DISCRETE ADJOINT SCHRÖDINGER BRIDGE SAMPLER

Wei Guo[*,1], **Yuchen Zhu**[‡,1], **Xiaochen Du**[‡,2], **Juno Nam**[‡,2],
Yongxin Chen[†,1], **Rafael Gómez-Bombarelli**[†,2], **Guan-Horng Liu**[†,3], **Molei Tao**[†,1],
Jaemoo Choi[*,1]
[1]Georgia Institute of Technology [2]Massachusetts Institute of Technology [3]FAIR at Meta

ABSTRACT

Learning discrete neural samplers is challenging due to the lack of gradients and combinatorial complexity. While stochastic optimal control (SOC) and Schrödinger bridge (SB) provide principled solutions, efficient SOC solvers like adjoint matching (AM), which excel in continuous domains, remain unexplored for discrete spaces. We bridge this gap by revealing that the core mechanism of AM is *state-space agnostic*, and introduce **discrete ASBS**, a unified framework that extends AM and adjoint Schrödinger bridge sampler (ASBS) to discrete spaces. Theoretically, we analyze the optimality conditions of the discrete SB problem and its connection to SOC, identifying a necessary cyclic group structure on the state space to enable this extension. Empirically, discrete ASBS achieves competitive sample quality with significant advantages in training efficiency and scalability.

1 INTRODUCTION

Sampling from unnormalized distributions is a fundamental problem across computational statistics (Liu, 2008; Brooks et al., 2011), Bayesian inference (Gelman et al., 2013), and statistical mechanics (Landau & Binder, 2014). The goal is to draw samples from a target distribution ν on a state space $\mathcal{X}$, where $\nu(x) \propto \mathrm{e}^{-\beta E(x)}$ is specified through an energy function $E : \mathcal{X} \to \mathbb{R}$ and inverse temperature $\beta > 0$. In high-dimensional settings with complex energy landscapes, traditional methods such as Markov chain Monte Carlo (MCMC) suffer from slow mixing and poor scalability. Recent advances in **neural samplers** address these challenges by parameterizing sampling dynamics with deep neural networks, enabling efficient learning even without i.i.d. samples from the target. Crucially, they provide amortized inference, replacing costly MCMC with rapid generation from a pretrained model.

For continuous state space $\mathcal{X} = \mathbb{R}^D$, neural samplers have achieved remarkable success through diverse methodologies (He et al., 2025; Sanokowski et al., 2025b), including sequential Monte Carlo (Phillips et al., 2024), escorted transport (Vargas et al., 2024; Albergo & Vanden-Eijnden, 2025; Chen et al., 2025; Blessing et al., 2025b; Du et al., 2025), stochastic optimal control (SOC, Zhang & Chen (2022); Vargas et al. (2023); Richter & Berner (2024)), parallel tempering (Rissanen et al., 2025; Akhound-Sadegh et al., 2025; Zhang et al., 2026), etc. Among SOC approaches, **adjoint matching** (**AM**, Domingo-Enrich et al. (2025a); Havens et al. (2025)) has emerged as a powerful solver, enabling the framework of **adjoint Schrödinger bridge sampler** (**ASBS**, Liu et al. (2025a)), which offers fast convergence, scalability, and flexibility in reference dynamics. However, AM relies fundamentally on continuous calculus, making its extension to discrete state spaces highly nontrivial.

Motivated by these successes, recent work has explored discrete neural samplers based on continuous-time Markov chains (CTMCs), proposing similar methods based on escorted transport (Holderrieth et al., 2025; Ou et al., 2025b) and SOC formulations (Zhu et al., 2025a; Guo et al., 2026). Despite this progress, the Schrödinger Bridge (SB) problem – a distributionally constrained optimization framework central to optimal transport (Léonard, 2014; Chen et al., 2016b; 2021) and continuous diffusion models (Chen et al., 2021; De Bortoli et al., 2021; Chen et al., 2022; Liu et al., 2022; Shi et al., 2023) – remains underdeveloped for discrete spaces (Ksenofontov & Korotin, 2025; Kim et al., 2025). The theoretical connection between discrete SB and SOC is not fully established, and no discrete SOC solver achieves the efficiency of continuous AM. While a recent work So et al.

[*]Core contributors, equal contribution; [‡]co-second authors, equal contribution; [†] equal advising, alphabetical order. Correspondance to: {wei.guo, jchoi843}@gatech.edu.

Table 1: Conceptual comparison between denoising matching and adjoint matching in continuous and discrete state spaces. Denoising matching holds for general reference dynamics while adjoint matching requires $\boxed{\text{additive}}$ noise.

	Continuous state space $\mathcal{X} = \mathbb{R}^D$	Discrete state space $\mathcal{X} = \mathbb{Z}_N^D$
Ref. dyn. p^r	$\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t,\ X_0 \sim \mu$	CTMC (X_t) with transition rate (r_t), $X_0 \sim \mu$
Ctrl. dyn. p^u	$\mathrm{d}X_t = (b_t + \sigma_t u_t)(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t,\ X_0 \sim \mu$	CTMC (X_t) with transition rate (u_t), $X_0 \sim \mu$
SB Prob.	$\min_{u \text{ s.t. } X_1 \sim \nu} \mathbb{E}_{X \sim p^u}\Big[\int_0^1 \frac{1}{2}\lVert u_t(X_t)\rVert^2 \mathrm{d}t\Big]$	$\min_{u \text{ s.t. } X_1 \sim \nu} \mathbb{E}_{X \sim p^u}\Big[\int_0^1 \sum_{y \neq X_t} \big(u_t \log \frac{u_t}{r_t} + r_t - u_t\big)(y, X_t)\mathrm{d}t\Big]$
SOC Prob.	$\min_u \mathbb{E}_{X \sim p^u}\Big[\int_0^1 \frac{1}{2}\lVert u_t(X_t)\rVert^2 \mathrm{d}t + \log \frac{\widehat{\varphi}_1}{\nu}(X_1)\Big]$	$\min_u \mathbb{E}_{X \sim p^u}\Big[\int_0^1 \sum_{y \neq X_t} \big(u_t \log \frac{u_t}{r_t} + r_t - u_t\big)(y, X_t)\mathrm{d}t + \log \frac{\widehat{\varphi}_1}{\nu}(X_1)\Big]$
Opt. Ctrl.	$u_t^\star(x) = \sigma_t \nabla \log \varphi_t(x)$	$u_t^\star(y, x) = r_t(y, x)\frac{\varphi_t(y)}{\varphi_t(x)}$
Corrector	$\nabla \log \widehat{\varphi}_1(x)$	$\frac{\widehat{\varphi}_1(y)}{\widehat{\varphi}_1(x)}$
Example of $\boxed{\text{Additive}}$ Noise	$b_t \equiv 0 \Rightarrow \boxed{p^r_{1\mid t}(x_1 \mid x) = q_t(x_1 - x)}$ $q_t(\varepsilon) = \mathcal{N}(\varepsilon; 0, \widetilde{\sigma}_t^2 I)$	$r_t(y, x) = \frac{\gamma_t}{N} 1_{d_{\mathrm{H}}(x,y)=1} \Rightarrow \boxed{p^r_{1\mid t}(x_1 \mid x) = q_t(x_1 - x)}$ $q_t(\varepsilon) = A(t,1)^{d_{\mathrm{H}}(\varepsilon,0)} B(t,1)^{D - d_{\mathrm{H}}(\varepsilon,0)}$
Denoising Matching	$\nabla \log \varphi_t(x) = \mathbb{E}_{p^\star_{1\mid t}(x_1 \mid x)} \nabla_x \log p^r_{1\mid t}(x_1 \mid x)$ (1) $\nabla \log \widehat{\varphi}_1(x) = \mathbb{E}_{p^\star_{0\mid 1}(x_0 \mid x)} \nabla_x \log p^r_{1\mid 0}(x \mid x_0)$ (2)	$\frac{\varphi_t(y)}{\varphi_t(x)} = \mathbb{E}_{p^\star_{1\mid t}(x_1 \mid x)} \frac{p^r_{1\mid t}(x_1 \mid y)}{p^r_{1\mid t}(x_1 \mid x)}$ $\frac{\widehat{\varphi}_1(y)}{\widehat{\varphi}_1(x)} = \mathbb{E}_{p^\star_{0\mid 1}(x_0 \mid x)} \frac{p^r_{1\mid 0}(y \mid x_0)}{p^r_{1\mid 0}(x \mid x_0)}$
Adjoint Matching	$\nabla \log \varphi_t(x) = \mathbb{E}_{p^\star_{1\mid t}(x_1 \mid x)} \nabla \log \varphi_1(x_1)$ (3) $\nabla \log \widehat{\varphi}_1(x) = \mathbb{E}_{p^\star_{0\mid 1}(x_0 \mid x)} \nabla \log \widehat{\varphi}_0(x_0)$ (4)	$\frac{\varphi_t(y)}{\varphi_t(x)} = \mathbb{E}_{p^\star_{1\mid t}(x_1 \mid x)} \frac{\varphi_1(y + x_1 - x)}{\varphi_1(x_1)}$ $\frac{\widehat{\varphi}_1(y)}{\widehat{\varphi}_1(x)} = \mathbb{E}_{p^\star_{0\mid 1}(x_0 \mid x)} \frac{\widehat{\varphi}_0(y + x_0 - x)}{\widehat{\varphi}_0(x_0)}$

(2026) attempted to adapt AM to masked discrete diffusion, their approach yields complex training objectives, leaving a principled and efficient discrete AM framework an open challenge.

In this work, we develop a unified theoretical framework for SB and SOC on discrete state spaces, introducing the **discrete adjoint Schrödinger bridge sampler (DASBS)**. We derive optimality conditions for the discrete SB problem that structurally mirror the continuous setting, and extend AM to discrete domains through novel controller and corrector learning objectives. Our approach reveals that a group structure on the state space and uniform reference dynamics are essential – requirements that parallel the additive noise assumption in continuous AM.

Our **key contributions** are: **I. Unified Discrete Framework**: We formalize discrete neural sampling as an SB problem for CTMCs and establish its equivalent SOC formulation. **II. Generalizing Adjoint Matching**: We identify state-space agnostic principles underlying AM, providing conceptual clarity for extensions beyond continuous spaces. **III. Discrete ASBS**: We introduce discrete ASBS, a principled algorithm that alternates between adjoint and corrector matching, derived from variational characterizations of the learning objectives. **IV. Empirical Validation**: We validate DASBS on high-dimensional discrete benchmarks, demonstrating competitive performance and efficiency.

2 Preliminaries

Diffusion Samplers on Continuous State Space $\mathcal{X} = \mathbb{R}^D$ MCMC sampling based on equilibrium dynamics is typically slow to mix when the target distribution is complex. Recent progress in non-equilibrium measure transport for generative modeling such as diffusion models (Song et al., 2021) has motivated sampling methods based on controlled stochastic differential equations (SDEs), commonly referred to as diffusion neural samplers. The sampling dynamics are described by SDE:

$$p^u:\ \mathrm{d}X_t = (b_t + \sigma_t u_t)(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t,\ X_0 \sim \mu, \tag{5}$$

where $b : [0,1] \times \mathcal{X} \to \mathcal{X}$ is the base drift, $\sigma : [0,1] \to \mathbb{R}_{>0}$ is the noise schedule, and μ is the initial source distribution. Given $(b_\cdot, \sigma_\cdot, \mu)$, the goal is to learn a parameterized control $u : [0,1] \times \mathcal{X} \to \mathcal{X}$ such that the marginal distribution of X_1 matches the target distribution ν. We use p^u to denote the **path measure** induced by this SDE with control u. Formally, $p^u(X_{[0,1]})$ can be viewed as the limit of the joint distribution of $(X_{t_0}, X_{t_1}, ..., X_{t_K})$ as the partition $0 = t_0 < t_1 < ... < t_K = 1$ becomes finer and finer, and a rigorous definition is through the Radon-Nikodým derivative (RND).

Schrödinger Bridge Problem on Continuous State Spaces One way to formulate the learning of diffusion samplers is through a distributionally constrained optimal transport problem known as the **Schrödinger bridge (SB)** problem (Léonard, 2014; Chen et al., 2016b):

$$\min_{u \text{ s.t. } X_1 \sim \nu} \Big\{ \mathrm{KL}(p^u \| p^r) = \mathop{\mathbb{E}}_{X \sim p^u} \int_0^1 \frac{1}{2} \lVert u_t(X_t) \rVert^2 \mathrm{d}t \Big\}, \tag{6}$$

where p^u is the controlled path measure (5) and p^r is the path measure of a *reference dynamics* with zero control ($u \equiv 0$). The optimal control can be characterized by

$$u_t^\star(x) = \sigma_t \nabla \log \varphi_t(x), \tag{7}$$

where the **SB potentials** $(\varphi_t, \widehat{\varphi}_t)$ are defined (up to multiplicative constants) through time integrations with respect to the reference transition kernel $p^r_{t|s}(y|x)$:

$$\varphi_t(x) = \int p^r_{1|t}(y|x)\varphi_1(y)\mathrm{d}y, \quad \varphi_0(x)\widehat{\varphi}_0(x) = \mu(x);$$
$$\widehat{\varphi}_t(x) = \int p^r_{t|0}(x|y)\widehat{\varphi}_0(y)\mathrm{d}y, \quad \varphi_1(x)\widehat{\varphi}_1(x) = \nu(x).$$

Stochastic Optimal Control Characteristics of SB An interesting connection between the SB formulation and **stochastic optimal control (SOC)** is revealed by the characterization of the optimal control $u^\star$. As shown in Dai Pra (1991); Chen et al. (2016b); Liu et al. (2025a), $u^\star$ (7) is also the solution to the following SOC problem:

$$\min_u \mathop{\mathbb{E}}_{X\sim p^u}\Big[\int_0^1 \frac{1}{2}\|u_t(X_t)\|^2\mathrm{d}t + \log\frac{\widehat{\varphi}_1(X_1)}{\nu(X_1)}\Big]. \tag{8}$$

(8) highlights that SB (6) admits an SOC interpretation, where the terminal marginal constraint is encoded through the terminal cost $\log\frac{\widehat{\varphi}_1}{\nu}$. This perspective provides a useful bridge between SB theory and SOC formulations.

Adjoint Schrödinger Bridge Sampler (ASBS, Liu et al. (2025a)) When $b_t \equiv 0$, the reference dynamics reduces to Brownian motion. In this setting, the reference transition kernel is **additive** in the sense that sampling from $p^r_{1|t}(\cdot|x)$ can be achieved by adding a noise $\epsilon \sim q_t$ onto x, i.e.,

$$p^r_{1|t}(y|x) = q_t(y-x), \quad q_t = \mathcal{N}(0, \widetilde{\sigma}_t^2 I), \tag{9}$$

where $\widetilde{\sigma}_t^2 = \int_t^1 \sigma_s^2 \mathrm{d}s$. Importantly, this additive property is central in simplifying the associated SOC problem (8); it yields tractable conditional distributions $p^r_{1|t}$ and enables explicit expressions for the quantities appearing in the optimality conditions (Havens et al., 2025). Exploiting this structure, Liu et al. (2025a) derived the **adjoint matching (AM)** identities (3) and (4), as well as the corresponding **denoising matching (DM)** identities (1) and (2), and proposed an alternating procedure based on (2) and (3) to learn the controller and the corrector, which converges under suitable regularity conditions. See the left part of Tab. 1 for details.

Connection to the Memoryless Case (Adjoint Sampling) An earlier work, adjoint sampling (Havens et al., 2025), considered a special case in which the reference path measure p^r is **memoryless**, i.e. $p^r_{0,1}(x,y) = p^r_0(x)p^r_1(y)$, under which one can show that $\widehat{\varphi}_1 \propto p^r_1$. As a consequence, the corrector $\nabla \log \widehat{\varphi}_1$ is known and the learning problem reduces to a single regression objective for the controller. From this perspective, ASBS is a generalization of adjoint sampling that relaxes the memoryless assumption, allowing for nontrivial boundary coupling. It is also observed in Liu et al. (2025a) that non-memoryless reference dynamics enable reduced noise levels compared with memoryless ones, thus offering improved performance.

3 SB and SOC Theory for CTMC

In this section, we introduce the SB problem for continuous-time Markov chains (CTMCs), serving as a discrete analog of (6). We then extend the corresponding SB and SOC theory from $\mathbb{R}^D$ to the discrete CTMC (Léonard, 2014). In particular, we derive optimality conditions for the optimal transition rate $u^\star$, analogous to (7), and formulate an associated SOC problem in the spirit of (8).

3.1 Problem Setting

Continuous-time Markov chain Throughout this paper, we will consider the discrete state space $\mathcal{X}$ as the set of length-D sequences with N possible states $[N] := \{1, 2, ..., N\}$, i.e., $\mathcal{X} = [N]^D$. A

continuous-time Markov chain (CTMC) $(X_t)_{t\in[0,1]}$ is a stochastic process taking values in $\mathcal{X}$ and characterized by its **transition rate** $r = (r_t(y,x))_{t\in[0,1]}^{x,y\in\mathcal{X}}$, defined by

$$r_t(y,x) = \lim_{h\to 0} \frac{\Pr(X_{t+h} = y | X_t = x) - 1_{y=x}}{h},$$

where $1_A \in \{0,1\}$ is the indicator of a statement A.

SB Problem on Discrete State Spaces Following Sec. 2, we address sampling problem by formulating a SB problem for CTMCs, whose optimal path measure $p^\star$ satisfies the boundary marginal constraints $p_0^\star = \mu$ and $p_1^\star = \nu$. Let $r_t(y,x)$ and $u_t(y,x)$ denote the reference and controlled transition rates, and write p^r and p^u for their induced **path measures**. We assume a common initial distribution $p_0^r = p_0^u = \mu$. Consider the following SB problem for CTMCs:

$$\min_{u \text{ s.t. } p_1^u=\nu} \Big\{ \mathrm{KL}(p^u \| p^r) = \mathop{\mathbb{E}}_{X\sim p^u} \int_0^1 \sum_{y\neq X_t} \Big(u_t \log \frac{u_t}{r_t} + r_t - u_t \Big)(y, X_t) \mathrm{d}t \Big\}. \tag{SB}$$

Next, we will make explicit the connection between this SB problem and a corresponding SOC formulation, in direct analogy with (8).

3.2 SB AND SOC THEORY FOR CTMC

Characterization of the Optimal Transition Rate We now characterize the optimal transition rate solving (SB). The following result provides an explicit description of the optimal transition rate in terms of a pair of time-dependent potentials, which play a role analogous to the SB potentials in the continuous-state setting. See App. B.2 for the proof.

Theorem 3.1. *The optimal transition rate $u^\star$ for* (SB) *can be expressed as*

$$u_t^\star(y,x) = \frac{\varphi_t(y)}{\varphi_t(x)} r_t(y,x), \ \forall y \neq x, \tag{10}$$

*where the **SB potentials** $(\varphi_t, \widehat{\varphi}_t)$ satisfy*

$$\forall 0 \le s < t \le 1 : \begin{cases} \varphi_s(x) = \sum_y p_{t|s}^r(y|x)\varphi_t(y), & (11) \\ \widehat{\varphi}_t(x) = \sum_y p_{t|s}^r(x|y)\widehat{\varphi}_s(y), & (12) \end{cases}$$

and the optimal path measure $p^\star$ satisfies

$$p_t^\star(x) = \varphi_t(x)\widehat{\varphi}_t(x), \qquad \frac{p_{t|s}^\star(x|y)}{p_{t|s}^r(x|y)} = \frac{\varphi_t(x)}{\varphi_s(y)}, \qquad \frac{p_{s|t}^\star(y|x)}{p_{t|s}^r(x|y)} = \frac{\widehat{\varphi}_s(y)}{\widehat{\varphi}_t(x)}. \tag{13}$$

(11) and (12) is a forward-backward representation with respect to the reference transition kernel, and the boundary marginal constraints are encoded through the coupling conditions at boundary times: $\varphi_0\widehat{\varphi}_0 = \mu$, $\varphi_1\widehat{\varphi}_1 = \nu$.

SOC Problem on Discrete State Spaces The SOC problem with terminal cost $g : \mathcal{X} \to \mathbb{R}$ is

$$\min_{u \text{ s.t. } p_0^u=\mu} \Big\{ \mathrm{KL}(p^u \| p^r) + \mathop{\mathbb{E}}_{X\sim p^u} g(X_1) \Big\} \tag{SOC}$$

SOC Characteristics of SB While the optimality conditions in (10) provide an explicit characterization of $u^\star$, they are challenging to solve in practice. The main difficulties are twofold: the coupled boundary constraints at times $t = 0$ and $t = 1$, and the need to evaluate expectations with respect to the reference transition kernels. The SOC formulation circumvents these by avoiding the direct solution of coupled equations. The following is an SOC reinterpretation of the SB problem:

Theorem 3.2. *The optimal transition rate $u_t^\star$* (10) *for* (6) *solves* (SOC) *with terminal cost $g \leftarrow \log \frac{\widehat{\varphi}_1}{\nu}$.*

Sketch of proof. The optimal path measures of (SB) and (SOC) can be respectively written as

$$\frac{p^\star_{\text{SB}}(X_{[0,1]})}{p^r(X_{[0,1]})} = \frac{\widehat{\varphi}_0(X_0)}{\mu(X_0)}\frac{\nu(X_1)}{\widehat{\varphi}_1(X_1)}, \tag{14}$$

$$\frac{p^\star_{\text{SOC}}(X_{[0,1]})}{p^r(X_{[0,1]})} = \frac{\mathrm{e}^{-g(X_1)}}{Z(X_0)}, \quad Z(x) = \mathbb{E}_{p^r_{1|0}(y|x)}\mathrm{e}^{-g(y)}. \tag{15}$$

See App. B.4 for the full proof. Thm. 3.2 shows that (SB) admits an equivalent SOC formulation where the terminal marginal constraint $p^u_1 = \nu$ is incorporated through the terminal cost $g = \log \frac{\widehat{\varphi}_1}{\nu}$. This will be instrumental for developing tractable learning objectives in the discrete setting.

Connection to Memoryless Reference Dynamics If further assume p^r is **memoryless**, then (11) implies $\varphi_0(x) = \mathbb{E}_{p^r_1}\mathrm{e}^{-g} = \text{const}$, and (12) implies $\widehat{\varphi}_1(x) = p^r_1(x)\sum_y \widehat{\varphi}_0(y) \propto p^r_1(x)$. Therefore, $g = \log \frac{p^r_1}{\nu} + \text{const}$.

Relation to Continuous State Spaces Finally, we note that the SOC and SB theory developed above for discrete state spaces closely parallels its continuous counterpart introduced in Sec. 2. A detailed comparison is provided in the upper part of Tab. 1.

4 Discrete Adjoint Schrödinger Bridge Sampler (DASBS)

In this section, we introduce a principled theory and algorithm for learning the optimal transition rates in the discrete SB problem. Building on the SOC and SB formulation for CTMCs developed in the previous section, we propose a discrete analogue of ASBS (Liu et al., 2025a) by developing discrete versions of adjoint matching and denoising matching for controller and corrector.

A key challenge in the discrete setting is the absence of additive noise structure (9) that plays a central role in developing AM framework. To address this issue, we will adopt a *cyclic group structure* on the discrete state space, which allows discrete transitions to be interpreted in an additive form. Under this perspective, a uniform reference transition rate emerges as a natural choice for inducing tractable and symmetric transition kernels.

Choice of the Reference Path Measure (10) implies that it suffices to learn the ratio of φ_t at all pairs of $x, y \in \mathcal{X}$ such that $r_t(y, x) > 0$. We follow the typical strategy in discrete diffusion models (Campbell et al., 2022; Lou et al., 2024; Schiff et al., 2025) to restrict the transition to pairs of x, y with Hamming distance $d_{\text{H}}(x, y) = 1$, i.e., x and y differ at exactly one entry. Throughout this paper, we consider the reference path measure p^r starting from an arbitrary tractable initial distribution $p^r_0 = \mu$ and induced by the following **uniform transition rate** r that keeps the uniform distribution on $\mathcal{X}$ ($p_{\text{unif}} = \frac{1}{N^D}$) invariant:

$$r_t(y, x) = \begin{cases} \frac{\gamma_t}{N}, & \text{if } d_{\text{H}}(y, x) = 1, \\ -\gamma_t D\left(1 - \frac{1}{N}\right), & \text{if } y = x, \\ 0, & \text{if otherwise,} \end{cases} \tag{16}$$

where $\gamma_\cdot : [0, 1] \to \mathbb{R}_+$ is a noise schedule. We remark that for two given time steps $0 \le s < t \le 1$ and states $x, y \in \mathcal{X}$, $p^r_{t|s}(y|x)$ can be written in a closed form (Prop. B.5), and so is $p^r_{t|0,1}(x|x_0, x_1)$ (Prop. B.6). See App. B.5 for details. Thus, it suffices to learn $\frac{\varphi_t(y)}{\varphi_t(x)}$ for all $d_{\text{H}}(x, y) = 1$. Define the **controller** matrix $\Phi^\star_t(x) \in \mathbb{R}^{D\times N}$ whose (d, n)-th element is $\Phi^\star_t(x)_{d,n} = \frac{\varphi_t(x^{d\leftarrow n})}{\varphi_t(x)}$, where $x^{d\leftarrow n}$ denotes the vector obtained by replacing the d-th entry of x with n.

Cyclic Group Structure for the State Space To extend AM (3) to discrete space, we treat the state space $\mathcal{X} = [N]^D$ as $\mathbb{Z}^D_N$, where $\mathbb{Z}_N$ is the cyclic group of integers modulo N. Thus, the sum and difference of any two elements in $\mathcal{X}$ are still in $\mathcal{X}$. The key benefit of doing so is that the transition kernel $p^r_{1|t}(\cdot|x)$ is again **additive**:

$$\exists q_t \ \text{(see Prop. B.5)} \qquad \text{s.t.} \quad p^r_{1|t}(y|x) = q_t(y - x). \tag{17}$$

Controller Adjoint Matching With an additive noise, we can thus consider a target score matching (De Bortoli et al., 2024; Zhang et al., 2025) objective like in the continuous AM (Domingo-Enrich et al., 2025a; Havens et al., 2025). Applying (17), we can refactor (11) as follows:

$$\varphi_t(y) \underset{(11)}{=} \sum_{x_1 \in \mathbb{Z}_N^D} p^r_{1|t}(x_1|y)\varphi_1(x_1) \underset{(17)}{=} \sum_{x_1 \in \mathbb{Z}_N^D} p^r_{1|t}(x_1+\Delta|y+\Delta)\varphi_1(x_1)$$

$$\underset{x_1' \leftarrow x_1+\Delta}{=} \sum_{x_1' \in \mathbb{Z}_N^D} p^r_{1|t}(x_1'|y+\Delta)\varphi_1(x_1'-\Delta) \underset{\Delta \leftarrow x-y}{=} \sum_{x_1' \in \mathbb{Z}_N^D} p^r_{1|t}(x_1'|x)\varphi_1(x_1'-x+y), \tag{18}$$

$$\implies \frac{\varphi_t(y)}{\varphi_t(x)} = \mathbb{E}_{p^r_{1|t}(x_1|x)} \frac{\varphi_1(x_1+y-x)}{\varphi_t(x)} \underset{(13)}{=} \mathbb{E}_{p^\star_{1|t}(x_1|x)} \frac{\varphi_1(x_1+y-x)}{\varphi_1(x_1)}. \tag{19}$$

When $y \leftarrow x^{d\leftarrow n}$, since $x^{d\leftarrow n} - x$ has at most one non-zero coordinate, x_1 and $x_1 + x^{d\leftarrow n} - x = x_1^{d \leftarrow x_1^d + n - x^d}$ differ in at most one entry. Let $D_f(a\|b) = f(a) - f(b) - (a-b)f'(b) \geq 0$ be the Bregman divergence induced by a strictly convex and differentiable function f, and note that $\mathbb{E}\xi = \operatorname{argmin}_{c\in\mathbb{R}} \mathbb{E}D_f(\xi\|c)$ for any random variable ξ (Lou et al., 2024). We thus have the following variational characterization of the controller $\Phi^\star$:

$$\Phi^\star = \operatorname*{argmin}_{\Phi} \mathbb{E}_t \mathbb{E}_{p^\star_{t,1}(x,x_1)} \sum_{d=1}^{D} \sum_{n \neq x^d} D_f\left(\frac{\varphi_1(x_1^{d \leftarrow x_1^d + n - x^d})}{\varphi_1(x_1)} \middle\| \Phi_t(x)_{d,n} \right), \tag{20}$$

where t is a random variable in $(0,1)$ and one can sample $p^\star_{t,1}(x,x_1)$ by

$$p^\star_{0,t,1}(x_0,x,x_1) = p^\star_{0,1}(x_0,x_1)p^\star_{t|0,1}(x|x_0,x_1) \underset{(26)}{=} p^\star_{0,1}(x_0,x_1)p^r_{t|0,1}(x|x_0,x_1).$$

We can further rewrite the ratio in (20) as follows: [0]

$$\frac{\varphi_1(x_1^{d\leftarrow\triangle})}{\varphi_1(x_1)} \underset{(13)}{=} \frac{\nu(x_1^{d\leftarrow\triangle})}{\nu(x_1)} \Bigg/ \underbrace{\frac{\widehat{\varphi}_1(x_1^{d\leftarrow\triangle})}{\widehat{\varphi}_1(x_1)}}_{=\widehat{\Phi}^\star(x_1)_{d,\triangle}}. \tag{21}$$

Notably, here, we require the **discrete score** $x \mapsto \left(\frac{\nu(x^{d\leftarrow n})}{\nu(x)}\right)_{d,n}$ of the target distribution, which is a *first-order* oracle similar to the score of the target distribution in continuous AM.

Corrector Adjoint Matching From (20) and (21), to learn the controller $\Phi^\star_t$, we require estimating $\frac{\widehat{\varphi}_1(y)}{\widehat{\varphi}_1(x)}$ for all $d_{\mathrm{H}}(x,y) = 1$. Let the **corrector** matrix $\widehat{\Phi}^\star(x) \in \mathbb{R}^{D\times N}$ be defined by its entries $\widehat{\Phi}^\star(x)_{d,n} = \frac{\widehat{\varphi}_1(x^{d\leftarrow n})}{\widehat{\varphi}_1(x)}$. Following the spirit of (18), we can derive the following identity and variational characterization of $\widehat{\Phi}^\star$ (see App. B.7 for proof):

$$\frac{\widehat{\varphi}_1(z)}{\widehat{\varphi}_1(y)} = \sum_x p^\star_{t|1}(x|y) \frac{\widehat{\varphi}_t(x-y+z)}{\widehat{\varphi}_t(x)}, \ \forall t \in [0,1), \tag{22}$$

$$\implies \widehat{\Phi}^\star = \operatorname*{argmin}_{\widehat{\Phi}} \mathbb{E}_{p^\star_{t,1}(x,x_1)} \sum_{d=1}^{D} \sum_{n\neq x^d} D_f\left(\frac{\widehat{\varphi}_t(x^{d\leftarrow x^d+n-x_1^d})}{\widehat{\varphi}_t(x)} \middle\| \widehat{\Phi}(x_1)_{d,n} \right). \tag{23}$$

However, if $t \neq 0$, leveraging the relation (13) requires knowing the intractable $p^\star_t$; otherwise, as μ is known, and assume it is *fully supported* on $\mathcal{X}$, we have[0]

$$\frac{\widehat{\varphi}_0(x^{d\leftarrow\square})}{\widehat{\varphi}_0(x)} \underset{(13)}{=} \frac{\mu(x^{d\leftarrow\square})}{\mu(x)} \Bigg/ \underbrace{\frac{\varphi_0(x^{d\leftarrow\square})}{\varphi_0(x)}}_{=\Phi^\star_0(x)_{d,\square}}. \tag{24}$$

[0]For conciseness, $\triangle := x_1^d + n - x^d$ and $\square := x^d + n - x_1^d$.

Corrector Denoising Matching A limitation of (22) and (24) is that they require the explicit density of μ to be positive everywhere. To circumvent this, one can leverage the following identity:

$$\frac{\widehat{\varphi}_1(z)}{\widehat{\varphi}_1(y)} \underset{(12)}{=} \sum_x p^r_{1|t}(z|x)\frac{\widehat{\varphi}_t(x)}{\widehat{\varphi}_1(y)} \underset{(13)}{=} \sum_x \frac{p^r_{1|t}(z|x)}{p^r_{1|t}(y|x)} p^\star_{t|1}(x|y). \tag{25}$$

Notably, while (22) relies on the additive noise (17), (25) holds under general $p^r_{1|t}$. Thus, one can obtain a similar variational characterization of the corrector $\widehat{\Phi}^\star$ by replacing the regression target (i.e., the ratio) in (23) with $\frac{p^r_{1|t}(x_1^{d\leftarrow n}|x)}{p^r_{1|t}(x_1|x)}$. As this is related to the discrete score of the transition kernel $p^r_{1|t}(\cdot|x)$, we borrow the terminology in the continuous domain and call it **denoising matching (DM)**.

Alternating Update We can thus arrive at the core algorithm of **discrete ASBS**, following the practice in ASBS to learn $\Phi \approx \Phi^\star$ and $\widehat{\Phi} \approx \widehat{\Phi}^\star$ with an alternating update. We initialize $\widehat{\Phi}^{(0)}$ to be all-one following the practice in Liu et al. (2025a), and for stage $k = 1, 2, ...$, we solve the following two problems sequentially:

$$\Phi^{(k)} := \underset{\Phi}{\operatorname{argmin}}\ \mathbb{E}_t w_t \mathbb{E}_{\substack{p_{0,1}^{\mathrm{sg}(r\Phi)}(x_0,x_1)\\ p^r_{t|0,1}(x|x_0,x_1)}} \sum_{d=1}^{D}\sum_{n\neq x^d} D_f\left(\frac{\varphi_1(x_1^{d\leftarrow x_1^d+n-x^d})}{\varphi_1(x_1)}\middle\| \Phi_t(x)_{d,n}\right), \tag{ctrl-AM}$$

$$\widehat{\Phi}^{(k)} := \underset{\widehat{\Phi}}{\operatorname{argmin}}\ \mathbb{E}_{p_{0,1}^{\mathrm{sg}(r\Phi^{(k)})}(x_0,x_1)} \sum_{d=1}^{D}\sum_{n\neq x_1^d} D_f\left(\frac{\widehat{\varphi}_0(x_0^{d\leftarrow x_0^d+n-x_1^d})}{\widehat{\varphi}_0(x_0)}\middle\| \widehat{\Phi}(x_1)_{d,n}\right), \tag{corr-AM}$$

$$\widehat{\Phi}^{(k)} := \underset{\widehat{\Phi}}{\operatorname{argmin}}\ \mathbb{E}_t w_t \mathbb{E}_{\substack{p_{0,1}^{\mathrm{sg}(r\Phi^{(k)})}(x_0,x_1)\\ p^r_{t|0,1}(x|x_0,x_1)}} \sum_{d=1}^{D}\sum_{n\neq x_1^d} D_f\left(\frac{p^r_{1|t}(x_1^{d\leftarrow n}|x)}{p^r_{1|t}(x_1|x)}\middle\| \widehat{\Phi}(x_1)_{d,n}\right). \tag{corr-DM}$$

The stop gradient operator $\mathrm{sg}(\cdot)$ applied onto a model means not tracking the gradient when querying the model. In (ctrl-AM) and (corr-DM), $t \sim \mathrm{Unif}(0,1)$, $w_\cdot : [0,1] \to \mathbb{R}_+$ is a time weight function. For the two AM losses, the ratio in (ctrl-AM) is computed via (21) by replacing $\widehat{\Phi}^\star$ with the current $\mathrm{sg}(\widehat{\Phi}^{(k-1)})$, and the ratio in (corr-AM) is computed via (24) by replacing $\Phi^\star$ with the current $\mathrm{sg}(\Phi^{(k)})$. In all three losses, $p^{\mathrm{sg}(r\Phi)}$ means sampling from the CTMC with transition rate $u_t(x^{d\leftarrow n}, x) = r_t(x^{d\leftarrow n}, x)\,\mathrm{sg}(\Phi_t(x)_{d,n})$, $n \neq x^d$, which replaces $p^\star_{0,1}$ in the expectation with the detached non-optimal path measure. The theoretical justification of the validity of this replacement will be discussed in Thm. 5.2, and in practice, an optional **trajectory importance reweighting** using the RND $p^\star(x_{[0,1]})/p^{\mathrm{sg}(r\Phi)}(x_{[0,1]})$ can be incorporated (App. B.9).

Initialization Following ASBS (Liu et al., 2025a), one can initialize either the controller or the corrector to be non-informative (i.e., output all ones). In Prop. B.12, we prove that under the reference transition rate (16) and uniform initialization $\mu = p_{\mathrm{unif}}$, these two approaches are equivalent, and hence we always start with the all-one corrector.

Inference We use the **τ-leaping** method (Gillespie, 2001; Campbell et al., 2022; Lou et al., 2024) to sample each dimension's transition independently. Since $u_t(x^{d\leftarrow n}, x) = \frac{\gamma_t}{N}\Phi_t(x)_{d,n}$, $n \neq x^d$, this means we fix $\Phi_\tau(X_\tau)_{d,n}$ on the interval $\tau \in [t, t+h]$ for a small step size $h > 0$, and assume each dimension evolves independently. We defer the details to App. B.6.

5 Additional Theory and Insights of DASBS

In this section, we provide further theory and insights to the DASBS framework.

AM v.s. DM for Corrector Recall that our derivation of the AM losses (ctrl-AM) and (corr-AM) leverages the relations (19) and (22), which rely explicitly on the additiveness of the reference transition kernel $p^r_{1|t}(\cdot|x)$ (17). In contrast, the DM loss (corr-DM) does not rely on this condition.

AM v.s. DM for Controller A parallel relationship holds for the *controller*, which yields a different way of the alternating update by replacing the ratio in (ctrl-AM) with $\frac{p^r_{1|t}(x_1|x^{d\leftarrow n})}{p^r_{1|t}(x_1|x)}$ (see App. B.8 and (ctrl-DM) for details). However, though theoretically grounded, this formulation may suffer

from a weak *mutual supervisory signal* during alternating updates: we rely on the trained $\widehat{\Phi}^{(k-1)}$ to supervise the training of $\Phi^{(k)}$, but such supervisory information only comes in from the *implicit boundary relation* $\Phi_1(x)_{d,n}\widehat{\Phi}(x)_{d,n} = \frac{\nu(x^{d\leftarrow n})}{\nu(x)}$, which is not explicitly enforced in the loss. Even when the reference dynamics is *memoryless* and there is no need to learn the corrector, we discover in Fig. 1 that AM (ctrl-AM) works significantly better than its DM counterpart (ctrl-DM).

Further Connection to Target Matching Such contrast resembles the distinction between **target matching (TM)** and **denoising matching (DM)** in the literature of learning scores of continuous probability distributions (De Bortoli et al., 2024; Kahouli et al., 2025): for two continuous random vectors $x, y \sim p(x, y)$, we can express the score as

$$\underbrace{\mathbb{E}_{p(x|y)}\nabla_x \log p(x)}_{\text{for additive } p(y|x)} = \nabla_y \log p(y) = \underbrace{\mathbb{E}_{p(x|y)}\nabla_y \log p(y|x)}_{\text{for general } p(y|x)},$$

where **additive** means $p(y|x) = q(y - x)$ for some distribution q. In other words, TM regresses onto the target score, whereas DM regresses onto the score of the transition kernel. TM leverages a more meaningful learning signal and avoids the numerical instability of DM when the noise level is very small (where $\nabla_y \log p(y|x)$ becomes singular), thereby providing faster convergence of training, a benefit we observe directly in our discrete experiments (Fig. 1).

Unified View of Adjoint Matching **Adjoint matching** (**AM**, Domingo-Enrich et al. (2025a)) is originally derived by analyzing the ODE of the *adjoint state* – the gradient of the cost-to-go from $X_t = x$ with respect to x. This formulation connects the adjoint state to the gradient of the objective with respect to control parameters, yielding a learning objective whose **unique fixed-point** corresponds to the optimal control. However, such gradient-based perspective is not directly generalizable to discrete domains. Recently, So et al. (2026) made a first attempt by expressing the optimal control as an expectation under the optimal path measure; however, their reliance on *masked* transition rate resulted in a complex training objective, due to the lack of the additive property. In contrast, our derivation identifies that the **additive reference noise** is the key structural requirement that enables a TM-like loss in discrete domains. We refer readers to the lower half of Tab. 1 for a side-by-side comparison of DM and AM formulations across continuous and discrete settings, and conclude with a unified characterization of the intrinsic nature of AM:

*AM: a **fixed-point iteration** driven by a **target matching** objective converging to the **optimal** $\boldsymbol{p^\star}$.*

DASBS Unifies Existing Memoryless SOC Solvers We further establish the connection between DASBS and the **weighted denoising cross-entropy (WDCE)** loss for solving *memoryless* SOC problems (Zhu et al., 2025a;b) (see Props. B.9 and B.10 for full statement and proof):

Proposition 5.1. *Under a* memoryless *reference path measure* p^r *(for both uniform* (16) *and masked* (39) *discrete diffusion), the denoising loss for the controller* (ctrl-DM)*, with trajectory importance reweighting, generalized KL divergence,* [1] *and time weight* $w_t \leftarrow \frac{\gamma_t}{N}$*, is equivalent to the WDCE loss.*

Convergence Analysis Finally, following the continuous arguments (Liu et al., 2025a, Thm. 4), we establish the following convergence guarantee of DASBS.

Theorem 5.2. ***(1)*** *The path measure induced by the unique fixed-point of* (ctrl-AM) *($\Phi^{(k)}$) solves a **forward half bridge problem*** $\min_{p \text{ s.t. } p_0=\mu} \mathrm{KL}(p\|q^{\widehat{\varphi}_1^{(k-1)}})$ *for some path measure* $q^{\widehat{\varphi}_1^{(k-1)}}$ *induced by* $\widehat{\varphi}_1^{(k-1)}$ *(Def. B.11).* ***(2)*** *The unique fixed-point of* (corr-DM) *is the same as the unique fixed-point of* (corr-DM)*. Denote it as* $\widehat{\Phi}^{(k)}$*, which induces a path measure* $q^{\widehat{\varphi}_1^{(k)}}$ *that solves a **backward half bridge problem*** $\min_{q \text{ s.t. } q_1=\nu} \mathrm{KL}(p^{r\Phi^{(k)}}\|q)$.

Consequently, convergence is guaranteed by the theory of iterative proportional fitting (Rüschendorf, 1995; Chen et al., 2016a; De Bortoli et al., 2021). See App. B.11 for the proof.

Experiments We postpone the experiments to App. C.

[1] $f(t) = t\log t \implies D_f(a\|b) = a\log\frac{a}{b} - a + b$.

6 Conclusion and Future Work

In this work, we introduced a unified framework for discrete SB and SOC, proposing DASBS as an optimal-policy fixed-point iteration. By leveraging an additive noise scheme based on group structure, DASBS effectively extends AM to discrete domains. Several limitations remain: our evaluation is currently restricted to synthetic benchmarks, and performance on more complex distributions remains unknown. Additionally, the first-order nature of DASBS may be costly to implement if computing energy is expensive or not parallelizable. Future directions include extending the framework to discrete SOC with running costs (Domingo-Enrich et al., 2025a), exploring non-uniform reference dynamics like the Ehrenfest process (Winkler et al., 2024), and broadening our theoretical insights of AM to general state (e.g., Park et al. (2024); Woo et al. (2025); Park et al. (2025)).

References

Tara Akhound-Sadegh, Jungyoon Lee, Joey Bose, Valentin De Bortoli, Arnaud Doucet, Michael M. Bronstein, Dominique Beaini, Siamak Ravanbakhsh, Kirill Neklyudov, and Alexander Tong. Progressive inference-time annealing of diffusion models for sampling from Boltzmann densities. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=vf2GHcxzMV.

Michael Samuel Albergo and Eric Vanden-Eijnden. NETS: A non-equilibrium transport sampler. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=QqGw9StPbQ.

Vincent Beffara and Hugo Duminil-Copin. The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. *Probability Theory and Related Fields*, 153(3):511–542, 2012. doi: 10.1007/s00440-011-0353-8.

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 *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a. URL https://openreview.net/forum?id=6RlbOEcOS4.

Denis Blessing, Julius Berner, Lorenz Richter, and Gerhard Neumann. Underdamped diffusion bridges with applications to sampling. In *The Thirteenth International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=Q1QTxFm0Is.

Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov chain Monte Carlo*. CRC press, 2011.

Andrew Campbell, Joe Benton, Valentin De Bortoli, Thomas Rainforth, George Deligiannidis, and Arnaud Doucet. A continuous time framework for discrete denoising models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 28266–28279. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/b5b528767aa35f5b1a60fe0aaeca0563-Paper-Conference.pdf.

Hongrui Chen and Lexing Ying. Convergence analysis of discrete diffusion model: Exact implementation through uniformization. *Journal of Machine Learning*, 2025. doi: 10.4208/jml.240812.

Junhua Chen, Lorenz Richter, Julius Berner, Denis Blessing, Gerhard Neumann, and Anima Anandkumar. Sequential controlled Langevin diffusions. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=dImD2sgy86.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Tianrong Chen, Guan-Horng Liu, and Evangelos Theodorou. Likelihood training of Schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nioAdKCEdXB.

Yongxin Chen, Tryphon Georgiou, and Michele Pavon. Entropic and displacement interpolation: a computational approach using the Hilbert metric. *SIAM Journal on Applied Mathematics*, 76(6): 2375–2396, 2016a. doi: 10.1137/16M1061382. URL https://doi.org/10.1137/16M1061382.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169:671–691, 2016b. doi: 10.1007/s10957-015-0803-z.

Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger bridge. *SIAM Review*, 63(2):249–313, 2021. doi: 10.1137/20M1339982. URL https://doi.org/10.1137/20M1339982.

Jaemoo Choi, Yongxin Chen, Molei Tao, and Guan-Horng Liu. Non-equilibrium annealed adjoint sampler. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=ay7WDSq0Kb.

Paolo Dai Pra. A stochastic control approach to reciprocal diffusion processes. *Applied mathematics and Optimization*, 23(1):313–329, 1991. doi: 10.1007/BF01442404.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=9BnCwiXB0ty.

Valentin De Bortoli, Michael Hutchinson, Peter Wirnsberger, and Arnaud Doucet. Target score matching. *arXiv preprint arXiv:2402.08667*, 2024.

Carles Domingo-Enrich, Jiequn Han, Brandon Amos, Joan Bruna, and Ricky T. Q. Chen. Stochastic optimal control matching. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=wfU2CdgmWt.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=xQBRrtQM8u.

Carles Domingo-Enrich, Yuanqi Du, and Michael Samuel Albergo. A unified perspective on fine-tuning and sampling with diffusion and flow models, 2025b. URL https://openreview.net/forum?id=SEv6F8hxMC.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Yuanqi Du, Jiajun He, Francisco Vargas, Yuanqing Wang, Carla P Gomes, José Miguel Hernández-Lobato, and Eric Vanden-Eijnden. FEAT: Free energy estimators with adaptive transport. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=GQXeLGYMda.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021. URL http://jmlr.org/papers/v22/20-451.html.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 3 edition, 2013.

Daniel T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 07 2001. ISSN 0021-9606. doi: 10.1063/1.1378322. URL https://doi.org/10.1063/1.1378322.

Wei Guo, Jaemoo Choi, Yuchen Zhu, Molei Tao, and Yongxin Chen. Proximal diffusion neural sampler. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=XTHQqS7ObC.

Jiequn Han and Weinan E. Deep learning approximation for stochastic control problems. *arXiv preprint arXiv:1611.07422*, 2016.

Aaron J Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Daniel S. Levine, Brandon M Wood, Bin Hu, Brandon Amos, Brian Karrer, Xiang Fu, Guan-Horng Liu, and Ricky T. Q. Chen. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=6Eg1OrHmg2.

Jiajun He, Yuanqi Du, Francisco Vargas, Dinghuai Zhang, Shreyas Padhy, RuiKang OuYang, Carla Gomes, and José Miguel Hernández-Lobato. No trick, no treat: Pursuits and challenges towards simulation-free training of neural samplers. *arXiv preprint arXiv:2502.06685*, 2025.

Byeongho Heo, Song Park, Dongyoon Han, and Sangdoo Yun. Rotary position embedding for vision transformer. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision – ECCV 2024*, pp. 289–305, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-72684-2.

Peter Holderrieth, Michael Samuel Albergo, and Tommi Jaakkola. LEAPS: A discrete neural sampler via locally equivariant networks. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=Hq2RniQAET.

Samuel Howard, Nikolas Nüsken, and Jakiw Pidstrigach. Control consistency losses for diffusion bridges. *arXiv preprint arXiv:2512.05070*, 2025.

Khaled Kahouli, Romuald Elie, Klaus-Robert Müller, Quentin Berthet, Oliver T Unke, and Arnaud Doucet. Control variate score matching for diffusion models. *arXiv preprint arXiv:2512.20003*, 2025.

Frank P Kelly. *Reversibility and stochastic networks*. Cambridge University Press, 2011.

Sergei Kholkin, Francisco Vargas, and Alexander Korotin. Sampling from energy distributions with target concrete score identity. *arXiv preprint arXiv:2510.23106*, 2025.

Jun Hyeong Kim, Seonghwan Kim, Seokhyun Moon, Hyeongwoo Kim, Jeheon Woo, and Woo Youn Kim. Discrete diffusion Schrödinger bridge matching for graph transformation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=tQyh0gnfqW.

Grigoriy Ksenofontov and Alexander Korotin. Categorical Schrödinger bridge matching. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=RBly0nOr2h.

David P. Landau and Kurt Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 4 edition, 2014.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. ISSN 1078-0947. doi: 10.3934/dcds.2014.34.1533. URL https://www.aimsciences.org/article/id/d5bcf817-901d-4104-b7da-eade7847c53e.

Xuechen Li, Ting-Kam Leonard Wong, Ricky T. Q. Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 3870–3882. PMLR, 26–28 Aug 2020. URL https://proceedings.mlr.press/v108/li20i.html.

Guan-Horng Liu, Tianrong Chen, Oswin So, and Evangelos Theodorou. Deep generalized Schrödinger bridge. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=fp33Nsh0O5.

Guan-Horng Liu, Jaemoo Choi, Yongxin Chen, Benjamin Kurt Miller, and Ricky T. Q. Chen. Adjoint Schrödinger bridge sampler. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a. URL https://openreview.net/forum?id=rMhQBlhh4c.

Jun S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer Publishing Company, Incorporated, 2008. ISBN 0387763694.

Zhen Liu, Tim Z. Xiao, Carles Domingo-Enrich, Weiyang Liu, and Dinghuai Zhang. Value gradient guidance for flow matching alignment. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL https://openreview.net/forum?id=6MmOy2Ji8V.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32819–32848. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/lou24a.html.

Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, pp. 23–40. Springer, 2024. doi: 10.1007/978-3-031-72980-5_2.

Lars Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65:117–149, Feb 1944. doi: 10.1103/PhysRev.65.117. URL https://link.aps.org/doi/10.1103/PhysRev.65.117.

Jingyang Ou, Shen Nie, Kaiwen Xue, Fengqi Zhu, Jiacheng Sun, Zhenguo Li, and Chongxuan Li. Your absorbing discrete diffusion secretly models the conditional distributions of clean data. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=sMyXP8Tanm.

Zijing Ou, Ruixiang Zhang, and Yingzhen Li. Discrete neural flow samplers with locally equivariant transformer. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL https://openreview.net/forum?id=Wk65okms3T.

Byoungwoo Park, Jungwon Choi, Sungbin Lim, and Juho Lee. Stochastic optimal control for diffusion bridges in function spaces. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=WyQW4G57Zd.

Byoungwoo Park, Juho Lee, and Guan-Horng Liu. Functional adjoint sampler: Scalable sampling on infinite dimensional spaces. *arXiv preprint arXiv:2511.06239*, 2025.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182, 2023. doi: 10.1109/ICCV51070.2023.00387.

Angus Phillips, Hai-Dang Dau, Michael John Hutchinson, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Particle denoising diffusion sampler. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=vMUnnS4OWC.

Jakiw Pidstrigach, Elizabeth Louise Baker, Carles Domingo-Enrich, George Deligiannidis, and Nikolas Nüsken. Conditioning diffusions using Malliavin calculus. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=0A4JSAU3FD.

Lev Semenovich Pontryagin. *Mathematical theory of optimal processes*. Routledge, 1987. doi: 10.1201/9780203749319.

Yinuo Ren, Haoxuan Chen, Grant M. Rotskoff, and Lexing Ying. How discrete and continuous diffusion meet: Comprehensive analysis of discrete diffusion models via a stochastic integral framework. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=6awxwQEI82.

Yinuo Ren, Haoxuan Chen, Yuchen Zhu, Wei Guo, Yongxin Chen, Grant M. Rotskoff, Molei Tao, and Lexing Ying. Fast solvers for discrete diffusion models: Theory and applications of high-order algorithms. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL https://openreview.net/forum?id=OuklL6Q3sO.

Lorenz Richter and Julius Berner. Improved sampling via learned diffusions. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=h4pNROsO06.

Severi Rissanen, RuiKang OuYang, Jiajun He, Wenlin Chen, Markus Heinonen, Arno Solin, and José Miguel Hernández-Lobato. Progressive tempering sampler with diffusion. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=uBMnbCBEtZ.

Ludger Rüschendorf. Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, 23(4):1160 – 1174, 1995. doi: 10.1214/aos/1176324703. URL https://doi.org/10.1214/aos/1176324703.

Sebastian Sanokowski, Wilhelm Franz Berghammer, Sepp Hochreiter, and Sebastian Lehner. Variational annealing on graphs for combinatorial optimization. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=SLx7paoaTU.

Sebastian Sanokowski, Sepp Hochreiter, and Sebastian Lehner. A diffusion model framework for unsupervised neural combinatorial optimization. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 43346–43367. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/sanokowski24a.html.

Sebastian Sanokowski, Wilhelm Franz Berghammer, Haoyu Peter Wang, Martin Ennemoser, Sepp Hochreiter, and Sebastian Lehner. Scalable discrete diffusion samplers: Combinatorial optimization and statistical physics. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=peNgxpbdxB.

Sebastian Sanokowski, Lukas Gruber, Christoph Bartmann, Sepp Hochreiter, and Sebastian Lehner. Rethinking losses for diffusion bridge samplers. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL https://openreview.net/forum?id=O58KDUfB4x.

Yair Schiff, Subham Sekhar Sahoo, Hao Phung, Guanghan Wang, Sam Boshar, Hugo Dalla-torre, Bernardo P de Almeida, Alexander M Rush, Thomas PIERROT, and Volodymyr Kuleshov. Simple guidance mechanisms for discrete diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=i5MrJ6g5G1.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=qy07OHsJT5.

Oswin So, Brian Karrer, Chuchu Fan, Ricky T. Q. Chen, and Guan-Horng Liu. Discrete adjoint matching. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=VXB4xxAgOf.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Robert H. Swendsen and Jian-Sheng Wang. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett.*, 57:2607–2609, Nov 1986. doi: 10.1103/PhysRevLett.57.2607. URL https://link.aps.org/doi/10.1103/PhysRevLett.57.2607.

Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, Jan 1987. doi: 10.1103/PhysRevLett.58.86. URL https://link.aps.org/doi/10.1103/PhysRevLett.58.86.

Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Herve Jegou. Training data-efficient image transformers & distillation through attention. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10347–10357. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/touvron21a.html.

Francisco Vargas, Will Sussman Grathwohl, and Arnaud Doucet. Denoising diffusion samplers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=8pvnfTAbu1f.

Francisco Vargas, Shreyas Padhy, Denis Blessing, and Nikolas Nüsken. Transport meets variational inference: Controlled Monte Carlo diffusions. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=PP1rudnxiW.

Ludwig Winkler, Lorenz Richter, and Manfred Opper. Bridging discrete and continuous state spaces: Exploring the Ehrenfest process in time-continuous diffusion models. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53017–53038. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/winkler24a.html.

Dongyeop Woo, Minsu Kim, Minkyu Kim, Kiyoung Seong, and Sungsoo Ahn. Energy-based generator matching: A neural sampler for general state space. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=UKJkad8aUF.

Leo Zhang, Peter Potaptchik, Jiajun He, Yuanqi Du, Arnaud Doucet, Francisco Vargas, Hai-Dang Dau, and Saifuddin Syed. Accelerated parallel tempering via neural transports. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=CODnlyYUli.

Qinsheng Zhang and Yongxin Chen. Path integral sampler: A stochastic control approach for sampling. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=_uCb2ynRu7Y.

Ruixiang Zhang, Shuangfei Zhai, Yizhe Zhang, James Thornton, Zijing Ou, Joshua M. Susskind, and Navdeep Jaitly. Target concrete score matching: A holistic framework for discrete diffusion. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=ZMrdvSm7xi.

Yuchen Zhu, Wei Guo, Jaemoo Choi, Guan-Horng Liu, Yongxin Chen, and Molei Tao. MDNS: Masked diffusion neural sampler via stochastic optimal control. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a. URL https://openreview.net/forum?id=xIH95kXNR2.

Yuchen Zhu, Wei Guo, Jaemoo Choi, Petr Molodyk, Bo Yuan, Molei Tao, and Yongxin Chen. Enhancing reasoning for diffusion LLMs via distribution matching policy optimization. *arXiv preprint arXiv:2510.08233*, 2025b.

A Related Works

1. Adjoint Matching The idea of AM (Domingo-Enrich et al., 2025a) can be traced back to the adjoint method in optimal control theory (Pontryagin, 1987), and a few of its earlier applications in machine learning (Han & E, 2016; Chen et al., 2018; Li et al., 2020; Domingo-Enrich et al., 2024). The core idea is to define the adjoint state $a_t(X; u)$ associated with a control u and trajectory $X = (X_t)_{t\in[0,1]}$ as $\nabla_{X_t} J_t(X)$, where $J_t(X)$ the cost-to-go along the trajectory starting from X_t. Then, one can establish the dynamics of the adjoint state as an ODE backward in time starting from a_1 being the gradient of the terminal cost. Furthermore, the gradient of the full cost-to-go at the initial time $t = 0$ with respect to the control parameter can be written as an integral associated with the adjoint state, thus leading to a matching objective using the stop-gradient operator. It is shown in Domingo-Enrich et al. (2025a) that the optimal control is the unique fixed-point of the learning objective. AM-based methods have been applied to multiple tasks including continuous neural sampler training (Havens et al., 2025; Liu et al., 2025a; Choi et al., 2025; Blessing et al., 2025a), fine-tuning diffusion/flow-based models (Blessing et al., 2025a; Liu et al., 2025b; Domingo-Enrich et al., 2025b), and transition path sampling (Pidstrigach et al., 2025; Howard et al., 2025), showing competitive performance.

While AM is popularly applied in continuous state spaces, its generalization to discrete state spaces remains unexplored. Recently, So et al. (2026) made a first attempt by defining the discrete adjoint state as an estimator of the exponential of the value difference, and establishing its connection with the optimal transition rate, leading to a matching objective whose unique fixed-point is the optimal transition rate. However, their training loss requires row-out from intermediate states given the samples in the buffer, which is complicated.

2. Discrete Diffusion Neural Samplers The study of continuous neural samplers has inspired the construction of similar algorithms for learning discrete distributions. For uniform-based discrete diffusion, Holderrieth et al. (2025) employed the discrete Jarzynski equality and learns the transport via locally equivariant neural networks, Ou et al. (2025b) followed and refined this line of study by proposing a different loss formulation, and Kholkin et al. (2025) proposed a neural sampler based on target concrete score identity. For mask-based discrete diffusion, Zhu et al. (2025a) formulated the sampling problem as an SOC problem similar to Zhang & Chen (2022), and proposed samplers based on masked and uniform discrete diffusion. Finally, discrete diffusion neural samplers are also used for solving combinatorial optimization problems via sampling from a low-temperature target distribution, e.g., Sanokowski et al. (2023; 2024; 2025a); Ou et al. (2025b); Guo et al. (2026).

B Theoretical Derivation

B.1 Kolmogorov Equations for Continuous-time Markov Chains

Proposition B.1. *Suppose $p_{t|s}(y|x)$, $0 \le s < t \le 1$ are the transition kernels of a CTMC with transition rate r. Then the following equations hold:*

- ***Kolmogorov forward equation (KFE):***

$$\partial_t p_{t|s}(y|x) = \sum_z r_t(y, z) p_{t|s}(z|x), \tag{KFE}$$

- ***Kolmogorov backward equation (KBE):***

$$\partial_s p_{t|s}(y|x) = -\sum_z r_s(z, x) p_{t|s}(y|z). \tag{KBE}$$

These are standard results in the theory of CTMC and can be proved by leveraging the definition of the transition rate.

B.2 Schrödinger Bridge: Proof of Thm. 3.1

We refer readers to Léonard (2014) for a comprehensive review of the SB theory. For completeness, we provide a self-contained proof here.

Lemma B.2. *There exists some non-negative functions ϕ, φ_1 such that the optimal path measure $p^\star$ satisfies*

$$p^\star_{(0,1)|0,1} = p^r_{(0,1)|0,1}, \tag{26}$$

$$p^\star_{0,1}(x,y) = p^r_{0,1}(x,y)\phi(x)\varphi_1(y), \quad \textit{s.t. } p^\star_0 = \mu,\ p^\star_1 = \nu. \tag{27}$$

In other words, $p^\star(X_{[0,1]}) = p^r(X_{[0,1]})\phi(X_0)\varphi_1(X_1)$.

Proof. By the chain rule of KL divergence, we have

$$\mathrm{KL}(p^u \| p^r) = \mathrm{KL}(p^u_{0,1} \| p^r_{0,1}) + \mathbb{E}_{p^u_{0,1}(x_0,x_1)} \mathrm{KL}(p^u_{(0,1)|0,1}(\cdot|x_0,x_1) \| p^r_{(0,1)|0,1}(\cdot|x_0,x_1)).$$

Therefore, the optimal $p^u_{(0,1)|0,1}$ is $p^r_{(0,1)|0,1}$, and we only need to solve the following *static* SB problem:

$$\begin{aligned} &\min_{p_{0,1}} \ \mathrm{KL}(p_{0,1} \| p^r_{0,1}) \\ &\text{s.t. } p_0 = \mu,\ p_1 = \nu \iff \sum_y p_{0,1}(x,y) = \mu(x),\ \sum_x p_{0,1}(x,y) = \nu(y). \end{aligned}$$

Let the Lagrangian multiplier functions for the above constraints be $\alpha(x)$ and $\beta(y)$, respectively. The Lagrangian function is

$$\begin{aligned} L(p_{0,1}, \alpha, \beta) = &\sum_{x,y} p_{0,1}(x,y) \log \frac{p_{0,1}(x,y)}{p^r_{0,1}(x,y)} \\ &+ \sum_x \alpha(x) \left(\sum_y p_{0,1}(x,y) - \mu(x) \right) + \sum_y \beta(y) \left(\sum_x p_{0,1}(x,y) - \nu(y) \right). \end{aligned}$$

Set the partial derivative with respect to $p_{0,1}(x,y)$ to zero, we have

$$\log \frac{p_{0,1}(x,y)}{p^r_{0,1}(x,y)} + 1 + \alpha(x) + \beta(y) = 0 \implies p^\star_{0,1}(x,y) = p^r_{0,1}(x,y) \mathrm{e}^{-1-\alpha(x)-\beta(y)}.$$

Therefore, by defining $\phi(x) := \mathrm{e}^{-1-\alpha(x)}$ and $\varphi_1(y) := \mathrm{e}^{-\beta(y)}$, we complete the proof. Note that ϕ and φ_1 are defined up to a constant scaling factor. □

Lemma B.3. *Define $\widehat{\varphi}_0 := \phi p^r_0$, and define the SB potentials φ_t and $\widehat{\varphi}_t$ through the following relation:*

$$\varphi_t(x) = \sum_y p^r_{1|t}(y|x)\varphi_1(y), \quad \widehat{\varphi}_t(x) = \sum_y p^r_{t|0}(x|y)\widehat{\varphi}_0(y). \tag{28}$$

Then,

$$\partial_t \varphi_t(x) = -\sum_y \varphi_t(y) r_t(y,x), \tag{29}$$

$$\partial_t \widehat{\varphi}_t(x) = \sum_y \widehat{\varphi}_t(y) r_t(x,y), \tag{30}$$

and furthermore, (11) *and* (12) *hold.*

Proof. First, we prove (29) and (30):

$$\begin{aligned} \partial_t \varphi_t(x) &= \sum_y \partial_t p^r_{1|t}(y|x)\varphi_1(y) \underset{\text{(KBE)}}{=} -\sum_y \sum_z r_t(z,x) p^r_{1|t}(y|z)\varphi_1(y) = -\sum_z r_t(z,x)\varphi_t(z), \\ \partial_t \widehat{\varphi}_t(x) &= \sum_y \partial_t p^r_{t|0}(x|y)\widehat{\varphi}_0(y) \underset{\text{(KFE)}}{=} \sum_y \sum_z r_t(x,z) p^r_{t|0}(z|y)\widehat{\varphi}_0(y) = \sum_z r_t(x,z)\widehat{\varphi}_t(z). \end{aligned}$$

Next, we prove (11) and (12) by verifying that the partial derivatives to t in (11) and to s in (12) are zero:

$$\sum_y \partial_t(p^r_{t|s}(y|x)\varphi_t(y)) = \sum_y\sum_z r_t(y,z)p^r_{t|s}(z|x)\varphi_t(y) - \sum_y\sum_z p^r_{t|s}(y|x)r_t(z,y)\varphi_t(z) = 0,$$

$$\sum_y \partial_s(p^r_{t|s}(x|y)\widehat{\varphi}_s(y)) = -\sum_y\sum_z r_s(z,y)p^r_{t|s}(x|z)\widehat{\varphi}_s(y) + \sum_y\sum_z p^r_{t|s}(x|y)r_s(y,z)\widehat{\varphi}_s(z) = 0.$$

□

Proof of (13). We first study the marginal distribution at an arbitrary time point $t \in [0,1]$:

$$\begin{aligned} p^\star_t(x) &= \sum_{x_0,x_1} p^\star_{0,1}(x_0,x_1)p^\star_{t|0,1}(x|x_0,x_1) \underset{\text{(26) and (27)}}{=} \sum_{x_0,x_1} p^r_{0,1}(x_0,x_1)\phi(x_0)\varphi_1(x_1)p^r_{t|0,1}(x|x_0,x_1) \\ &= \sum_{x_0} p^r_0(x_0)p^r_{t|0}(x|x_0)\phi(x_0)\sum_{x_1} p^r_{1|t}(x_1|x)\varphi_1(x_1) \underset{(28)}{=} \widehat{\varphi}_t(x)\varphi_t(x). \end{aligned}$$

Next, we study the joint distribution at two arbitrary time points $0 \le s < t \le 1$:

$$\begin{aligned} p^\star_{s,t}(x,y) &= \sum_{x_0,x_1} p^\star_{0,1}(x_0,x_1)p^\star_{s,t|0,1}(x,y|x_0,x_1) \\ &\underset{\text{(26) and (27)}}{=} \sum_{x_0,x_1} p^r_{0,1}(x_0,x_1)\phi(x_0)\varphi_1(x_1)p^r_{s,t|0,1}(x,y|x_0,x_1) \\ &= \sum_{x_0} p^r_{0,s,t,1}(x_0,x,y,x_1)\phi(x_0)\varphi_1(x_1) \\ &= \sum_{x_0} p^r_0(x_0)p^r_{s|0}(x|x_0)\phi(x_0)\sum_{x_1} p^r_{t|s}(y|x)p^r_{1|t}(x_1|y)\varphi_1(x_1) \\ &\underset{(28)}{=} \widehat{\varphi}_s(x)p^r_{t|s}(y|x)\varphi_t(y). \end{aligned}$$

Therefore, the second and third equalities in (13) follow immediately.

Proof of (14). This is obvious from Lem. B.2 and (13):

$$p^\star(X_{[0,1]}) = p^r(X_{[0,1]})\phi(X_0)\varphi_1(X_1) = p^r(X_{[0,1]})\frac{\widehat{\varphi}_0(X_0)}{\mu(X_0)}\varphi_1(X_1) = p^r(X_{[0,1]})\frac{\widehat{\varphi}_0(X_0)}{\mu(X_0)}\frac{\nu(X_1)}{\widehat{\varphi}_1(X_1)}.$$

□

Proof of (10). From (13), for $y \neq x$,

$$u^\star_t(y,x) = \lim_{h\to 0}\frac{p^\star_{t+h|t}(y|x)}{h} = \lim_{h\to 0}\frac{p^r_{t+h|t}(y|x)}{h}\frac{\varphi_{t+h}(y)}{\varphi_t(x)} = \frac{\varphi_t(y)}{\varphi_t(x)}\lim_{h\to 0}\frac{p^r_{t+h|t}(y|x)}{h} = \frac{\varphi_t(y)}{\varphi_t(x)}r_t(y,x).$$

□

B.3 Stochastic Optimal Control: Proof of (15)

Proof. Using the chain rule of KL divergence and as $p^u_0 = p^r_0 = \mu$, we have

$$\mathrm{KL}(p^u\|p^r) + \mathbb{E}_{X\sim p^u} g(X_1) = \mathbb{E}_{\mu(X_0)}\left[\mathbb{E}_{p^u(X_{(0,1]}|X_0)}\log\frac{p^u(X_{(0,1]}|X_0)}{p^r(X_{(0,1]}|X_0)\mathrm{e}^{-g(X_1)}}\right].$$

Therefore, for any X_0, the optimal $p^u(X_{(0,1]}|X_0)$ is

$$p^\star(X_{(0,1]}|X_0) = \frac{1}{Z(X_0)}p^r(X_{(0,1]}|X_0)\mathrm{e}^{-g(X_1)},$$

where

$$Z(X_0) = \mathbb{E}_{p^r(X_{(0,1]}|X_0)\mathrm{e}^{-g(X_1)}} = \mathbb{E}_{p^r_{1|0}(y|X_0)}\mathrm{e}^{-g(y)}.$$

Finally, combining this with $p^\star_0 = p^r_0 = \mu$ completes the proof. □

B.4 Proof of Thm. 3.2

Proof. See Apps. B.2 and B.3 for the proofs of (14) and (15), respectively. To conclude the proof, it suffices to show the equivalence when $g \leftarrow \log \frac{\widehat{\varphi}_1}{\nu}$:

$$Z(x) \underset{(15)}{=} \sum_y \frac{p^r_{1|0}(y|x)}{\widehat{\varphi}_1(y)}\nu(y) \underset{(13)}{=} \sum_y \frac{p^\star_{0|1}(x|y)}{\widehat{\varphi}_0(x)}\nu(y) = \sum_y \frac{p^\star_{0,1}(x,y)}{\widehat{\varphi}_0(x)} = \frac{\mu(x)}{\widehat{\varphi}_0(x)}.$$

□

Corollary B.4. *The problem* (SOC) *with terminal cost g is equivalent to the problem* (SB) *with terminal distribution*

$$\nu(x) = \mathrm{e}^{-g(x)} \sum_y p^r_{1|0}(x|y)\frac{\mu(y)}{Z(y)}, \qquad \textit{where } Z(y) = \mathbb{E}_{p^r_{1|0}(\cdot|y)}\mathrm{e}^{-g}.$$

Proof. It suffices to express ν by g by comparing (14) and (15). We have $\mu = \widehat{\varphi}_0 Z$ and $\nu = \widehat{\varphi}_1 \mathrm{e}^{-g}$. From (12), $\widehat{\varphi}_1(x) = \sum_y p^r_{1|0}(x|y)\widehat{\varphi}_0(y)$. Combining all these three equations completes the proof. □

B.5 Uniform Reference Dynamics

We write $r_t(y,x) = \gamma_t r(y,x)$ where

$$r(y,x) = \begin{cases} \frac{1}{N}, & \text{if } d_{\mathrm{H}}(y,x) = 1, \\ -D\left(1 - \frac{1}{N}\right), & \text{if } y = x, \\ 0, & \text{if otherwise.} \end{cases}$$

By verifying the detailed balance condition, one is easy to see that this transition rate keeps $p_{\mathrm{unif}} = \mathrm{Unif}(\mathcal{X})$ invariant:

$$p_{\mathrm{unif}}(x)r_t(y,x) = p_{\mathrm{unif}}(y)r_t(x,y) = \frac{\gamma_t}{N^{D+1}}1_{d_{\mathrm{H}}(y,x)=1}, \ \forall x \neq y.$$

Proposition B.5. *Define $\overline{\gamma}_{s,t} := \int_s^t \gamma_u \mathrm{d}u$. Then for $x, y \in \mathcal{X}$ and $0 \le s < t \le 1$, the reference transition probability is*

$$p^r_{t|s}(y|x) = A(s,t)^{d_{\mathrm{H}}(x,y)}B(s,t)^{D-d_{\mathrm{H}}(x,y)}, \textit{ where } A(s,t) := \frac{1-\mathrm{e}^{-\overline{\gamma}_{s,t}}}{N}, \ B(s,t) := \frac{1+(N-1)\mathrm{e}^{-\overline{\gamma}_{s,t}}}{N}. \tag{31}$$

Remark. As a corollary, $p^r_{1|t}(y|x) = q_t(y-x)$ where $q_t(\varepsilon) = A(t,1)^{d_{\mathrm{H}}(\varepsilon,0)}B(t,1)^{D-d_{\mathrm{H}}(\varepsilon,0)}$.

Proof. Note that each dimension evolves independently with transition rate matrix $\gamma_t \boldsymbol{R}$ where $\boldsymbol{R} \in \mathbb{R}^{N\times N}$ has off-diagonal entries $\frac{1}{N}$ and diagonal entries $\frac{1}{N}-1$, i.e., $\boldsymbol{R} = \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}} - \boldsymbol{I}$. For $x \in [N]$, the vector $\boldsymbol{p}^r_{t|s}(\cdot|x) := (p^r_{t|s}(y|x))_{y\in[N]}$ satisfies the Kolmogorov forward equation $\partial_t \boldsymbol{p}^r_{t|s}(\cdot|x) = \gamma_t \boldsymbol{R}\boldsymbol{p}^r_{t|s}(\cdot|x)$ with initial condition $\boldsymbol{p}^r_{s|s}(\cdot|x) = \boldsymbol{e}^x$, i.e., the one-hot vector with 1 at the x-th entry. Therefore, we have $\boldsymbol{p}^r_{t|s}(\cdot|x) = \mathrm{e}^{\overline{\gamma}_{s,t}\boldsymbol{R}}\boldsymbol{e}^x$.

The matrix exponential can be computed as follows:

$$\mathrm{e}^{s\mathbf{1}\mathbf{1}^{\mathrm{T}}} = \sum_{k=0}^{\infty}\frac{s^k}{k!}(\mathbf{1}\mathbf{1}^{\mathrm{T}})^k = \boldsymbol{I} + \sum_{k=1}^{\infty}\frac{s^k}{k!}N^{k-1}\mathbf{1}\mathbf{1}^{\mathrm{T}} = \boldsymbol{I} + \frac{\mathrm{e}^{sN}-1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}},$$

$$\Longrightarrow \mathrm{e}^{\lambda\boldsymbol{R}} = \mathrm{e}^{\frac{\lambda}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}-\lambda\boldsymbol{I}} = \mathrm{e}^{-\lambda}\mathrm{e}^{\frac{\lambda}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}} = \mathrm{e}^{-\lambda}\left(\boldsymbol{I} + \frac{\mathrm{e}^{\lambda}-1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right) = \mathrm{e}^{-\lambda}\boldsymbol{I} + \frac{1-\mathrm{e}^{-\lambda}}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}.$$

Therefore, the per-dimension transition probability for $x, y \in [N]$ is

$$p^r_{t|s}(y|x) = \begin{cases} \frac{1-\mathrm{e}^{-\overline{\gamma}_{s,t}}}{N} =: A(s,t) & \text{if } y \neq x, \\ \frac{1+(N-1)\mathrm{e}^{-\overline{\gamma}_{s,t}}}{N} =: B(s,t) & \text{if } y = x, \end{cases} \tag{32}$$

which implies the full transition probability for $x, y \in \mathcal{X}$ is (31).

□

Remark. Note that when $\overline{\gamma}_{0,1} = \infty$, we have $p^r_{1|0}(y|x) = \frac{1}{N^D}$ for all $x, y \in \mathcal{X}$, i.e., the reference dynamics is memoryless. Under the assumption that $p^r_0 = p_{\text{unif}}$, $p^r_t = p_{\text{unif}}$ and thus $p^r_{t|1}(\cdot|y)$ can be implemented as follows: for each entry of y, independently, with probability $1 - \mathrm{e}^{-\overline{\gamma}_{t,1}}$, replace it with a random state from $\mathrm{Unif}[N]$.

Proposition B.6. *Under the reference dynamics p^r, each dimension of $p^r_{t|0,1}(x|x_0, x_1)$ for $x, x_0, x_1 \in \mathcal{X}$ is independent, and the per-dimensional distribution is given by*

$$p^r_{t|0,1}(x|x_0, x_1) = \begin{cases} \text{if } x_0 \neq x_1: & \begin{cases} \frac{A(0,t)A(t,1)}{A(0,1)}, & \text{if } x \notin \{x_0, x_1\}, \\ \frac{B(0,t)A(t,1)}{A(0,1)}, & \text{if } x = x_0, \\ \frac{A(0,t)B(t,1)}{A(0,1)}, & \text{if } x = x_1, \end{cases} \\ \text{if } x_0 = x_1: & \begin{cases} \frac{A(0,t)A(t,1)}{B(0,1)}, & \text{if } x \neq x_0 = x_1, \\ \frac{B(0,t)B(t,1)}{B(0,1)}, & \text{if } x = x_0 = x_1, \end{cases} \end{cases} \tag{33}$$

for $x, x_0, x_1 \in [N]$.

Proof. Due to the Markov property of p^r, we have

$$p^r_{t|0,1}(x|x_0, x_1) = \frac{p^r_{0,t,1}(x_0, x, x_1)}{p^r_{0,1}(x_0, x_1)} = \frac{p^r_{t|0}(x|x_0)p^r_{1|t}(x_1|x)}{p^r_{1|0}(x_1|x_0)}.$$

Note that again, each dimension is independent under p^r. Using the per-dimension transition probability (32), one can easily obtain the desired result. □

Choice of Noise Schedule γ_t We consider the following noise schedules:

- Constant schedule: $\gamma_t \equiv \gamma > 0$, $\overline{\gamma}_{s,t} = \gamma(t - s)$.
- Modified log-linear schedule: $\gamma_t = \frac{\gamma}{t+\alpha}$ for some $\gamma, \alpha > 0$, $\overline{\gamma}_{s,t} = \gamma \log \frac{t+\alpha}{s+\alpha}$. A larger α means a stronger memory effect, and $\alpha = 0$ recovers the memoryless case.

The ablation studies of the noise schedules can be found at Figs. 2 and 3.

B.6 Inference via τ-leaping

Proposition B.7. *Using τ-leaping to discretize the controlled CTMC, the transition probability for any dimension d is approximated as*

$$\Pr(X^d_{t+h} = n | X_t = x) \approx \begin{cases} \frac{\overline{\gamma}_{t,t+h}}{N} \Phi_t(x)_{d,n}, & \text{if } n \neq x^d, \\ 1 - \frac{\overline{\gamma}_{t,t+h}}{N} \sum_{n' \neq x^d} \Phi_t(x)_{d,n'}, & \text{if } n = x^d. \end{cases}$$

The approximate transition probability can be computed as

$$\begin{aligned} p^u_{t+h|t}(y|x) &\approx \prod_{d=1}^{D} \Pr(X^d_{t+h} = y^d | X_t = x) \\ &= \prod_{d=1}^{D} \left(\frac{\overline{\gamma}_{t,t+h}}{N} \Phi_t(x)_{d,y^d} \right)^{1_{y^d \neq x^d}} \left(1 - \frac{\overline{\gamma}_{t,t+h}}{N} \sum_{n' \neq x^d} \Phi_t(x)_{d,n'} \right)^{1_{y^d = x^d}}. \end{aligned} \tag{34}$$

We summarize the sampling procedure in Alg. 1. Note that for uniform discrete diffusion models, other samplers such as uniformization (Chen & Ying, 2025; Ren et al., 2025a) and higher-order τ-leaping (Ren et al., 2025b) can also be applied, which we leave as future works.

Algorithm 1 Sampling of discrete adjoint Schrödinger bridge sampler (DASBS) via τ-leaping

Require: Model Φ, initial distribution μ, time discretization $0 = t_0 < t_1 < ... < t_M = 1$.
1: Sample initial state $x_0 \sim \mu$.
2: **for** $i = 0$ **to** $M - 1$ **do**
3: Query $\Phi_{t_i}(x_i) \in \mathbb{R}_+^{D\times N}$ and compute $\overline{\gamma}_{t_i,t_{i+1}}$.
4: For each $d \in [D]$ (in parallel), independently sample

$$\Pr(x_{i+1}^d = n) = \begin{cases} \frac{\overline{\gamma}_{t_i,t_{i+1}}}{N}\Phi_{t_i}(x_i)_{d,n}, & \text{if } n \neq x_i^d, \\ 1 - \frac{\overline{\gamma}_{t_i,t_{i+1}}}{N}\sum_{n'\neq x_i^d}\Phi_{t_i}(x_i)_{d,n'}, & \text{if } n = x_i^d. \end{cases}$$

5: **end for**
output x_M as the generated sample.

B.7 Target Matching Loss

Proof of (23).

$$\widehat{\varphi}_1(z) \underset{(12)}{=} \sum_{x\in\mathbb{Z}_N^D} p^r_{1|t}(z|x)\widehat{\varphi}_t(x) \underset{(17)}{=} \sum_{x\in\mathbb{Z}_N^D} p^r_{1|t}(z+\Delta|x+\Delta)\widehat{\varphi}_t(x)$$

$$\underset{x'\leftarrow x+\Delta}{=} \sum_{x'\in\mathbb{Z}_N^D} p^r_{1|t}(z+\Delta|x')\widehat{\varphi}_t(x'-\Delta) \underset{\Delta\leftarrow y-z}{=} \sum_{x'\in\mathbb{Z}_N^D} p^r_{1|t}(y|x')\widehat{\varphi}_t(x'-y+z),$$

$$\implies \frac{\widehat{\varphi}_1(z)}{\widehat{\varphi}_1(y)} = \sum_x p^r_{1|t}(y|x)\frac{\widehat{\varphi}_t(x-y+z)}{\widehat{\varphi}_1(y)} \underset{(13)}{=} \sum_x p^\star_{t|1}(x|y)\frac{\widehat{\varphi}_t(x-y+z)}{\widehat{\varphi}_t(x)}.$$

□

B.8 Denoising Matching Loss

We first prove the denoising matching characterization of the controller:

$$\frac{\varphi_t(y)}{\varphi_t(x)} \underset{(11)}{=} \sum_z p^r_{1|t}(z|y)\frac{\varphi_1(z)}{\varphi_t(x)} = \sum_z \frac{p^r_{1|t}(z|y)}{p^r_{1|t}(z|x)}\frac{p^r_{1|t}(z|x)\varphi_1(z)}{\varphi_t(x)} \underset{(13)}{=} \sum_z \frac{p^r_{1|t}(z|y)}{p^r_{1|t}(z|x)}p^\star_{1|t}(z|x), \tag{35}$$

$$\implies \Phi^\star = \operatorname*{argmin}_{\Phi} \mathbb{E}_t\mathbb{E}_{p^\star_{t,1}(x,x_1)}\sum_{d=1}^D\sum_{n\neq x^d} D_f\left(\frac{p^r_{1|t}(x_1|x^{d\leftarrow n})}{p^r_{1|t}(x_1|x)}\middle\|\Phi_t(x)_{d,n}\right). \tag{36}$$

Thus, the denoising matching loss for the controller reads

$$\Phi^{(k)} := \operatorname*{argmin}_{\Phi} \mathbb{E}_t w_t \mathbb{E}_{\substack{p_{0,1}^{\mathrm{sg}(r\Phi)}(x_0,x_1)\\ p^r_{t|0,1}(x|x_0,x_1)}}\sum_{d=1}^D\sum_{n\neq x^d} D_f\left(\frac{p^r_{1|t}(x_1|x^{d\leftarrow n})}{p^r_{1|t}(x_1|x)}\middle\|\Phi_t(x)_{d,n}\right). \tag{ctrl-DM}$$

Moreover, the variational characterization of $\widehat{\Phi}^\star$ in the style of denoising matching is as follows:

$$\widehat{\Phi}^\star = \operatorname*{argmin}_{\widehat{\Phi}} \mathbb{E}_t\mathbb{E}_{p^\star_{t,1}(x,x_1)}\sum_{d=1}^D\sum_{n\neq x^d} D_f\left(\frac{p^r_{1|t}(x_1^{d\leftarrow n}|x)}{p^r_{1|t}(x_1|x)}\middle\|\widehat{\Phi}(x_1)_{d,n}\right).$$

B.9 Trajectory Importance Reweighting

By **trajectory importance reweighting**, we refer to the practice when we use the RND $\frac{p^\star(x_{[0,1]})}{p^u(x_{[0,1]})}$ in computing the losses (ctrl-AM) to (corr-DM). For instance, (corr-AM) becomes the following form:

$$\Phi^{(k)} := \operatorname*{argmin}_{\Phi}\ \mathbb{E}_t w_t \boxed{\mathbb{E}_{p^{\mathrm{sg}(r\Phi)}(x_{[0,1]})}\frac{p^\star(x_{[0,1]})}{p^{\mathrm{sg}(r\Phi)}(x_{[0,1]})}}\mathbb{E}_{p^r_{t|0,1}(x|x_0,x_1)}\sum_{d=1}^D\sum_{n\neq x^d} D_f\left(\frac{\varphi_1(x_1^{d\leftarrow x_1^d+n-x^d})}{\varphi_1(x_1)}\middle\|\Phi_t(x)_{d,n}\right).$$

In practice, for stability, after obtaining the log RND $\log \frac{p^\star(x_{[0,1]})}{p^{\mathrm{sg}(r\Phi)}(x_{[0,1]})}$ up to an additive constant over a batch of trajectories, one typically apply *softmax* to normalize the sum of weights to one.

Leveraging Props. B.5 and B.7, we can approximately compute $\log \frac{p^u(x_{[0,1]})}{p^r(x_{[0,1]})}$. With the approximation in the following Lem. B.8, we can approximate the log weights for importance sampling, $\log \frac{p^\star(x_{[0,1]})}{p^u(x_{[0,1]})}$.

Lemma B.8. *When the reference dynamics is memoryless,* $\log \frac{p^\star(x_{[0,1]})}{p^r(x_{[0,1]})} = \log \frac{\nu(x_1)}{p_1^r(x_1)} + \mathrm{const}$*; otherwise, the log RND can be computed and approximated as follows:*

$$\log \frac{p^\star(x_{[0,1]})}{p^r(x_{[0,1]})} = \int_0^1 \sum_{y \neq x_t} \left(1 - \frac{\varphi_t(y)}{\varphi_t(x_t)}\right) r_t(y, x_t)\mathrm{d}t + \sum_{t: x_{t_-} \neq x_t} \log \frac{\varphi_t(x_t)}{\varphi_t(x_{t_-})} \tag{37}$$

$$\approx \sum_{i=0}^{M-1} \left[\sum_{d=1}^{D} \sum_{n \neq x_{t_i}^d} \frac{\overline{\gamma}_{t_i, t_{i+1}}}{N} \left(1 - \Phi_{t_i}(x_{t_i})_{d,n}\right) + \sum_{d: x_{t_i}^d \neq x_{t_{i+1}}^d} \log \Phi_{t_i}(x_{t_i})_{d, x_{t_{i+1}}^d} \right], \tag{38}$$

where $0 = t_0 < ... < t_M = 1$ *are the discretized time points.*

Proof. From (14), one can find $\log \frac{p^\star(x_{[0,1]})}{p^r(x_{[0,1]})} = \log \frac{\varphi_1(x_1)}{\varphi_0(x_0)}$. When memoryless, following the argument in Sec. 3, $\varphi_0 = \mathrm{const}$, $\log \varphi_1 = \log \frac{\nu}{\widehat{\varphi}_1} = \log \frac{\nu}{p_1^r} + \mathrm{const}$, which yields the desired result.

For general cases, inspired by the discussion in Liu et al. (2025a, App. D.4): (29) implies

$$\partial_t \varphi_t(x) = \sum_{y \neq x} (\varphi_t(x) - \varphi_t(y)) r_t(y, x) \implies \partial_t \log \varphi_t(x) = \sum_{y \neq x} \left(1 - \frac{\varphi_t(y)}{\varphi_t(x)}\right) r_t(y, x).$$

Hence, we have

$$\begin{aligned} \log \frac{\varphi_1(x_1)}{\varphi_0(x_0)} &= \int_0^1 \partial_t \log \varphi_t(x_t)\mathrm{d}t + \sum_{t: x_{t_-} \neq x_t} \log \frac{\varphi_t(x_t)}{\varphi_t(x_{t_-})} \\ &= \int_0^1 \sum_{y \neq x_t} \left(1 - \frac{\varphi_t(y)}{\varphi_t(x_t)}\right) r_t(y, x_t)\mathrm{d}t + \sum_{t: x_{t_-} \neq x_t} \log \frac{\varphi_t(x_t)}{\varphi_t(x_{t_-})}. \end{aligned}$$

Under our setting, the summation in the first term can be easily calculated:

$$\sum_{y \neq x_t} \left(1 - \frac{\varphi_t(y)}{\varphi_t(x_t)}\right) r_t(y, x_t) = \sum_{d=1}^{D} \sum_{n \neq x_t^d} \left(1 - \frac{\varphi_t(x_t^{d \leftarrow n})}{\varphi_t(x_t)}\right) r_t(x_t^{d \leftarrow n}, x_t) = \sum_{d=1}^{D} \sum_{n \neq x_t^d} \left(1 - \Phi_t^\star(x_t)_{d,n}\right) \frac{\gamma_t}{N}.$$

For the second term, in theory, x_{t_-} and x_t differ by one dimension only, but during discretization, multiple dimensions may change simultaneously. A heuristic approximation is to decompose it into multiple single-dimension changes:

$$\log \frac{\varphi_t(x_t)}{\varphi_t(x_{t_-})} \approx \sum_{d: x_{t_-}^d \neq x_t^d} \log \frac{\varphi_t(x_{t_-}^{d \leftarrow x_t^d})}{\varphi_t(x_{t_-})} = \sum_{d: x_{t_-}^d \neq x_t^d} \log \Phi_t^\star(x_{t_-})_{d, x_t^d}.$$

We can thus summarize the approximate calculation of $\frac{p^\star(x_{[0,1]})}{p^r(x_{[0,1]})} = \frac{\varphi_1(x_1)}{\varphi_0(x_0)}$ on the time-discretized trajectory $(x_{t_0}, x_{t_1}, ..., x_{t_M})$ as follows:

$$\log \frac{\varphi_1(x_1)}{\varphi_0(x_0)} \approx \sum_{i=0}^{M-1} \left[\sum_{d=1}^{D} \sum_{n \neq x_{t_i}^d} \frac{\overline{\gamma}_{t_i, t_{i+1}}}{N} \left(1 - \Phi_{t_i}^\star(x_{t_i})_{d,n}\right) + \sum_{d: x_{t_i}^d \neq x_{t_{i+1}}^d} \log \Phi_{t_i}^\star(x_{t_i})_{d, x_{t_{i+1}}^d} \right].$$

Finally, as the ground-truth $\Phi^\star$ is unavailable, we use Φ to approximate its value. This concludes the proof.

□

Remark. Unlike the method proposed in Liu et al. (2025a, App. D.4), here we don't need to train $\frac{\widehat{\varphi}_t(y)}{\widehat{\varphi}_t(x)}$ along the trajectory. This is because here we derive through $\frac{\varphi_1(x_1)}{\varphi_0(x_0)}$ instead of $\frac{\widehat{\varphi}_1(x_1)}{\widehat{\varphi}_0(x_0)}$.

Remark. In the case of *masked* diffusion, the path measure p^u can be *exactly* sampled and the log RND on a give trajectory can be computed *precisely*; however, here, the estimated $\log \frac{p^\star(x_{[0,1]})}{p^u(x_{[0,1]})}$ involves *time discretization error*, and furthermore, *learning error* if we use (38) under non-memoryless cases. This possibly explains the low ESS observed in Fig. 1 even after convergence.

B.10 Connection between DASBS and WDCE

Proposition B.9. *Assume $\overline{\gamma}_{0,1} = \infty$, i.e., the reference path measure p^r is memoryless. Then the denoising loss for training the controller* (ctrl-DM) *with trajectory importance reweighting, D_f being the generalized KL divergence,*[1] *and time-weight $w_t \leftarrow \frac{\gamma_t}{N}$ reduces the denoising cross-entropy (WDCE) loss in UDNS (Zhu et al., 2025a, App. F), which is equal to* $\mathrm{KL}(p^\star \| p^u) + \mathrm{const}$.

Proof. The first equation on Zhu et al. (2025a, Page 39) reads (using the notation in this paper)

$$\mathrm{KL}(p^\star \| p^u) = \mathbb{E}_{p^{\mathrm{sg}(u)}(\overline{x}_{[0,1]})} \frac{p^\star(\overline{x}_{[0,1]})}{p^{\mathrm{sg}(u)}(\overline{x}_{[0,1]})} \mathbb{E}_t \frac{\gamma_t}{N} \mathbb{E}_{p^\star_{t|1}(x|\overline{x}_1)} \sum_{d=1}^{D} \sum_{n \neq x^d} D_f \left(\frac{p^\star_{t|1}(x^{d\leftarrow n}|\overline{x}_1)}{p^\star_{t|1}(x|\overline{x}_1)} \middle\| \Phi_t(x)_{d,n} \right) + \mathrm{const},$$

where D_f is the generalized KL divergence. Under the memoryless assumption, we have $p^\star_{t|1}(x|x_1) = p^r_{t|1}(x|x_1)$. Due to the symmetry $p^r_{t|1}(x|x_1) = p^r_{t|1}(x_1|x)$, the equivalence to (ctrl-DM) is obvious.

□

Proposition B.10. *Consider the mask-augmented state space $\mathcal{X} = \{1, ..., N, \mathbf{M}\}^D$. Let the reference transition rate be*

$$r_t(y, x) = \begin{cases} \frac{\gamma_t}{N}, & \text{if } y = x^{d\leftarrow n},\ x^d = \mathbf{M},\ n \in [N], \\ -\gamma_t |\{d : x^d = \mathbf{M}\}|, & \text{if } y = x, \\ 0, & \text{otherwise,} \end{cases} \tag{39}$$

for some noise schedule $\gamma_\cdot : [0,1] \to \mathbb{R}_+$. Let $\overline{\gamma}_{s,t} := \int_s^t \gamma_u \mathrm{d}u$ for $0 \le s < t \le 1$, and assume $\overline{\gamma}_{t,1} = \infty$ for any $0 \le t < 1$. Then, the denoising loss for training the controller (ctrl-DM) *with trajectory importance weighting, D_f being the generalized KL divergence,*[1] *and time-weight $\frac{\gamma_t}{N}$ reduces to the WDCE loss in MDNS (Zhu et al., 2025a), which is also equal to* $\mathrm{KL}(p^\star \| p^u) + \mathrm{const}$.

Proof. Throughout the proof, we always assume $n \in [N]$ is a non-mask state. From Zhu et al. (2025a), we have the following results:

$$\begin{aligned}
u^\star_t(x^{d\leftarrow n}, x) &= \gamma_t \Pr_{X\sim\nu}(X^d = n | X^{\mathrm{UM}} = x^{\mathrm{UM}}) 1_{x^d = \mathbf{M}} \\
\Phi^\star_t(x)_{d,n} = \frac{\varphi_t(x^{d\leftarrow n})}{\varphi_t(x)} &= N \Pr_{X\sim\nu}(X^d = n | X^{\mathrm{UM}} = x^{\mathrm{UM}}) 1_{x^d = \mathbf{M}} \\
p^r_{t|s}(y|x) &= \prod_{d: x^d = \mathbf{M}} \left(\frac{1 - \mathrm{e}^{-\overline{\gamma}_{s,t}}}{N} \right)^{1_{y^d \neq \mathbf{M}}} \left(\mathrm{e}^{-\overline{\gamma}_{s,t}} \right)^{1_{y^d = \mathbf{M}}} \cdot \prod_{d: x^d \neq \mathbf{M}} 1_{x^d = y^d},\ 0 \le s < t \le 1 \\
\implies \frac{p^r_{1|t}(x_1 | x^{d\leftarrow n})}{p^r_{1|t}(x_1|x)} &= N 1_{x_1^d = n} 1_{x^d = \mathbf{M}} \text{ (suppose } \forall d \text{ s.t. } x^d \neq \mathbf{M}, x_1^d = x^d)
\end{aligned}$$

Let $s_\theta : \mathcal{X} \to \mathbb{R}_{\ge 0}^{D \times N}$ be the neural network to learn the conditional distribution in ν, i.e., $s_\theta(x)_{d,n} \approx \Pr_{X\sim\nu}(X^d = n | X^{\mathrm{UM}} = x^{\mathrm{UM}})$ for $x^d = \mathbf{M}$, and $s_\theta(x)_{d,n} = 1_{x^d = n}$ for $x^d \neq \mathbf{M}$. We assume $\sum_{n=1}^N s_\theta(x)_{d,n} = 1$ for all d.

Therefore, with generalized KL divergence, the loss (ctrl-DM) with trajectory importance reweighting becomes

$$\begin{aligned}
&\mathbb{E}_t w_t \mathbb{E}_{p^{\mathrm{sg}(\theta)}(x_{[0,1]})} \frac{\mathrm{d}p^\star}{\mathrm{d}p^{\mathrm{sg}(\theta)}}(x_{[0,1]}) \mathbb{E}_{p^r_{t|0,1}(x|x_0,x_1)} \sum_{d=1}^{D} \sum_{n \neq x^d} D_f(N 1_{x_1^d=n} 1_{x^d=\mathbf{M}} \| N s_\theta(x)_{d,n} 1_{x^d=\mathbf{M}} + 1_{x^d=n}) \\
&= \mathbb{E}_t w_t \mathbb{E}_{p^{\mathrm{sg}(\theta)}(x_{[0,1]})} \frac{\mathrm{d}p^\star}{\mathrm{d}p^{\mathrm{sg}(\theta)}}(x_{[0,1]}) \mathbb{E}_{p^r_{t|1}(x|x_1)} \sum_{d:x^d=\mathbf{M}} \sum_{n=1}^{N} D_f(N 1_{x_1^d=n} \| N s_\theta(x)_{d,n}) \\
&= \mathbb{E}_t w_t \mathbb{E}_{p^{\mathrm{sg}(\theta)}(x_{[0,1]})} \frac{\mathrm{d}p^\star}{\mathrm{d}p^{\mathrm{sg}(\theta)}}(x_{[0,1]}) \mathbb{E}_{p^r_{t|1}(x|x_1)} \sum_{d:x^d=\mathbf{M}} \sum_{n=1}^{N} \left(N 1_{x_1^d=n} \log \frac{1_{x_1^d=n}}{s_\theta(x)_{d,n}} - N 1_{x_1^d=n} + N s_\theta(x)_{d,n} \right) \\
&= \mathbb{E}_t w_t \mathbb{E}_{p^{\mathrm{sg}(\theta)}(x_{[0,1]})} \frac{\mathrm{d}p^\star}{\mathrm{d}p^{\mathrm{sg}(\theta)}}(x_{[0,1]}) \mathbb{E}_{p^r_{t|1}(x|x_1)} \sum_{d:x^d=\mathbf{M}} -N \log s_\theta(x)_{d,x_1^d} + \mathrm{const.}
\end{aligned}$$

Thus, with $t \sim \mathrm{Unif}(0,1)$ and time weights $\frac{\gamma_t}{N}$, the loss

$$\min_\theta \mathbb{E}_{p^{\mathrm{sg}(\theta)}(x_{[0,1]})} \frac{\mathrm{d}p^\star}{\mathrm{d}p^{\mathrm{sg}(\theta)}}(x_{[0,1]}) \mathbb{E}_{t\sim\mathrm{Unif}(0,1)} \gamma_t \mathbb{E}_{p^r_{t|1}(x|x_1)} \sum_{d:x^d=\mathbf{M}} -\log s_\theta(x)_{d,x_1^d},$$

which is exactly the same as the WDCE loss in Zhu et al. (2025a) if we choose the canonical noise schedule in masked diffusion model: $\gamma_t = \frac{1}{t}$. □

B.11 Convergence of Alternating Update: Proof of Thm. 5.2

Definition B.11. *For a function $\psi : \mathcal{X} \to \mathbb{R}_+$, we use q^ψ to denote the path measure of a CTMC $(Y_t)_{t\in[0,1]}$ induced by the* backward *transition rate $u_t^{\leftarrow}(y,x) = \frac{\psi_t(y)}{\psi_t(x)} r_t(x,y)$, $y \neq x$ and initialized at $Y_1 \sim \nu$, where ψ_t satisfies $\psi_t(x) = \sum_y p^r_{t|0}(x|y)\psi_0(y)$, $\psi_1 = \psi$. In other words,*

$$u_t^{\leftarrow}(y,x) = \lim_{h\to 0} \frac{\Pr(Y_{t-h} = y | Y_t = x) - 1_{x=y}}{h}.$$

Remark. By the Bayes formula, one can obtain its equivalent forward transition rate (Kelly, 2011): for $y \neq x$,

$$\begin{aligned}
u_t^{\rightarrow}(y,x) &= \lim_{h\to 0} \frac{1}{h} \Pr(Y_{t+h} = y | Y_t = x) = \lim_{h\to 0} \frac{1}{h} \frac{\Pr(Y_t = x | Y_{t+h} = y) \Pr(Y_{t+h} = y)}{\Pr(Y_t = x)} \\
&= \lim_{h\to 0} \frac{1}{h} \frac{q^\psi_{t+h}(y)}{q^\psi_t(x)} (u^{\leftarrow}_{t+h}(x,y)h + o(h)) = \frac{q^\psi_t(y)}{q^\psi_t(x)} u_t^{\leftarrow}(x,y) = \frac{(q^\psi_t/\psi_t)(y)}{(q^\psi_t/\psi_t)(x)} r_t(y,x).
\end{aligned}$$

B.11.1 Proof of Part (1) of Thm. 5.2

Proof. Let $(Y_t)_{t\in[0,1]} \sim q^{\widehat{\varphi}_1^{(k-1)}} =: q$. By (KFE),

$$\begin{aligned}
\partial_t q_t(x) &= \sum_y q_t(y) u_t^{\rightarrow}(x,y) = \sum_{y\neq x} (q_t(y) u_t^{\rightarrow}(x,y) - q_t(x) u_t^{\rightarrow}(y,x)) \\
&= \sum_{y\neq x} \Big(\frac{(q_t/\psi_t)(x)}{(q_t/\psi_t)(y)} r_t(x,y) q_t(y) - \frac{(q_t/\psi_t)(y)}{(q_t/\psi_t)(x)} r_t(y,x) q_t(x) \Big), \\
\implies \partial_t \log q_t(x) &= \sum_{y\neq x} \Big(\frac{\psi_t(y)}{\psi_t(x)} r_t(x,y) - \frac{(q_t/\psi_t)(y)}{(q_t/\psi_t)(x)} r_t(y,x) \Big).
\end{aligned}$$

On the other hand, using (KFE) again,

$$\begin{aligned}\partial_t \psi_t(x) &= \sum_y \partial_t p^r_{t|0}(x|y)\psi_0(y) = \sum_y \sum_z r_t(x,z) p^r_{t|0}(z|y)\psi_0(y) = \sum_z r_t(x,z)\psi_t(z) \\ &= \sum_{y\neq x} (r_t(x,y)\psi_t(y) - r_t(y,x)\psi_t(x)), \\ \implies \partial_t \log \psi_t(x) &= \sum_{y\neq x} \Big(r_t(x,y)\frac{\psi_t(y)}{\psi_t(x)} - r_t(y,x)\Big).\end{aligned}$$

Now we compute $\mathrm{KL}(p^u \| q)$ where p^u is the path measure of a CTMC $(X_t)_{t\in[0,1]}$ induced by transition rate u_t and initial distribution μ:

$$\begin{aligned}\mathrm{KL}(p^u\|q) &= \mathrm{KL}(\mu\|q_0) + \mathbb{E}_{p^u(X)} \int_0^1 \sum_{y\neq X_t} \Big(u_t \log \frac{u_t}{u_t^{\rightarrow}} + u_t^{\rightarrow} - u_t \Big)(y, X_t)\mathrm{d}t \\ &= \mathrm{KL}(\mu\|q_0) + \mathbb{E}_{p^u(X)} \int_0^1 \sum_{y\neq X_t} \Big[\Big(u_t \log \frac{u_t}{r_t} + r_t - u_t \Big)(y, X_t) \\ &\quad + \Big(- u_t(y,X_t) \log \frac{(q_t/\psi_t)(y)}{(q_t/\psi_t)(X_t)} + \frac{(q_t/\psi_t)(y)}{(q_t/\psi_t)(X_t)} r_t(y,X_t) - r_t(y,X_t) \Big)\Big]\mathrm{d}t.\end{aligned}$$

The second term is $\mathrm{KL}(p^u\|p^r)$. To deal with the third term, we leverage Lem. B.13:

$$\begin{aligned}\mathbb{E}_{p^u(X)} \log \frac{(\psi_1/q_1)(X_1)}{(\psi_0/q_0)(X_0)} &= \mathbb{E}_{p^u(X)} \Big[\int_0^1 (\partial_t \log \psi_t(X_t) - \partial_t \log q_t(X_t))\mathrm{d}t + \sum_{t: X_{t_-}\neq X_t} \log \frac{(\psi_t/q_t)(X_t)}{(\psi_t/q_t)(X_{t_-})} \Big] \\ &= \mathbb{E}_{p^u(X)} \int_0^1 \sum_{y\neq X_t} \Big(\frac{(q_t/\psi_t)(y)}{(q_t/\psi_t)(X_t)} r_t(y,X_t) - r_t(y,X_t) \Big)\mathrm{d}t \\ &\quad + \mathbb{E}_{p^u(X)} \int_0^1 \sum_{y\neq X_t} u_t(y,X_t) \log \frac{(\psi_t/q_t)(y)}{(\psi_t/q_t)(X_t)}\mathrm{d}t.\end{aligned}$$

Therefore, we conclude that

$$\begin{aligned}\mathrm{KL}(p^u\|q) &= \mathrm{KL}(p^u\|p^r) + \mathbb{E}_{p^u(X)} \log \frac{(\psi_1/q_1)(X_1)}{(\psi_0/q_0)(X_0)} + \mathrm{const} \\ &= \mathrm{KL}(p^u\|p^r) + \mathbb{E}_{p^u(X)} \log \frac{\psi_1}{q_1}(X_1) + \mathrm{const} \\ &= \mathrm{KL}(p^u\|p^r) + \mathbb{E}_{p^u(X)} \log \frac{\widehat{\varphi}_1^{(k-1)}}{\nu}(X_1) + \mathrm{const},\end{aligned}$$

where const does not depend on u. Therefore, this is an SOC problem with terminal cost $g \leftarrow \log \frac{\widehat{\varphi}_1^{(k-1)}}{\nu}$. Let the optimal path measure to this SOC problem be $p^{(k)}$ and, by relating this SOC problem with an equivalent SB problem through Cor. B.4, let the SB potentials to this problem be $(\varphi_t^{(k)}, \widehat{\varphi}_t^{(k)})$. Then, $\varphi_1^{(k)} = \mathrm{e}^{-g} = \frac{\nu}{\widehat{\varphi}_1^{(k-1)}}$ by Cor. B.4. From App. B.2 and similar argument as (19), we can leverage the additive noise and obtain

$$\frac{\varphi_t^{(k)}(y)}{\varphi_t^{(k)}(x)} = \mathbb{E}_{p^{(k)}_{1|t}(x_1|x)} \frac{\varphi_1^{(k)}(x_1+y-x)}{\varphi_1^{(k)}(x_1)} = \mathbb{E}_{p^{(k)}_{1|t}(x_1|x)} \frac{(\nu/\widehat{\varphi}_1^{(k-1)})(x_1+y-x)}{(\nu/\widehat{\varphi}_1^{(k-1)})(x_1)}.$$

On the other hand, from the property of Bregman divergence, the unique fixed-point of (ctrl-AM), $\Phi^{(k)}$, satisfies

$$\Phi_t^{(k)}(x)_{d,n} = \mathbb{E}_{p^{r_{\Phi^{(k)}}}_{1|t}(x_1|x)} \frac{(\nu/\widehat{\varphi}_1^{(k-1)})(x^{d\leftarrow x_1^d + n - x^d})}{(\nu/\widehat{\varphi}_1^{(k-1)})(x)}.$$

Thus, we conclude that $\Phi_t^{(k)}(x)_{d,n} = \frac{\varphi_t^{(k)}(x^{d\leftarrow n})}{\varphi_t^{(k)}(x)}$. □

B.11.2 Proof of Part (2) of Thm. 5.2

Proof. By similar arguments using Bayes formula, we can write $p^{r\Phi^{(k)}} =: p$ as a backward CTMC initialized at p_1 with backward transition rate

$$u_t^{\leftarrow(k)}(y,x) = \frac{(p_t^{(k)}/\varphi_t^{(k)})(y)}{(p_t^{(k)}/\varphi_t^{(k)})(x)} r_t(x,y) = \frac{\widehat{\varphi}_t^{(k)}(y)}{\widehat{\varphi}_t^{(k)}(x)} r_t(x,y),$$

where $\varphi_t^{(k)}$ and $\widehat{\varphi}_t^{(k)}$ are the SB potentials to the SOC problem in the proof of the first part. Then, for any backward CTMC $(Y_t)_{t\in[0,1]} \sim q$ initialized at $Y_1 \sim \nu$ with backward transition rate $u_t^{\leftarrow}$, we have

$$\mathrm{KL}(p\|q) = \mathrm{KL}(p_1\|\nu) + \mathbb{E}_{X\sim p}\int_0^1 \sum_{y\neq X_t} \Big(u_t^{\leftarrow(k)} \log \frac{u_t^{\leftarrow(k)}}{u_t^{\leftarrow}} + u_t^{\leftarrow} - u_t^{\leftarrow(k)} \Big)(y, X_t)\mathrm{d}t.$$

Therefore, it is obvious that the optimal $u_t^{\leftarrow}(y,x)$ is equal to $u_t^{\leftarrow(k)}(y,x) = \frac{\widehat{\varphi}_t^{(k)}(y)}{\widehat{\varphi}_t^{(k)}(x)} r_t(x,y)$, i.e., the optimal q is $q^{\widehat{\varphi}_1^{(k)}}$ (Def. B.11). By similar arguments as in the proof of (12), we have $\widehat{\varphi}_t(x) = \sum_y p_{t|0}^r(x|y)\widehat{\varphi}_0(y)$, and hence by definition the optimal q to the backward half bridge problem is $q^{\widehat{\varphi}_1^{(k)}}$. On the other hand, by similar arguments as (22) and (25),

$$\frac{\widehat{\varphi}_1^{(k)}(z)}{\widehat{\varphi}_1^{(k)}(y)} = \mathbb{E}_{p_{t|1}^{(k)}(x|y)} \frac{p_{1|t}^r(z|x)}{p_{1|t}^r(y|x)} = \mathbb{E}_{p_{t|1}^{(k)}(x|y)} \frac{\widehat{\varphi}_0^{(k)}(x-y+z)}{\widehat{\varphi}_0^{(k)}(x)} = \frac{(\mu/\varphi_0^{(k)})(x-y+z)}{(\mu/\varphi_0^{(k)})(x)}.$$

On the other hand, using the property of Bregman divergence, the unique fixed-point of (corr-DM) satisfies

$$\frac{\widehat{\varphi}_1^{(k)}(x^{d\leftarrow n})}{\widehat{\varphi}_1^{(k)}(x)} = \mathbb{E}_{p_{t|1}^{(k)}(x_0|x)} \frac{p_{1|t}^r(x_1^{d\leftarrow n}|x)}{p_{1|t}^r(x_1|x)},$$

while the unique fixed-point of (corr-AM) satisfies

$$\frac{\widehat{\varphi}_1^{(k)}(x^{d\leftarrow n})}{\widehat{\varphi}_1^{(k)}(x)} = \mathbb{E}_{p_{t|1}^{(k)}(x_0|x)} \frac{(\mu/\varphi_0^{(k)})(x_0^{d\leftarrow x_0^d+n-x_1^d})}{(\mu/\varphi_0^{(k)})(x_0)}.$$

By comparing the three equations above, the proof is complete. □

B.12 Other Omitted Results and Proofs

Proposition B.12. *Assume $\mu = p_{\mathrm{unif}}$. Using the reference transition rate* (16)*, initializing the controller to be one (i.e., let $\Phi_t^{(0)}(x)_{d,n} = 1$ for all $n \neq x^d$) and alternatively training the corrector and the controller is equivalent to initializing the corrector to be one (i.e., let $\widehat{\Phi}^{(0)}(x)_{d,n} = 1$ for all $n \neq x^d$) and alternatively training the controller and the corrector.*

Proof. Suppose we let $\Phi_t^{(0)}(x)_{d,n} = 1$ for all $n \neq x^d$. Then the optimal corrector can be computed as follows:

$$\begin{aligned}\widehat{\Phi}^{(1)}(x_1)_{d,n} &= \mathbb{E}_{p_{t|1}^r(x|x_1)} \frac{p_{1|t}^r(x_1^{d\leftarrow n}|x)}{p_{1|t}^r(x_1|x)} = \sum_x \frac{p_{t|1}^r(x|x_1)}{p_{1|t}^r(x_1|x)} p_{1|t}^r(x_1^{d\leftarrow n}|x) \\ &= \sum_x \frac{p_t^r(x)}{p_1^r(x_1)} p_{1|t}^r(x_1^{d\leftarrow n}|x) = \frac{1}{p_1^r(x_1)} \sum_x p_t^r(x) p_{1|t}^r(x_1^{d\leftarrow n}|x) = \frac{p_1^r(x_1^{d\leftarrow n})}{p_1^r(x_1)}.\end{aligned}$$

Thus, under (16), when $p_0^r = \mu = p_{\mathrm{unif}}$, one has $p_1^r = p_{\mathrm{unif}}$, so $\widehat{\Phi}^{(1)}(x_1)_{d,n} = 1$ for all $n \neq x_1^d$. □

Remark. We remark that this choice coincides the optimal corrector when assuming p^r is memoryless, since under memoryless condition $\widehat{\varphi}_1 \propto p_1^r$.

Lemma B.13. *For a CTMC $(X_t)_{t\in[0,1]} \sim p^u$ with transition rate u_t and any function $g : [0,1] \times \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, we have*

$$\mathbb{E}_{p^u(X)} \sum_{t:X_{t_-} \neq X_t} g(t, X_{t_-}, X_t) = \mathbb{E}_{p^u(X)} \int_0^1 \sum_{y \neq X_t} g(t, X_t, y) u_t(y, X_t) \mathrm{d}t.$$

Proof. Consider the time-discretization: $\Delta t = \frac{1}{N}$ and $t_n = n\Delta t$. Then,

$$\begin{aligned}
\mathbb{E}_{p^u(X)} \sum_{t:X_{t_-} \neq X_t} g(t, X_{t_-}, X_t) &= \mathbb{E}_{p^u(X)} \sum_{n=0}^{N-1} 1_{X_{t_n} \neq X_{t_{n+1}}} g(t_n, X_{t_n}, X_{t_{n+1}}) + O(\Delta t) \\
&= \sum_{n=0}^{N-1} \sum_{x,y} p^u_{t_n}(x) p^u_{t_{n+1}|t_n}(y|x) 1_{x \neq y} g(t_n, x, y) + O(\Delta t) \\
&= \sum_{n=0}^{N-1} \sum_{x} p^u_{t_n}(x) \sum_{y \neq x} u_{t_n}(y, x) g(t_n, x, y) \Delta t + O(\Delta t) \\
&= \mathbb{E}_{p^u(X)} \int_0^1 \sum_{y \neq X_t} g(t, X_t, y) u_t(y, X_t) \mathrm{d}t.
\end{aligned}$$

□

C EXPERIMENTS

Table 2: Sampling from lattice Ising models with $L = 24$. **Best** and second best results among all *uniform-based* discrete neural samplers are highlighted. $*$: Measured on one A6000 GPU with largest feasible batch size for β_{high}. †: For β_{critical} and β_{low}, using warm-up strategy in PDNS (Guo et al., 2026). ‡: Failed to converge to meaningful distributions at β_{critical} and β_{low} even with warm-up.

	Inv. Temp.			$\beta_{\text{high}} = 0.28$			$\beta_{\text{critical}} = 0.4407$			$\beta_{\text{low}} = 0.6$		
Type	Metrics ↓	Steps (×1e3)*	Runtime (h)*	ΔMag.	ΔCorr.	EW_2	ΔMag.	ΔCorr.	EW_2	ΔMag.	ΔCorr.	EW_2
Uniform	**DASBS**	**3.75**	**0.5**	2.7e − 3	1.8e − 3	8.6	**5.7e − 2**	**4.7e − 2**	**12.1**	**2.0e − 2**	**1.8e − 3**	**2.0**
	LEAPS	30	8.4	**1.8e − 3**	**9.2e − 4**	**3.1**	5.9e − 2	2.8e − 1	96.5	3.0e − 2	5.5e − 1	176.6
	UDNS‡	50	11.9	9.0e − 3	8.7e − 3	23.6	–	–	–	–	–	–
	DFNS‡	50	2.1	9.3e − 1	8.0e − 1	661.6	–	–	–	–	–	–
Masked	MDNS†	50	16.8	3.9e − 3	7.4e − 4	0.1	1.1e − 2	5.6e − 3	5.1	9.0e − 3	4.7e − 3	5.3
MCMC	MH	–	–	8.9e − 4	2.9e − 4	1.2	2.5e − 2	3.7e − 3	293.3	4.0e − 2	6.6e − 4	109.9

Table 3: Learning to sample from lattice Potts models with $L = 16$ and $N = 4$. **Best** result among all *uniform-based* discrete neural samplers are highlighted. †: For β_{critical} and β_{low}, using warm-up strategy in PDNS (Guo et al., 2026).

	Inv. Temp.	$\beta_{\text{high}} = 0.9$			$\beta_{\text{critical}} = 1.0986$			$\beta_{\text{low}} = 1.3$		
Type	Metrics ↓	ΔMag.	ΔCorr.	EW_2	ΔMag.	ΔCorr.	EW_2	ΔMag.	ΔCorr.	EW_2
Uniform	**DASBS**	8.7e − 3	6.6e − 3	**5.3**	**4.1e − 3**	**2.9e − 2**	**20.0**	**4.2e − 3**	**7.3e − 3**	**3.6**
	LEAPS	**5.7e − 3**	**5.0e − 3**	8.2	3.2e − 1	2.6e − 1	79.9	3.6e − 1	3.5e − 1	90.5
Mask	MDNS	6.5e − 3	5.1e − 3	6.4	5.2e − 3	4.6e − 3	1.9	8.4e − 4	6.1e − 4	0.7
MCMC	MH	5.2e − 2	4.7e − 2	98.6	4.9e − 1	4.0e − 1	273.2	6.8e − 1	6.4e − 1	313.1

C.1 MAIN RESULTS

In this section, we evaluate the effectiveness and efficiency of the proposed DASBS algorithm. We demonstrate that it achieves competitive sample quality across standard benchmarks while offering significant advantages in training speed. We further investigate the benefits of AM and non-memoryless reference dynamics through ablation studies.

Experimental Setup We test our method on two discrete distributions from statistical physics: the Ising and Potts models on square lattices. They are well known for exhibiting phase transitions as a function of the inverse temperature (Onsager, 1944; Beffara & Duminil-Copin, 2012) and serve as standard benchmarks for discrete sampling. For the Ising model, we use a lattice of size $L = 24$ (i.e., sequence length $D = L^2 = 576$) with states $\{\pm 1\}$, and consider inverse temperatures $\beta_{\text{high}} = 0.28$, $\beta_{\text{critical}} = \log(1 + \sqrt{2})/2 \approx 0.4407$, and $\beta_{\text{low}} = 0.6$. For the Potts model, we use $L = 16$ ($D = L^2 = 256$) with $N = 4$ states, and consider $\beta_{\text{high}} = 0.9$, $\beta_{\text{critical}} = \log(1 + \sqrt{q}) \approx 1.0986$, and $\beta = 1.3$. We benchmark DASBS against leading discrete neural samplers, including LEAPS (Holderrieth et al., 2025), DFNS (Ou et al., 2025b), UDNS (Zhu et al., 2025a, App. F), and MDNS (Zhu et al., 2025a), as well as the classical Metropolis-Hastings (MH) algorithm.

Results and Discussion Quantitative results for the Ising and Potts models are summarized in Tabs. 2 and 3, respectively. We assess sample quality using three metrics: the deviation of the magnetization (ΔMag.), the deviation of the 2-point correlation (ΔCorr.), and the energy Wasserstein-2 distance (EW_2), all relative to ground-truth samples from the Swendsen-Wang (SW, Swendsen & Wang (1986; 1987)) algorithm. For Ising model with high temperature, we also report the number of training steps and runtime on a single A6000 GPU, utilizing the maximum feasible batch size. Across all regimes, DASBS delivers **satisfactory sample fidelity** that is highly competitive with state-of-the-art uniform-based discrete neural samplers. Notably, DASBS confers a distinct advantage in **training efficiency**. We attribute this speed-up to three factors: (1) the use of the discrete score as a highly informative first-order oracle; (2) a memory-efficient design that avoids storing full roll-outs or computing loss over full trajectories; and (3) the simplicity of the matching loss objective.

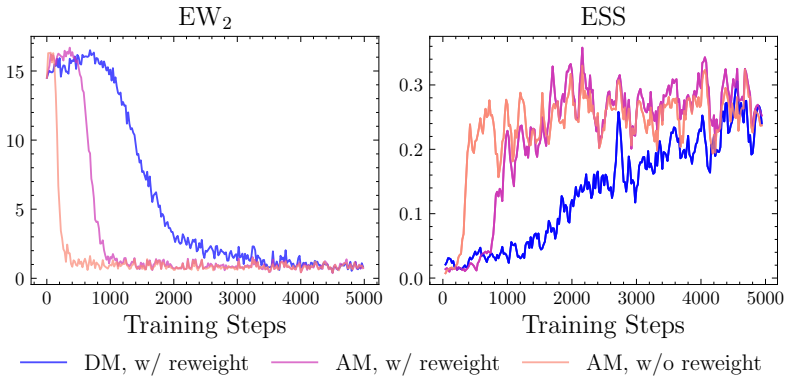

Figure 1: Ablation study of the adjoint matching (AM) and denoising matching (DM) training losses for the *memoryless* noise schedule $\gamma_t = \frac{1}{t}$ on Potts model with $L = 8$, $N = 3$, $\beta_{\text{high}} = 0.5$. Reweighting means using trajectory importance weight $p^\star / p^{\text{sg}(u)}$ in the training losses. DM with reweighting corresponds to UDNS (Zhu et al., 2025a, App. F). *Left*: Energy Wasserstein-2 distance to the ground-truth samples from SW algorithm. *Right*: Effective sample size computed from the trajectory importance weights.

Ablation Study: AM v.s. DM in Learning Controller In Fig. 2, we empirically validate the claim that AM is more efficient than DM. We focus on an 8×8 Potts model ($N = 3$, $\beta_{\text{high}} = 0.5$). We choose the standard *memoryless* noise schedule $\gamma_t = \frac{1}{t}$ (Ou et al., 2025a; Zhu et al., 2025a) to isolate the impact of the training objectives (ctrl-AM) v.s. (ctrl-DM). Fig. 1 shows that, while all three methods eventually converge, AM exhibits **significantly faster convergence** than DM when trajectory importance reweighting is enabled. Furthermore, we observe that the importance reweighting can inadvertently slow down the convergence due to small effective sample sizes (even after convergence), likely stemming from variance of the estimated RND.

Ablation Study: Memoryless v.s. Non-Memoryless We investigate the impact of the noise schedule γ_t in Fig. 2, and in particular, we show that non-memoryless noise schedule performs better than the memoryless one. We adopt the *modified log-linear schedule* $\gamma_t = \frac{\gamma}{t+\alpha}$ on the 24×24 Ising model with $\beta_{\text{high}} = 0.28$, which is memoryless when $\alpha = 0$. Fig. 2 shows that when either

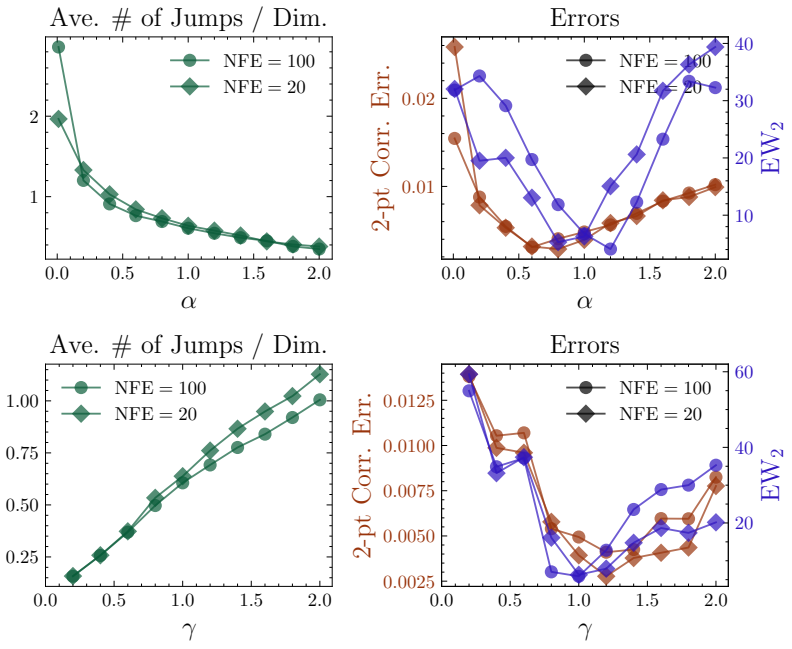

Figure 2: Ablation study of the hyperparameters α and γ for the modified log-linear noise schedule $\gamma_t = \frac{\gamma}{t+\alpha}$ on Ising model with $L = 24$ and $\beta_{\text{high}} = 0.28$. The case of $\alpha = 0$ is *memoryless*. NFE is the number of function evaluations during generation for both training and inference. *Top*: fix $\gamma = 1$ and vary α. *Bottom*: fix $\alpha = 1$ and vary γ. *Left*: average number of jumps for each dimension during generation. *Right*: 2-point correlation error and energy Wasserstein-2 distance to ground-truth samples drawn from SW algorithm.

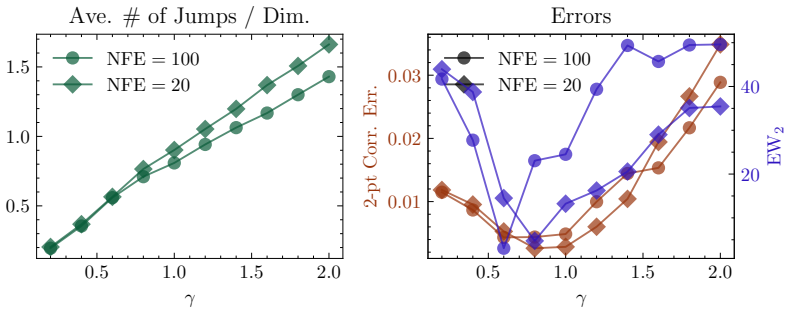

Figure 3: Ablation study of the hyperparameter γ for the constant noise schedule $\gamma_t \equiv \gamma$ on Ising model with $L = 24$ and $\beta_{\text{high}} = 0.28$. NFE is the number of function evaluations during generation for both training and inference. *Left*: average number of jumps for each dimension during generation. *Right*: 2-point correlation error and energy Wasserstein-2 distance to ground-truth samples drawn from SW algorithm.

hyperparameter is fixed at 1, the error metrics exhibit a distinct U-shaped curve as a function of the other parameter. Deviating from the optimal regime leads to degradation in sample quality: small γ or large α reduces the transition rate magnitude and may hinder exploration, while large γ or small α (particularly the memoryless case $\alpha = 0$) induces excessive jumps during generation, also degrading performance. Notably, these trends remain consistent across different computational budgets, demonstrating the robustness of the optimal configuration. A similar study for the constant noise schedule $\gamma_t \equiv \gamma$ is provided in Fig. 3, and a similar trend happens as the modified log-linear schedule: the generated samples reach the best quality at around $\gamma \in [0.5, 1]$.

C.2 TARGET DISTRIBUTIONS

We consider Ising and Potts models on a square lattice $\Lambda = [L]^2$ with L sites per dimension. We write $i \sim j$ if $i, j \in \Lambda$ are adjacent on the lattice. For simplicity, we impose periodic boundary conditions in both the horizontal and vertical directions. Both target distributions can be written in the form of

$$\nu(x) = \frac{1}{Z} \mathrm{e}^{-\beta E(x)},$$

where E is the energy function (Hamiltonian), $\beta > 0$ is the inverse temperature, and $Z = \sum_{x \in \mathcal{X}} \mathrm{e}^{-\beta E(x)}$ is the partition function.

For Ising model with interaction parameter $J \in \mathbb{R}$ and external magnetic field $h \in \mathbb{R}$, the energy is

$$E_{\text{Ising}}(x) = -J \sum_{i \sim j} x^i x^j - h \sum_i x^i, \; x \in \{\pm 1\}^\Lambda.$$

We keep $J = 1$ and $h = 0$ through out the experiments. For Potts model with N states and interaction parameter $J \in \mathbb{R}$, the energy is

$$H_{\text{Potts}}(x) = -J \sum_{i \sim j} 1_{x^i = x^j}, \; x \in [N]^\Lambda.$$

We keep $J = 1$ through out the experiments. Finally, we can compute the discrete score of these two distributions as follows:

$$\frac{\nu(x^{i \leftarrow n})}{\nu(x)} = \begin{cases} \exp\left(\beta(n - x^i)\left(J \sum_{j:\, j \sim i} x^j + h\right)\right), \; \forall n \in \{\pm 1\}, & \text{Ising;} \\ \exp\left(\beta J \left(\sum_{j:\, j \sim i} (1_{x^j = n} - 1_{x^j = x^i})\right)\right), \; \forall n \in [N], & \text{Potts.} \end{cases}$$

C.3 IMPLEMENTATION DETAILS

Model Backbone We follow the implementation of MDNS (Zhu et al., 2025a) to use vision transformers (ViT, Dosovitskiy et al. (2021)) to serve as the backbone for the discrete diffusion model. In particular, we use the DeiT (Data-efficient image Transformers) framework (Touvron et al., 2021) with 2-dimensional rotary position embedding (Heo et al., 2025), which better captures the 2-dimensional spatial structure of the Ising and Potts models. While MDNS's model only requires the position input $x \in [N]^D$, in our learning objectives, the controller also receives time input $t \in [0, 1]$. Hence, we adopt the adaptive layer normalization (adaLN) mechanism in the DiT (Peebles & Xie, 2023) and SiT (Ma et al., 2024) to deal with the time conditioning. For learning 24×24 Ising model and 16×16 Potts model with 4 states, we use a model with 6 blocks, hidden dimensions 32 and 4 heads. The total number of parameters in the model is around 144k (controller, with time conditioning) or 95k (corrector, without time conditioning).

Training Among all the training tasks, we use the AdamW optimizer (Loshchilov & Hutter, 2019) with a constant learning rate of $1\mathrm{e}-3$ (β_{high}) and $5\mathrm{e}-4$ ($\beta_{\text{critical}}, \beta_{\text{low}}$). We always use exponential moving average (EMA) to stabilize the training, with a decay rate of 0.9999. All experiments are trained on an NVIDIA RTX A6000 GPU. For all distributions, we always use the generalized KL divergence[1] as the Bregman divergence, always use uniform time weight $w_t \equiv 1$, and adpot the modified log-linear noise schedule with $\gamma = 1$ and $\alpha = 0.5$. We train for 5 stages with 500 steps for controller and 250 steps for corrector. We keep a running buffer of size 512 (β_{high}) or 4096 ($\beta_{\text{critical}}, \beta_{\text{low}}$), and resampling a batch of 128 pairs of (x_0, x_1) using the current controller every 20 (β_{high}) or 10 ($\beta_{\text{critical}}, \beta_{\text{low}}$) gradient updates. Finally, for β_{high}, we use $\mu = p_{\text{unif}}$ as an initialization and adopt AM loss (corr-AM) for the corrector; for $\beta_{\text{critical}}, \beta_{\text{low}}$, we observe that uniform initialization does not lead to good performance, and thus initialize μ as the zero-temperature distribution (i.e., $\mathrm{Unif}\{\pm \mathbf{1}\}$ for Ising and $\mathrm{Unif}\{\mathbf{1}, 2\mathbf{1}, ..., N\mathbf{1}\}$ for Potts, where $\mathbf{1}$ is the all-one vector). We use DM loss (corr-DM) for training the corrector as μ is not positive everywhere.

Generating Baseline and Ground Truth Samples We follow the implementation in existing literature (Guo et al., 2026) for baselines. For the learning-based baseline, we train LEAPS and MDNS on 24×24 Ising model and 16×16 Potts model for up to 150k steps for each temperature. For MDNS, we apply the warm-up strategy (Zhu et al., 2025a) to initialize the training under β_{critical}

from the pretrained checkpoint for β_{high}, and initialize the training under β_{low} from the pretrained checkpoint for β_{critical}. The Metropolis-Hastings (MH) and Swendsen-Wang (SW) sampling exactly follows the implementation in Guo et al. (2026) that ensures sufficient mixing.

Evaluation We follow the procedure detailed in Zhu et al. (2025a, App. D.3, E.2) for the computation of **magnetization** and **2-point correlation error**. To compute the empirical **energy Wasserstein-2 distance** for two set of samples $\{x_i\}$ and $\{y_j\}$, we obtain the energies of the datasets $\mathcal{E}_1 = \{E(x_i)\}$ and $\mathcal{E}_2 = \{E(y_j)\}$, and use the function $\texttt{np.sqrt}(\texttt{ot.wasserstein_1d}(\mathcal{E}_1,\ \mathcal{E}_2,\ \texttt{p} = 2))$ from the POT package (Flamary et al., 2021).

Effective Sample Size (ESS) We follow the practice in existing literature on neural samples (Zhang & Chen, 2022; Holderrieth et al., 2025; Zhu et al., 2025a). For a batch of i.i.d. samples $\{x_i\}_{i\in[B]}$ from p, suppose we associate each sample x_i with a weight $w_i = \frac{\widehat{q}(x_i)}{p(x_i)}$, where $\widehat{q}$ is the unnormalized probability density / mass function of a probability distribution q, then the (normalized) **effective sample size (ESS)** of the samples with respect to q is defined as $\frac{(\frac{1}{B}\sum_i w_i)^2}{\frac{1}{B}\sum_i w_i^2} \in \left[\frac{1}{B}, 1\right]$.

Ablation Studies on TM v.s. DM (Fig. 1) We use a model with 4 blocks, hidden dimension 32 and 4 heads. We use a batch size of 256 and in (ctrl-DM) and (ctrl-AM), we sample 32 t's following $\text{Unif}[0, 1]$ for each pair of (x_0, x_1). All three cases are trained under the same random seed for 5000 steps, with a buffer size of 1024 and resampling frequency 20 with generation NFE 100.

Ablation Studies on Noise Schedules (Figs. 2 and 3) We use a model with 4 blocks, hidden dimension 32 and 4 heads. We use a batch size of 64 and in (ctrl-AM), we sample 8 t's following $\text{Unif}[0, 1]$ for each pair of (x_0, x_1). All runs are trained under the same random seed for 5 stages with 200 controller update steps and 200 corrector update steps. The buffer size is 256 and resampling frequency is 20.