DISTRIBUTION MATCHING DISTILLATION FOR CONTINUOUS DIFFUSION LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Continuous diffusion language models are becoming competitive with their discrete counterparts, but generation still requires many sequential network evaluations. Trajectory-based distillation methods have been developed to reduce this cost, while distributional distillation remains underexplored in this setting, particularly in exploiting the discrete structure of language and enabling multi-step generation. We develop a unified formulation of existing distributional distillation approaches that clarifies how the student's probabilistic parameterization determines the available gradient estimators. Building on this formulation, we introduce two methods that share the same student architecture and reverse-KL objective: Simplex-DMD uses a continuous token relaxation and pathwise gradients, while Reinforce-DMD uses categorical token sampling and REINFORCE with a learned density ratio. We extend both methods to multi-step generation and investigate the training and sampling choices associated with each parameterization. On OpenWebText, using a sequence length of 1,024 tokens, Simplex-DMD and Reinforce-DMD improve the generative perplexity-entropy frontier over competing methods at complementary sampling budgets. In particular, at comparable entropy, Simplex-DMD reduces generative perplexity by 49% at 4 network evaluations, while Reinforce-DMD achieves a 20% reduction at 256 evaluations, each relative to the best evaluated baseline at the same budget.

1 INTRODUCTION

Diffusion language models operate either directly on discrete token space (Austin et al., 2021; Lou et al., 2024; Sahoo et al., 2024) or in a continuous embedding space (Dieleman et al., 2022), with recent continuous models achieving competitive performance (Chen et al., 2026; Yang et al., 2026; Hu et al., 2026). Both approaches update all sequence positions in parallel, yet generation still requires many sequential network evaluations. Reducing this cost while preserving generation quality is the goal of diffusion distillation.

Distillation trains a student to generate samples in fewer steps using a pretrained teacher. Trajectory-based approaches, including consistency and flow map methods, supervise student predictions or transitions along the teacher's dynamics (Song et al., 2023; Boffi et al., 2025b). Distributional approaches instead match generated and target distributions without prescribing the trajectories that connect them. They have enabled one-step and multi-step generation for continuous data (Luo et al., 2023; Yin et al., 2024; Salimans et al., 2024). Recent work has also explored distributional distillation for discrete diffusion (Hoogeboom et al., 2026; Li et al., 2026) and trajectory-based flow maps for continuous diffusion on discrete data (Roos et al., 2026; Potaptchik et al., 2026; Lee et al., 2026).

We study distributional distillation for Continuous Diffusion Models on discrete data (CDMd). In these CDMd, a denoising network estimates the tokenwise posterior distributions of the clean sequence given the noisy embedding sequence and its noise level. We retain this output parameterization for the student. The challenge is to derive tractable training procedures that use a pretrained teacher to match the student's generated distribution to the data distribution in fewer sampling steps. We compare student and data distributions after noising in embedding space and consider two uses of the student's probability vectors: as continuous token representations, or as categorical distributions from which tokens are sampled. These constructions induce different student output distributions.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The continuous relaxation permits differentiation through the generated representations, whereas we optimize categorical sampling with a score-function estimator.

Our contributions are as follows:

- We formulate a unified distribution-matching framework in which noised student and data distributions are compared through a generic discrepancy. Different choices of discrepancy recover distribution-matching objectives used in prior work. We further extend this formulation to multi-step generation.
- We specialize our work to reverse-KL matching, and derive two training procedures. Simplex Distribution Matching Distillation (Simplex-DMD) uses continuous token relaxations and pathwise gradients, with an auxiliary denoiser estimating the score of the noised student distribution, following Diff-Instruct and DMD (Luo et al., 2023; Yin et al., 2024). REINFORCE Distribution Matching Distillation (Reinforce-DMD) samples discrete tokens and uses a score-function estimator (Williams, 1992), with a discriminator estimating the density ratio between noised student and data distributions. Reinforce-DMD additionally uses a KL regularizer toward the teacher's categorical predictions.
- We evaluate both methods on OpenWebText, distilling the same LangFlow teacher (Chen et al., 2026) to generate sequences of 1,024 tokens. We assess quality and diversity jointly through generative perplexity–entropy frontiers under temperature sweeps (Pynadath et al., 2026). The two methods improve these frontiers in complementary sampling regimes: Simplex-DMD at few-step budgets of 4–16 network evaluations, and Reinforce-DMD at larger multi-step budgets of 128–256 evaluations. (Figure 1). Ablations examine the training and sampling choices supporting these gains.

Notation. Let K denote the vocabulary size, let $\mathsf{V} = \{e^1, \dots, e^K\}$ be the canonical basis of $\mathbb{R}^K$ and $\mathsf{X} = \mathsf{V}^L$ the space of sequences of length L. Capital letters denote random variables and lower-case letters their realizations. We represent a discrete sample by a matrix $\mathbf{x} \in \mathbb{R}^{L\times K}$ with rows $\mathbf{x}^\ell \in \mathsf{V}$ and entries $\mathbf{x}^{\ell k}$. We write $\Delta_K = \{u \in \mathbb{R}^K : u^k \geq 0,\ \sum_{k=1}^K u^k = 1\}$ for the probability simplex. For $\mathsf{S} \subseteq \mathbb{R}^n$, S^L denotes the matrices with L rows in S. The embedding matrix $\mathbf{E} \in \mathbb{R}^{K\times d}$ maps $\mathbf{x}$ to $\mathbf{x}\mathbf{E}$. We denote the stop-gradient operator by $\mathrm{sg}(\cdot)$.

2 Preliminaries: Continuous Diffusion for Discrete Data

Let p_{data} be a distribution on X. Continuous Diffusion Models for discrete data (CDMd) embed each discrete sample $\mathbf{x}$ as $\mathbf{x}\mathbf{E} \in \mathbb{R}^{L\times d}$ and learn to reverse a Gaussian corruption process in embedding space. Generation starts from Gaussian noise, proceeds through learned denoising transitions, and ends with a categorical decoder that produces a token sequence. We describe this construction within the Variational Diffusion Model (VDM) framework (Kingma et al., 2021). The embedding matrix $\mathbf{E}$ may be fixed or learned jointly with the model (Strudel et al., 2022; Li et al., 2022).

Let $(\alpha_t)_{t\in[0,1]}$ be a decreasing noise schedule with $\alpha_0 = 1$ and $\alpha_1 = 0$, and set $\sigma_t^2 = 1 - \alpha_t^2$. Given $\mathbf{X} \sim p_{\mathrm{data}}$, the forward process $(Z_t)_{t\in[0,1]}$ is the Markov process in $\mathbb{R}^{L\times d}$ with conditional laws $q_{t|\mathbf{x}}(\cdot \mid \mathbf{x}) = \mathrm{N}(\alpha_t\mathbf{x}\mathbf{E}, \sigma_t^2\,\mathrm{Id})$ and transition kernels $q_{t|s}(\cdot \mid z_s) = \mathrm{N}(\alpha_{t|s}z_s, \sigma_{t|s}^2\,\mathrm{Id})$ for $0 \leq s < t \leq 1$, where $\alpha_{t|s} = \alpha_t/\alpha_s$, $\sigma_{t|s}^2 = \sigma_t^2 - \alpha_{t|s}^2\sigma_s^2$, and Id is the identity on vectorized latent matrices. We denote by p_t the marginal density of Z_t.

Fix a grid $0 = t_0 < \cdots < t_n = 1$ and write $s_i = t_{i-1}$. The reverse generative model is given by

$$p^\theta(\mathbf{x}) = \int p_1(z_1)\, p^\theta_{\mathbf{x}|0}(\mathbf{x} \mid z_0) \prod_{i=1}^n p^\theta_{s_i|t_i}(z_{s_i} \mid z_{t_i})\, \mathrm{d}z_{t_0:t_n}\,,$$

The transitions in this model are given using a denoising network $\mathbf{x}^\theta(z_t, t) \in \Delta_K^L$ by

$$p^\theta_{s|t}(z_s \mid z_t) = q_{s|t,\mathbf{x}}\big(z_s \mid z_t,\, \mathbf{x}^\theta(z_t, t)\big)\,,$$

where $q_{s|t,\mathbf{x}}$ is the conditional density of Z_s given $Z_t = z_t$ and $\mathbf{X} = \mathbf{x}$, that we extend to simplex-valued inputs. By Bayes' rule,

$$q_{s|t,\mathbf{x}}(z_s \mid z_t, \mathbf{x}) = \frac{q_{t|s}(z_t \mid z_s)\, q_{s|\mathbf{x}}(z_s \mid \mathbf{x})}{q_{t|\mathbf{x}}(z_t \mid \mathbf{x})} = \mathrm{N}\left(z_s;\ \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2} z_t + \frac{\alpha_s\sigma_{t|s}^2}{\sigma_t^2}\mathbf{x}\mathbf{E},\ \frac{\sigma_s^2\sigma_{t|s}^2}{\sigma_t^2}\,\mathrm{Id}\right). \quad (1)$$

The same network sets the conditional likelihood as $p^{\theta}_{\mathbf{x}|0}(\mathbf{x} \mid z_0) = \prod_{\ell=1}^{L} \langle \mathbf{x}^{\theta,\ell}(z_0, 0), \mathbf{x}^{\ell} \rangle$. The generative model is fitted by minimizing the negative evidence lower bound (NELBO); as the time-grid mesh tends to zero, it takes the form (Kingma et al., 2021; Gulrajani & Hashimoto, 2023):

$$\mathsf{L}_{\mathrm{c}}(\mathbf{x}; \theta) = \mathsf{L}_{\mathrm{prior}}(\mathbf{x}) + \mathbb{E}_{Z_0 \sim q_{0|\mathbf{x}}(\cdot|\mathbf{x})} \left[-\sum_{\ell=1}^{L} \log \langle \mathbf{x}^{\theta,\ell}(Z_0, 0), \mathbf{x}^{\ell} \rangle \right] + \frac{1}{2} \int_0^1 [-\mathrm{SNR}'(t)] \, \mathbb{E}_{Z_t \sim q_{t|\mathbf{x}}(\cdot|\mathbf{x})} \left[\left\| \big(\mathbf{x}^{\theta}(Z_t, t) - \mathbf{x}\big)\mathbf{E} \right\|_{\mathrm{F}}^2 \right] \mathrm{d}t \, ,$$

where $\mathrm{SNR}(t) = \alpha_t^2 / \sigma_t^2$, and $\mathsf{L}_{\mathrm{prior}}$ denote the endpoint prior (see Equation (14)). The diffusion loss is such that, when $t > 0$, the embedded denoiser output $\mathbf{x}^{\theta}\mathbf{E}$ targets the posterior mean in embedding space $(t, z_t) \mapsto m_t(z_t)\mathbf{E}$, where $m_t(z_t) = \mathbb{E}[\mathbf{X} | Z_t = z_t] \in \Delta_K^L$ is the conditional mean of $\mathbf{X}$ given Z_t. For $t = 0$, the reconstruction term ensures that denoiser network's target is m_0. Alternatively, the denoiser network can be trained directly by cross-entropy (Dieleman et al., 2022; Chen et al., 2026):

$$\mathsf{L}_{\mathrm{CE}}(\mathbf{x}; \theta) = \mathbb{E}_{\substack{t \sim \mathrm{Unif}(0,1) \\ Z_t \sim q_{t|\mathbf{x}}(\cdot|\mathbf{x})}} \left[-\sum_{\ell=1}^{L} \log \langle \mathbf{x}^{\theta,\ell}(Z_t, t), \mathbf{x}^{\ell} \rangle \right] \, ,$$

for which the target at optimum is m_t. Although not equivalent to the NELBO above, the cross-entropy loss can yield a likelihood bound with a suitable weighting (Davis et al., 2026). Embedding normalization, self-conditioning, and learned noise schedules are further design choices detailed in Appendix B.1.

3 Distribution Matching Distillation for Continuous Diffusion on Discrete Data

Distributional distillation trains a student by matching the distribution of its generated samples to the data distribution. In our setting, the student takes as input a latent Z_T at noise level $T \in (0, 1]$ and produces a distribution over clean token representations. These outputs are then renoised with the Gaussian forward process of Section 2, and their distribution are matched with the corresponding data marginals p_t at lower noise levels $t < T$. For the approaches considered, the matching objective involves the fixed pretrained CDMd p^{θ}, used as the distillation teacher.

More precisely, given $Z_T = z_T$, the student network outputs $\mathbf{x}^{\eta}(z_T, T) \in \Delta_K^L$, which parameterizes a conditional kernel $\mathsf{k}_T^{\eta}(\cdot \mid z_T)$. We consider two parameterizations of this kernel. The first uses the probability vectors themselves as simplex-valued token representations, while the second samples one categorical token at each sequence position:

$$\mathsf{k}_T^{\eta}(\mathrm{d}\mathbf{x} \mid z_T) = \delta_{\mathbf{x}^{\eta}(z_T, T)}(\mathrm{d}\mathbf{x}), \quad \mathbf{x} \in \Delta_K^L, \qquad \text{(continuous relaxation)}, \tag{2}$$

$$\mathsf{k}_T^{\eta}(\mathbf{x} \mid z_T) = \prod_{\ell=1}^{L} \left\langle \mathbf{x}^{\ell}, \mathbf{x}^{\eta,\ell}(z_T, T) \right\rangle, \quad \mathbf{x} \in \mathsf{X}, \qquad \text{(discrete sampling)}. \tag{3}$$

These two parameterizations share the same distribution-matching objective but lead to different gradient estimators. We formulate this objective for one-step generation in a framework unifying existing methods by considering a general class of matching objectives, extend it to multi-step generation and derive Simplex-DMD for the continuous relaxation and Reinforce-DMD for discrete sampling.

3.1 A unified view of one-step distribution matching

We first consider one-step generation and isolate a common formulation of distributional distillation. Existing distribution-matching methods can be viewed as comparing the noised student and data distributions across noise levels, while differing in the discrepancy used to compare these marginals. We formalize this view through a generic discrepancy D, before specializing it to the reverse KL used by our methods.

In the one-step setting, $T = 1$ and the student starts from $Z_1 \sim p_1 = \mathrm{N}(0, \mathrm{Id})$. Its clean output $\mathbf{X}^{\eta}$ is sampled according to $\mathsf{k}_1^{\eta}(\cdot \mid Z_1)$, inducing the marginal distribution $p_{\mathbf{x}}^{\eta} = \int \mathsf{k}_1^{\eta}(\cdot \mid z_1) \, p_1(z_1) \, \mathrm{d}z_1$.

Applying the Gaussian forward process of Section 2 to $\mathbf{X}^\eta \sim p^\eta_{\mathbf{x}}$ gives a latent process $(Z^\eta_t)_{t\in[0,1]}$. We denote its marginals by p^η_t, with the corresponding data marginal being p_t. Comparing $p^\eta_{\mathbf{x}}$ with p_{data} is difficult; Gaussian noising smoothes both distributions in the same embedding space, where the diffusion teacher p^θ provides quantities that can be leveraged by the loss objectives. Matching these marginals across noise levels defines the generic one-step distribution-matching objective:

$$\mathsf{L}_{\text{distill}}(\eta) = \int_0^1 \mathbb{E}_{Z^\eta_t \sim p^\eta_t}[\mathsf{D}(p^\eta_t, p_t, Z^\eta_t)]\ \mathrm{d}t\ , \tag{4}$$

where D denote a local discrepancy between student and data marginals evaluated at z_t. Different choices of D recover distribution-matching objectives used in prior work, as summarized in Table 1; full objectives and derivations are given in Appendix C. The gradient of (4) has two contributions: (A) from the dependence of the sampling law on η, and (B) from the dependence of D on p^η_t:

$$\nabla_\eta \mathsf{L}_{\text{distill}}(\eta) = \int_0^1 \Bigg[\underbrace{\int \mathsf{D}(p^\eta_t, p_t, z_t)\, \nabla_\eta p^\eta_t(z_t)\, \mathrm{d}z_t}_{(A)} + \underbrace{\int p^\eta_t(z_t)\, \nabla_\eta \mathsf{D}(p^\eta_t, p_t, z_t)\, \mathrm{d}z_t}_{(B)} \Bigg] \mathrm{d}t. \tag{5}$$

Criterion	Local discrepancy D	Representative methods
Continuous diffusion		
Reverse KL	$\log p^\eta_t(z_t) - \log p_t(z_t)$	Diff-Instruct (Luo et al., 2023), DMD (Yin et al., 2024)
Score / Velocity / Moment matching	$\left\|\nabla_{z_t} \log p^\eta_t(z_t) - \nabla_{z_t} \log p_t(z_t)\right\|^2$	SiD (Zhou et al., 2024); FGM (Huang et al., 2024); Moment matching (Salimans et al., 2024)
Discrete diffusion		
Bregman divergence	$\sum_{d_{\mathrm{H}}(z',z_t)=1} d_F\left(\frac{p^\eta_t(z')}{p^\eta_t(z_t)}, \frac{p_t(z')}{p_t(z_t)}\right)$	IDLM (Li et al., 2026); D-MMD (Hoogeboom et al., 2026)
Posterior f-divergence	$\sum_\ell D_f\left(p^{\eta,\ell}_{\mathbf{x}|t}(\cdot \mid z_t),\, p^\ell_{\mathbf{x}|t}(\cdot \mid z_t)\right)$	DiMO (Zhu et al., 2025)

Table 1: Local discrepancies for distributional distillation. Score, velocity and moment discrepancies agree up to time-dependent factors. Definitions and derivations are given in Appendix C.

Following Diff-Instruct and DMD (Luo et al., 2023; Yin et al., 2024), we focus in this work on the reverse KL, choosing $\mathsf{D}(p^\eta_t, p_t, z_t) = \log\left(p^\eta_t(z_t)/p_t(z_t)\right)$. The loss thus becomes:

$$\mathsf{L}_{\text{DMD}}(\eta) = \int_0^1 \mathrm{KL}(p^\eta_t \,\|\, p_t)\ \mathrm{d}t. \tag{6}$$

In this case, (B) integrates to zero (Appendix D.1), leaving only the dependence of the sampled student marginal on η in (A). Estimating this term depends on the student parameterization: the continuous relaxation admits a pathwise gradient, whereas discrete sampling requires a score function estimator. We derive the two estimators in Sections 3.3 and 3.4.

The previous objective (4) matches the student and data distributions only after embedding and Gaussian noising. It is therefore not immediate that exact matching of the noised marginals recovers the original discrete data distribution. The following result shows that this is nevertheless the case, under a simple geometric condition on the embedding matrix that we satisfy.

Theorem 1. *Assume that the rows of* $\mathbf{E}$ *are distinct and have a common Euclidean norm. Then*

$$\mathsf{L}_{\text{DMD}}(\eta) = 0 \qquad \Longleftrightarrow \qquad p^\eta_{\mathbf{x}} = p_{\text{data}}.$$

We give the proof in Appendix D.2.

3.2 Multi-step generation

The one-step construction of Section 3.1 learns to transport the terminal marginal p_1 to the data distribution. In order to extend the method to multi-step generation, we adapt this construction to arbitrary starting noise levels $T \in (0, 1]$: given $Z_T \sim p_T$, the student, conditioned on T, produces

a clean sample $\mathbf{X}^{\eta,T} \sim \mathsf{k}_T^\eta(\cdot \mid Z_T)$, and we denote its marginal by $p_{\mathbf{x}}^{\eta,T}$. Renoising this sample to a noise level $t < T$ with the forward process of Section 2 yields the latent states $(Z_t^{\eta,T})$, inducing marginals $p_t^{\eta,T}$, which we match to the data marginal p_t using the same reverse-KL objective as in the one-step case (6). Varying T yields the multi-step objective:

$$\mathsf{L}_{\mathrm{DMD}}^{\mathrm{ms}}(\eta) = \mathbb{E}_{(t,T)\sim\nu}\left[\mathrm{KL}\left(p_t^{\eta,T} \,\middle\|\, p_t\right)\right] \ , \tag{7}$$

where the joint distribution ν over (t,T) is defined as $\nu(\mathrm{d}t,\mathrm{d}T) = \nu_T(\mathrm{d}t)\,\nu_1(\mathrm{d}T)$, with $\nu_s = \mathrm{Unif}(0,s)$ for $s \in (0,1]$. Learning these transports for all starting levels enables multi-step generation by composing them along decreasing noise levels $1 = T_n > T_{n-1} > \cdots > T_0 = 0$. Starting from $\widehat{Z}_{T_n} \sim p_1$, at each level T_i the student draws $\widehat{\mathbf{X}}^{\eta,T_i} \sim \mathsf{k}_{T_i}^\eta(\cdot \mid \widehat{Z}_{T_i})$ and maps it to the next level T_{i-1}. We explore the following family of DDIM-inspired transitions (Song et al., 2021), a standard approach for exploring multi-step performance, which controls the stochasticity of the sampling procedure:

$$p_{t|T}^\eta(z_t \mid z_T) = \int q_{t|\mathbf{x},T}^\zeta(z_t \mid \mathbf{x}, z_T)\,\mathsf{k}_T^\eta(\mathrm{d}\mathbf{x} \mid z_T) \ , \qquad t < T,$$

where, for $0 \le \zeta_{t,T} \le \sigma_t$,

$$q_{t|\mathbf{x},T}^\zeta(z_t \mid \mathbf{x}, z_T) = \mathrm{N}\left(z_t;\, \alpha_t \mathbf{x}\mathbf{E} + \sqrt{\sigma_t^2 - \zeta_{t,T}^2}\,\frac{z_T - \alpha_T \mathbf{x}\mathbf{E}}{\sigma_T}, \zeta_{t,T}^2\,\mathrm{Id}\right) \ . \tag{8}$$

The parameter $\zeta_{t,T}$ interpolates between retaining the noise inferred from the current latent and injecting fresh Gaussian noise. At $\zeta_{t,T} = 0$, the update is deterministic conditional on $\mathbf{x}$. At $\zeta_{t,T} = \sigma_t$, it discards the current latent noise and recovers the forward-renoising transition used to define $Z_t^{\eta,T}$, whereas $\zeta_{t,T} = \sigma_t\sigma_{T|t}/\sigma_T$ recovers the Gaussian bridge (1). Finally, the generative model takes a final step from $\widehat{Z}_0$ by applying the categorical decoder parameterized by $\mathbf{x}^\eta(\widehat{Z}_0, 0)$, yielding $n+1$ NFE. In the forward renoising $\zeta_{t,T} = \sigma_t$ case, these transitions recover the correct diffusion marginals p_t if $\widehat{Z}_T \sim p_T$ and the induced clean student marginal equals p_{data}; in general, the diffusion marginals are also recovered if the student kernel equals the exact posterior $p_{\mathbf{x}|T}(\cdot \mid z_T)$, see Appendix D.5.

3.3 Simplex-DMD: continuous relaxation

Under the continuous relaxation (2), the student's probability vectors are used as the clean sample: $\mathbf{X}^{\eta,T} = \mathbf{x}^\eta(Z_T, T)$. The noising process is differentiable with respect to the student parameters, allowing us to optimize the multi-step distribution-matching objective with a pathwise gradient.

Student gradient. We first derive the gradient of the matching objective (7). Define the reparameterized forward-noising map $\pi_t : (\mathbf{x}, \varepsilon) \mapsto \alpha_t \mathbf{x}\mathbf{E} + \sigma_t\varepsilon$, such that $Z_t^{\eta,T} \overset{d}{=} \pi_t(\mathbf{x}^\eta(Z_T,T),\varepsilon)$ for $Z_T \sim p_T$ and $\varepsilon \sim \mathrm{N}(0,\mathrm{Id})$. This explicits the dependence of the noised latent on η and how the reparameterization trick (Kingma & Welling, 2014) applies to (7), similarly to Diff-Instruct and DMD (Luo et al., 2023; Yin et al., 2024):

$$\nabla_\eta \mathsf{L}_{\mathrm{DMD}}^{\mathrm{fs}}(\eta) = \mathbb{E}_{(t,T)\sim\nu,\, Z_T\sim p_T,\, \varepsilon\sim\mathrm{N}(0,\mathrm{Id})}\left[\left(s^{\eta,T}(Z_t^{\eta,T},t) - s(Z_t^{\eta,T},t)\right)^\top \frac{\partial Z_t^{\eta,T}}{\partial \eta}\right] \ , \tag{9}$$

where $s^{\eta,T}(z,t) = \nabla_z \log p_t^{\eta,T}(z)$ and $s(z,t) = \nabla_z \log p_t(z)$ are respectively the student and data scores (see Appendix D.1). Thus, the pathwise update requires evaluating the difference between the score of the noised student distribution and that of the data distribution.

Auxiliary score estimation. Neither score in (9) is available exactly, but under Gaussian noising, each is determined by the posterior mean of its corresponding clean representation. This relation is given by Tweedie's formula (Efron, 2011):

$$s(z,t) = \frac{\alpha_t m_t(z)\mathbf{E} - z}{\sigma_t^2} \ , \qquad s^{\eta,T}(z,t) = \frac{\alpha_t m_t^{\eta,T}(z)\mathbf{E} - z}{\sigma_t^2} \ , \tag{10}$$

where $m_t^{\eta,T}(z) = \mathbb{E}\Big[\mathbf{X}^{\eta,T} \mid Z_t^{\eta,T} = z\Big]$ is the posterior mean of the student distribution. The frozen teacher denoiser $\mathbf{x}^\theta(z,t)$ provides an estimate of the data posterior mean, and hence an estimate $s^\theta(z,t)$ of $s(z,t)$. We train an auxiliary denoiser $\mathbf{x}^\phi(z,t,T) \in \Delta_K^L$ to target the posterior mean $m_t^{\eta,T}$ of the student distribution, by minimizing the following soft-target cross-entropy:

$$\mathsf{L}_{\text{aux}}(\phi) = \mathbb{E}_{(t,T)\sim\nu,\, Z_T\sim p_T,\, \varepsilon\sim\mathrm{N}(0,\mathrm{Id})}\left[\sum_{\ell=1}^{L} \mathrm{CE}\Big(\mathbf{X}^{\eta,T,\ell}, \mathbf{x}^{\phi,\ell}(Z_t^{\eta,T}, t, T)\Big)\right]. \tag{11}$$

We alternate n_{aux} auxiliary updates on student outputs detached with one student update.

3.4 REINFORCE-DMD: DISCRETE SAMPLING

The student clean sample $\mathbf{X}^{\eta,T}$ is now drawn from the discrete distribution $\mathsf{k}_T^\eta(\cdot \mid Z_T) = \mathbf{x}^\eta(Z_T, T)$ of (3), preventing the direct pathwise differentiation used by Simplex-DMD.

Student gradient. The gradient of the multi-step objective (7) admits the same decomposition as that of the one-step derived in (5). In our reverse KL matching setting, contribution (B) vanishes, so it remains to differentiate the conditional law of $\mathbf{X}^{\eta,T}$ which gives the exact score-function identity (Williams, 1992):

$$\nabla_\eta \mathsf{L}_{\text{DMD}}^{\text{ms}}(\eta) = \mathbb{E}_{\substack{T\sim\nu_1, Z_T\sim p_T \\ \mathbf{X}^{\eta,T}\sim \mathsf{k}_T^\eta(\cdot\mid Z_T)}} \left[C_T(\mathbf{X}^{\eta,T})\, \nabla_\eta \log \mathsf{k}_T^\eta(\mathbf{X}^{\eta,T} \mid Z_T)\right],$$

$$C_T(\mathbf{X}^{\eta,T}) = \mathbb{E}_{\substack{t\sim\nu_T \\ Z_t^{\eta,T}\sim q_{t|\mathbf{x}}(\cdot\mid\mathbf{X}^{\eta,T})}} \left[\log \frac{p_t^{\eta,T}(Z_t^{\eta,T})}{p_t(Z_t^{\eta,T})}\right].$$

The cost C_T is the expected log-ratio of the student and data marginals at time t. Unlike the scores in the Simplex-DMD gradient, which diffusion denoisers can estimate, this log-ratio is not directly available and requires another estimator.

Auxiliary estimation. In order to estimate this log-ratio of interest, we train a discriminator $D_\psi(z_t, t, T) \in (0,1)$ to distinguish noised data from noised student samples:

$$\mathsf{L}_{\text{disc}}(\psi) = -\int \left[\mathbb{E}_{Z_t\sim p_t} \log D_\psi(Z_t,t,T) + \mathbb{E}_{Z_t^{\eta,T}\sim p_t^{\eta,T}} \log\left(1 - D_\psi(Z_t^{\eta,T},t,T)\right)\right] \nu(\mathrm{d}t, \mathrm{d}T). \tag{12}$$

For a fixed student, the population optimum is $D^\star(z,t,T) = p_t(z)/(p_t(z) + p_t^{\eta,T}(z))$, whose negative logit equals $\log(p_t^{\eta,T}/p_t)$ (Appendix D.3). For a pair $0 < t < T \le 1$ and independent noise $\varepsilon \sim \mathrm{N}(0, \mathrm{Id})$, the pointwise cost estimate is:

$$\widehat{C}_\psi(\mathbf{X}^{\eta,T}; t, T, \varepsilon) = \log \frac{1 - D_\psi(\pi_t(\mathbf{X}^{\eta,T}, \varepsilon), t, T)}{D_\psi(\pi_t(\mathbf{X}^{\eta,T}, \varepsilon), t, T)},$$

Variance reduction and teacher supervision. For each (Z_T, T), we draw $G \ge 2$ independent triplets $(\mathbf{X}_g^{\eta,T}, t_g, \varepsilon_g)$ with $\mathbf{X}_g^{\eta,T} \sim \mathsf{k}_T^\eta(\cdot \mid Z_T)$, $t_g \sim \nu_T$, and $\varepsilon_g \sim \mathrm{N}(0, \mathrm{Id})$. We compute the costs $\widehat{C}_g = \widehat{C}_\psi(\mathbf{X}_g^{\eta,T}; t_g, T, \varepsilon_g)$ and the leave-one-out baselines $\widehat{b}_g = (G-1)^{-1} \sum_{g'\neq g} \widehat{C}_{g'}$ (Kool et al., 2019). Inspired by the group-based baseline in GRPO, we center each cost by subtracting $\widehat{b}_g$ (Shao et al., 2024), and include a KL anchor required for stable training in our experiments (Section 4.2), yielding the per-group surrogate:

$$\widehat{\mathsf{L}}_{\text{R-DMD}}(\eta) = \frac{1}{G}\sum_{g=1}^{G} \mathrm{sg}(\widehat{C}_g - \widehat{b}_g) \log \mathsf{k}_T^\eta(\mathbf{X}_g^{\eta,T} \mid Z_T) + \beta \sum_{\ell=1}^{L} \mathrm{KL}\big(\mathbf{x}^{\eta,\ell}(Z_T,T) \,\big\|\, \mathbf{x}^{\theta,\ell}(Z_T,T)\big), \tag{13}$$

where $\beta \ge 0$. We average this surrogate over sampled (Z_T, T) and their groups. The leave-one-out baseline preserves the expected cost-based update. We use $G = 4$, requiring one student evaluation and G cost evaluations per group. This, given this anchor, any improvements over the base teacher model is only due to the matching objective.

We alternate n_{aux} discriminator updates on detached student samples with one student update using (13). During student updates, the discriminator, costs and baselines are detached.

3.5 Related work

Continuous diffusion for discrete data. Diffusion-LM jointly learns embeddings and a diffusion model (Li et al., 2022), SED uses self-conditioning to embedding diffusion (Strudel et al., 2022), Plaid develops a likelihood-based training framework (Gulrajani & Hashimoto, 2023). Alongside these developments, CDCD introduced score interpolation, parameterizing the score through an estimate of the posterior mean (Dieleman et al., 2022). More recently, LangFlow (Chen et al., 2026) and RePlaid (Yang et al., 2026) refined noise scheduling and training, achieving language modeling performance competitive with that of discrete diffusion models.

Trajectory distillation. *Consistency models* learn to map states along a probability-flow trajectory to a common endpoint, enabling generation with one or a few model evaluations (Song et al., 2023; Song & Dhariwal, 2024). *Flow map matching* generalizes this construction to maps between arbitrary times, with objectives for distilling pretrained models as well as training directly from data through self-distillation (Boffi et al., 2025a;b). *Categorical flow maps*, including CFM (Roos et al., 2026), DFM (Potaptchik et al., 2026), and FMLM (Lee et al., 2026), build on CDMd to extend these ideas to discrete data, enabling one- and few-step generation. For discrete diffusion, SDTT matches predictions along teacher rollouts (Deschenaux & Gulcehre, 2025), while Di4C distills transition compositions into mixture models that capture correlations across dimensions (Hayakawa et al., 2025). A complementary approach is taken by Rectified Flow (Liu et al., 2023) and a discrete adaptation ReDi (Yoo et al., 2025), which iteratively rectifies the source-target coupling in flow matching to accelerate sampling.

Distributional distillation. Distributional distillation methods differ primarily in their choice of local discrepancy and how they estimate the loss gradient. Table 1 summarizes these methods, and Appendix C provides their derivations.

4 Experiments

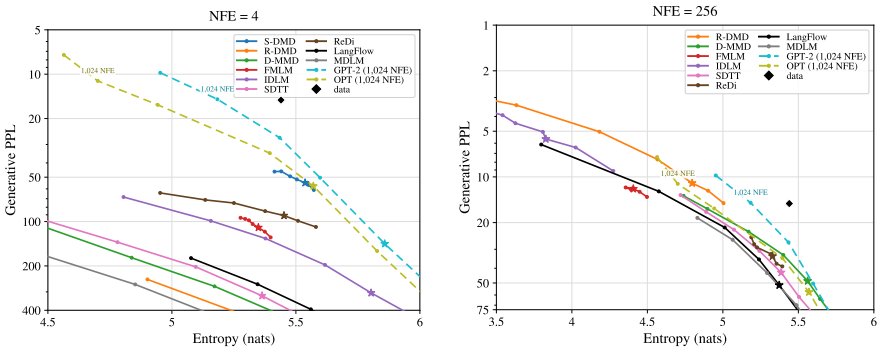
We evaluate whether Simplex-DMD and Reinforce-DMD preserve generation quality and diversity when distilling a continuous diffusion language model to reduced sampling budgets. We compare them with diffusion and distillation baselines across network evaluation budgets, and ablate the main training and sampling choices.

We distill the released LangFlow model (Chen et al., 2026) on OpenWebText, using sequences of length 1024 over the GPT-2 vocabulary ($K = 50{,}257$). We keep the teacher embedding matrix $\mathbf{E}$ fixed and initialize both students and their auxiliary networks from the teacher. Training runs for $10{,}000$ total iterations with $n_{\text{aux}} = 4$ auxiliary updates per student update, i.e., $2{,}000$ student updates, requiring approximately 40 H100 GPU hours.

We evaluate generation quality with generative perplexity (Gen PPL, lower is better), computed by GPT-2 Large, and diversity with unigram entropy (higher is better). Since lower perplexity can result from reduced diversity, we compare Gen PPL–entropy frontiers obtained by varying the sampling temperature (Pynadath et al., 2026). Baselines include the LangFlow teacher, the continuous-diffusion distillation method FMLM, the discrete diffusion models and distillation methods MDLM, D-MMD, IDLM, SDTT, and ReDi (see Section 3.5), and the autoregressive references GPT-2 (124M) and OPT (125M). All methods use a common evaluation protocol. Table 3 reports model sizes and training budgets, while Appendix E and table 4 provide further evaluation and configuration details.

4.1 Generation quality across sampling budgets

Figure 1 compares generative frontiers at representative low and high sampling budgets. At 4 network evaluations, Simplex-DMD achieves lower Gen PPL than the evaluated diffusion baselines throughout the overlapping entropy range. At 256 evaluations, Reinforce-DMD achieves the lowest Gen PPL among the evaluated methods at matched entropy around 5.0. The two methods therefore improve generation quality in complementary sampling regimes: Simplex-DMD at low budgets and Reinforce-DMD at larger budgets.



(a) 4 network evaluations.

(b) 256 network evaluations.

Figure 1: Generative frontiers on OpenWebText at NFE $= 4$ and 256. Temperature varies from 0.8 to 1.1; stars mark temperature 1.0 and diamonds the data. Gen PPL uses an inverted log scale. Dashed GPT-2 and OPT use 1,024 evaluations. Simplex-DMD is omitted at NFE $= 256$ due to low diversity. Full sweeps are in Appendix E.2.

	Network evaluations					
Method	2	4	8	16	128	256
H_{NFE}	5.65	5.44	5.38	5.20	5.27	5.00
LangFlow (Chen et al., 2026)	1077	317	147	58.6	39.1	21.3
MDLM (Sahoo et al., 2024)	1909	678	267	72.8	44.3	23.8
FMLM (Lee et al., 2026)	142 (5.32)	128 (5.40)	94.7	55.2	22.6 (4.87)	13.5 (4.50)
D-MMD (Hoogeboom et al., 2026)	1752	431	135	41.7	27.3	18.6
IDLM (Li et al., 2026)	955	145	48.4	24.3	11.6 (4.75)	9.14 (4.27)
SDTT (Deschenaux & Gulcehre, 2025)	1496	372	103	39.4	32.6	20.1
ReDi (Yoo et al., 2025)	261 (5.46)	90.2	49.4	32.4 (5.32)	30.7	25.0 (5.19)
Simplex-DMD	**178**	**45.6**	**26.7**	**17.0**	3.60 (3.82)	2.51 (3.31)
Reinforce-DMD	1112	556	347	68.8	**26.9**	**14.9**

Table 2: Gen PPL on OpenWebText at matched unigram entropy H_{NFE}; lower is better. H_{NFE} is the Simplex-DMD entropy closest to the data for NFE $=$ 2–16, and the corresponding Reinforce-DMD entropy for NFE $= 128, 256$. Log Gen PPL is linearly interpolated at H_{NFE}. If a curve does not reach the target entropy, its closest point is reported with entropy in parentheses. Bold indicates the lowest matched value.

At NFE $= 4$, matching the entropy at 5.44, Simplex-DMD reaches a Gen PPL of 45.6, compared with 90.2 for ReDi, corresponding to a 49% reduction. At NFE $= 256$ and matched entropy 5.00, Reinforce-DMD reaches a Gen PPL of 14.9, compared with 18.6 for D-MMD, a 20% reduction.

With just four network evaluations, Simplex-DMD approaches the quality-diversity frontier of OPT-125M which uses 1,024 NFEs, bringing autoregressive-level performance within reach of few-step diffusion generation. At matched unigram entropy $H \approx 5.54$, their Gen PPL values are 54.6 and 52.1, respectively. Reinforce-DMD also compares favorably with OPT: at $H \approx 5.00$, it achieves Gen PPL 14.9 versus 17.9, using 256 evaluations.

To compare methods across sampling budgets more systematically, Table 2 reports Gen PPL at a budget-specific matched-entropy target H_{NFE}. We linearly interpolate log Gen PPL between measured frontier points that bracket each target. Simplex-DMD obtains the lowest matched-entropy Gen PPL at 2, 4, 8, and 16 network evaluations, whereas Reinforce-DMD obtains the lowest values at 128 and 256 evaluations. At NFE $= 2$, FMLM reaches a lower raw Gen PPL than Simplex-DMD (142 versus 178), but only at a substantially lower entropy (5.32 versus 5.65), and is therefore not a matched-diversity comparison.

4.2 Training and Sampling Choices

We explicit the design choices made to obtain the results presented in the previous subsection, and the related ablations validating them.

Teacher supervision stabilizes Reinforce-DMD. Figure 2 studies the KL anchor toward the teacher predictions introduced in (13). Removing this term ($\beta = 0$) leads to collapse, with the generated distribution reaching essentially zero unigram entropy. In contrast, the tested positive values of β, as well as an annealed schedule, produce similar generative frontiers. Thus, teacher supervision is necessary for stable Reinforce-DMD training in our experiments, while performance is relatively insensitive to its precise weight within the tested range. We use $\beta = 0.5$ in the default configuration.

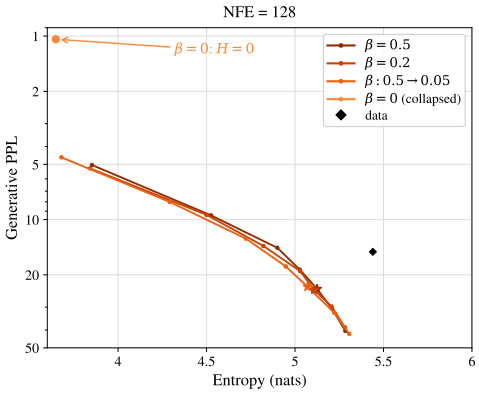


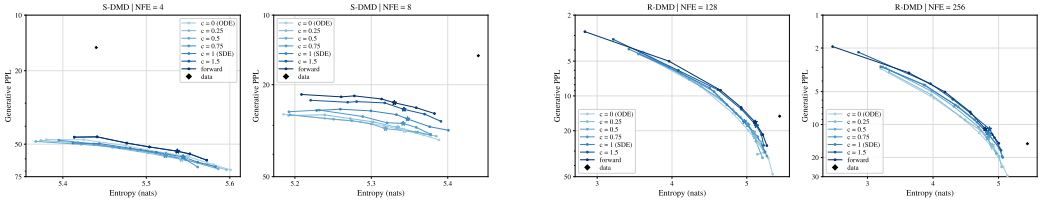
Figure 2: Effect of the KL anchor weight β on Reinforce-DMD at NFE $= 128$. Without the anchor, generation collapses ($H = 0$, off-scale).

Forward renoising performs best among the tested samplers. The multi-step construction of Section 3.2 defines a family of sampling transitions through the noise coefficient $\zeta_{t,T}$. We compare

$$\zeta_{t,T} = c\,\frac{\sigma_t \sigma_{T|t}}{\sigma_T}, \qquad c \in \{0, 0.25, 0.5, 0.75, 1, 1.5\},$$

clipping c to respect the constraints on $\zeta_{t,T}$ if necessary, with forward renoising, $\zeta_{t,T} = \sigma_t$, where a maximum amount of fresh Gaussian noise is added at each sampling step. Across the swept budgets, forward renoising gives the best observed Gen PPL–entropy frontier for both Simplex-DMD and Reinforce-DMD (Figure 3).

Additional training choices. Further ablations in Appendices E.3 and E.5 support the remaining choices of our default configuration. Initializing the student from scratch leads to collapse in the tested configurations, motivating initialization from the teacher. Self-conditioning improves Simplex-DMD but provides no benefit for Reinforce-DMD in our experiments, so we use it only for Simplex-DMD. For Simplex-DMD, a bridge-based training construction inspired by moment matching (Salimans et al., 2024) underperforms the forward-renoising training objective; we discuss its auxiliary score approximation in Appendix D.4.



(a) Simplex-DMD, NFE $= 4$ and 8

(b) Reinforce-DMD, NFE $= 128$ and 256

Figure 3: Sampler comparison for $c \in \{0, 0.25, 0.5, 0.75, 1, 1.5\}$ and forward renoising.

5 Conclusion

We introduced a distribution-matching framework for distilling continuous diffusion models on discrete data and derived two methods from a common reverse-KL objective: Simplex-DMD, based on simplex-valued outputs and pathwise gradients, and Reinforce-DMD, based on categorical sampling and score-function estimation. On OpenWebText, they improve generation at complementary sampling budgets, with Simplex-DMD strongest at low budgets and Reinforce-DMD at larger budgets among the evaluated distillation methods. Like other distribution-matching approaches, our methods require a pretrained teacher and an auxiliary model during training. Our experiments are also limited to academic model scale; scaling to substantially larger language models will require studying this training overhead and extending evaluation beyond the Gen PPL–entropy protocol used here.

AI USE STATEMENT

Generative AI tools assisted with writing and editing the manuscript, including improving clarity and structure and providing feedback on the presentation and interpretation of experimental results. They also assisted with coordinating experiment scripts, cluster jobs, and data collection. The authors reviewed and revised all AI-assisted suggestions and take full responsibility for the scientific claims and final content of the paper.

REFERENCES

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. In *Advances in Neural Information Processing Systems*, volume 38, 2025a. URL https://arxiv.org/abs/2505.18825.

Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. *Transactions on Machine Learning Research*, 2025b. URL https://openreview.net/forum?id=cqDH0e6ak2.

Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. Analog bits: Generating discrete data using diffusion models with self-conditioning. In *International Conference on Learning Representations (ICLR)*, 2023.

Yuxin Chen, Chumeng Liang, Hangke Sui, Ruihan Guo, Chaoran Cheng, Jiaxuan You, and Ge Liu. Langflow: Continuous diffusion rivals discrete in language modeling. *arXiv preprint arXiv:2604.11748*, 2026.

Oscar Davis, Anastasiia Filippova, Pierre Ablin, Victor Turrisi, Amitis Shidani, Marco Cuturi, and Louis Béthune. Scaling categorical flow maps. *arXiv preprint arXiv:2605.07820*, 2026.

Justin Deschenaux and Caglar Gulcehre. Beyond autoregression: Fast LLMs via self-distillation through time. In *International Conference on Learning Representations (ICLR)*, 2025.

Sander Dieleman, Laurent Sartran, Arman Roshannai, Nikolay Savinov, Yaroslav Ganin, Pierre H. Richemond, Arnaud Doucet, Robin Strudel, Chris Dyer, Conor Durkan, Curtis Hawthorne, Rémi Leblond, Will Grathwohl, and Jonas Adler. Continuous diffusion for categorical data. *arXiv preprint arXiv:2211.15089*, 2022.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

Samson Gourevitch, Yazid Janati, Dario Shariatian, Umut Simsekli, Eric Moulines, Eric P. Xing, and Alain Durmus. Uniform diffusion models revisited: Leave-one-out denoiser and absorbing state reformulation. *arXiv preprint arXiv:2605.22765*, 2026.

Ishaan Gulrajani and Tatsunori B. Hashimoto. Likelihood-based diffusion language models. *arXiv preprint arXiv:2305.18619*, 2023.

Satoshi Hayakawa, Yuhta Takida, Masaaki Imaizumi, Hiromi Wakaki, and Yuki Mitsufuji. Distillation of discrete diffusion through dimensional correlations. In *International Conference on Machine Learning (ICML)*, 2025.

Emiel Hoogeboom, David Ruhe, Jonathan Heek, Thomas Mensink, and Tim Salimans. Beyond single tokens: Distilling discrete diffusion models via discrete MMD. *arXiv preprint arXiv:2603.20155*, 2026.

Keya Hu, Linlu Qiu, Yiyang Lu, Hanhong Zhao, Tianhong Li, Yoon Kim, Jacob Andreas, and Kaiming He. ELF: Embedded language flows. *arXiv preprint arXiv:2605.10938*, 2026. URL https://arxiv.org/abs/2605.10938.

Zemin Huang, Zhengyang Geng, Weijian Luo, and Guo-jun Qi. Flow generator matching. *arXiv preprint arXiv:2410.19310*, 2024.

Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations*, 2014. URL https://arxiv.org/abs/1312.6114.

Diederik P. Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Wouter Kool, Herke van Hoof, and Max Welling. Buy 4 reinforce samples, get a baseline for free! 2019.

Chanhyuk Lee, Jaehoon Yoo, Manan Agarwal, Sheel Shah, Jerry Huang, Aditi Raghunathan, Seunghoon Hong, Nicholas M. Boffi, and Jinwoo Kim. Flow map language models: One-step language modeling via continuous denoising. *arXiv preprint arXiv:2602.16813*, 2026.

David Li, Nikita Gushchin, Dmitry Abulkhanov, Eric Moulines, Ivan Oseledets, Maxim Panov, and Alexander Korotin. IDLM: Inverse-distilled diffusion language models. *arXiv preprint arXiv:2602.19066*, 2026.

Xiang Lisa Li, John Thickstun, Ishaan Gulrajani, Percy Liang, and Tatsunori B. Hashimoto. Diffusion-LM improves controllable text generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with Rectified Flow. In *International Conference on Learning Representations*, 2023.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *International Conference on Machine Learning (ICML)*, 2024.

Weijian Luo, Tianyang Hu, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhihua Zhang. Diff-instruct: A universal approach for transferring knowledge from pre-trained diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Peter Potaptchik, Jason Yim, Adhi Saravanan, Peter Holderrieth, Eric Vanden-Eijnden, and Michael S. Albergo. Discrete flow maps. *arXiv preprint arXiv:2604.09784*, 2026.

Patrick Pynadath, Jiaxin Shi, and Ruqi Zhang. Generative frontiers: Why evaluation matters for diffusion language models, 2026. URL https://arxiv.org/abs/2604.02718.

Daan Roos, Oscar Davis, Floor Eijkelboom, Michael Bronstein, Max Welling, İsmail İlkan Ceylan, Luca Ambrogioni, and Jan-Willem van de Meent. Categorical flow maps. *arXiv preprint arXiv:2602.12233*, 2026.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T. Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Subham Sekhar Sahoo, Justin Deschenaux, Aaron Gokaslan, Guanghan Wang, Justin Chiu, and Volodymyr Kuleshov. The diffusion duality. In *International Conference on Machine Learning (ICML)*, 2025.

Tim Salimans, Thomas Mensink, Jonathan Heek, and Emiel Hoogeboom. Multistep distillation of diffusion models via moment matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Yair Schiff, Subham Sekhar Sahoo, Hao Phung, Guanghan Wang, Alexander Rush, Volodymyr Kuleshov, Hugo Dalla-Torre, Sam Boshar, Bernardo P. de Almeida, and Thomas Pierrot. Simple guidance mechanisms for discrete diffusion models. In *International Conference on Learning Representations (ICLR)*, 2025.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2310.14189.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning (ICML)*, 2023.

Robin Strudel, Corentin Tallec, Florent Altché, Yilun Du, Yaroslav Ganin, Arthur Mensch, Will Grathwohl, Nikolay Savinov, Sander Dieleman, Laurent Sifre, and Rémi Leblond. Self-conditioned embedding diffusion for text generation. *arXiv preprint arXiv:2211.04236*, 2022.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992. doi: 10.1007/BF00992696. URL https://doi.org/10.1007/BF00992696.

Zhihan Yang, Wei Guo, Shuibai Zhang, Subham Sekhar Sahoo, Yongxin Chen, Arash Vahdat, Morteza Mardani, and John Thickstun. Continuous diffusion scales competitively with discrete diffusion for language. *arXiv preprint arXiv:2605.18530*, 2026.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Frédo Durand, William T. Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Jaehoon Yoo, Wonjung Kim, and Seunghoon Hong. ReDi: Rectified discrete flow. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Mingyuan Zhou, Huangjie Zheng, Zhendong Wang, Mingzhang Yin, and Hai Huang. Score identity distillation: Exponentially fast distillation of pretrained diffusion models for one-step generation. In *International Conference on Machine Learning (ICML)*, 2024.

Yuanzhi Zhu, Xi Wang, Stéphane Lathuilière, and Vicky Kalogeiton. Di[m]o: Distilling masked diffusion models into a one-step generator. *arXiv preprint arXiv:2503.15457*, 2025.

CONTENTS

A Notation **14**

B Background: objectives and parametrizations **15**

B.1 Continuous diffusion on discrete data: objective and design choices 15

B.2 Discrete diffusion: the bound, the plug-in and the concrete score 17

C Distributional distillation: discrepancies and gradients **18**

C.1 Continuous distillation baselines: SiD, FGM and moment matching 19

C.2 Discrete distillation as a Bregman divergence between concrete scores 20

C.3 Masked Diffusion Distillation: collapse to a posterior KL 21

D Method: derivations and proofs **21**

D.1 Distribution matching distillation: gradient derivation 21

D.2 Recovering the discrete data distribution 22

D.3 The optimal discriminator and the log-ratio 24

D.4 Auxiliary score estimation with bridge renoising 24

D.5 Multi-step generation 25

E Supplementary experiments **25**

E.1 Implementation details 25

E.2 Full generative frontiers 27

E.3 Design space 28

E.4 Auxiliary loss 28

E.5 Bridge renoising for Simplex-DMD 28

E.6 Generated samples 30

A NOTATION

Data, representations, and dimensions. The vocabulary is V with K tokens, and $\mathsf{X} = \mathsf{V}^L$ is the space of sequences of length L. We identify token k with the canonical basis vector $\mathbf{e}_k \in \mathbb{R}^K$ and discrete sequences with matrices of one-hot rows. The probability simplex is $\Delta_K = \{u \in \mathbb{R}^K : u_k \geq 0,\ \sum_k u_k = 1\}$; relaxed sequences and collections of token probabilities belong to Δ_K^L. We write $\mathbf{x}^\ell$ for row ℓ and x_k^ℓ for its k-th coordinate. A collection of token marginals does not by itself specify a joint sequence distribution. The clean data law is p_{data}. The embedding matrix $\mathbf{E} \in \mathbb{R}^{K \times d}$ has rows E_k, where d is the embedding dimension, and maps $\mathbf{x}$ to $\mathbf{xE} \in \mathbb{R}^{L \times d}$. Uppercase symbols such as $\mathbf{X}$ and Z_t denote random variables; lowercase symbols denote their realizations.

Gaussian diffusion and noise levels. The data forward process has marginals p_t and conditional laws

$$Z_t = \alpha_t \mathbf{XE} + \sigma_t \varepsilon, \qquad q_{t|\mathbf{x}}(\cdot \mid \mathbf{x}) = \mathrm{N}(\alpha_t \mathbf{xE}, \sigma_t^2 \,\mathrm{Id}), \qquad \varepsilon \sim \mathrm{N}(0, \mathrm{Id}).$$

Here Id is the identity on vectorized latent matrices, $\sigma_t^2 = 1 - \alpha_t^2$, and the idealized endpoints are $(\alpha_0, \sigma_0) = (1, 0)$ and $(\alpha_1, \sigma_1) = (0, 1)$, so $p_1 = \mathrm{N}(0, \mathrm{Id})$. The signal-to-noise ratio is $\mathrm{SNR}(t) = \alpha_t^2/\sigma_t^2$, and $\gamma(t) = -\log \mathrm{SNR}(t)$ is the noise-level parameter supplied to the networks in practice. For $s < t$, $q_{t|s}$ denotes the forward transition, with $\alpha_{t|s} = \alpha_t/\alpha_s$ and $\sigma_{t|s}^2 = \sigma_t^2 - \alpha_{t|s}^2 \sigma_s^2$. The Gaussian bridge $q_{s|t,\mathbf{x}}$ conditions on both the later latent and the clean representation; $p_{s|t}^\theta$ denotes a learned teacher reverse transition.

Posteriors, denoisers, and scores. The joint data posterior is $p_{\mathbf{x}|t}(\cdot \mid z_t)$, with token marginal $p_{\mathbf{x}|t}^\ell(\cdot \mid z_t)$ and mean $m_t(z_t) = \mathbb{E}[\mathbf{X} \mid Z_t = z_t] \in \Delta_K^L$. The pretrained teacher has parameters θ, denoiser $\mathbf{x}^\theta(z_t, t)$, and generative law p^θ; p_t always denotes the data forward marginal. Student parameters are η, auxiliary-denoiser parameters are ϕ, and discriminator parameters are ψ. The auxiliary denoiser $\mathbf{x}^\phi(z, t, T)$ targets $m_t^{\eta,T}(z) = \mathbb{E}[\mathbf{X}^{\eta,T} \mid Z_t^{\eta,T} = z]$. The data and student scores are $s(z,t) = \nabla_z \log p_t(z)$ and $s^{\eta,T}(z,t) = \nabla_z \log p_t^{\eta,T}(z)$; s^θ denotes the teacher estimate of the data score. The scalar time s in a transition is distinct from the score function $s(z,t)$.

Student outputs and training marginals. Given Z_T, the student predicts $\mathbf{x}^\eta(Z_T, T) \in \Delta_K^L$, which parameterizes the clean-output kernel $\mathsf{k}_T^\eta(\cdot \mid Z_T)$. For Simplex-DMD this kernel is the point mass $\delta_{\mathbf{x}^\eta(Z_T,T)}$; for Reinforce-DMD it is the product of categorical laws with these token probabilities. We write $\mathbf{X}^{\eta,T} \sim \mathsf{k}_T^\eta(\cdot \mid Z_T)$ and $p_{\mathbf{x}}^{\eta,T}$ for its marginal when $Z_T \sim p_T$. Forward noising gives $Z_t^{\eta,T}$ with law $p_t^{\eta,T}$, where $\pi_t(\mathbf{x}, \varepsilon) = \alpha_t \mathbf{xE} + \sigma_t \varepsilon$ denotes the reparameterized noising map. The one-step notation $\mathbf{X}^\eta$, $p_{\mathbf{x}}^\eta$, and p_t^η omits the fixed starting level $T = 1$.

Time sampling and matching objectives. In the multi-step objective, T is the starting noise level and $t < T$ is the level at which noised distributions are compared. We sample $(t, T) \sim \nu$, with $\nu(\mathrm{d}t, \mathrm{d}T) = \nu_T(\mathrm{d}t)\nu_1(\mathrm{d}T)$ and $\nu_s = \mathrm{Unif}(0, s)$. The local discrepancy is D, and $\mathsf{L}_{\mathrm{distill}}$ denotes the generic distribution-matching objective. For reverse KL, the one-step and multi-step losses are

$$\mathsf{L}_{\mathrm{DMD}}(\eta) = \int_0^1 \mathrm{KL}(p_t^\eta \,\|\, p_t)\ \mathrm{d}t, \qquad \mathsf{L}_{\mathrm{DMD}}^{\mathrm{ms}}(\eta) = \mathbb{E}_{(t,T)\sim\nu}\left[\mathrm{KL}\left(p_t^{\eta,T} \,\middle\|\, p_t\right)\right].$$

The auxiliary-denoiser and discriminator objectives are $\mathsf{L}_{\mathrm{aux}}$ and $\mathsf{L}_{\mathrm{disc}}$, respectively; n_{aux} is the number of auxiliary updates per student update.

Discrete-sampling cost and variance reduction. The discriminator $D_\psi(z, t, T) \in (0, 1)$ estimates the probability that a noised sample comes from data rather than the student. Its negative logit, $\log((1 - D_\psi)/D_\psi)$, estimates the log-ratio $\log(p_t^{\eta,T}/p_t)$. The expected log-ratio cost is C_T, and $\widehat{C}_\psi$ denotes its pointwise estimate using a sampled time and Gaussian noise. For a group of G samples sharing (Z_T, T), g indexes a sample, $\widehat{C}_g$ is its estimated cost, and $\widehat{b}_g = (G-1)^{-1} \sum_{g' \neq g} \widehat{C}_{g'}$ is its leave-one-out baseline. The coefficient $\beta \geq 0$ weights the tokenwise KL anchor from the student to the frozen teacher.

Multi-step sampling and evaluation. The inference grid is $1 = T_n > T_{n-1} > \cdots > T_0 = 0$. Hatted variables $\widehat{Z}_{T_i}$ and $\widehat{\mathbf{X}}^{\eta,T_i}$ denote states and clean proposals along the composed sampler; their laws need not equal the corresponding training marginals. The student transition is $p^{\eta}_{t|T}$, formed by composing k^{η}_T with the DDIM-inspired kernel $q^{\zeta}_{t|\mathbf{x},T}$. The noise standard deviation $\zeta_{t,T}$ controls fresh noise: 0 gives a deterministic update conditional on the clean proposal, $\sigma_t\sigma_{T|t}/\sigma_T$ gives the Gaussian bridge, and σ_t gives forward renoising. The sampler experiments use the multiplier c in $\zeta_{t,T} = c\sigma_t\sigma_{T|t}/\sigma_T$. Sampling temperature rescales output logits and is separate from the starting noise level T (the sample appendix also uses T locally for temperature). NFE denotes the number of network evaluations per generated sample. Gen PPL is generative perplexity under GPT-2-large; entropy is the mean per-sequence Shannon entropy of empirical token frequencies, measured in nats.

Operators and appendix conventions. We use $\mathrm{KL}(P \,\|\, Q)$ for Kullback–Leibler divergence, $\mathrm{CE}(u, v) = -\sum_k u_k \log v_k$ for cross-entropy, $\mathrm{sg}(\cdot)$ for stop-gradient, δ_x for a point mass, and $\overset{d}{=}$ for equality in distribution. In Appendix D.2, V is the set of one-hot sequences, $\mathcal{E}(\mathbf{x}) = \mathbf{x}\mathbf{E}$, ρ is a clean law, and $Q_t\rho$ is its embedded, noised law; $f_{\#}\rho$ denotes the law of $f(X)$ for $X \sim \rho$. In the comparison of continuous distillation objectives in Appendix C, clean samples are written directly in Euclidean space, so m_t there is a mean in that space; s_t and v_t denote the score and velocity fields. Other locally defined symbols retain the meanings specified in their respective derivations.

B Background: Objectives and Parametrizations

We expand the continuous and discrete formulations recalled in Section 2.

B.1 Continuous Diffusion on Discrete Data: Objective and Design Choices

We review four design choices for continuous diffusion on token embeddings: training objectives, embedding normalization, self-conditioning and noise schedules, highlighting how existing models implement them.

The objective. We distinguish likelihood-based training, used by Diffusion-LM (Li et al., 2022), Plaid (Gulrajani & Hashimoto, 2023), and RePlaid (Yang et al., 2026), from direct cross-entropy training. In the notation of Section 2, the continuous-time negative evidence lower bound consists of prior, reconstruction, and diffusion terms:

$$\mathsf{L}_{\mathrm{c}}(\mathbf{x};\theta) = \mathsf{L}_{\mathrm{prior}}(\mathbf{x}) + \mathbb{E}_{Z_0\sim q_{0|\mathbf{x}}(\cdot|\mathbf{x})}\left[-\sum_{\ell=1}^{L}\log\langle \mathbf{x}^{\theta,\ell}(Z_0,0), \mathbf{x}^{\ell}\rangle\right] + \frac{1}{2}\int_0^1 [-\mathrm{SNR}'(t)]\,\mathbb{E}_{Z_t\sim q_{t|\mathbf{x}}(\cdot|\mathbf{x})}\left[\left\|\left(\mathbf{x}^{\theta}(Z_t,t)-\mathbf{x}\right)\mathbf{E}\right\|_{\mathrm{F}}^2\right]\,\mathrm{d}t,$$

where $\mathrm{SNR}(t) = \alpha_t^2/\sigma_t^2$. For a standard Gaussian generative prior, the prior term is explicit:

$$\mathsf{L}_{\mathrm{prior}}(\mathbf{x}) = \mathrm{KL}\big(q_{1|\mathbf{x}}(\cdot \mid \mathbf{x}) \,\big\|\, \mathrm{N}(0,\mathrm{Id})\big)\ . \tag{14}$$

Since $\alpha_1 = 0$ and $\sigma_1 = 1$ in Section 2, $q_{1|\mathbf{x}}(\cdot \mid \mathbf{x}) = \mathrm{N}(0,\mathrm{Id})$ and this term vanishes. It is nonzero only for schedules with a finite terminal signal-to-noise ratio. The reconstruction term trains the categorical decoder at time zero, while the diffusion term penalizes denoising errors in embedding space.

Alternatively, direct cross-entropy training supervises the clean token at every sampled noise level:

$$\mathsf{L}_{\mathrm{CE}}(\mathbf{x};\theta) = \mathbb{E}_{\substack{t\sim\mathrm{Unif}(0,1)\\ Z_t\sim q_{t|\mathbf{x}}(\cdot|\mathbf{x})}}\left[-\sum_{\ell=1}^{L}\log\langle \mathbf{x}^{\theta,\ell}(Z_t,t), \mathbf{x}^{\ell}\rangle\right].$$

This approach is used by CDCD (Dieleman et al., 2022) and LangFlow (Chen et al., 2026), with model-specific representations and parameterizations. Analog Bits uses squared-error regression by default, but also investigates sigmoid and softmax cross-entropy variants (Chen et al., 2023, Appendix B). SED instead combines a mean-squared error diffusion loss in a fixed embedding space with

a separate reconstruction cross-entropy loss for its trainable readout (Strudel et al., 2022, Section 3, Eqs. (4)–(5)). Both objectives above are averaged over $\mathbf{x} \sim p_{\text{data}}$ during training. Unlike the diffusion term of the NELBO, this cross-entropy directly supervises token probabilities rather than their embedded mean; the unweighted objective above is not the NELBO.

Embedding normalization. These objectives also exert different pressures on learned embeddings. The diffusion term alone vanishes if all token embeddings coincide, whereas cross-entropy favors distinguishable tokens and can encourage increasing their separation relative to the noise. To control this scale, embeddings are constrained to a sphere, with each row of $\mathbf{E}$ having norm 1 or $\sqrt{d}$, where d is the embedding dimension (Dieleman et al., 2022; Chen et al., 2026; Yang et al., 2026). This prevents the embedding norms from vanishing or diverging, although normalization alone does not prevent different tokens from sharing the same embedding.

Self-conditioning. Self-conditioning lets the denoiser refine an earlier prediction of the clean sample alongside the current noisy latent. Introduced in Analog Bits (Chen et al., 2023) and adopted for token embeddings in SED (Strudel et al., 2022), it conditions on predicted analog-bit vectors and continuous embeddings, respectively. In our token-probability parameterization, we express this mechanism by augmenting the network input as $\mathbf{x}^\theta(z_t, t, \hat{\mathbf{x}})$, where $\hat{\mathbf{x}}$ is a previous prediction in Δ_K^L or a zero matrix indicating that no prediction is available. At sampling time, the prediction from the preceding denoising step is reused without an additional network evaluation. Along the descending grid $t_n > \cdots > t_0$, this gives

$$\hat{\mathbf{x}}^{(i)} = \mathbf{x}^\theta\big(z_{t_i}, t_i, \hat{\mathbf{x}}^{(i+1)}\big), \qquad \hat{\mathbf{x}}^{(n+1)} = 0.$$

For $i \geq 1$, this prediction parameterizes the transition from t_i to t_{i-1} and is passed to the next network evaluation.

During training, a noisy state is sampled directly, so no preceding prediction is available. On a randomly selected subset of training examples, a preliminary forward pass supplies the conditioning input:

$$\hat{\mathbf{x}} = \mathrm{sg}(\mathbf{x}^\theta(z_t, t, 0)).$$

The training loss is then evaluated using $\mathbf{x}^\theta(z_t, t, \hat{\mathbf{x}})$, with gradients flowing only through this second pass. The remaining examples use $\hat{\mathbf{x}} = 0$, so the network also learns to denoise without a previous prediction, as required at the first sampling step. The fraction of self-conditioned training examples is a model-specific choice.

The noise schedule. The noise schedule determines how quickly the forward process removes information from the clean embeddings. We parameterize it by the increasing negative log signal-to-noise ratio $\gamma(t) = -\log \mathrm{SNR}(t)$, so that

$$\alpha_t = \sqrt{\mathrm{sigmoid}(-\gamma(t))}, \qquad \sigma_t = \sqrt{\mathrm{sigmoid}(\gamma(t))}.$$

Thus increasing $\gamma(t)$ decreases the signal and increases the noise in $Z_t = \alpha_t \mathbf{X}\mathbf{E} + \sigma_t \varepsilon$. The endpoints $\alpha_0 = 1$ and $\alpha_1 = 0$ of Section 2 correspond to $\gamma(0) = -\infty$ and $\gamma(1) = +\infty$; practical schedules clip them to finite values. In practice, the networks receive the noise level as $\gamma(t)$ rather than t: the notation $\mathbf{x}^\theta(z_t, t)$ of the main text stands for $\mathbf{x}^\theta(z_t, \gamma(t))$, and likewise for the student and auxiliary networks.

LangFlow (Chen et al., 2026) learns a schedule that aims to distribute information loss evenly over time. It fits a scaled Gumbel cumulative distribution function to the observed token cross-entropy as a function of γ, learning its location μ, scale $b > 0$, and amplitude by squared-error regression. The corresponding quantile schedule is

$$\gamma(t) = \mu - b \log(-\log t), \qquad 0 < t < 1. \tag{15}$$

This allocates more noise levels to regions where the fitted cross-entropy changes rapidly; in practice, the noise range is clipped.

RePlaid (Yang et al., 2026) instead learns two endpoints and a monotone interpolation between them:

$$\gamma(t) = \gamma_0 + (\gamma_1 - \gamma_0)\tilde{\gamma}(t), \qquad \tilde{\gamma}(0) = 0, \quad \tilde{\gamma}(1) = 1.$$

The endpoints are optimized using the diffusion loss, while the monotone network $\tilde{\gamma}$ is trained to reduce the Monte Carlo variance of its estimator. For fixed endpoints and a denoiser taking $\gamma(t)$ as input, the continuous-time bound is invariant to the interior schedule, so its shape can be adjusted to improve estimation efficiency (Kingma et al., 2021).

In our experiments, we reuse the teacher's pretrained noise schedule and keep it fixed throughout distillation and sampling.

B.2 Discrete diffusion: the bound, the plug-in and the concrete score

Discrete diffusion progressively corrupts a token sequence and learns to reverse this process to generate data. We consider two choices of the replacement distribution π: a point mass on an additional absorbing [MASK] token for masked diffusion, or the uniform distribution over the vocabulary for uniform diffusion. The process starts directly from the clean sequence, $Z_0 = \mathbf{X}$, so $p_0 = p_{\mathrm{data}}$. We describe the forward process, the reverse model and its training bound, then explain how denoiser predictions parameterize reverse transitions and how this relates to concrete scores in continuous time.

Forward process. The forward process corrupts each position independently by replacing its token with a draw from π. Let $(\alpha_t)_{t\in[0,1]}$ be a decreasing schedule with $\alpha_0 = 1$ and $\alpha_1 = 0$. Between times $s < t$, a token is retained with probability α_t/α_s and otherwise resampled from π, giving the transition

$$q^{\ell}_{t|s}(z^{\ell}_t \mid z^{\ell}_s) \;=\; \mathrm{Cat}\Big(z^{\ell}_t;\; \tfrac{\alpha_t}{\alpha_s}\, z^{\ell}_s \;+\; \big(1 - \tfrac{\alpha_t}{\alpha_s}\big)\pi\Big)\,, \qquad 0 \le s < t \le 1\,. \tag{16}$$

In particular, setting $s = 0$ gives the conditional law given the clean token:

$$q^{\ell}_{t|\mathbf{x}}(z^{\ell}_t \mid \mathbf{x}^{\ell}) \;=\; \mathrm{Cat}\big(z^{\ell}_t;\; \alpha_t \mathbf{x}^{\ell} + (1-\alpha_t)\pi\big)\,, \tag{17}$$

These conditional laws and the forward transitions determine the bridges $q_{s|t,\mathbf{x}}$ by Bayes' rule. At $t = 1$, every position follows $\mathrm{Cat}(\pi)$ independently of the clean sequence, so the terminal distribution is the product prior $p_1 = \mathrm{Cat}(\pi)^{\otimes L}$.

Generative model and KL bound. Generation starts from this prior and applies learned reverse transitions along a grid $0 = t_0 < \cdots < t_n = 1$, with $s_i = t_{i-1}$. Each transition predicts the positions independently conditional on the full current sequence, yielding

$$p^{\theta}(\mathbf{x}) \;=\; \sum_{z_{t_1:t_n}} p_1(z_1) \prod_{i=1}^{n} p^{\theta}_{s_i|t_i}(z_{s_i} \mid z_{t_i})\,, \qquad p^{\theta}_{s|t}(z_s \mid z_t) = \prod_{\ell=1}^{L} p^{\theta,\ell}_{s|t}(z^{\ell}_s \mid z_t)\,, \tag{18}$$

where $z_{s_1} = \mathbf{x}$ is the generated sequence at time zero. To relate the generated distribution to the data distribution, let $p_{s|t}$ denote the exact reverse transition induced by the forward process with $\mathbf{X} \sim p_{\mathrm{data}}$. The exact and learned reverse chains share the same terminal prior. Applying the chain rule for KL divergence to their trajectory laws, then marginalizing to time zero, bounds the discrepancy between data and generated samples by the accumulated transition errors:

$$\mathrm{KL}\big(p_{\mathrm{data}} \,\big\|\, p^{\theta}\big) \;\le\; \sum_{i=1}^{n} \mathbb{E}_{Z_{t_i}\sim p_{t_i}} \Big[\mathrm{KL}\Big(p_{s_i|t_i}(.\mid Z_{t_i}) \,\Big\|\, p^{\theta}_{s_i|t_i}(.\mid Z_{t_i})\Big)\Big]\,. \tag{19}$$

In general, the exact reverse transition contains dependencies between positions that the factorized model cannot represent. For a fixed z_t, each transition error separates into tokenwise marginal errors and a term accounting for these dependencies:

$$\mathrm{KL}\Big(p_{s|t}(.\mid z_t) \,\Big\|\, p^{\theta}_{s|t}(.\mid z_t)\Big) \;=\; \sum_{\ell=1}^{L} \mathrm{KL}\Big(p^{\ell}_{s|t}(.\mid z_t) \,\Big\|\, p^{\theta,\ell}_{s|t}(.\mid z_t)\Big) \;+\; \mathcal{I}_{s|t}(z_t)\,, \tag{20}$$

where $p^{\ell}_{s|t}(.\mid z_t) = \sum_{z_s^{-\ell}} p_{s|t}(z_s \mid z_t)$ is the ℓ-th token marginal of the exact transition and $\mathcal{I}_{s|t}(z_t) = \mathrm{KL}\Big(p_{s|t}(.\mid z_t) \,\Big\|\, \prod_{\ell} p^{\ell}_{s|t}(.\mid z_t)\Big)$ measures the conditional dependence between positions under the exact reverse transition. Only the marginal errors depend on θ. Thus, within the factorized family, the optimal transition matches each exact token marginal; the remaining term $\mathcal{I}_{s|t}(z_t)$ is the irreducible error due to factorization at this time step. This motivates parameterizing the model through predictions that recover these marginals.

The bridge plug-in. The bound asks for a reverse transition, whereas the network produces a point of Δ_K^L. Nearly every work joins the two by *plugging in*, evaluating the bridge as though the prediction were the clean token (Sahoo et al., 2024; Schiff et al., 2025),

$$p_{s|t}^{\theta,\ell}(. \mid z_t) \;=\; q_{s|t,\mathbf{x}}^{\ell}\big(. \mid z_t^{\ell},\, \mathbf{x}^{\theta,\ell}(z_t,t)\big)\,. \tag{21}$$

Whether this costs anything is the question of whether the token marginals of (20) are themselves of that form, and they are.

Proposition 1 (Gourevitch et al., 2026)**.** *Let $\mathbf{x}^{\mathrm{loo},\ell}(z_t,t) = \mathbb{E}[\mathbf{x}^{\ell} \mid z_t^{-\ell}]$ be the leave-one-out denoiser, and let the bridge be extended to a simplex-valued first argument by its Bayes-ratio form, as both processes admit. Then, for every position ℓ,*

$$p_{s|t}^{\ell}(. \mid z_t) \;=\; q_{s|t,\mathbf{x}}^{\ell}\big(. \mid z_t^{\ell},\, \mathbf{x}^{\mathrm{loo},\ell}(z_t,t)\big)\,.$$

When π is uniform this is the only simplex-valued field with that property; when π is the mask, the leave-one-out denoiser agrees with the denoiser $\mathbb{E}[\mathbf{x}^{\ell} \mid z_t]$ at every masked position, and the latter has it too.

The plug-in at $\mathbf{x}^{\theta} = \mathbf{x}^{\mathrm{loo}}$ therefore cancels the sum in (20), leaving only the factorization error. The object it fits is the leave-one-out denoiser, which does not read the position it predicts.

Continuous time. Refining the grid removes the factorization error, two positions both jumping within one interval being a second-order event, so (19) becomes tight over the family (18) in the limit. That limit is written on ratios rather than on transitions. For a state z_t, a position ℓ and a token k, write $z_t^{\ell\to k}$ for z_t with its ℓ-th token replaced by k; these are the states at Hamming distance one from z_t. The *concrete score* of the marginal p_t collects the corresponding ratios,

$$s_t^{\ell}(z_t)_k \;=\; \frac{p_t(z_t^{\ell\to k})}{p_t(z_t)}\,, \tag{22}$$

and the same ratios for $q_{t|\mathbf{x}}$ are closed-form, only the ℓ-th factor of (17) changing,

$$s_t^{\ell}(z_t \mid \mathbf{x})_k \;=\; \frac{\langle \alpha_t \mathbf{x}^{\ell} + (1-\alpha_t)\pi,\, e_k\rangle}{\langle \alpha_t \mathbf{x}^{\ell} + (1-\alpha_t)\pi,\, z_t^{\ell}\rangle}\,. \tag{23}$$

The second is defined for any point of Δ_K^L in place of $\mathbf{x}$, so the plug-in of (21) carries over to it,

$$s_t^{\theta,\ell}(z_t) \;=\; s_t^{\ell}\big(z_t \mid \mathbf{x}^{\theta}(z_t,t)\big)\,, \tag{24}$$

and returns (22) exactly at $\mathbf{x}^{\theta} = \mathbf{x}^{\mathrm{loo}}$, as in Proposition 1 (Gourevitch et al., 2026). The limit of (19) compares the two scores through a Bregman divergence, weighted by the rate at which the reverse process leaves Z_t^{ℓ} (Lou et al., 2024),

$$\mathsf{L}_{\mathrm{d}}(\theta) \;=\; \int_0^1 R_t\, \mathbb{E}_{Z_t\sim p_t}\Big[\sum_{\ell=1}^{L} \langle \pi,\, Z_t^{\ell}\rangle \sum_{e_k \neq Z_t^{\ell}} d_F\Big(s_t^{\ell}(Z_t)_k\,,\; s_t^{\theta,\ell}(Z_t)_k\Big)\Big]\,\mathrm{d}t\,, \tag{25}$$

with $R_t = -\alpha_t'/\alpha_t$ the substitution rate of (16), the reverse rate towards $z_t^{\ell\to k}$ being $R_t\langle \pi, z_t^{\ell}\rangle\, s_t^{\ell}(z_t)_k$, generated by $F(u) = u\log u - u$, so that $d_F(a,b) = a\log(a/b) - a + b$. The data score (22) is not available, but it is the average of (23) over the denoising posterior, and a Bregman divergence is insensitive to that substitution up to a constant in θ: drawing $\mathbf{X} \sim p_{\mathbf{x}|t}(. \mid Z_t)$ and scoring against $s_t^{\ell}(Z_t \mid \mathbf{X})$ is the form actually minimized. Nothing in (25) requires s_t^{θ} to come from a denoiser: parametrizing the concrete score directly by a network is SEDD (Lou et al., 2024).

C Distributional distillation: discrepancies and gradients

This appendix derives the objectives summarized in Table 1, the gradient each method descends, and the identities relating them.

C.1 Continuous distillation baselines: SiD, FGM and moment matching

Score identity distillation, flow generator matching and moment matching distil continuous diffusion models on continuous data, the case $\mathbf{E} = \mathrm{Id}$ of Section 2. Each is stated in a different field — scores, velocities and conditional means — and each estimates the gradient of the same objective differently. We restate them on one path and identify what each one computes.

Path and fields. The student is a deterministic one-step generator: $\xi \sim p_{\mathrm{prior}}$ and $\mathbf{x}^\eta(\xi) \in \mathbb{R}^d$, corrupted along the path of Section 2,

$$Z_t^\eta = \alpha_t\, \mathbf{x}^\eta(\xi) + \sigma_t \varepsilon\,, \qquad \varepsilon \sim \mathrm{N}(0, \mathrm{Id})\,,$$

with p_t^η the law of Z_t^η and p_t that of the data under the same kernel $q_{t|\mathbf{x}}$. At each noise level the student law is described by any one of three fields, its conditional mean, its score and its velocity,

$$\begin{aligned} m_t^\eta(z_t) &= \mathbb{E}\big[\mathbf{x}^\eta(\xi) \mid Z_t^\eta = z_t\big], & s_t^\eta(z_t) &= \nabla_{z_t} \log p_t^\eta(z_t), \\ v_t^\eta(z_t) &= \mathbb{E}\big[\dot\alpha_t \mathbf{x}^\eta(\xi) + \dot\sigma_t \varepsilon \mid Z_t^\eta = z_t\big], \end{aligned}$$

and m_t, s_t, v_t denote the same fields under the data law. Tweedie's identity, $s_t^\eta = (\alpha_t m_t^\eta - z_t)/\sigma_t^2$, and its velocity counterpart, $v_t^\eta = \dot\alpha_t m_t^\eta + \dot\sigma_t (z_t - \alpha_t m_t^\eta)/\sigma_t$, make the three differences proportional at each noise level,

$$s_t^\eta - s_t = \frac{\alpha_t}{\sigma_t^2}\big(m_t^\eta - m_t\big), \qquad v_t^\eta - v_t = \Big(\dot\alpha_t - \alpha_t \frac{\dot\sigma_t}{\sigma_t}\Big)\big(m_t^\eta - m_t\big),$$

so matching scores, velocities or conditional means is one criterion under three names, up to a weight in t. We work with the score.

The objective and its two gradient terms. The squared discrepancy between the two fields, integrated over noise levels, is

$$\mathsf{L}(\eta) = \int_0^1 w(t)\, \mathbb{E}_{Z_t \sim p_t^\eta}\Big[\big\|s_t^\eta(Z_t) - s_t(Z_t)\big\|^2\Big]\, \mathrm{d}t\,. \tag{26}$$

Here η enters twice, through the law of the state and through the field evaluated at it. Freezing one at a time gives two losses,

$$\begin{aligned} \mathsf{L}_1(\eta) &= \int_0^1 w(t)\, \mathbb{E}_{\xi,\varepsilon}\Big[\big\|\mathrm{sg}(s_t^\eta)(Z_t^\eta) - s_t(Z_t^\eta)\big\|^2\Big]\, \mathrm{d}t\,, \\ \mathsf{L}_2(\eta) &= \int_0^1 w(t)\, \mathbb{E}_{Z_t \sim \mathrm{sg}(p_t^\eta)}\Big[\big\|s_t^\eta(Z_t) - s_t(Z_t)\big\|^2\Big]\, \mathrm{d}t\,, \end{aligned} \tag{27}$$

where $\mathrm{sg}(\cdot)$ freezes the parameters of a field, or of the law a state is drawn from, but not the value at which the field is read. Both equal $\mathsf{L}(\eta)$, and their gradients are the two terms of (5),

$$\nabla_\eta \mathsf{L}_1 = (A)\,, \qquad \nabla_\eta \mathsf{L}_2 = (B)\,, \qquad \nabla_\eta \mathsf{L} = (A) + (B)\,.$$

For the reverse KL the second vanishes, $\mathbb{E}_{p_t^\eta}[\nabla_\eta \log p_t^\eta] = \nabla_\eta \int p_t^\eta = 0$. Here it does not, and L_2 cannot be differentiated as it stands: the auxiliary network estimates s_t^η at the current parameters, not its derivative $\nabla_\eta s_t^\eta$.

Trading the student field for the sample that produced it. The unknown field is removed from one side of the squared norm by an identity on the student's own path.

Lemma 1. *Let $\sigma_t > 0$ on the support of w. Then $\mathsf{L}(\eta) = \mathsf{L}^{\mathrm{proj}}(\eta)$, where*

$$\mathsf{L}^{\mathrm{proj}}(\eta) = \int_0^1 w(t)\, \mathbb{E}_{\xi,\varepsilon}\Big[\big(s_t^\eta - s_t\big)^\top \big(s_t(\cdot \mid \mathbf{x}^\eta(\xi)) - s_t\big)(Z_t^\eta)\Big]\, \mathrm{d}t\,, \tag{28}$$

and $s_t(z_t \mid x) = \nabla_{z_t} \log q_{t|\mathbf{x}}(z_t \mid x) = (\alpha_t x - z_t)/\sigma_t^2$ is the conditional score.

Proof. $p_t^\eta(z_t)\, s_t^\eta(z_t) = \nabla_{z_t} p_t^\eta(z_t) = \int \nabla_{z_t} q_{t|\mathbf{x}}(z_t \mid x)\, p^\eta(x)\, \mathrm{d}x = \int q_{t|\mathbf{x}}(z_t \mid x)\, s_t(z_t \mid x)\, p^\eta(x)\, \mathrm{d}x$, so that $\mathbb{E}_{Z_t \sim p_t^\eta}[f(Z_t)^\top s_t^\eta(Z_t)] = \mathbb{E}_{\xi,\varepsilon}[f(Z_t^\eta)^\top s_t(Z_t^\eta \mid \mathbf{x}^\eta(\xi))]$ for any square-integrable f. Take $f = s_t^\eta - s_t$ and subtract $\mathbb{E}[f^\top s_t]$ from both sides. □

The marginal field on the left is unknown; the conditional field on the right is explicit, and differentiable in η through the sample that produced the state. Equation (28) is the form SiD calls projected score matching (Zhou et al., 2024). Freezing the field in it, as in (27), isolates the part of its gradient due to the sampling operation,

$$\mathsf{L}_1^{\mathrm{proj}}(\eta) = \int_0^1 w(t)\, \mathbb{E}_{\xi,\varepsilon}\Big[\big(\mathrm{sg}(s_t^\eta) - s_t\big)^\top \big(s_t(\cdot \mid \mathbf{x}^\eta(\xi)) - s_t\big)(Z_t^\eta)\Big]\, \mathrm{d}t\,, \tag{29}$$

where η now reaches the objective twice over, through the state and through $\mathbf{x}^\eta(\xi)$ inside the conditional score. That second route is what the swap has bought: $\mathsf{L}_1^{\mathrm{proj}}$ and L_1 agree in value, both being $\mathsf{L}(\eta)$, but not in gradient, part of (B) having moved into the sampling operation. The gap between them is exactly the missing term.

Proposition 2 (Huang et al., 2024). $\nabla_\eta \mathsf{L}_2(\eta) = 2\,\nabla_\eta\big(\mathsf{L}_1^{\mathrm{proj}}(\eta) - \mathsf{L}_1(\eta)\big)$.

FGM states this in the velocity field and proves it from the identity of Lemma 1 (Huang et al., 2024). The exact gradient of (26) is therefore the gradient of a computable loss,

$$\nabla_\eta \mathsf{L}(\eta) \;=\; \nabla_\eta\big(2\,\mathsf{L}_1^{\mathrm{proj}}(\eta) - \mathsf{L}_1(\eta)\big)\,, \tag{30}$$

in which the student field appears only frozen, at the current parameters, which is what the auxiliary network provides.

What each method computes. The three baselines descend $\mathsf{L}_1^{\mathrm{proj}}$ and L_1 in different proportions. FGM takes the combination (30) and so descends the exact gradient (Huang et al., 2024). SiD descends $\mathsf{L}_1^{\mathrm{proj}} - \alpha\mathsf{L}_1$ for a scalar α (Zhou et al., 2024), whose gradient is, by Proposition 2,

$$\nabla_\eta\big(\mathsf{L}_1^{\mathrm{proj}} - \alpha\,\mathsf{L}_1\big) \;=\; (1-\alpha)\,(A) \;+\; \tfrac{1}{2}\,(B)\,.$$

FGM is therefore SiD at $\alpha = 1/2$, doubled, the factor being absorbed in the step size. The values reported as best for SiD, $\alpha \in [0.75, 1.2]$ (Zhou et al., 2024), sit past that point; at $\alpha = 1$ the pathwise term disappears and only $\frac{1}{2}(B)$ is left, and beyond it the pathwise term is descended with the opposite sign. The one-step version of moment matching descends (29), the case $\alpha = 0$, with Z_t^η frozen in addition (Salimans et al., 2024). The three are one family, read off a single pair of losses.

C.2 Discrete distillation as a Bregman divergence between concrete scores

D-MMD (Hoogeboom et al., 2026) and IDLM (Li et al., 2026) distil discrete diffusion models, and their few-step constructions differ: D-MMD draws the next state from the bridge, IDLM rather uses the forward process. At one step they descend the same objective, derived once here.

The objective. The student draws $\xi \sim p_{\mathrm{prior}}$ and emits $\mathbf{x}^\eta(\xi)$, of sequence law p^η, and the criterion is the reverse Kullback–Leibler divergence $\mathrm{KL}(p^\eta \,\|\, p_{\mathrm{data}})$. Diffuse both laws along (16). Their reverse chains are Markov on the same grid and agree at $t = 1$, so (19) holds with the two exchanged,

$$\mathrm{KL}(p^\eta \,\|\, p_{\mathrm{data}}) \;\le\; \sum_{i=1}^{n} \mathbb{E}_{Z_{t_i}\sim p^\eta_{t_i}}\Big[\mathrm{KL}\Big(p^\eta_{s_i|t_i}(\,.\mid Z_{t_i}) \,\Big\|\, p_{s_i|t_i}(\,.\mid Z_{t_i})\Big)\Big]\,,$$

and Proposition 1 applies to both sides, each token marginal being a bridge with a leave-one-out denoiser plugged in — the student's on the left, the data's on the right. The passage to continuous time of Appendix B.2 runs unchanged and returns a Bregman divergence between the two concrete scores, read at the states the student visits,

$$\mathrm{KL}(p^\eta \,\|\, p_{\mathrm{data}}) \;\le\; \int_0^1 R_t\, \mathbb{E}_{Z_t\sim p_t^\eta}\Big[\sum_{\ell=1}^{L} \langle \pi,\, Z_t^\ell\rangle \sum_{e_k \neq Z_t^\ell} d_F\Big(s_t^{\eta,\ell}(Z_t)_k\,,\; s_t^\ell(Z_t)_k\Big)\Big]\,\mathrm{d}t\,. \tag{31}$$

Neither score is available in distillation: the teacher stands in for s_t, and an auxiliary model, fitted on the student's own samples, for s_t^η.

Why a difference of two training losses computes it. Neither method evaluates (31). Both subtract two per-sample training losses of the kind Appendix B.2 ends on, evaluated on student samples. For a simplex-valued field and the concrete score $\hat{s}$ it induces through (24), that loss is (Lou et al., 2024)

$$\ell_t(\hat{s}; z_t, \mathbf{x}) \;=\; \sum_{\ell=1}^{L} \big\langle \pi,\, z_t^{\ell} \big\rangle \sum_{e_k \neq z_t^{\ell}} \Big[\hat{s}^{\ell}(z_t)_k \;-\; s_t^{\ell}(z_t \mid \mathbf{x})_k \, \log \hat{s}^{\ell}(z_t)_k \Big] ,$$

the integrand of (25) up to a term free of $\hat{s}$: an evidence bound rather than a divergence, minimized over $\mathbf{X} \sim p^{\eta}_{\mathbf{x}|t}(. \mid z_t)$ at $\hat{s} = s_t^{\eta}$. Subtracting two of them removes that free term and leaves the divergence,

$$\mathbb{E}_{\mathbf{X} \sim p^{\eta}_{\mathbf{x}|t}(.|z_t)} \Big[\ell_t(s_t^{\theta}; z_t, \mathbf{X}) - \ell_t(s_t^{\eta}; z_t, \mathbf{X}) \Big] \;=\; \sum_{\ell=1}^{L} \big\langle \pi,\, z_t^{\ell} \big\rangle \sum_{e_k \neq z_t^{\ell}} d_F\Big(s_t^{\eta,\ell}(z_t)_k \,,\; s_t^{\theta,\ell}(z_t)_k \Big) ,$$

which is the integrand of (31) with the teacher in place of the data. IDLM writes it as the teacher's loss on student samples minus its minimum over auxiliary fields (Li et al., 2026); D-MMD adds a term pulling the auxiliary towards the teacher, which leaves the fixed point where it is (Hoogeboom et al., 2026). Under the mask this divergence takes a simpler form still, derived in Appendix C.3.

C.3 Masked Diffusion Distillation: collapse to a posterior KL

Take $\pi = e_m$, the point mass on the mask symbol. The weight $\langle \pi, z_t^{\ell} \rangle$ of (31) is then 1 at masked positions and 0 elsewhere, so only masked positions are scored. At one of them the head puts no mass on m, so the denominator of (23) is $1 - \alpha_t$ and, for every $k \neq m$,

$$s_t^{\ell}\big(z_t \mid \hat{\mathbf{x}}\big)_k \;=\; \frac{\alpha_t}{1 - \alpha_t}\, \hat{\mathbf{x}}_k^{\ell} ,$$

the concrete score being the head itself, up to a factor that depends on t alone. Since d_F is positively homogeneous, $d_F(ca, cb) = c\, d_F(a, b)$, that factor comes out of the sum; and since both heads sum to one over $k \neq m$, the linear part of $d_F(a, b) = a \log(a/b) - a + b$ cancels, leaving

$$\sum_{e_k \neq z_t^{\ell}} d_F\Big(s_t^{\eta,\ell}(z_t)_k \,,\; s_t^{\theta,\ell}(z_t)_k \Big) \;=\; \frac{\alpha_t}{1 - \alpha_t}\, \mathrm{KL}\big(\mathbf{x}^{\eta,\ell}(z_t, t) \,\big\|\, \mathbf{x}^{\theta,\ell}(z_t, t) \big) \;.$$

The bound (31) is therefore a sum of posterior divergences over masked positions,

$$\mathrm{KL}(p^{\eta} \,\|\, p_{\mathrm{data}}) \;\leq\; \int_0^1 \frac{-\alpha_t'}{1 - \alpha_t}\, \mathbb{E}_{Z_t \sim p_t^{\eta}} \Bigg[\sum_{\ell \,:\, Z_t^{\ell} = e_m} \mathrm{KL}\big(\mathbf{x}^{\eta,\ell}(Z_t, t) \,\big\|\, \mathbf{x}^{\theta,\ell}(Z_t, t) \big) \Bigg] \, \mathrm{d}t \,,$$

which is the discrepancy Table 1 lists for DiMO (Zhu et al., 2025) and the loss D-MMD implements in its masked instance (Hoogeboom et al., 2026). Nothing of the sort happens under the uniform law, where every position is scored and the denominator of (23) keeps the head in it.

D Method: derivations and proofs

The derivations behind Section 3: the gradient the objective is trained with, the proof that the vertex constraint is recovered, the discriminator that supplies the log-ratio, and the alternative objective we distil as a baseline.

D.1 Distribution matching distillation: gradient derivation

We derive the pathwise gradient used by Simplex-DMD in (9). The key observation is that differentiating the reverse KL produces both a contribution through the generated latent and an explicit density derivative; the latter has zero expectation. Throughout, the embedding matrix $\mathbf{E}$, the noise schedule, and the sampling distribution ν are fixed. We assume sufficient regularity to interchange differentiation and integration.

Reparameterized student samples. For $(t, T) \sim \nu$, draw $Z_T \sim p_T$ and independent noise $\varepsilon \sim \mathrm{N}(0, \mathrm{Id})$. Under the continuous relaxation,

$$\mathbf{X}^{\eta,T} = \mathbf{x}^{\eta}(Z_T, T), \qquad Z_t^{\eta,T} = \pi_t(\mathbf{X}^{\eta,T}, \varepsilon) = \alpha_t \mathbf{x}^{\eta}(Z_T, T)\mathbf{E} + \sigma_t \varepsilon.$$

This construction realizes the marginal $p_t^{\eta,T}$ on a sampling space whose law does not depend on η. The multi-step objective is therefore

$$\mathsf{L}_{\mathrm{DMD}}^{\mathrm{ms}}(\eta) = \mathbb{E}_{\substack{(t,T)\sim\nu \\ Z_T\sim p_T,\, \varepsilon\sim\mathrm{N}(0,\mathrm{Id})}} \left[\log \frac{p_t^{\eta,T}(Z_t^{\eta,T})}{p_t(Z_t^{\eta,T})} \right].$$

Differentiating the reverse KL. Fix (t, T) and write $s^{\eta,T}(z,t) = \nabla_z \log p_t^{\eta,T}(z)$ and $s(z,t) = \nabla_z \log p_t(z)$. Holding Z_T and ε fixed, the chain rule gives

$$\nabla_\eta \log \frac{p_t^{\eta,T}(Z_t^{\eta,T})}{p_t(Z_t^{\eta,T})} = \left(s^{\eta,T}(Z_t^{\eta,T}, t) - s(Z_t^{\eta,T}, t) \right)^\top \frac{\partial Z_t^{\eta,T}}{\partial \eta} + \nabla_\eta \log p_t^{\eta,T}(z)\Big|_{z=Z_t^{\eta,T}}.$$

In the last term, the density is differentiated at a fixed argument z. Its expectation vanishes because $p_t^{\eta,T}$ integrates to one:

$$\mathbb{E}_{Z_t^{\eta,T}\sim p_t^{\eta,T}} \left[\nabla_\eta \log p_t^{\eta,T}(z)\Big|_{z=Z_t^{\eta,T}} \right] = \int_{\mathbb{R}^{L\times d}} \nabla_\eta p_t^{\eta,T}(z)\, \mathrm{d}z = \nabla_\eta 1 = 0.$$

This is the cancellation of contribution (B) in (5); it also holds for the categorical student, although that parameterization does not admit the pathwise derivative.

Averaging the remaining term yields

$$\nabla_\eta \mathsf{L}_{\mathrm{DMD}}^{\mathrm{ms}}(\eta) = \mathbb{E}_{\substack{(t,T)\sim\nu \\ Z_T\sim p_T,\, \varepsilon\sim\mathrm{N}(0,\mathrm{Id})}} \left[\left(s^{\eta,T}(Z_t^{\eta,T}, t) - s(Z_t^{\eta,T}, t) \right)^\top \frac{\partial Z_t^{\eta,T}}{\partial \eta} \right],$$

which is (9). Latent coordinates are flattened when forming the score–Jacobian product. The latent Jacobian includes the embedding map and the signal coefficient:

$$\frac{\partial Z_t^{\eta,T}}{\partial \eta} = \alpha_t \frac{\partial [\mathbf{x}^{\eta}(Z_T, T)\mathbf{E}]}{\partial \eta}.$$

The one-step gradient follows by fixing $T = 1$, sampling t uniformly, and identifying $p_t^{\eta,1}$ with p_t^{η}.

Estimated scores. The identity above uses exact scores. In practice, the frozen teacher and the auxiliary denoiser provide their estimates through (10). During a student update, the estimated score difference is detached and gradients flow only through the reparameterized latent. Thus the implemented update approximates the exact pathwise gradient, with accuracy determined by the score estimates.

D.2 Recovering the discrete data distribution

This subsection proves Theorem 1: matching the embedded, noised marginals recovers the clean data distribution, even when the student outputs continuous probability vectors. We then explain how the same argument identifies the transports from noisy data marginals to the clean data distribution used for multi-step generation.

Clean and noised distributions. We use the clean student law $p_{\mathbf{X}}^{\eta}$ from Section 3.1 and identify discrete sequences with the one-hot vertex set $V = \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}^L$ of Δ_K^L. The data law p_{data} is supported on V, whereas a relaxed student can place mass throughout Δ_K^L. For a clean distribution ρ on this simplex, let $Q_t\rho$ denote the law of

$$Z_t = \alpha_t \mathbf{X}\mathbf{E} + \sigma_t \varepsilon, \qquad \mathbf{X} \sim \rho, \quad \varepsilon \sim \mathrm{N}(0, \mathrm{Id}),$$

with independent Gaussian noise. Thus $Q_t p_{\mathbf{x}}^{\eta} = p_t^{\eta}$ and $Q_t p_{\text{data}} = p_t$. Writing the objective as a function of the clean law gives

$$\mathsf{L}_{\text{DMD}}(\rho) = \int_0^1 \text{KL}(Q_t\rho \,\|\, p_t)\, \mathrm{d}t, \qquad \mathsf{L}_{\text{DMD}}(\eta) = \mathsf{L}_{\text{DMD}}(p_{\mathbf{x}}^{\eta}),$$

which is exactly (6) for the student law. Here ρ denotes a clean distribution; ν retains its meaning as the time-sampling distribution in the multi-step objective. For the proofs, write $\mathcal{E}(\mathbf{x}) = \mathbf{x}\mathbf{E}$ for the position-wise embedding map. We assume that α_t and σ_t are measurable, finite, and strictly positive for $t \in (0, 1)$; endpoint values do not affect the integral, and the loss may be infinite.

Why equal-norm embeddings suffice. Constraining the token embeddings to a common sphere is standard in the lineage of Section 2: CDCD L_2-normalizes them before every use and backpropagates through the normalization (Dieleman et al., 2022), LangFlow places them on a sphere of radius $\sqrt{d}$ (Chen et al., 2026), and RePlaid constrains each row to unit length (Yang et al., 2026). If the rows are also distinct, this normalization ensures that no token embedding is a convex combination of the others, as the following proposition shows.

Proposition 3. *Distinct rows sharing a common Euclidean norm are in convex position.*

Proof. Let $\|E_k\| = R$ for all k and suppose $E_i = \sum_{k\neq i} a_k E_k$ with $a \in \Delta_{K-1}$. For $k \neq i$, $\|E_i - E_k\|^2 = 2R^2 - 2\langle E_i, E_k\rangle$ is positive by distinctness, so $\langle E_i, E_k\rangle < R^2$. Taking the inner product of the supposed combination with E_i gives $R^2 = \sum_{k\neq i} a_k \langle E_i, E_k\rangle < R^2$. □

More generally, the argument below only requires distinct rows in convex position; it does not require the embedding map to be injective on the whole simplex.

Proof of the theorem. We first show that a token embedding can only come from its one-hot vector. We then undo Gaussian noising to identify the clean distribution.

Lemma 2. *Under the assumption of Theorem 1, $x \in \Delta_K$ and $x\mathbf{E} = E_i$ imply $x = \mathbf{e}_i$. Hence $\mathcal{E}^{-1}(\mathcal{E}(V)) = V$, and $\mathcal{E}$ is injective on V.*

Proof. From $E_i = x_i E_i + \sum_{k\neq i} x_k E_k$, if $x_i < 1$ then dividing by $1 - x_i$ gives $E_i = \sum_{k\neq i} \frac{x_k}{1-x_i} E_k$, a convex combination of the other rows, contradicting convex position. Hence $x_i = 1$. Applying this at every position gives the second claim, and injectivity on V follows from distinctness of the rows. □

Lemma 3 (Identification at one noise level)**.** *Under the assumption of Theorem 1, for any probability measure ρ on Δ_K^L and any $t \in (0, 1)$, $Q_t\rho = Q_t p_{\text{data}}$ holds if and only if $\rho = p_{\text{data}}$.*

Proof. Only the forward implication needs an argument. Both measures are Gaussian smoothings, so their characteristic functions factor as $\widehat{(\alpha_t\mathcal{E})_{\#}\rho} \cdot \hat{g}$ with $\hat{g}(u) = e^{-\sigma_t^2\|u\|^2/2}$ nowhere zero. Dividing by $\hat{g}$ and using $\alpha_t \neq 0$ gives $\mathcal{E}_{\#}\rho = \mathcal{E}_{\#}p_{\text{data}}$. That measure is carried by $\mathcal{E}(V)$, so ρ is carried by $\mathcal{E}^{-1}(\mathcal{E}(V)) = V$ by Lemma 2. On that finite set $\mathcal{E}$ is injective, so equality of the embedded laws is equality of the mass at every vertex sequence. □

Proof of Theorem 1. We prove that $\mathsf{L}_{\text{DMD}}(\rho) = 0$ if and only if $\rho = p_{\text{data}}$, for any probability measure ρ on Δ_K^L; the theorem is the case $\rho = p_{\mathbf{x}}^{\eta}$. The reverse implication is immediate, $Q_t\rho$ being a function of ρ alone. Conversely, nonnegativity of the KL divergence forces $Q_t\rho = Q_t p_{\text{data}}$ for almost every $t \in (0, 1)$, and any such t with Lemma 3 closes the argument. □

Implications for multi-step generation. In Section 3.2, the student starts from $Z_T \sim p_T$ and produces a clean sample with law $p_{\mathbf{x}}^{\eta,T}$. For each fixed $T > 0$, Lemma 3 applies to this law just as in the one-step case: matching $p_t^{\eta,T}$ to p_t at any noise level $0 < t < T$ implies $p_{\mathbf{x}}^{\eta,T} = p_{\text{data}}$. Consequently, a zero multi-step objective (7) gives, for ν_1-almost every starting level T,

$$\int \mathsf{k}_T^{\eta}(\mathrm{d}\mathbf{x} \mid z_T)\, p_T(\mathrm{d}z_T) = p_{\text{data}}(\mathrm{d}\mathbf{x}).$$

Indeed, nonnegativity of the KL divergence implies matching for ν_T-almost every $t < T$, and any such level identifies the clean law. Thus the objective learns a family of transports from p_T to p_{data}, indexed by the starting noise level.

D.3 The optimal discriminator and the log-ratio

We show that the population optimum of (12) recovers the log-density ratio required by Reinforce-DMD. This is the standard optimal discriminator argument of Goodfellow et al. (2014), applied to the noised data and student distributions at each pair (t, T).

Proposition 4. *Fix the student parameters η and a pair $0 < t < T \leq 1$ with $\sigma_t > 0$. Over measurable functions $D : \mathbb{R}^{L\times d} \to (0, 1)$, the objective*

$$\ell_{t,T}(D) = -\mathbb{E}_{Z_t\sim p_t}[\log D(Z_t)] - \mathbb{E}_{Z_t^{\eta,T}\sim p_t^{\eta,T}}\left[\log\left(1 - D(Z_t^{\eta,T})\right)\right]$$

has the unique minimizer, up to sets of Lebesgue measure zero,

$$D^\star(z, t, T) = \frac{p_t(z)}{p_t(z) + p_t^{\eta,T}(z)}.$$

Consequently,

$$\log\frac{1 - D^\star(z, t, T)}{D^\star(z, t, T)} = \log\frac{p_t^{\eta,T}(z)}{p_t(z)}.$$

Proof. Gaussian noising with $\sigma_t > 0$ makes both densities strictly positive. Writing the objective as

$$\ell_{t,T}(D) = \int_{\mathbb{R}^{L\times d}}\left[-p_t(z)\log D(z) - p_t^{\eta,T}(z)\log\left(1 - D(z)\right)\right]\mathrm{d}z,$$

we can minimize the integrand separately at each z. Set $a = p_t(z) > 0$ and $b = p_t^{\eta,T}(z) > 0$. For $u \in (0, 1)$, let $f(u) = -a\log u - b\log(1 - u)$. Its derivatives are

$$f'(u) = -\frac{a}{u} + \frac{b}{1 - u}, \qquad f''(u) = \frac{a}{u^2} + \frac{b}{(1 - u)^2} > 0.$$

Thus f is strictly convex, and its unique minimizer solves $f'(u) = 0$, giving $u = a/(a + b)$. Substitution yields $D^\star$ and $(1 - D^\star)/D^\star = b/a$, which proves the log-ratio identity. □

Averaging these objectives over $(t, T) \sim \nu$ gives (12). Hence an unrestricted discriminator conditioned on both noise levels has the stated optimum for ν-almost every pair.

Leave-one-out baseline. For a fixed discriminator, the leave-one-out baseline $\widehat{b}_g$ is independent of $\mathbf{X}_g^{\eta,T}$ conditionally on (Z_T, T). Since the conditional score function has zero expectation, subtracting this baseline preserves the expected cost-based update (Kool et al., 2019).

D.4 Auxiliary score estimation with bridge renoising

The multi-step formulation in Section 3.2 can also use the bridge $q_{t|T,\mathbf{x}}$ in place of forward noising. For Simplex-DMD, consider $Z_T \sim p_T$ and $Z_t \sim q_{t|T,\mathbf{x}}(\cdot \mid Z_T, \mathbf{x}^\eta(Z_T, T))$, and denote the resulting marginal by $p_t^{\eta,T\to t}$. Taking the conditional expectation of the Gaussian bridge score, with the coefficients of (1), gives

$$\nabla_{z_t}\log p_t^{\eta,T\to t}(z_t) = \frac{\alpha_t\sigma_{T|t}^2\mathbb{E}[\mathbf{x}^\eta(Z_T, T) \mid Z_t = z_t]\mathbf{E} + \alpha_{T|t}\sigma_t^2\mathbb{E}[Z_T \mid Z_t = z_t] - \sigma_T^2 z_t}{\sigma_t^2\sigma_{T|t}^2},$$

for positive bridge variance. Both expectations are under this joint law. The additional conditional mean $\mathbb{E}[Z_T \mid Z_t]$ is not estimated by the auxiliary token denoiser. Estimating it would introduce another prediction task, for example through a separate network or output head; alternatively, one could train an auxiliary model to predict the combined bridge mean directly. Forward noising retains the auxiliary token-denoising task of the one-step method. We leave auxiliary estimators tailored to the bridge to future work. Related experiments on continuous data by Salimans et al. (2024, Appendix D) reported poor results when fitting the exact bridge score. We do not evaluate bridge renoising during Reinforce-DMD training.

D.5 MULTI-STEP GENERATION

We justify here the DDIM-inspired family of transitions introduced in Section 3.2. The following proposition gives two conditions under which a transition preserves the exact diffusion marginals.

Proposition 5. *Fix $0 \le t < T \le 1$ and suppose that $\widehat{Z}_T \sim p_T$. Given $\widehat{Z}_T$, sample*

$$\widehat{\mathbf{X}} \sim \mathsf{k}_T^{\eta}(\cdot \mid \widehat{Z}_T),$$

and then

$$\widehat{Z}_t \sim q^{\zeta}_{t|\mathbf{x},T}(\cdot \mid \widehat{\mathbf{X}}, \widehat{Z}_T).$$

Then either of the following conditions implies $\widehat{Z}_t \sim p_t$:

1. *$\zeta_{t,T} = \sigma_t$ and the marginal distribution of $\widehat{\mathbf{X}}$ is p_{data};*
2. *$\mathsf{k}_T^{\eta}(\cdot \mid z_T) = p_{\mathbf{x}|T}(\cdot \mid z_T)$ for p_T-almost every z_T. In this case the conclusion holds for every $0 \le \zeta_{t,T} \le \sigma_t$.*

Proof. For $\zeta_{t,T} = \sigma_t$, the coefficient multiplying the noise inferred from $\widehat{Z}_T$ vanishes, and the transition becomes

$$\widehat{Z}_t = \alpha_t \widehat{\mathbf{X}}\mathbf{E} + \sigma_t \varepsilon, \qquad \varepsilon \sim \mathrm{N}(0, \mathrm{Id}).$$

If $\widehat{\mathbf{X}} \sim p_{\mathrm{data}}$, this is exactly the forward construction defining p_t.

For the second statement, sampling $\widehat{\mathbf{X}}$ from the exact posterior given $\widehat{Z}_T \sim p_T$ reconstructs the exact joint law of $(\mathbf{X}, Z_T)$ under the forward process. Hence we may write

$$\widehat{Z}_T = \alpha_T \widehat{\mathbf{X}}\mathbf{E} + \sigma_T \varepsilon_T, \qquad \varepsilon_T \sim \mathrm{N}(0, \mathrm{Id}),$$

with ε_T independent of $\widehat{\mathbf{X}}$. A draw from (8) can therefore be written as

$$\widehat{Z}_t = \alpha_t \widehat{\mathbf{X}}\mathbf{E} + \sqrt{\sigma_t^2 - \zeta_{t,T}^2}\, \varepsilon_T + \zeta_{t,T} \varepsilon,$$

where $\varepsilon \sim \mathrm{N}(0, \mathrm{Id})$ is independent of $(\widehat{\mathbf{X}}, \varepsilon_T)$. The two Gaussian noise terms combine into a $\mathrm{N}(0, \sigma_t^2 \,\mathrm{Id})$ variable independent of $\widehat{\mathbf{X}}$, so that

$$\widehat{Z}_t \mid \widehat{\mathbf{X}} \sim \mathrm{N}(\alpha_t \widehat{\mathbf{X}}\mathbf{E}, \sigma_t^2 \,\mathrm{Id}).$$

Since $\widehat{\mathbf{X}} \sim p_{\mathrm{data}}$, marginalizing over $\widehat{\mathbf{X}}$ gives $\widehat{Z}_t \sim p_t$. □

Applying Proposition 5 recursively along a decreasing grid shows that, if its assumptions hold at every transition, then

$$\widehat{Z}_{T_i} \sim p_{T_i}, \qquad i = 0, \dots, n.$$

In particular, for forward renoising, exact distribution matching at each starting level implies the required clean-marginal condition under the geometric assumption of Theorem 1.

E SUPPLEMENTARY EXPERIMENTS

E.1 IMPLEMENTATION DETAILS

Training budgets and model sizes. Table 3 separates teacher pretraining from the additional updates used to train the distilled students. Both of our students retain the LangFlow architecture (Chen et al., 2026); their auxiliary networks are used only during training and are excluded from the student parameter counts. Optimization-step counts alone do not imply equal training compute: batch sizes and the number of network evaluations per update can differ, and our methods alternate auxiliary and student updates. Each student is trained for $10{,}000$ optimizer updates in total, comprising $8{,}000$ auxiliary updates and $2{,}000$ student updates. Our training GPU budget comprises four nodes with eight NVIDIA H100 GPUs each (32 GPUs in total).

Method	Teacher	Teacher updates	Distillation updates	Student params.
SDTT (Deschenaux & Gulcehre, 2025)	MDLM	1M	$r \times 10$k $r = 1, \dots, 7$	≈170M
IDLM (Li et al., 2026)	MDLM	1M	1M	169.63M
FMLM (Lee et al., 2026)	FLM	1.5M	1M	≈170M
ReDi (Yoo et al., 2025)	DUO	∼500k (+500k)	≤1M	≈165M
Simplex-DMD (ours)	LangFlow	1M	10k	≈170M
Reinforce-DMD (ours)	LangFlow	1M	10k	≈170M

Table 3: Training budgets and model sizes for the distilled checkpoints. Teacher updates are pretraining steps; distillation updates are additional optimizer updates. For our methods, the 10k total includes 8k auxiliary and 2k student updates. SDTT uses successive round checkpoints, indexed by r. The ReDi student budget is an upper bound. Parameter counts refer to one sampling model, including embeddings, and exclude auxiliary training networks.

For SDTT, the teacher is `sdtt@teacher_1M_sm` (`mdlm-owt`); IDLM uses `kuleshov-group/mdlm-owt`, both based on MDLM (Sahoo et al., 2024). The FMLM checkpoint is initialized from the local FLM `flm_owt` checkpoint and trained with PSD; IDLM uses a learning rate of 10^{-6}. ReDi uses the DUO (Sahoo et al., 2025) checkpoint `s-sahoo/duo` (`duo.ckpt`). Both of our students use the same released LangFlow checkpoint (Chen et al., 2026). For all the baselines, we use their default samplers presented in their associated paper.

Training and sampling configuration. Table 4 summarizes the training and sampling configuration of each student; the ablations of Section 4.2 and below vary its components. At inference, we divide the temporal interval $[0, 1]$ uniformly, using $T_i = i/n$ for $i = 0, \dots, n$, and traverse this grid from 1 to 0. Sampling temperature is applied at the network output: we divide the token logits by the temperature before the softmax that produces the clean token probabilities.

Component	Simplex-DMD	Reinforce-DMD
Student initialization	teacher	teacher
Auxiliary initialization	teacher	teacher
Renoising kernel	$q_{t\|\mathbf{x}}$	$q_{t\|\mathbf{x}}$
Self-conditioning	on	off
KL anchor β	–	0.5
Sampler	forward	forward

Table 4: Training and sampling configuration of each student. A dash marks a component the objective does not have. The last row is applied at inference.

Discriminator architecture. The Reinforce-DMD discriminator uses the same transformer backbone as the teacher, including rotary positional embeddings and noise-level conditioning through adaptive layer normalization. It operates directly on noised latents in the teacher's embedding space, omitting the token-embedding layer and replacing the vocabulary prediction head with a scalar head. We initialize the backbone from the teacher checkpoint and initialize the scalar head separately. The head applies noise-conditioned adaptive layer normalization followed by a projection to one logit per position; averaging these logits over sequence positions yields a single scalar logit per sequence.

Optimization. We use a global training batch size of 256 and AdamW at a constant learning rate of 3×10^{-4} after 2500 warmup steps, without weight decay, in `bf16`, with gradients clipped at 1.0 and an exponential moving average of decay 0.9999. We use $n_{\text{aux}} = 4$: four auxiliary updates on detached student samples, followed by one student update. At each optimization step, we sample the time pair $(t, T) \sim \nu$.

Evaluation harness. We evaluate on OpenWebText with sequences of length 1024, sweeping sampling temperatures $\{0.80, 0.85, 0.90, 0.95, 1.00, 1.05, 1.10\}$ applied to the output logits and sampling budgets $\{1, 2, 4, \ldots, 1024\}$, with the autoregressive references evaluated only at 1024 NFE. Generative perplexity is computed by exponentiating the mean token negative log-likelihood under GPT-2-large. We evaluate with a single random seed, as in the default experimental setup of Gourevitch et al. (2026, Appendix I.1). We pair perplexity with diversity measured as the mean per-sequence Shannon entropy of empirical GPT-2 token frequencies, in nats, to obtain a quality–diversity frontier at each sampling budget, following Pynadath et al. (2026).

E.2 Full generative frontiers

Figure 4 extends the method comparison of Figure 1 to all evaluated sampling budgets.

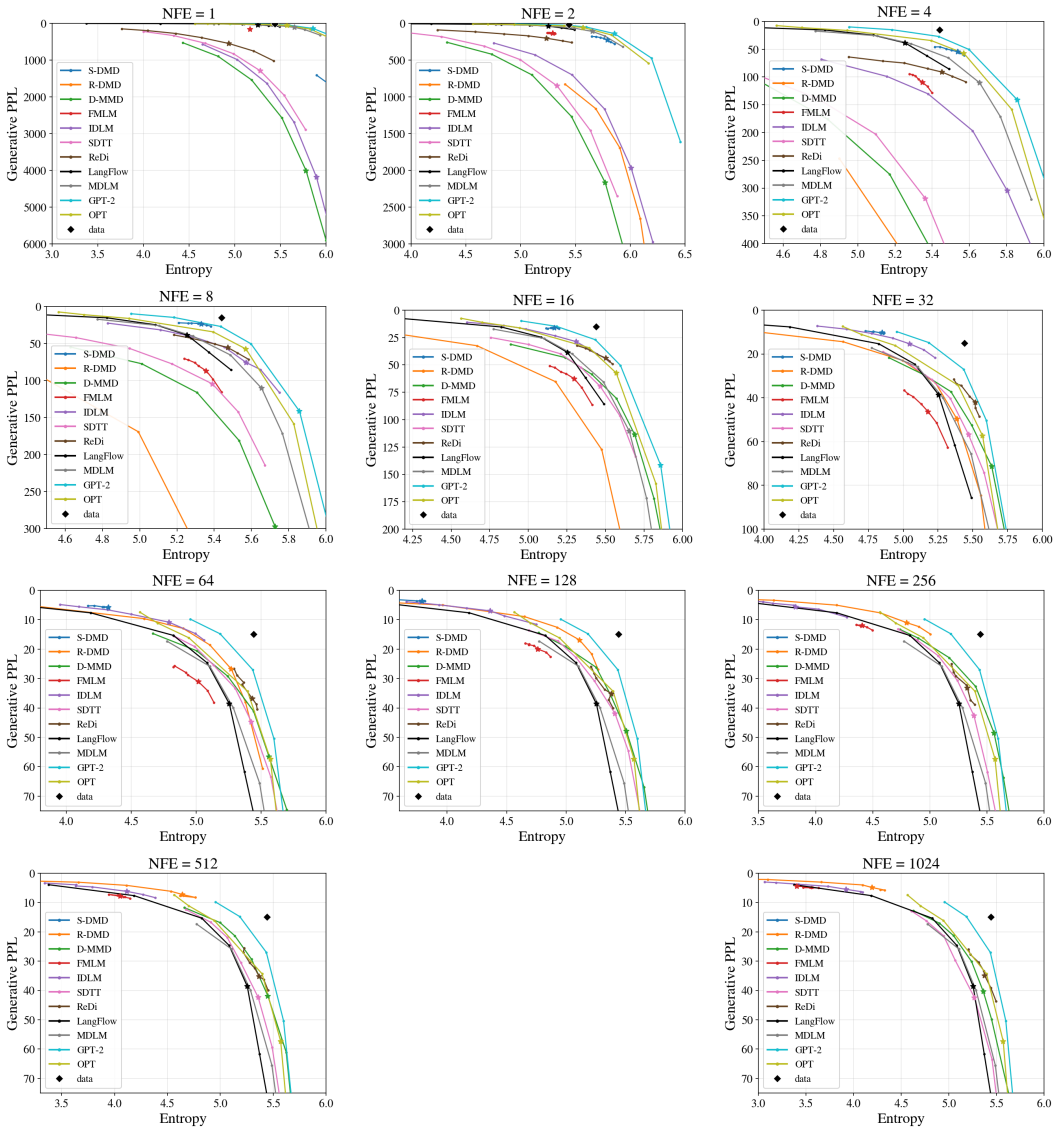


Figure 4: Generative frontier on OpenWebText at every step budget, read as in Figure 1.

E.3 Design space

Figure 5 places the variants of each student on the frontier, every variant differing from Table 4 in a single component. The bridge-renoising ablation for Simplex-DMD is reported separately in Appendix E.5.

Simplex-DMD variants. The top row compares the recipe with three variants. *No sc* trains the student without self-conditioning, the reuse of the previous clean prediction as an extra network input described in Appendix B.1. *No tsched* replaces the teacher's learned noise schedule, which the recipe reuses (Appendix B.1, (15)), with a default schedule γ. *No student init* initializes the student from scratch instead of from the teacher. The no-tsched variant lies on the recipe's curve and is told apart only by where along the curve its temperature-1.0 point sits; without self-conditioning, the curve reaches a slightly higher Gen PPL at the same entropy. The student trained from scratch collapses to near-uniform tokens, with unigram entropy about 6.9 and Gen PPL above 6×10^4; it is marked in the bottom-right corner, outside the plotted range.

Reinforce-DMD variants. The bottom row varies the weight β of the KL anchor toward the teacher in (13): $\beta \in \{0, 0.2, 0.5\}$, and a schedule annealing β from 0.5 to 0.05 over training. An additional variant trains Reinforce-DMD with self-conditioning, which the recipe does not use for this student. With $\beta = 0$, the student collapses to zero unigram entropy; this variant is marked in the top-left corner, outside the plotted range. The positive weights are separated by little, and $\beta = 0.5$ is the best of them. Self-conditioning moves the frontier to a higher Gen PPL at every entropy.

E.4 Auxiliary loss

Equation (11) trains the auxiliary denoiser with cross-entropy against the student's relaxed output. Two alternatives have the same population minimizer, the conditional mean $m_t^{\eta,T}$, and so leave the gradient identity (9) unchanged. The first regresses the auxiliary on that output under squared error,

$$\mathsf{L}_{\text{aux}}^{\text{L2}}(\phi) = \mathbb{E}_{(t,T)\sim\nu,\, Z_T\sim p_T,\, \varepsilon}\left[\sum_{\ell=1}^{L} \left\|\mathbf{X}^{\eta,T,\ell} - \mathbf{x}^{\phi,\ell}(Z_t^{\eta,T}, t, T)\right\|^2\right], \tag{32}$$

and the second keeps cross-entropy but replaces the simplex target by a token drawn at each position, $Y^\ell \sim \mathrm{Cat}(\mathbf{X}^{\eta,T,\ell})$, written $\mathbf{e}_{Y^\ell}$ for its one-hot vector,

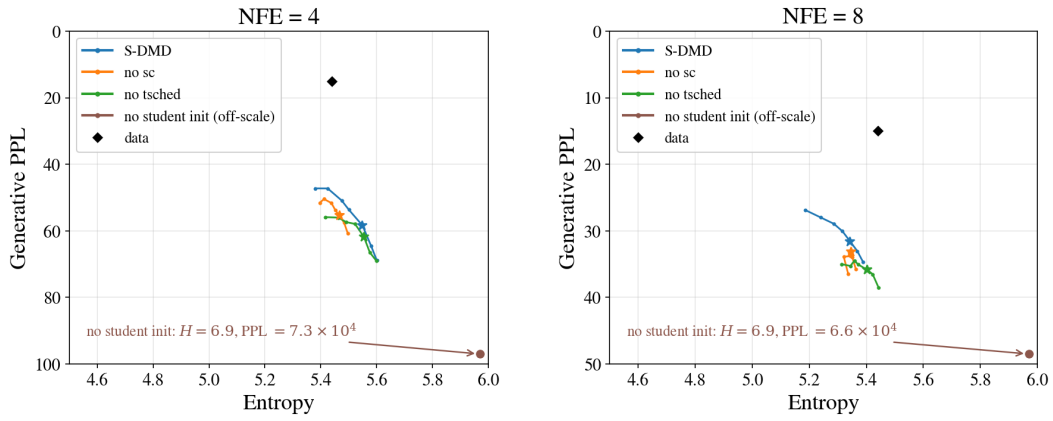
$$\mathsf{L}_{\text{aux}}^{\text{hard}}(\phi) = \mathbb{E}_{(t,T)\sim\nu,\, Z_T\sim p_T,\, \varepsilon,\, Y}\left[\sum_{\ell=1}^{L} \mathrm{CE}\Big(\mathbf{e}_{Y^\ell}, \mathbf{x}^{\phi,\ell}(Z_t^{\eta,T}, t, T)\Big)\right]. \tag{33}$$

Since $\mathbb{E}[\mathbf{e}_{Y^\ell} \mid Z_t^{\eta,T}] = \mathbb{E}[\mathbf{X}^{\eta,T,\ell} \mid Z_t^{\eta,T}]$, (33) estimates the same quantity as (11) with the extra variance of the categorical draw.

Figure 6 compares the three as Simplex-DMD training variants. They do not move the frontier; they move the temperature-1.0 point along it. The sampled-token target reaches the lowest Gen PPL at every budget, but always at the lowest entropy of the three, and squared error sits at the opposite end, more diverse and less fluent, with the soft target between them. The variants separate as the budget grows, and they separate by how fast entropy erodes: from $\mathrm{NFE} = 8$ to $\mathrm{NFE} = 1024$ the soft target loses 1.85 nats of unigram entropy against 2.24 for squared error and 2.36 for the sampled-token target, so at $\mathrm{NFE} = 1024$ it retains 3.5 nats against 3.1 and 3.0. Since Gen PPL falls with entropy along this frontier, a variant that collapses faster reports the better perplexity while generating less diverse text. We keep the soft target for that reason. These three variants share a training recipe that differs from Table 4 in two components, and each is a single run, so the comparison is indicative of the loss rather than a clean ablation of the default configuration.

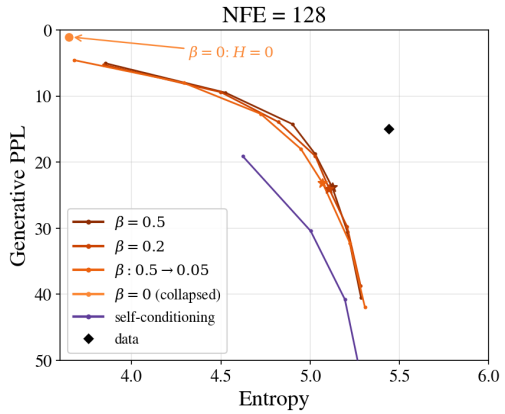
E.5 Bridge renoising for Simplex-DMD

We compare forward noising $q_{t|\mathbf{x}}$ with the bridge $q_{t|T,\mathbf{x}}$ during Simplex-DMD training (Figure 7). Each variant is a separate training run; forward noising is the default in Table 4. At every displayed budget, forward noising reaches a lower Gen PPL than the bridge at the same entropy, with the gap narrowing as the budget grows. Appendix D.4 discusses the auxiliary score approximation used for the bridge variant. We do not evaluate bridge renoising during Reinforce-DMD training.

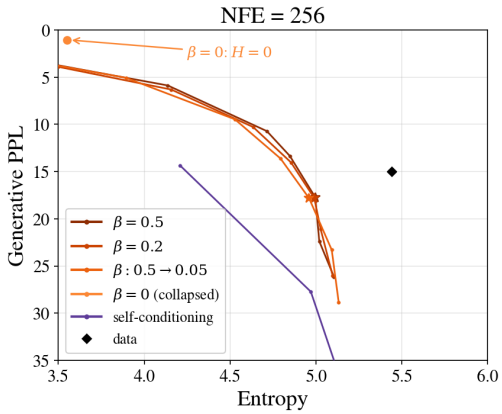


(a) Simplex-DMD, NFE $= 4$

(b) Simplex-DMD, NFE $= 8$

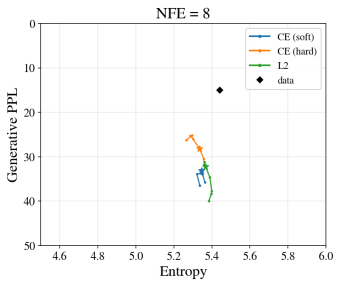


(c) Reinforce-DMD, NFE $= 128$

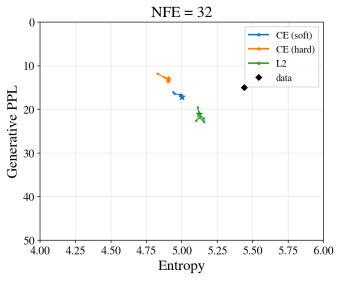


(d) Reinforce-DMD, NFE $= 256$

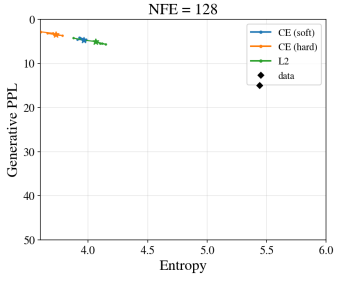
Figure 5: Design space on OpenWebText, one component changed per variant, read as in Figure 1. Each student is shown at the two budgets where it is competitive.



(a) NFE $= 8$

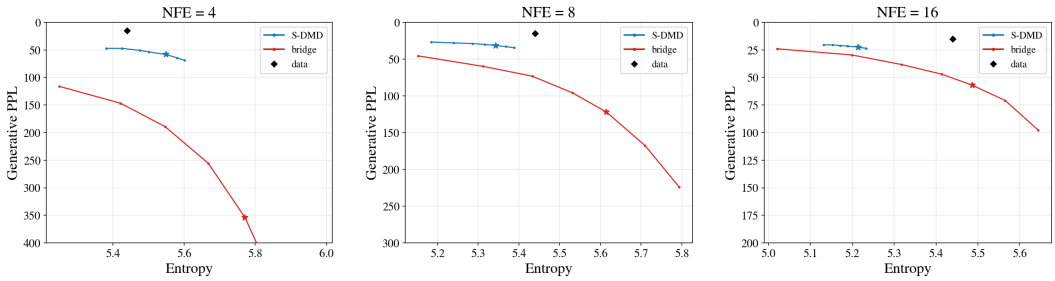


(b) NFE $= 32$



(c) NFE $= 128$

Figure 6: Auxiliary loss for Simplex-DMD, read as in Figure 1: the soft target of (11) (CE soft), the squared error of (32) (L2) and the sampled-token target of (33) (CE hard). The three variants trace one frontier and differ in where their temperature-1.0 point sits on it, the sampled-token target lowest in entropy and squared error highest. Each variant is a separate training run; all are sampled with the ODE step ($c = 0$).



(a) NFE $= 4$ (b) NFE $= 8$ (c) NFE $= 16$

Figure 7: Forward noising $q_{t|\mathbf{x}}$ against the bridge $q_{t|T,\mathbf{x}}$ as the training renoising kernel for Simplex-DMD, read as in Figure 1. Each variant is a separate training run.

E.6 Generated samples

We show three unconditional samples from each student, Simplex-DMD at 4 and Reinforce-DMD at 128 network evaluations, and from the LangFlow teacher and MDLM at 1024 network evaluations. For each model, we use the sampling temperature T whose unigram entropy is closest to that of the data (5.44), and show the first three stored samples of that evaluation point, without selection and truncated to their first 180 words; <eos> marks an end-of-sequence token inside the 1024-token sequence. The header of each box reports the Gen PPL and unigram entropy of the whole evaluation point.

Simplex-DMD NFE $= 4, T = 0.85$ · Gen PPL 45.6, entropy 5.44

Sample 1

game, and they are much more powerful and terrifying than ones you play."

All of that. That's why elves fight against the ash, and why they do in combat and express their interests. The most baffest answer to this question, though, is whether cooperative action can do anything to an otherwise beautiful species.

There are powerful heroes like Balondakland. Magic's heroes are beautiful and powerful, and they make them so much better. Attending resistance to powerful heroes, like almost about every single character, defeated or defeated in the game, is mind-shacking. Magic is about strategy, not a competitive strategy that is solely focused on how you play and fight. You have to win powerful battles. Why isn't anyone going to tell you that advice? <eos> Imagine walking through every major venue in the United States every weekend and being told that British music is revered and charing kids to believe. You know how they treat kids? You just want to tell their kids, because they will always play. Are you kidding?

It is your job to do it like that: Your kids …

Sample 2

Times. You may opt-out at any time. agree agree to receive occasional updates and special offers for The New York Times's products and services. Thank you for subscribing. An error has occurred. Please try again later. View all New York Times newsletters. <eos> The increasingly technical aspect of chemical attack has started to wreak havoc in the United States. Even as the amount of bombs dropped and vast amounts of gas has soared, military officials said the problem was pervasive. And chemical attack threats aren't new in the United States, said of physicist Dr. Mark Col. Salff of Columbia University. Experts believe they are common in countries that don't yet know much — and tend to detonate civilian infrastructure behind a chemical attack.

Indeed, the United States' government conducted a simulated chemical attack attack last year at Fort McMourstal, Va., where although it was the Pentagon's first interagency Air Air Weapons Center (IAF) in the early 1990s, the airborne alert didn't require the Bush administration to start experimenting with chemical attacks. Now, experts say the chemical attack's capabilities isn't compared with many …

Sample 3

about Hillary Clinton — and Trump refused to vilify her.

It's pretty rare example of a woman who donated to Trump's re-election campaign. But She Doesn't Really Support The Public

Continue Reading Below Below Bagale Morinas

Handousee-stick speech.

But this woman's lawyer is of no: She's giving public officials a history of her work to the presidency.

Nobody without dating a woman has a certain course of conduct — there's one reason why divorce lawyers find it very difficult: Anyone using public money to attack his or her spouse just gets a de-fated.

The fact is, there's a good chance that this divorce or the whole thing could be worse if that it makes Donald Trump look worse. But supporting Trump is just another candidate or vice candidate.

We aren't so sure if this has anything to do with Melania. But to be sure, it doesn't mean Melania the mind-blung or most stack of any political conversations we've ever seen. When Melania gets on jumpballs, automated drivers crash into Melania's emails, phone numbers, phone numbers, and private apps.

Of course, . . .

Reinforce-DMD NFE $= 128$, $T = 1.10$ · Gen PPL 26.9, entropy 5.27

Sample 1

decreasing; changes in pattern of hacerenial events do not improve after use by patients without required DCI criteria; changes in markers of hacerenial events do not improve after use in patients with nonmenidiients; lesora conditions do not differentiate as an increase in patients with hacerenial events over time (9%); GRAs do not affect as an increase in patients with hacerenial events over time (9%); there has been an increase in GRAs and high blood pressure in patients with hacerenial events regardless of BMI (10%); there has been an increase in MII after signs of hypertension were reported in patients with high blood pressure (10%); in patients with hacerenial events over time (15%); GRAs have been either decreased in patients with hacerenial events in baseline and without exercise (16); there has been an increase in obesity and high blood pressure in diabetes (15) in patients with hacerenial events without diarrhea (15%); in all cases, there was no detectable cardiovascular symptoms after onset of autoimmune events; patients with inflammatory bowel sensitivity were found to show a reduction in patients with hacerenial events . . .

Sample 2

emio_build=true <class# rm appemio.init" - create "/tmp/retockbyC" JAX 15.0 appemio.init" - create "/tmp/retockbyC" JAX 15.0

1 2 3 4 5 6 7 8 rm appemio_build=true <class# remove docker appemio.init" - create "/tmp/retockbyC" JAX 15.0 appemio.init" - create "/tmp/retockbyC" JAX 15.0 <class# remove docker./ appemio.init" - create "/tmp/retockbyC" JAX 14.0 <class# remove docker appemio.init" - create "/tmp/retockbyC" JAX 15.0 <class# rm appemio_build=true

This won't work automatically; it still works by default inside your 'r80 in your build directory before it actually works. So we want to live everything on your 'r80.

Setting Up into the Appemio Environments

To start running appemio, you need to set up a new, new build inside make all configuration files.

You will start with appemio/esuse-build, which will create a new configurationfile that is called appemio/esuse, but it's already available automatically:

1 cp /var/.share/appemio/esuse

which will build inside add appemio/esuse, but that's located in lib/build/esuse.module.conf :

1 cp /var/.share/appemio/esuse

will create a new configurationfile so also download appemio/esuse that's already available. It creates appemio/esuse.conf, which is file appemio-build and

which is located in lib/build/appemio/esuse.conf :

1 2 . . .

Sample 3

ands, we will continue to seek new coal sources," Bowman said in his budget. "As a rural country, we have a great need for our provinces, our bodies, and our provinces and communities to have the edge in demonstrating this wave of growth."

Story continues below advertisement

Under Mr. Trudeau's plan, each province has limited greenhouse gas emissions by nearly 50 megawatts. More than 50 states in the U.S. have set rules to address fossil fuels. The two largest jurisdictions in the states of Vermont, North Dakota and Washington, D.C. have reduced greenhouse gas emissions by 20 per cent year-per-year.

Those reductions overlap far beyond Prime Minister Justin Trudeau's pledge to spend billions of dollars on the province's dirty tape.

Low-emissions costs of coal projects increase by more than 15 per cent year-per-year, while reductions now account for 85 per cent of the cost of low-carbon infrastructure, according to a Bloomberg Institute analysis Monday. More than half of those costs will come from low-emissions and natural gas.

In July, Mr. Trudeau unveiled plans to erect 30 megawatts of new oilsands . . .

LangFlow (teacher) NFE $= 1024$, $T = 1.10$ · Gen PPL 85.6, entropy 5.49

Sample 1

2 hits were not (corrected)

8/9/25: Meat (@Hot Alert) has been picked up for three hours.. 2 kills were not (corrected) 11 hits were not (corrected)

2/7/29/25: Burn from Asgard (@Hot Alert) has been picked up for six hours. 4 hits were not (corrected)

OTHER ERROR/INCOMPLETE: Lowened 13 stars on my playlist 3/17/03/27: Las Mountains (@Hot Alert) has been picked up for eight hours.. 3 hits were not (corrected) 5/3/11/30: The Future Problem (@Hot Alert) was picked up for over a hour. 2 hits were not (corrected) <eos> If Golson turns home and you come there, you can put him on the roster. A wanna-play clown in your system? — leke Moslon (@Soslon_ESPN) November 19, 2015

AFNDLEVILLE – The life of recruiting draft ranting is a nightmare.

Luckily for a core scorer who accepted draft year after drowning school had spent its highest semester, Tymontise Golson is secretly living in town for another day.

The former Ohio State guard, who appeared on Charlie Troke's draft earlier this month, took to that point Wednesday night meeting with a group of league reporters.

"Of . . .

Sample 2

of skill and education plans for secondary sailing across Europe. Of the 352 charter boats there are 3,200 different applications available around the world with each one recognised in quantity.

David McConnell, national accommodation on the invaluable Rolic Drum Local Focus, which provides qualification guidance for all single luxury consultation applications, said: "We have seen sailing fishing stocks increase shortages due to the growing level of ocean education in demand."

Sinking expert Roger Dixon said: "These results show that the UK fishing market has a powerful programme of skills and education initiatives available to children and kids around the world. <eos> [Editor Side Note: The following first marks our former September 16th Guest's honour.]

Even today, one of music's most evocative terms is Tavement Line's "See That I Grived." A Athopper-singing blend of lyrics that were inherited from his nearly haphazard and petulous live records in 1987, Six Flags Alegs eight years later. In celebration of it allure, Tavement Line has nicely prime prop reissues but part of their numerous compilation. After associating their last album of single "Kids I Still Walk" . . .

Sample 3

a weakness," he told breaking-well for six minutes in his dissertation, "and sometimes you good both very quickly." <eos> RAR (IAM)) [LANAP] – Israeli forces launched three separate missile attacks on the country's border towards Israel on Wednesday, including what Palestinian targeted witnesses say is the first retaliatory clashes.

The raids took place at a checkpoint on both sides of the Palestinian town of Ash Sabh, where Todi Palm flew heavy bombs on Lod al Sala, but were identified by a border insurgent, according to local residents.

A Palestinian militant investigating Outside Gaza said civilians died from the attack Wednesday, but others gave vague of other missile intercepts at the sight. No casualties could be immediately reported.

The missile bombardations by Israel came less than two weeks after its security forces raided a heavily populated fence and fence carrying anti-tank rockets. Two people were killed and an Israeli Israeli plane was intercepted near Ramte.

Anshara was seized by a five-day truce mediated by Israel and Iranian-led rebels, during which the rebels wanted in their area to be backed with militant terror groups . . .

MDLM NFE $= 1024$, $T = 0.95$ · Gen PPL 65.5, entropy 5.49

Sample 1

civil rights profession say the allegations are "serious and troubling."

"It's clear that very little of the jurors have acknowledged wrongdoing or wrongdoing," Toohey said, echoing a refrain common in the civil rights movement. "These are very serious allegations."

Chris Douglass, a United States Department of National Security Policy lawyer at United States Witness in Florida, explained that the case often disadvantaged millennials who are pressured not to take part in the investigation and their reporting completed.

"They're exploited," Toohey said.

United States Witness names 105 officers involved in Miami civil rights cases that have been identified.

Miami is a unique for civil rights attorneys all across America because corruption, says Lee Grenner, a former federal prosecutor of color, makes it difficult for attorneys who can initiate cases in foreign countries before the Supreme Court (also known as the American "African Supreme Court") can call upon outside legal counsel, thus making it more effective and expensive to investigate black rights cases. <eos> Update: A population center, 18 for Malta, released a paper encouraging more condomless pills. The paper, which was authored as . . .

Sample 2

for money. In fact, the whole system will regulate our feelings in ways that only make sense for us to really address how it treats our selves equally. For a start, some studies point out that one system away from we altogether can actually allow some people to feel more about loss than what might go into it, though the system is likely also a poor measure of value, resulting in a lack of emotional profits. Although it does not do any better than normal savings accounts, it offers some possible benefits.

By the way, history shows that if we misheat ourselves we can do the impossible for ourselves, whether everything is on going. But if we more emphasize the concern of diminishing return, we may simply drop everything, and we "blow up a cascade of misery and despair. And indeed the path to collapse teaches our system of the most intricate ways we manage for money: How do we make joy, dispassion, love, compassion, and even loss easy? Here's what happens. <eos> McGlorady's husband and her room throw the weight of . . .

Sample 3

.2 million repeats) now in its third quarter, which would be tied with the primetime telecast for longest-running programming in network history.

Disney's ABC came in at 5.6 million repeats, possibly 17% higher than NBC's previous 5.6– and close to breaking even with cable history in the share of live programming on network television.

10 p.m. ET/6'30 ABC was tickllled this week, and the network is expected to become the seventh ABC network to win post WWF slot at 10 p.m. ET/7'30.

Chuck Licht missed out

In an appearance at the 2011 Republican National Convention last Sunday, Roberts said his network drew the least share of the Los Angeles market in all-day coverage. The 2010 Comcast MediaStar convention averaged 1.7 million viewers on channel 88, the network's highest number among all telecasts, up more than the network's 4,5,000 watched by Comcast cable network Fox's $2.9 million teleday draw, and NBC.

Cost Mario Roberts said the Tampa Bay Times dressing is the direct result of the rise of cable "puttinging people for the great particular class of being rebuffed by the . . .