}_k)\Delta t + \|\boldsymbol{b}_\theta\|^2(\Delta t)^2\big] \\ &= -(\boldsymbol{b}_\theta\cdot\bar{\boldsymbol{b}})\Delta t - (\boldsymbol{b}_\theta\cdot\Delta \boldsymbol{W}_k) + \frac{1}{2}\|\boldsymbol{b}_\theta\|^2\Delta t \end{aligned} \tag{73}$$

Summing over all K steps and adding $r(\boldsymbol{X}_K)$, we have

$$\begin{aligned} &\sum_{k=0}^{K-1}\Big[\log p^0(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k) - \log p^{b_\theta}(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k)\Big] + r(\boldsymbol{X}_K) \\ &\quad= \underbrace{\frac{1}{2}\sum_{k=0}^{K-1}\|\boldsymbol{b}_\theta(\boldsymbol{X}_k)\|^2\Delta t - \sum_{k=0}^{K-1}(\boldsymbol{b}_\theta\cdot\bar{\boldsymbol{b}})(\boldsymbol{X}_k)\Delta t - \sum_{k=0}^{K-1}\boldsymbol{b}_\theta(\boldsymbol{X}_k)\cdot\Delta\cdot\Delta \boldsymbol{W}_k}_{:=\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:K})} + r(\boldsymbol{X}_K) \end{aligned} \tag{74}$$

Therefore, we can write $\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:K}) + r(\boldsymbol{X}_K)$ as

$$\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}(\boldsymbol{X}_{0:K}) + r(\boldsymbol{X}_K) = \sum_{k=0}^{K-1}\log\frac{p^0(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k)}{p^{b_\theta}(\boldsymbol{X}_{k+1}|\boldsymbol{X}_k)} + r(\boldsymbol{X}_K) = \log\left(\frac{p^0(\boldsymbol{X}_{0:K})\exp(r(\boldsymbol{X}_K))}{p^{b_\theta}(\boldsymbol{X}_{0:K})}\right) \tag{75}$$

where $r(\boldsymbol{X}_K) := \log\pi_{\mathcal{B}}(\boldsymbol{R}_K)$. Since $r(\boldsymbol{X}_K)$ does not depend on θ, adding it to $\hat{\mathcal{F}}_{\boldsymbol{b}_\theta,\bar{\boldsymbol{b}}}$ leaves the gradient of the cross-entropy objective unchanged, which gives the stated loss. □

F Cell Perturbation Experiment Details

F.1 Experiment Setup

Data Processing For this experiment, we extract the cell perturbation data from the Tahoe-100M dataset, which consists of 50 cell lines and over 1000 different drug-dose conditions [Zhang et al., 2026]. Specifically, we use data on a single cell line (A-549) under two drug perturbation conditions: Clonidine at $5\mu M$ and Trametinib at $5\mu M$. The initial distribution is defined as the DMSO-treated

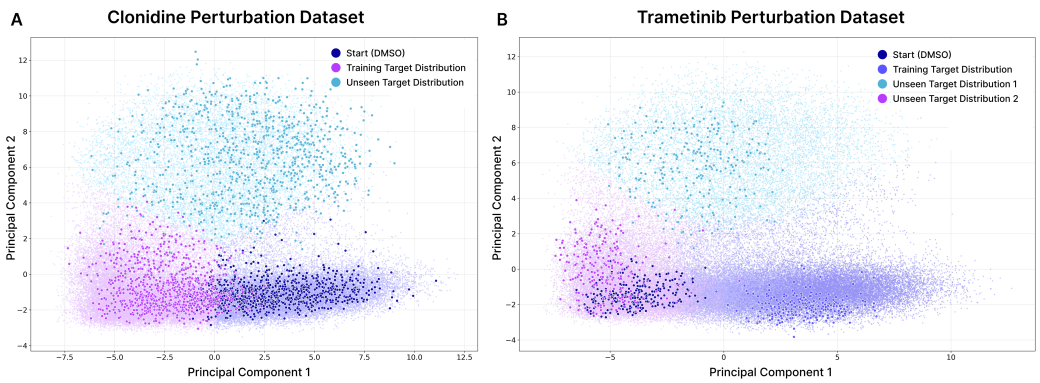

Figure 3: **Training and held-out cell clusters for the cell perturbation experiment.** All cells are plotted with the top 2 PCs. **(A)** Clonidine ($5\mu M$) perturbation data containing the initial DMSO-treated control cells (dark blue), perturbed cluster for training (magenta), and held-out validation cluster (turquoise). **(B)** Trametinib ($5\mu M$) perturbation data containing the initial DMSO-treated control cells (navy blue), the perturbed cluster for training (purple), and two held-out perturbed clusters (magenta and turquoise).

control cells. Using centroid-based sampling, we balance each experiment by keeping, from every cluster, the cells closest to its centroid until it matches the smallest cluster in that experiment, giving 1033 cells per cluster for Clonidine and 381 for Trametinib (original counts in Table 4). We split the cells into training and validation sets with a 0.9/0.1 ratio. All final visualizations are plotted with the first two principal components, but the cell state trajectories are simulated for $d \in \{50, 100, 150\}$ for Clonidine and $d = 50$ for Trametinib.

Table 4: **Cell counts per perturbed cluster for perturbation experiments.**

	Clonidine		**Trametinib**		
	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 3
Original Cell Count	1675	1033	1622	686	381

Model Architecture To learn the dependencies across the n particles, we use Transformer blocks to learn dependencies between the positions $\boldsymbol{R}_t \in \mathbb{R}^{n\times d}$ and velocities $\boldsymbol{V}_t \in \mathbb{R}^{n\times d}$ in the system. The input to the model is the concatenated position and velocity vectors $\texttt{Cat}(\boldsymbol{R}_t, \boldsymbol{V}_t, \boldsymbol{R}_{\mathcal{B}} - \boldsymbol{R}_t, \|\boldsymbol{R}_{\mathcal{B}} - \boldsymbol{R}_t\|) \in \mathbb{R}^{n\times(3d+1)}$. The architecture consists of an input projection layer that projects the input to $d_{\text{hidden}} = 256$, a 4-layer Transformer Encoder (`gelu` activation, feedforward dimension $d_{\text{ff}} = 512$, 8 attention heads, and dropout 0.1), and two MLP decoder heads that predict the scaling factor $\alpha_t(\boldsymbol{R}_t, \boldsymbol{V}_t)$ and correction vector $\boldsymbol{h}^i_t(\boldsymbol{R}_t, \boldsymbol{V}_t)$ per atom. To ensure positive scaling, the `softplus` activation is applied to the output of the scalar MLP head. The correction vector is projected onto the orthogonal plane of the vector pointing in the direction of the target $\hat{\boldsymbol{s}}_i$, and the output bias force is the sum of the scaled directional unit vector $\alpha^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)\hat{\boldsymbol{s}}_i$ and the projected correction vector $(\boldsymbol{I} - \hat{\boldsymbol{s}}_i\hat{\boldsymbol{s}}_i^\top)\boldsymbol{h}^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)$.

Terminal Reward To solve the EntangledSB problem, we defined the terminal reward to be the log-probability under the target density $\pi_{\mathcal{B}}$ given by $r(\boldsymbol{X}_T) = \log \pi_{\mathcal{B}}(\boldsymbol{R}_T)$. We set the target density as the Gaussian centered around $\boldsymbol{R}_{\mathcal{B}}$ with radius σ.

$$\pi_{\mathcal{B}}(\boldsymbol{R}_T) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left(\frac{-\|\boldsymbol{R}_T - \boldsymbol{R}_{\mathcal{B}}\|_2^2}{2\sigma^2}\right) \tag{76}$$

Therefore, to train the bias force to reconstruct the input target state of a cell cluster $\boldsymbol{R}_{\mathcal{B}} \in \mathbb{R}^{n\times d}$, we define $r : \mathcal{X} \to \mathbb{R}$ as:

$$r(\boldsymbol{X}_T) := \frac{-\|\boldsymbol{R}_T - \boldsymbol{R}_{\mathcal{B}}\|_2^2}{2\sigma^2} \tag{77}$$

The specific values for σ are provided in Table 5.

Simulating Trajectories on the Data Manifold To define an energy landscape on the data manifold, we use the RBF metric introduced in Kapuśniak et al. [2024], which is lower in magnitude when $\boldsymbol{x}$ is within the support of the data manifold and larger in magnitude when $\boldsymbol{x}$ moves away from the support of the dataset into sparse regions. First, we define the elements of a function h_j^{RBF} that satisfies $h_j^{\text{RBF}}(\boldsymbol{x}) \approx 1$ on the data manifold as

$$h_j^{\text{RBF}}(\boldsymbol{x}) = \sum_{m=1}^{N_c} \omega_{m,j}(\boldsymbol{x}) \exp\left(-\frac{\lambda_m}{2}\|\boldsymbol{x} - \hat{\boldsymbol{x}}_m\|^2\right) \tag{78}$$

$$\lambda_m = \frac{1}{2}\left(\frac{\kappa}{|C_m|}\sum_{\boldsymbol{x}\in C_m}\|\boldsymbol{x} - \hat{\boldsymbol{x}}_m\|^2\right)^{-2} \tag{79}$$

where $\{\omega_{m,j}\}_{m=1}^{N_c}$ for each cluster and each coordinate are learned given a dataset $\mathcal{D}$ to enforce $h_j^{\text{RBF}}(\boldsymbol{x}) \approx 1$ for all $\boldsymbol{x} \in \mathcal{D}$ with the following loss function

$$\mathcal{L}_{\text{RBF}}(\{\omega_{m,j}\}) = \sum_{\boldsymbol{x}_i \in \mathcal{D}} (1 - h_j^{\text{RBF}}(\boldsymbol{x}_i))^2 \tag{80}$$

Then, we define the potential energy function of the base dynamics as

$$U(\boldsymbol{x}) = -\sum_{j=1}^{d} \log(M_j(\boldsymbol{x}) + \varepsilon), \quad \text{where} \quad M_j(\boldsymbol{x}) = \frac{1}{(h_j^{\text{RBF}}(\boldsymbol{x}) + \varepsilon)^{\alpha}} \tag{81}$$

where $M_j(\boldsymbol{x})$ is small in regions of high data density and large in regions of low data density. The resulting force $-\nabla_{\boldsymbol{x}} U(\boldsymbol{x})$ is used by the simulator as the natural-gradient step.

Hyperparameters We present the hyperparameters used for the cell perturbation modeling experiment in Table 5. For each perturbation, we conducted ablations while keeping all other parameters constant on: **(1)** using the log-variance divergence with the learnable control variate $\mathcal{L}_{\text{LV}}$ described in App D.2 instead of the cross-entropy objective $\mathcal{L}_{\text{CE}}$ and **(2)** removing the dependency on velocities as a feature input to the bias force model $\boldsymbol{b}_\theta(\boldsymbol{R}_t)$. The hyperparameters of the Transformer architecture are given in App F.1 and are kept constant across all experiments. All models are trained with the Adam optimizer [Kingma and Ba, 2015] with learning rate $\eta = 0.0001$. Due to the small batch sizes used for training, we leverage the importance weights $w^\star(\boldsymbol{X}_{0:T})$ for categorical sampling from the replay buffer $\mathcal{R}$ as $\boldsymbol{X}_{0:T} \sim \texttt{Cat}(\texttt{softmax}_{\mathcal{R}}(w^\star(\boldsymbol{X}_{0:T})))$ to mimic the effect of the reweighting in the cross-entropy objective $\mathcal{L}_{\text{CE}}$ over a larger batch size.

Table 5: **Hyperparameter settings for cell perturbation experiment.** The Clonidine perturbation experiment is split into three columns for each of the three dimensions of principal components (PCs) used, $d \in \{50, 100, 150\}$.

Parameter	**Clonidine**			**Trametinib**	**Pancreatic β-Cell**
	50 PCs	100 PCs	150 PCs	50 PCs	30 PCs
number of rollouts N_{rollouts}	100	100	100	100	100
trains per rollout N_{epochs}	1000	1000	1000	1000	1000
step size Δt	0.01	0.01	0.01	0.01	0.01
total time steps T	100	100	100	100	100
number of samples M	64	64	64	64	64
number of particles n	16	16	16	16	16
batch size N_{batch}	64	64	64	64	64
buffer size $\lvert\mathcal{R}\rvert$	1000	1000	1000	1000	1000
radius σ	0.1	0.1	0.1	0.1	0.1
friction γ	2.0	2.0	2.0	2.0	1.0
learning rate	0.0001	0.0001	0.0001	0.0001	0.0001
RBF N_c	150	300	300	150	500
RBF κ	1.5	2.0	3.0	1.5	2.5

F.2 Evaluation Metrics

Maximum Mean Discrepancy (RBF-MMD) We evaluate the reconstruction accuracy of the target distribution and the distribution simulated with EntangledSBM using MMD with the RBF

kernel (RBF-MMD) on all d principal components used during training. For Clonidine, we evaluate $d \in \{50, 100, 150\}$ and for Trametinib, we evaluate $d = 50$. Given the simulated endpoints of M paths for M cell clusters $\{\boldsymbol{R}_T \in \mathbb{R}^{n\times d}\}_{j=1}^{M}$ and the target states $\{\boldsymbol{R}_\mathcal{B} \in \mathbb{R}^{n\times d}\}_{\ell=1}^{M}$, the RBF-MMD is calculated as

$$\text{RBF-MMD} = \frac{1}{M^2}\sum_{j=1}^{M}\sum_{\ell=1}^{M} k_{\text{mix}}(\boldsymbol{R}_T^j, \boldsymbol{R}_T^\ell) + \frac{1}{M^2}\sum_{j=1}^{M}\sum_{\ell=1}^{M} k_{\text{mix}}(\boldsymbol{R}_\mathcal{B}^j, \boldsymbol{R}_\mathcal{B}^\ell) - \frac{2}{M^2}\sum_{j=1}^{M}\sum_{\ell=1}^{M} k_{\text{mix}}(\boldsymbol{R}_T^j, \boldsymbol{R}_\mathcal{B}^\ell) \tag{82}$$

where we define the mixture of RBF kernel functions $k_{\text{mix}}(\cdot,\cdot)$ as

$$k_{\text{mix}}(\boldsymbol{R}, \boldsymbol{R}') = \frac{1}{|\Sigma|}\sum_{\sigma\in\Sigma} \exp\left(-\frac{\|\boldsymbol{R} - \boldsymbol{R}'\|^2}{2\sigma^2}\right) \tag{83}$$

for $\Sigma = \{0.01, 0.1, 1, 10, 100\}$.

1-Wasserstein ($\mathcal{W}_1$) and 2-Wasserstein ($\mathcal{W}_2$) Distances We further compute the $\mathcal{W}_1$ and $\mathcal{W}_2$ distances for the top two PCs of the simulated endpoints of M paths for M cell clusters $\{\boldsymbol{R}_T \in \mathbb{R}^{n\times d}\}_{j=1}^{M}$ which form the predicted distribution π_T and the *full distribution* that the target states are sampled from $\pi_\mathcal{B}$ since the $\mathcal{W}_1$ and $\mathcal{W}_2$ distances can be calculated for a pair of distributions with different sizes. Concretely, the $\mathcal{W}_1$ and $\mathcal{W}_2$ distances are calculated as

$$\mathcal{W}_1 = \left(\min_{\pi\in\Pi(\pi_T,\pi_\mathcal{B})} \int \|\boldsymbol{R}_T - \boldsymbol{R}_\mathcal{B}\|_2 d\pi(\boldsymbol{R}_T, \boldsymbol{R}_\mathcal{B})\right) \tag{84}$$

$$\mathcal{W}_2 = \left(\min_{\pi\in\Pi(\pi_T,\pi_\mathcal{B})} \int \|\boldsymbol{R}_T - \boldsymbol{R}_\mathcal{B}\|_2^2 d\pi(\boldsymbol{R}_T, \boldsymbol{R}_\mathcal{B})\right)^{1/2} \tag{85}$$

which quantify the minimal effort required to transform the simulated endpoint distribution into the true target distribution, demonstrating the ability of the model to capture the true perturbation dynamics.

G Pancreatic β-Cell Experiment Details

G.1 Experiment Setup

Dataset The pancreatic β-cell dataset from Veres et al. [2019] contains 51,274 cells over eight time points that evolve from human pluripotent stem cells to pancreatic β-like cells. Following Zhang et al. [2025a], the gene expression representations are projected to a 30-dimensional PCA space.

Model Architecture and Hyperparameters We leverage the same model architecture as the cell perturbation experiment described in App F, with hyperparameters provided in Table 5.

Evaluation Metrics We compute the 1-Wasserstein distance ($\mathcal{W}_1$) following Zhang et al. [2025a].

H Transition Path Sampling Experiment Details

H.1 Experiment Setup

Model Architecture We use the same model architecture as the cell perturbation experiment described in App F, with the addition of Kabsch alignment [Kabsch, 1976]. We define the **aligned frame** as the frame of the target coordinates and align the input positions and velocities of the *heavy atoms* (non-hydrogen atoms) to the aligned frame. The model then predicts the optimal bias force in the aligned frame, which we transform back to the original frame of the input positions.

Terminal Reward Following the cell perturbation experiment, given the coordinates of a single target state $\boldsymbol{R}_\mathcal{B}$, we define r as the log-probability under a Gaussian of radius σ centered at $\boldsymbol{R}_\mathcal{B}$, computed over all heavy atoms after Kabsch alignment:

$$r(\boldsymbol{X}_T) := \frac{-\|\boldsymbol{R}_T - \boldsymbol{R}_\mathcal{B}\|_2^2}{2\sigma^2} \tag{86}$$

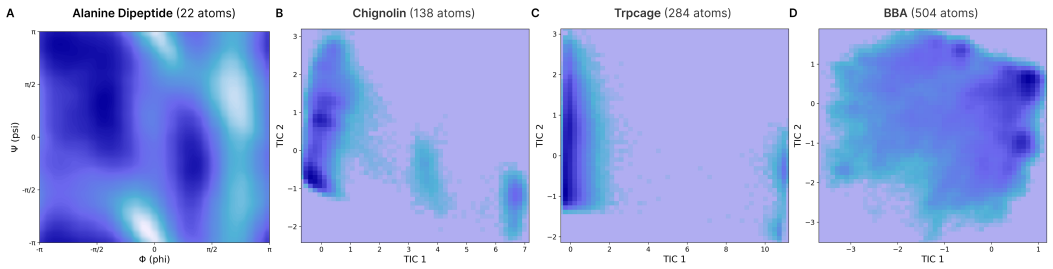

Figure 4: **Energy landscapes for transition path sampling experiments.** Potential energy plotted on dihedral angles (ϕ, ψ) for Alanine Dipeptide and the top two TICA components for fast-folding proteins.

No collective variable enters the reward or the model, and the backbone dihedrals (Alanine Dipeptide) and TICA components (fast-folding proteins) are used only to define a hit for the THP metric, following Seong et al. [2025].

The specific values for σ are provided in Table 6.

Molecular Dynamics Setup We closely follow the setup of Seong et al. [2025]. To simulate the position and velocity via the Langevin SDEs, we use the Velocity-Verlet with Velocity Randomization (VVVR) integrator [Sivak et al., 2014] in OpenMM [Eastman et al., 2017] that updates the state with a velocity-Verlet step for the deterministic forces and a stochastic Ornstein-Uhlenbeck step that randomizes velocities. We set the friction parameter to $\gamma = 1\text{ps}^{-1}$ and the time step size to $\Delta t = 1\text{fs}$ for a total time horizon of $T = 1000\text{fs}$ for Alanine Dipeptide and $T = 5000\text{fs}$ for the fast-folding proteins. Following Seong et al. [2025], we leverage temperature annealing with a starting temperature of $\tau_{\textbf{start}} = 600\text{K}$ and a final temperature of $\tau_{\textbf{end}} = 300\text{K}$ for Alanine Dipeptide and Chignolin and $\tau_{\textbf{end}} = 400\text{K}$ for the remaining fast-folding proteins.

To simulate the unconditional dynamics following the potential energy landscape $U(\boldsymbol{R}_t)$ visualized in Fig. 4 and the corresponding force field $F = -\nabla_{\boldsymbol{R}_t} U(\boldsymbol{R}_t)$, we use the AMBER99SB force field with the ILDN side-chain torsion corrections (`amber99sbildn`) [Lindorff-Larsen et al., 2010] for Alanine Dipeptide in a vacuum and the `ff14SBonlysc` force field [Maier et al., 2015] paired with the `gbn2` solvation model [Nguyen et al., 2013] for the fast-folding proteins.

Hyperparameters We present the hyperparameters used for the TPS experiment in Table 6. The hyperparameters of the Transformer architecture are given in App F.1 and are kept constant across all experiments. We run the benchmark against TPS-DPS [Seong et al., 2025] with the hyperparameters below to ensure fair comparison. All models are trained with the Adam optimizer [Kingma and Ba, 2015] with learning rate $\eta = 1 \times 10^{-4}$.

Table 6: **Hyperparameter settings for transition path sampling experiment.** The setup follows that of Seong et al. [2025] to ensure fair comparison.

Parameter	**Task**			
	Alanine Dipeptide	Chignolin	Trpcage	BBA
number of rollouts N_{rollouts}	100	100	100	100
trains per rollout N_{epochs}	1000	1000	1000	1000
step size Δt	1 fs	1 fs	1 fs	1 fs
total time steps T	1000	5000	5000	5000
number of samples M	64	64	64	64
number of particles n	22	138	284	504
batch size N_{batch}	16	1	1	1
buffer size $\lvert\mathcal{R}\rvert$	1000	100	100	100
starting temperature (Kelvin)	600	600	600	600
ending temperature (Kelvin)	300	300	400	400
radius σ	0.1	0.5	0.5	0.5
friction γ (fs^{-1})	0.001	0.001	0.001	0.001
learning rate	0.0001	0.0001	0.0001	0.0001

H.2 Evaluation Metrics

We follow the evaluations of Seong et al. [2025], Holdijk et al. [2023] and report three metrics: Root Mean Square Distance (RMSD), Target Hit Percentage (THP), and Energy of Transition State (ETS).

Root Mean Square Distance (RMSD) RMSD (↓) measures the distance between the *heavy atoms* (non-hydrogen atoms) of the final position of the system $\boldsymbol{R}_T$ and the target position $\boldsymbol{R}_{\mathcal{B}}$. To align the coordinate frames, we use the Kabsch algorithm [Kabsch, 1976], which determines the optimal rotation and translation that aligns the two sets of heavy atoms. The mean and standard deviation across all M paths are reported.

Target Hit Percentage (THP) THP (↑) measures the percentage of the total trajectories where the final position $\boldsymbol{R}_T$ reaches the vicinity of the target state $\boldsymbol{R}_{\mathcal{B}}$. Specifically, we consider a hit when the two backbone dihedral angles (ϕ, ψ) for Alanine Dipeptide or the first two TICA components for the fast-folding proteins (Chignolin, Trpcage, BBA), denoted $\xi(\boldsymbol{R})$, are within the 0.75-radius sphere around the target defined as $\pi_{\mathcal{B}} = \{\boldsymbol{R} \mid \|\xi(\boldsymbol{R}) - \xi(\boldsymbol{R}_{\mathcal{B}})\| < 0.75\}$. For M total paths, the THP is calculated as

$$\text{THP} = \frac{\sum_{i=1}^{M} \mathbf{1}[\boldsymbol{R}_T^i \in \pi_{\mathcal{B}}]}{M} \tag{87}$$

Energy of Transition State (ETS) ETS (↓) measures the maximum potential energy returned by $U : \mathcal{X} \to \mathbb{R}$ along the discrete transition path $\boldsymbol{R}_{0:K}$ in kJmol^{-1}. It is calculated only for trajectories that reach the vicinity of the target state ($\boldsymbol{R}_T \in \pi_{\mathcal{B}}$), which the THP metric classifies as hits.

$$\text{ETS}(\boldsymbol{R}_{0:K}) = \max_{k \in \{1,\dots,K\}} U(\boldsymbol{R}_k) \tag{88}$$

The mean and standard deviation across the hits for M paths are reported.

H.3 Compute Time

To assess the computational efficiency of our framework, we report the average rollout times of EntangledSBM alongside the closest baseline, TPS-DPS [Seong et al., 2025], across molecular dynamics systems of increasing atomic complexity in Table 7. EntangledSBM achieves runtimes that are comparable to TPS-DPS, with only modest overhead attributable to its Transformer-based interaction modeling. Rollout times for both methods increase with the number of atoms. The systems also differ in simulation length, so these times do not isolate the scaling with system size. For smaller systems such as Alanine Dipeptide (22 atoms), rollouts complete in well under a minute, while larger fast-folding proteins such as BBA (504 atoms) remain tractable within a few minutes per rollout. These results confirm that EntangledSBM preserves the practical scalability of transition path sampling methods while introducing the expressive capacity of Transformer-based interactions, making it well-suited for all-atom molecular simulations across diverse system sizes.

Table 7: **Runtime comparison between EntangledSBM and TPS-DPS for molecular dynamics experiments on a single NVIDIA B200 GPU**. We report rollout times against the closest baseline TPS-DPS [Seong et al., 2025] and show that EntangledSBM is comparable, with modest overhead due to Transformer-based interactions. Times are reported in seconds (mean ± standard deviation).

System	Method	Rollout Time (s)
Alanine Dipeptide (22 atoms)	TPS-DPS	$24.59_{\pm 0.23}$
	EntangledSBM	$31.47_{\pm 0.16}$
Chignolin (138 atoms)	TPS-DPS	$132.93_{\pm 1.05}$
	EntangledSBM	$160.84_{\pm 0.38}$
Trpcage (284 atoms)	TPS-DPS	$184.35_{\pm 2.01}$
	EntangledSBM	$196.62_{\pm 7.35}$
BBA (504 atoms)	TPS-DPS	$267.01_{\pm 1.85}$
	EntangledSBM	$291.68_{\pm 2.64}$

I Comparison to Log-Variance Divergence

In this section, we discuss the intuition behind the differences in cell state trajectories generated with the cross-entropy (CE) objective and log-variance (LV) objective for Clonidine, as shown in Fig. 5 and Table 8, and Trametinib, as shown in Fig. 6 and Table 9.

We observe that the log-variance (LV) objective fails to accurately simulate the intermediate dynamics of cells following perturbation and generates nearly straight, abrupt trajectories connecting the initial and terminal populations. The trajectories generated from the LV-trained bias forces ignore the gradual, curved progression of cell states through intermediate data manifolds, in contrast to the trajectories generated from the cross-entropy (CE)-trained bias forces (Fig. 5, 6). While the LV objective accurately reconstructs the target cell distribution, it fails to reconstruct the intermediate dynamics that arise from nonlinear coupling between position and velocity fields after perturbation (Fig. 5, 6).

While the LV objective aims to make the log–likelihood ratio $\log \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^{b_\theta}}(\boldsymbol{X}_{0:T})$ constant *in expectation* to minimize the variance, it is easiest to only maximize the terminal reward $r(\boldsymbol{X}_T)$ rather than intermediate path alignment with $\mathbb{P}^0$. In contrast, the CE objective is defined as the KL-divergence $D_{\mathrm{KL}}(\mathbb{P}^\star \| \mathbb{P}^{b_\theta})$ which expands into a path-integral action term $\frac{1}{2}\int_0^T \|\boldsymbol{b}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)\|^2 dt$ using Girsanov's theorem. This explicitly regularizes the whole trajectory, rewarding gradual, smooth transport through intermediate states.

Table 8: **Full comparisons for simulating cell cluster dynamics under Clonidine perturbation with EntangledSBM.** We report RBF-MMD for all d PCs with the target cluster and $\mathcal{W}_1$ and $\mathcal{W}_2$ distances of top 2 PCs against the full distribution of the perturbed cells for the seen and unseen populations after 100 simulation steps and cluster size set to $n = 16$. The mean and standard deviation of metrics from 5 independent simulations are reported. Comparisons include the base dynamics, without velocity conditioning, and EntangledSBM with the log-variance (LV) objective instead of cross-entropy (CE) for increasing principal component dimensions $d = \{50, 100, 150\}$.

	Seen Target Distribution			Unseen Target Distribution		
Model	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)
Base Dynamics (50 PCs)	$0.677_{\pm 0.001}$	$5.947_{\pm 0.005}$	$6.015_{\pm 0.005}$	$0.784_{\pm 0.001}$	$8.217_{\pm 0.005}$	$8.384_{\pm 0.005}$
EntangledSBM w/o Velocity Conditioning						
50 PCs	$0.440_{\pm 0.000}$	$1.741_{\pm 0.003}$	$1.857_{\pm 0.004}$	$0.478_{\pm 0.000}$	$2.907_{\pm 0.006}$	$3.022_{\pm 0.006}$
100 PCs	$0.494_{\pm 0.000}$	$2.315_{\pm 0.004}$	$2.423_{\pm 0.004}$	$0.539_{\pm 0.000}$	$4.110_{\pm 0.004}$	$4.249_{\pm 0.003}$
150 PCs	$0.510_{\pm 0.000}$	$2.497_{\pm 0.006}$	$2.620_{\pm 0.006}$	$0.560_{\pm 0.000}$	$4.573_{\pm 0.006}$	$4.716_{\pm 0.007}$
EntangledSBM w/ LV						
50 PCs	$0.294_{\pm 0.000}$	$0.205_{\pm 0.001}$	$0.285_{\pm 0.001}$	$0.290_{\pm 0.000}$	$0.230_{\pm 0.001}$	$0.382_{\pm 0.001}$
100 PCs	$0.330_{\pm 0.000}$	$0.277_{\pm 0.001}$	$0.332_{\pm 0.001}$	$0.346_{\pm 0.000}$	$0.358_{\pm 0.001}$	$0.492_{\pm 0.001}$
150 PCs	$0.323_{\pm 0.000}$	$0.128_{\pm 0.001}$	$0.278_{\pm 0.000}$	$0.327_{\pm 0.000}$	$0.193_{\pm 0.001}$	$0.425_{\pm 0.001}$
EntangledSBM w/ CE						
50 PCs	$0.401_{\pm 0.000}$	$0.342_{\pm 0.002}$	$0.400_{\pm 0.001}$	$0.419_{\pm 0.000}$	$0.538_{\pm 0.013}$	$0.705_{\pm 0.030}$
100 PCs	$0.455_{\pm 0.000}$	$0.953_{\pm 0.025}$	$1.015_{\pm 0.025}$	$0.500_{\pm 0.001}$	$0.899_{\pm 0.006}$	$1.055_{\pm 0.008}$
150 PCs	$0.478_{\pm 0.000}$	$0.753_{\pm 0.008}$	$0.826_{\pm 0.007}$	$0.506_{\pm 0.000}$	$0.700_{\pm 0.009}$	$0.811_{\pm 0.011}$

Table 9: **Full comparisons for simulating cell cluster dynamics under Trametinib perturbation with EntangledSBM.** We report RBF-MMD for all $d = 50$ PCs with the target cluster and $\mathcal{W}_1$ and $\mathcal{W}_2$ distances of the top 2 PCs against the full distribution of the perturbed cells for the seen and unseen populations after 100 simulation steps and cluster size set to $n = 16$. Mean and standard deviation of metrics from 5 independent simulations are reported. Comparisons include the base dynamics, without velocity conditioning, and EntangledSBM with the log-variance (LV) objective instead of cross-entropy (CE).

	Seen Target Distribution			Unseen Target Distribution 1			Unseen Target Distribution 2		
Method	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)	RBF-MMD (↓)	$\mathcal{W}_1$ (↓)	$\mathcal{W}_2$ (↓)
Base Dynamics	$0.938_{\pm 0.001}$	$7.637_{\pm 0.005}$	$7.653_{\pm 0.006}$	$0.900_{\pm 0.000}$	$7.766_{\pm 0.009}$	$7.877_{\pm 0.009}$	$0.754_{\pm 0.001}$	$1.201_{\pm 0.009}$	$1.455_{\pm 0.012}$
EntangledSBM w/o Velocity Conditioning	$0.449_{\pm 0.000}$	$1.506_{\pm 0.005}$	$1.544_{\pm 0.005}$	$0.476_{\pm 0.000}$	$2.116_{\pm 0.005}$	$2.197_{\pm 0.005}$	$0.480_{\pm 0.000}$	$0.505_{\pm 0.004}$	$0.627_{\pm 0.005}$
EntangledSBM w/ LV	$0.308_{\pm 0.000}$	$0.175_{\pm 0.002}$	$0.256_{\pm 0.001}$	$0.302_{\pm 0.000}$	$0.340_{\pm 0.001}$	$0.565_{\pm 0.000}$	$0.312_{\pm 0.000}$	$0.198_{\pm 0.001}$	$0.321_{\pm 0.001}$
EntangledSBM w/ CE	$0.428_{\pm 0.000}$	$0.392_{\pm 0.005}$	$0.434_{\pm 0.006}$	$0.409_{\pm 0.000}$	$0.453_{\pm 0.008}$	$0.561_{\pm 0.009}$	$0.451_{\pm 0.000}$	$0.394_{\pm 0.003}$	$0.469_{\pm 0.004}$

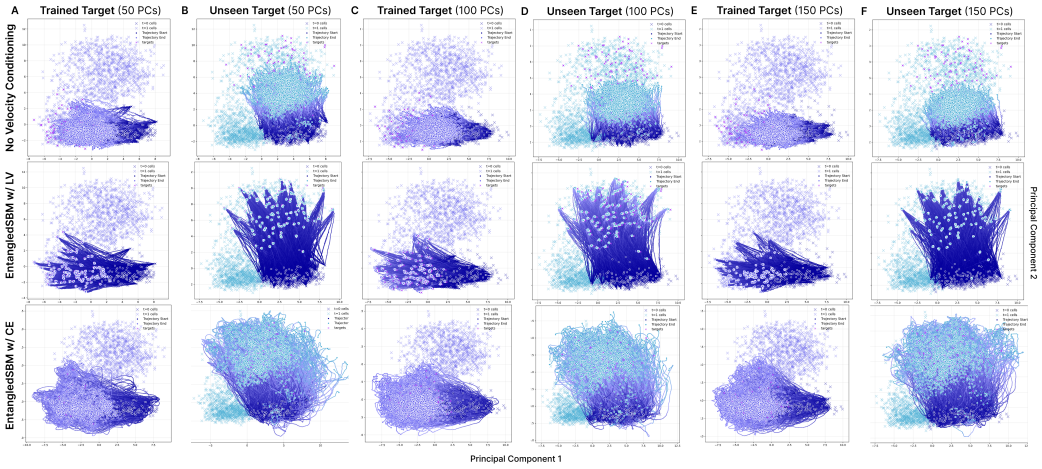

Figure 5: **Full visualization of simulated cell cluster dynamics with EntangledSBM under Clonidine perturbation.** Nearest neighbor cell clusters with $n = 16$ cells are simulated over 100 time steps. The gradient indicates the evolution of timesteps from the initial time $t = 0$ (navy) to the final time $t = T$ (purple or turquoise). Rows compare no velocity conditioning, the log-variance (LV) objective, and the cross-entropy (CE) objective. 50 PCs simulated to **(A)** trained target distribution and **(B)** unseen target. 100 PCs simulated to **(C)** trained target distribution and **(D)** unseen target. 150 PCs simulated to **(E)** trained target distribution and **(F)** unseen target.

J Trajectory Diversity Analysis

A natural concern when interpreting Energy of Transition State (ETS) values is whether higher ETS reflects *meaningful diversity* of folding pathways or simply a mismatch with the underlying energy landscape. To disentangle these two possibilities, we directly compare EntangledSBM against TPS-DPS along a suite of trajectory-level diversity metrics that probe energetic, mechanistic, and geometric variation across target-hitting trajectories.

J.1 Metrics

For each method, we restrict attention to trajectories that successfully reach the target state and compute the following quantities:

ETS range (kJ/mol) Difference between the maximum and minimum ETS values among target-hitting trajectories. A wider range indicates that the method explores transition states spanning a broader region of the energy landscape.

Contact correlation mean (lower is more diverse) For each pair of target-hitting trajectories, we extract the contact formation order vector (which native contacts form first, second, and so on along the path) and compute the Pearson correlation between the two vectors. We report the mean correlation across all pairs. Low mean correlation indicates that trajectories fold via mechanistically distinct routes, rather than reordering minor variants of the same pathway.

Energy effective dimensionality (higher is more diverse) Effective number of principal components required to describe the ensemble of per-trajectory energy profiles. Higher values indicate that energy profiles vary along many independent axes rather than along a single dominant mode.

R_g **midpoint standard deviation** (nm, higher is more diverse) Standard deviation of the radius of gyration evaluated at the midpoint of each trajectory. A higher value indicates that some trajectories pass through compact intermediates while others pass through expanded ones, providing a coarse geometric signature of distinct folding mechanisms.

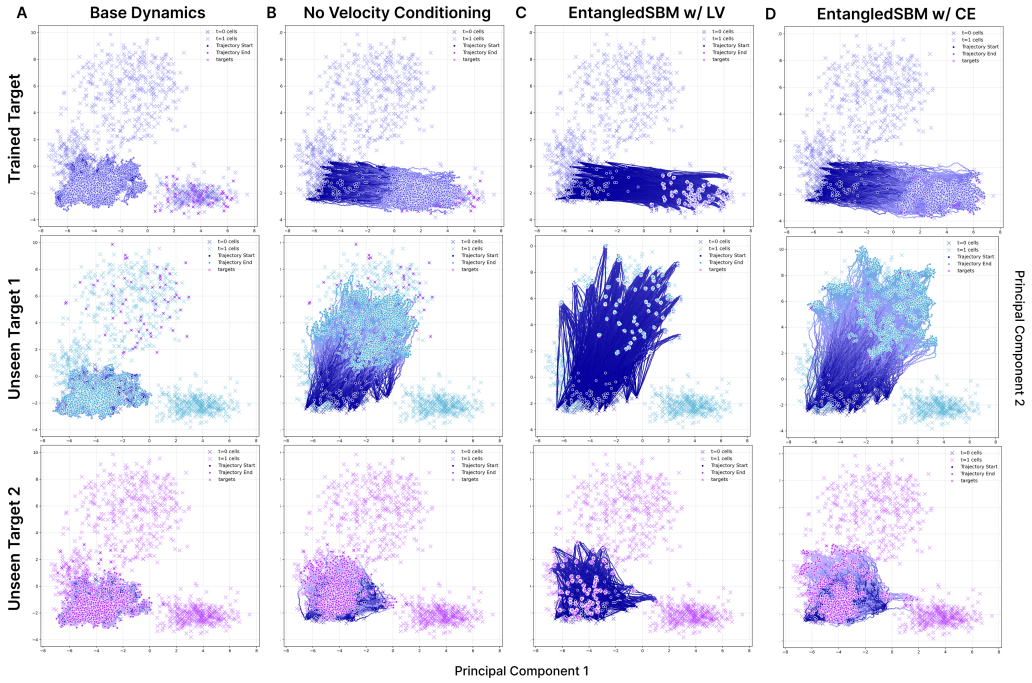

Figure 6: **Full visualization of simulated cell cluster dynamics with EntangledSBM under Trametinib perturbation.** All trajectories generated for $d = 50$ PCs. Nearest neighbor cell clusters with $n = 16$ cells are simulated over 100 time steps. The gradient indicates the evolution of timesteps from the initial time $t = 0$ (navy) to the final time $t = T$ (purple, turquoise, or magenta). **(A)** Base dynamics with no bias force. Dynamics with base and learned bias force trained with **(B)** no velocity conditioning, **(C)** log-variance (LV) objective, and **(D)** cross-entropy objective to the trained target distribution and two unseen target distributions.

J.2 Results

Table 10 reports the four diversity metrics on Chignolin and BBA, the two systems with the highest ETS values.

Table 10: **Trajectory diversity comparison between EntangledSBM and TPS-DPS on protein folding experiments.** Bold indicates the more diverse method on each metric. Contact correlation is reported as the mean ± standard deviation across trajectory pairs.

Protein	Method	ETS range (kJ/mol)	Contact corr. mean (↓)	Energy eff. dimensionality (↑)	R_g midpoint std (↑; nm)
Chignolin	TPS-DPS	433.45	$0.775_{\pm 0.024}$	2.17	0.02
	EntangledSBM	**535.58**	$\mathbf{0.698_{\pm 0.123}}$	**4.40**	**0.03**
BBA	TPS-DPS	539.52	$\mathbf{0.720_{\pm 0.018}}$	4.07	0.01
	EntangledSBM	**1097.26**	$0.726_{\pm 0.029}$	**8.35**	**0.02**

Energetic Diversity EntangledSBM achieves a substantially wider ETS range than TPS-DPS on both systems, with 535.58 vs. 433.45 kJ/mol on Chignolin and 1097.26 vs. 539.52 kJ/mol on BBA, the latter representing roughly a twofold increase. This indicates that EntangledSBM samples transition states spanning a broader region of the energy landscape, rather than concentrating on a single low-energy channel.

Mechanistic Diversity On Chignolin, EntangledSBM achieves a mean pairwise correlation of 0.698 compared with 0.775 for TPS-DPS, showing that pairs of EntangledSBM trajectories differ more in the *ordering* of native contact formation. On BBA, the two methods are essentially tied on this metric (0.726 vs. 0.720), but EntangledSBM substantially outperforms TPS-DPS on energy effective dimensionality (8.35 vs. 4.07) and R_g midpoint variability, indicating that even when contact ordering is comparable, the underlying energetic and geometric trajectories explore a richer set of independent modes.

These metrics show that the higher ETS values produced by EntangledSBM correspond to diverse and structured folding mechanisms rather than to a mismatch with the energy landscape. The combination of a wider ETS range, lower contact ordering correlation, and higher energy effective dimensionality indicates that the model produces interpretable variation in how the protein folds (i.e., different contact orders, different intermediate geometries, and different energetic profiles) while reaching the target state with higher THP and lower RMSD than TPS-DPS.

K Physical Feasibility of Generated Trajectories

To assess whether the trajectories generated by EntangledSBM are physically plausible, we evaluate each sampled path along three structural and dynamical metrics that together characterize local smoothness, steric feasibility, and structural consistency with the target fold.

K.1 Metrics

We compute the following quantities along generated trajectories:

Smoothness (nm/step, ↓) Mean per-step atomic displacement, averaged over all heavy atoms and integration steps. This quantifies the local continuity of the trajectory and detects abrupt discontinuities that would be incompatible with physical evolution.

Steric Clashes (count, ↓) Number of heavy-atom pairs whose pairwise distance falls below a clash threshold of 1.2 Å (0.12 nm), computed over all unordered heavy-atom pairs in each frame and summed across the trajectory. The threshold is chosen below typical heavy-atom bond lengths, so bonded neighbors are not counted as clashes.

Contact Fraction (↑) Fraction of native residue-residue contacts recovered along the path, where a native contact is defined by a Cα-Cα distance below the standard 8 Å cutoff in the reference structure. This captures structural correctness with respect to the target fold.

K.2 Results

Table 11 summarizes the three metrics for both **Chignolin** and **BBA**. Per-step displacements are on the order of 3×10^{-3} nm, indicating smooth evolution along the path. Clash counts represent about 0.13% (Chignolin) and 0.05% (BBA) of all heavy-atom pairs per frame, indicating that intermediate conformations remain largely free of non-physical atomic overlap, and contact fractions exceed 0.96, with BBA recovering essentially all native contacts (0.994). Together, these results indicate that EntangledSBM trajectories evolve continuously, remain sterically feasible, and are structurally consistent with the target fold throughout the transition.

Table 11: **Physical feasibility of intermediate trajectories for Chignolin and BBA.** Arrows indicate the preferred direction.

Protein	Smoothness (↓ nm/step)	Mean clashes (↓)	Contact fraction (↑)
Chignolin	0.00306	3.9	0.965
BBA	0.00296	14.8	0.994

L Algorithms

Here, we provide the pseudocode for training (Alg 2) and inference (Alg 3) with EntangledSBM.

Algorithm 2 EntangledSBM **Training**

Input: Parameterized networks $\alpha_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) : \mathbb{R}^{n\times d} \times \mathbb{R}^{n\times d} \to \mathbb{R}^n$ and $\boldsymbol{h}_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) : \mathbb{R}^{n\times d} \times \mathbb{R}^{n\times d} \to \mathbb{R}^{n\times d}$, potential energy function $U(\boldsymbol{R}_t) : \mathbb{R}^{n\times d} \to \mathbb{R}$, initial distribution $\pi_{\mathcal{A}}$, distribution of target states $\pi_{\mathcal{B}}$, buffer size $|\mathcal{R}|$, batch size N_{batch}, number of samples M, number of rollouts N_{rollouts}, training steps per rollout N_{epochs}, timestep size Δt

$\mathcal{R} \leftarrow \{\}$ ▷ *Initialize empty replay buffer*

for rollout in $1, \ldots, N_{\text{rollouts}}$ **do**

  for j in $1, \ldots, M$ **do**

    Sample initial state $\boldsymbol{X}_0 = (\boldsymbol{R}_0, \boldsymbol{V}_0)$ with $\boldsymbol{R}_0 \sim \pi_{\mathcal{A}}$

    for t in $0, \ldots, T-1$ **do**

      Predict $\alpha^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) \in \mathbb{R}$ and $\boldsymbol{h}^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) \in \mathbb{R}^d$ with parameterized neural network where $\alpha^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) \geq 0$ is enforced with `softplus` activation

      $\boldsymbol{s}_i \leftarrow \nabla_{\boldsymbol{r}^i_t} \log \pi_{\mathcal{B}}$, $\hat{\boldsymbol{s}}_i \leftarrow \boldsymbol{s}_i / \|\boldsymbol{s}_i\|$

      Compute bias force

$$\boldsymbol{b}^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t) \leftarrow \alpha^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)\hat{\boldsymbol{s}}_i + \left(\boldsymbol{I} - \hat{\boldsymbol{s}}_i\hat{\boldsymbol{s}}_i^\top\right)\boldsymbol{h}^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)$$

      Update each particle i following (6), shown here with Euler-Maruyama integration (the TPS experiments use the VVVR integrator, and the cell experiments use the first-order form)

$$\boldsymbol{r}^i_{t+1} = \boldsymbol{r}^i_t + \boldsymbol{v}^i_t \Delta t$$

$$\boldsymbol{v}^i_{t+1} = \boldsymbol{v}^i_t + \left(\frac{-\nabla_{\boldsymbol{r}^i_t} U(\boldsymbol{R}_t) + \boldsymbol{b}^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)}{m_i} - \gamma \boldsymbol{v}^i_t\right)\Delta t + \sqrt{\frac{2\gamma k_B \tau \Delta t}{m_i}}\,\boldsymbol{\epsilon}^i_t$$

    end for

  end for

  $\mathcal{R} \leftarrow \mathcal{R} \cup \{\boldsymbol{X}_{0:T}\}_{j=1}^{M}$ ▷ *update buffer with discrete trajectories*

  for step in $1, \ldots, N_{\text{epochs}}$ **do**

    Sample batch $\{\boldsymbol{X}_{0:T}\}_{j=1}^{N_{\text{batch}}}$ from buffer $\mathcal{R}$

    Compute cross-entropy objective $\mathcal{L}_{\text{CE}}$ with Prop. 4.4

    Update θ with $\nabla_\theta \mathcal{L}_{\text{CE}}$

  end for

end for

return parameterized $\alpha^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)$ and $\boldsymbol{h}^i_\theta(\boldsymbol{R}_t, \boldsymbol{V}_t)$

Algorithm 3 EntangledSBM **Inference**

Input: Trained networks $\alpha_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$ and $\boldsymbol{h}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$, potential energy function $U(\boldsymbol{R}_t)$, initial state $\boldsymbol{X}_0 = (\boldsymbol{R}_0, \boldsymbol{V}_0)$, target state $\boldsymbol{R}_\mathcal{B}$, time steps T, step size Δt, friction γ

$\boldsymbol{R}_t \leftarrow \boldsymbol{R}_0, \boldsymbol{V}_t \leftarrow \boldsymbol{V}_0$

$\mathcal{P} \leftarrow \{\boldsymbol{R}_0\}$ ▷ *initialize path*

for t in $0, \ldots, T-1$ **do**

  Predict $\alpha_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$ and $\boldsymbol{h}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$ with parameterized neural network where $\alpha_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t) \geq 0$ is enforced with `softplus` activation

  $\boldsymbol{s}_i \leftarrow \nabla_{\boldsymbol{r}_t^i} \log \pi_\mathcal{B}$, $\hat{\boldsymbol{s}}_i \leftarrow \boldsymbol{s}_i / \|\boldsymbol{s}_i\|$

  Compute bias force

$$\boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t) \leftarrow \alpha_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)\hat{\boldsymbol{s}}_i + \left(\boldsymbol{I} - \hat{\boldsymbol{s}}_i \hat{\boldsymbol{s}}_i^\top\right) \boldsymbol{h}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)$$

  Update each particle i following (6), shown here with Euler-Maruyama integration (the TPS experiments use the VVVR integrator, and the cell experiments use the first-order form)

$$\boldsymbol{r}_{t+1}^i = \boldsymbol{r}_t^i + \boldsymbol{v}_t^i \Delta t$$

$$\boldsymbol{v}_{t+1}^i = \boldsymbol{v}_t^i + \left(\frac{-\nabla_{\boldsymbol{r}_t^i} U(\boldsymbol{R}_t) + \boldsymbol{b}_\theta^i(\boldsymbol{R}_t, \boldsymbol{V}_t)}{m_i} - \gamma \boldsymbol{v}_t^i \right) \Delta t + \sqrt{\frac{2\gamma k_B \tau \Delta t}{m_i}}\, \boldsymbol{\epsilon}_t^i$$

  Append to path $\mathcal{P} \leftarrow \mathcal{P} \cup \{\boldsymbol{R}_{t+1}\}$

end for

return path $\mathcal{P}$

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We clearly state the contributions and scope in the abstract and the **main contributions** list in the introduction.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss limitations in the Limitations paragraph of Sec 7 and in more detail in Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Rigorous proofs for our theoretical results are provided in Appendix E.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide complete details to enable reproduction of our results. Experiment details for cell perturbation modeling, pancreatic β-cell differentiation modeling, and transition path sampling are provided in App F, G, and H. Algorithms 2 and 3 provide pseudocode for training and inference. Code and data are provided in the repository: https://github.com/sophtang/EntangledSBM. Together, these resources ensure full reproducibility.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code and data are provided in the repository https://github.com/sophtang/EntangledSBM and on Hugging Face at https://huggingface.co/ChatterjeeLab/EntangledSBM.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Full training and test details are provided in App F (cell perturbation), App G (pancreatic β-cell differentiation), and App H (transition path sampling). This includes data processing, model architecture (App F.1), terminal reward formulation, data manifold construction (App F.1), and the Adam optimizer configuration with learning rate $\eta = 1 \times 10^{-4}$. Full hyperparameter values are tabulated in Tables 5 and 6, and evaluation metric definitions are provided in App F.2 and H.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: All quantitative results in Tables 1, 2, 3, 8, and 9 report mean and standard deviation, except THP, which is a percentage of paths and has no error bar. Cell perturbation metrics (RBF-MMD, $\mathcal{W}_1$, $\mathcal{W}_2$) are computed across 5 independent simulations as described in

App F and App F.2. Transition path sampling metrics are computed from 64 paths per protein following App H.2: RMSD and THP use all paths, and ETS uses target-hitting paths. Runtime measurements in Table 7 also report mean and standard deviation across runs.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Compute resources are reported in App H for the transition path sampling experiments, with detailed rollout times on a single NVIDIA B200 GPU across all four protein systems (Alanine Dipeptide, Chignolin, Trp-cage, BBA) in Table 7. Hyperparameters that determine compute usage (number of rollouts, training steps per rollout, total time steps, batch size, buffer size) are documented for the cell perturbation experiments in Table 5 and for the transition path sampling experiments in Table 6.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The research conforms with the NeurIPS Code of Ethics. The work does not involve human subjects or sensitive personal data, all datasets and pre-trained force fields used are publicly available and properly cited.

   Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper includes a Broader Impact section in Appendix A that describes applications to protein folding and cell state perturbation modeling, and the limited dual-use risk of the framework given its methodological scope. The statement also notes the use of publicly available datasets (Tahoe-100M and the pancreatic β-cell dataset) under their original licenses, and the use of standard test proteins and publicly available force fields with no human or animal subjects.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release new pre-trained foundation models, datasets, or generators that pose a high risk of misuse. The released code provides a simulation framework for cellular and molecular dynamics that operates on publicly available data and standard force fields.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All existing assets used are properly credited via citations to the original works, including the Tahoe-100M dataset [Zhang et al., 2026] used in App F, the pancreatic β-cell dataset cited in App G, the AMBER99SB-ILDN and ff14SBonlysc force fields with the gbn2 solvation model used in App H, the OpenMM molecular dynamics engine, the VVVR integrator, and the TPS-DPS baseline whose hyperparameters and protocol we follow.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The new asset introduced is the EntangledSBM framework, which is documented through the methodological descriptions in Sec 3 and 4, the pseudo-code in Algorithms 1, 2, and 3 (App L), the model architecture details in App F.1, the hyperparameter tables in Tables 5 and 6, and the code repository at https://github.com/sophtang/EntangledSBM.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

Justification: The paper does not involve crowdsourcing or research with human subjects. All experiments use publicly available datasets and in silico simulations.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve human subjects research and therefore no IRB approval is required.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not used as an important, original, or non-standard component of the core methodology. The framework concerns Schrödinger bridge matching for interacting multi-particle dynamics with a Transformer-based bias-force parameterization, and does not employ language models in the methodology.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.