POSTERIOR SAMPLING BY SOURCE-SPACE MCMC VIA PRIOR-BASED FEW-STEP TRANSPORT MAPS

Anonymous authors
Paper under double-blind review

ABSTRACT

Bayesian inference increasingly uses informative but implicit priors represented only by samples, such as historical ensembles, simulator outputs, and pretrained generative models. The same computational problem appears in the test-time guidance task (generalized Bayes), where an explicit positive weight, e.g., an exponentiated reward, tilts an implicit prior. We develop a framework for source-space generalized Bayesian inference that combines inexpensive few-step prior transports with posterior stability guarantees. Specifically, we represent the prior using a one- or few-step improved MeanFlow (iMF) map and perform posterior sampling in its Gaussian source space. We establish Wasserstein error bounds between the exact and learned posteriors in terms of the joint population iMF and auxiliary-velocity loss, decomposed into training suboptimality and model-class approximation error. In the iMF source space, we adopt parallel tempering with preconditioned Crank–Nicolson updates and introduce a hybrid variant that incorporates split Hamiltonian Monte Carlo to improve sampling efficiency. Synthetic experiments show that the proposed framework can approximate posterior distributions accurately and efficiently, while CLIP-guided ImageNet experiments demonstrate its ability to steer a pretrained iMF image prior toward text-specified preferences.

1 INTRODUCTION

Bayesian inference combines prior knowledge with observations to characterize uncertainty about unknown quantities (Stuart, 2010). In many applications, however, prior information is available through samples, e.g., historical ensembles, simulator outputs, or pretrained generative models rather than an explicit density (Patel & Oberai, 2019; Chung et al., 2023; Lunz et al., 2018). Such prior samples can represent complex structures that are difficult to encode analytically, but their densities and scores are not directly available for posterior Markov chain Monte-Carlo (MCMC).

Formally, let $x \in \mathbb{R}^{d_x}$ be an unknown with prior π_0 available only through i.i.d. samples, and let y denote observations or a task specification. We encompass Bayesian and generalized Bayesian inference through the generalized posterior

$$\pi^y(\mathrm{d}x) = Z_y^{-1}\ell_y(x)\pi_0(\mathrm{d}x), \qquad Z_y = \int \ell_y(x)\pi_0(\mathrm{d}x), \tag{1}$$

where ℓ_y is an analytically given nonnegative weight and $0 < Z_y < \infty$. Standard Bayesian inference takes $\ell_y(x) = p(y|x)$, while generalized Bayes (Bissiri et al., 2016) uses $\ell_y(x) = \exp(-\tau L_y(x))$ for a task loss L_y and $\tau > 0$. Reward guidance (Korbak et al., 2022) fits the same formulation with $\ell_y(x) = \exp(\tau R_y(x))$, favoring samples with larger reward R_y. For example, y may be a text prompt, with $R_y(x) = \mathrm{CLIP}(x, y)$ measuring the cosine similarity between the CLIP embeddings of image x and prompt y (Radford et al., 2021). Henceforth, we use ℓ_y for either a likelihood or a general weight and refer to π^y simply as the posterior.

An approach is to learn the data-space prior score by denoising score matching (DSM) (Vincent, 2011) and combine it with the gradient of an analytic log-likelihood or reward to construct Langevin or Hamiltonian dynamics. These samplers are typically unadjusted because a learned score alone does not provide the density ratios needed for Metropolis correction. Moreover, sharp variation in the score near complex data manifolds can produce stiff dynamics, requiring small integration steps.

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.

Another approach is to first learn a deterministic generator $S_\theta : \mathbb{R}^{d_z} \to \mathbb{R}^{d_x}$ satisfying $(S_\theta)_{\#}\gamma \approx \pi_0$, where $\gamma = \mathcal{N}(0, I)$ denotes the source distribution, $\mathbb{R}^{d_z}$ is called the source space and $_{\#}$ denotes the pushforward. Under the learned model S_θ, the complex data-space prior is represented by the simple Gaussian reference measure γ. The corresponding *learned* posterior in the source space is

$$\nu_\theta^y(\mathrm{d}z) \propto \ell_y(S_\theta(z))\gamma(\mathrm{d}z). \tag{2}$$

We sample from ν_θ^y using MCMC and map the resulting samples to data space through S_θ. In contrast to the conventional data-space approach, ν_θ^y is available up to normalization, enabling Metropolis correction and thus an MCMC kernel that leaves ν_θ^y invariant. Furthermore, the source-space posterior is expected to have more favorable geometry for MCMC: its prior is Gaussian, while S_θ has already absorbed much of the curvature and anisotropy of π_0. Nevertheless, the likelihood/weight $\ell_y \circ S_\theta$ may reintroduce curvature, stiffness, anisotropy, or multimodality, particularly when the observation model is highly informative or nonlinear. Early work by Patel et al. (2022) used a Wasserstein GAN to construct S_θ and applied Hamiltonian Monte Carlo (HMC) with the No-U-Turn Sampler (Hoffman et al., 2014) to the resulting low-dimensional latent posterior. However, learning S_θ with a WGAN requires solving an adversarial min-max optimization problem, which can be challenging to stabilize. Flow-based models (Chen et al., 2018) offer an attractive alternative: flow matching learns a time-dependent vector field through a standard regression objective (Lipman et al., 2023; Tong et al., 2024). This provides a scalable approach to learning expressive transport maps for complex and multimodal priors. Under suitable regularity conditions on the vector field, the resulting ODE flow also defines a diffeomorphism between source and data spaces (Papamakarios et al., 2021). Source-space guidance for flow-matching models has been explored through both optimization and sampling. D-Flow (Ben-Hamu et al., 2024) optimizes the source variable, while Source-Guided Flow Matching (SGFM) (Wang et al., 2026b) samples from the source posterior distribution (2). Source Parallel Tempering (SPT) (Wang et al., 2026a) combines local preconditioned Crank-Nicolson (pCN) (Beskos et al., 2008) updates with replica exchanges (Earl & Deem, 2005) across tempered source distributions to address multimodality. These source-space methods are instances of *test-time guidance*, incorporating user-specified likelihoods or weights during inference.

For methods that represent the prior through a learned transport (Wang et al., 2026b;a), a central theoretical question is how prior-learning error controls the discrepancy between the learned posterior $(S_\theta)_{\#}\nu_\theta^y$ and the true posterior π^y in (1). We refer to this as *posterior stability*. Existing works have tried to address this question, but a gap remains. Wang et al. (2026b) assume a uniform vector-field approximation bound, $\|v_\theta(x,t) - v(x,t)\|_\infty \le \epsilon$ where $v(x,t)$ is the true velocity field and $v_\theta(x,t)$ is the learned velocity field. However, the flow-matching objective controls an average squared error rather than a uniform error. Wang et al. (2026a) address this mismatch by establishing posterior stability in terms of the L^2 transport-map error, i.e., $\mathbb{E}_{z\sim\gamma}\|S(z) - S_\theta(z)\|^2 \le \epsilon^2$, where S is the exact transport map induced by $v(x,t)$. Their connection to the flow matching objective takes S_θ to be the map induced by the *exactly integrated* learned flow, via Grönwall's inequality. In practice, however, numerical ODE solvers are employed, introducing an additional discretization error. Under standard regularity and stability assumptions, a numerical ODE solver of order k has global error $O(h^k)$ (Hairer et al., 1993). The training objective alone does not control this error. Reducing h requires more integration steps, increasing the cost of source-space MCMC. Minibatch optimal-transport coupling (Tong et al., 2024), as used by Wang et al. (2026b), can encourage straighter trajectories and facilitate few-step integration, but does not by itself provide a bound on this numerical error; their theoretical bound likewise concerns the *exactly integrated* learned flow. Thus, guarantees for the exactly integrated learned flow do not by themselves establish posterior stability for a fixed few-step implementation. Likewise, Purohit et al. (2025) use few-step consistency models but do not establish how the consistency-distillation training objective controls the resulting posterior error.

In this work, we pursue two complementary objectives: making source-space inference computationally efficient through few-step transports *and* establishing posterior stability guarantees for the implemented maps via the true population training objective. To this end, we use improved MeanFlow (iMF) (Geng et al., 2026b), which learns interval-average velocities without distillation and natively enables one-step or few-step generation without numerical ODE integration. We further explain why the original MeanFlow objective (Geng et al., 2026a) does not directly yield the same guarantee, highlighting that native few-step generation alone can fall short of establishing such stability results. Practically, the iMF maps make source-space gradients inexpensive to evaluate when available, allowing MCMC kernels to optionally exploit them to improve mixing.

Contributions We use iMF transport mappings to represent complex priors available only through samples in generalized Bayesian inference settings. Under stated regularity assumptions, we connect population training suboptimality and model-class approximation error for the joint iMF and auxiliary-velocity objective to posterior stability for the implemented one-step or few-step maps. Building on these transports, we adopt SPT with pCN updates (Wang et al., 2026a) to facilitate exploration across modes while exploiting the Gaussian reference structure of the high-dimensional source space. When likelihood or reward gradients are available, we additionally introduce Gaussian-reference split HMC (Beskos et al., 2011) at the coldest temperature of SPT to exploit local posterior geometry and potentially improve mixing. Both local kernels preserve the Gaussian reference measure in their reference dynamics, making them natural choices for inference in high-dimensional source spaces. We evaluate the proposed framework on synthetic posterior sampling and CLIP-guided ImageNet generation, demonstrating accurate and efficient posterior approximation in the synthetic settings and the ability to steer a pretrained image prior toward text-specified preferences.

Other related work

Transport MCMC Transport maps have long been used to improve MCMC by transforming difficult posterior target distributions into coordinates easier for sampling (Parno & Marzouk, 2018; Zhang et al., 2025). NeuTra HMC learns a variational transport tailored to posterior geometry, with an amortized extension that accommodates different observations within a specified probabilistic model (Hoffman et al., 2019).

Score-based and diffusion inverse solvers Diffusion-based inverse solvers offer another way to combine a learned sample-only prior with observations, typically by incorporating likelihood information into the reverse generative process. Diffusion posterior sampling (DPS) (Chung et al., 2023) is a prominent example, enabling posterior-oriented sampling. Its approximate guidance, however, does not generally preserve the desired posterior and can introduce *systematic* bias beyond that arising from prior learning and numerical discretization. Denoising-score MCMC (Ramzi et al., 2020) learns the prior score from samples using denoising score matching (Vincent, 2011) and combines it with an analytically available likelihood score to perform MCMC directly in data space.

Training-time posterior learning. Complementary to test-time guidance, amortized posterior learning trains observation-conditional generators to directly approximate posterior distributions. Neural posterior estimation and related methods learn from simulated parameter-observation pairs, enabling rapid inference for new observations (Greenberg et al., 2019; Radev et al., 2023; Cheng et al., 2026). The training cost is shared across subsequent inference tasks, but changes to the observation model beyond those covered during training may require retraining or adaptation.

2 Background

2.1 Flow matching

Flow matching (FM) (Lipman et al., 2023) learns a velocity field approximating the marginal ODE $\dot{w}_t = v(w_t, t)$, which transports a simple source distribution p_{src}, typically Gaussian, to a target distribution p_{tgt} available through samples. Let $X \sim p_{\mathrm{tgt}}$ and $Z \sim p_{\mathrm{src}}$ be independent target and source samples, and define $W_t = (1-t)X + tZ$ and the samplewise interpolation velocity $V = Z - X$. FM approximates the marginal instantaneous velocity $v(w,t) = \mathbb{E}[V \mid W_t = w]$ with a learned field v_θ, whose approximation error is measured by the marginal FM risk

$$\mathcal{R}_{\mathrm{FM}}(\theta) = \mathbb{E}\|v_\theta(W_T, T) - v(W_T, T)\|_2^2. \tag{3}$$

where $T \sim \mathrm{Unif}[0,1]$ independent of X, Z and the above expectation is taken over T, X, Z. Since v is intractable, we replace v by V, resulting in the population training loss

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}\,\|v_\theta(W_T, T) - V\|_2^2\,. \tag{4}$$

Minimizing $\mathcal{L}_{\mathrm{FM}}$ is equivalent to minimizing $\mathcal{R}_{\mathrm{FM}}$ since $\mathcal{L}_{\mathrm{FM}}(\theta) = \mathcal{R}_{\mathrm{FM}}(\theta) + C$ where $C = \mathbb{E}\|V - v(W_T, T)\|_2^2$ does not depend on θ (Lipman et al., 2023). Once trained, generation integrates $\dot{w}_t = v_\theta(w_t, t)$ from $t = 1$ to $t = 0$, using a numerical ODE solver.

2.2 MeanFlow and improved MeanFlow

MeanFlow (Geng et al., 2026a) instead parameterizes the *average velocity* $u(w_t, r, t) = \frac{1}{t-r}\int_r^t v(w_\tau, \tau)\, d\tau$. Fundamental calculus yields the so-called MeanFlow identity

$$u(w_t, r, t) = v(w_t, t) - (t - r)\left(\partial_t u + (\partial_w u)v(w_t, t)\right). \tag{5}$$

Based on this identity, the original MeanFlow objective regresses an average velocity u_θ to the right hand side of (5), and since $\partial_t u$ and $\partial_w u$ are unknown, it replaces those by $\partial_t u_\theta$ and $\partial_w u_\theta$ to get

$$\mathcal{R}_{\mathrm{MF}}(\theta) = \mathbb{E}\,\|u_\theta(W_T, R, T) - [v(W_T, T) - (T - R)\left(\partial_t u_\theta + (\partial_w u_\theta)v(W_T, T)\right)]\|_2^2, \tag{6}$$

where the expectation is taken over R, T, Z, X (recall $W_t = (1-t)X + tZ$). We omit the `stop-gradient` operator when defining population risks, since they leave objective values unchanged. Replacing v by V gives a tractable training loss,

$$\mathcal{L}_{\mathrm{MF}}(\theta) = \mathbb{E}\,\|u_\theta(W_T, R, T) - [V - (T - R)\left(\partial_t u_\theta + (\partial_w u_\theta)V\right)]\|_2^2. \tag{7}$$

Unlike flow matching, the gap $\mathcal{L}_{\mathrm{MF}}(\theta) - \mathcal{R}_{\mathrm{MF}}(\theta)$ generally depends on θ, making them inequivalent as scalar objectives[1]. Moreover, the training objective (7) is not standard since its effective target depends on the network itself. To address this, *improved MeanFlow (iMF)* (Geng et al., 2026b) replaces the interpolation velocity in $(\partial_w u_\theta)V$ in $\mathcal{L}_{\mathrm{MF}}$ by a learnable marginal velocity v_θ, defining

$$\mathcal{L}_{\mathrm{iMF}}(\theta) = \mathbb{E}\,\|u_\theta(W_T, R, T) + (T - R)[\partial_t u_\theta + (\partial_w u_\theta)v_\theta(W_T, T)] - V\|_2^2. \tag{8}$$

The corresponding marginal iMF risk is then

$$\mathcal{R}_{\mathrm{iMF}}(\theta) = \mathbb{E}\,\|u_\theta(W_T, R, T) + (T - R)[\partial_t u_\theta + (\partial_w u_\theta)v_\theta(W_T, T)] - v(W_T, T)\|_2^2 \tag{9}$$

Importantly, $\mathcal{L}_{\mathrm{iMF}}(\theta) - \mathcal{R}_{\mathrm{iMF}}(\theta)$ is independent of θ (Appendix A.1), recovering the equivalence between these objectives. $v_\theta(w_t, t)$ can be represented by the boundary value $u_\theta(w_t, t, t)$ or an auxiliary velocity head. In this work, we consider the latter, and the auxiliary head is explicitly supervised through FM loss (4), $\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}\,\|v_\theta(W_T, T) - V\|_2^2$, yielding the joint training loss

$$\mathcal{L}(\theta) = \mathcal{L}_{\mathrm{iMF}}(\theta) + \lambda\mathcal{L}_{\mathrm{FM}}(\theta), \qquad \lambda > 0. \tag{10}$$

We also denote the joint risk

$$\mathcal{R}(\theta) = \mathcal{R}_{\mathrm{iMF}}(\theta) + \lambda\mathcal{R}_{\mathrm{FM}}(\theta). \tag{11}$$

The auxiliary velocity field v_θ is used only during training; sampling relies solely on u_θ, with the *one-step* map given by

$$S_\theta(z) = z - u_\theta(z, 0, 1), \qquad z \sim p_{\mathrm{src}}, \tag{12}$$

More generally, a fixed schedule $1 = t_K > t_{K-1} > \cdots > t_0 = 0$ defines a *few-step* map through

$$w_K = z, \qquad w_{k-1} = w_k - (t_k - t_{k-1})u_\theta(w_k, t_{k-1}, t_k), \quad k = K, \ldots, 1, \qquad S_\theta(z) = w_0. \tag{13}$$

3 Source-space sampling via improved MeanFlow

3.1 Bayesian inference through a learned prior transport

Let π_0 be a prior accessed through samples and ℓ_y a nonnegative likelihood or weight associated with observations or a task y. We aim to sample from the posterior $\pi^y(dx) \propto \ell_y(x)\pi_0(dx)$. The idea of source-space sampling (Purohit et al., 2025; Wang et al., 2026b) is to learn a transport map $(S_\theta)_\#\gamma \approx \pi_0$ where $\gamma = \mathcal{N}(0, I)$ and perform posterior sampling in the source space of γ, targeting the pull-back posterior $\nu_\theta^y(dz) \propto \ell_y(S_\theta(z))\gamma(dz)$ then transport these source samples to the data space via S_θ. In case the transport map is exact, i.e., $(S_\theta)_\#\gamma = \pi_0$, an exact posterior source sample $z \sim \nu_\theta^y(dz)$ will give an exact sample, $S_\theta(z)$, from $\pi^y(dx)$ (Wang et al., 2026b, Thm. 1). It is then critical to learn a good prior transport S_θ and perform efficient sampling in the source space of S_θ.

When S_θ is not perfect, we seek to quantify the resulting posterior error. Let S denote the exact transport, i.e., $S_\#\gamma = \pi_0$, associated with the generative construction, assumed to exist under suitable regularity conditions. Define the ideal source posterior by $\nu^y(dz) \propto \ell_y(S(z))\gamma(dz)$. Our goal is to control the source-space discrepancy $W_2(\nu_\theta^y, \nu^y)$ and the data-space discrepancy $W_2((S_\theta)_\#\nu_\theta^y, \pi^y)$ in terms of the *training error associated with learning* S_θ (for iMF, it links to $\mathcal{R}(\theta)$, see next section). Here W_2 denotes the Wasserstein distance (Villani et al., 2009).

[1]The `stop-gradient` makes the two objectives share expected gradients (Lu & Wang, 2026), but this is an optimization device; our theory concerns objective values, not training dynamics.

3.2 IMPROVED MEANFLOW SOURCE AND ITS POSTERIOR STABILITY

We parameterize S_θ using improved MeanFlow (12) and (13), trained with the joint training loss $\mathcal{L}(\theta) = \mathcal{L}_{\mathrm{iMF}}(\theta) + \lambda \mathcal{L}_{\mathrm{FM}}(\theta)$ in (10). iMF enables few-step generation and ensures that $\mathcal{L}$ differs from the marginal risk $\mathcal{R}$ (11) only by an additive constant independent of θ. Consequently, for a fixed model class Θ,

$$\mathcal{R}(\theta) = \underbrace{\mathcal{L}(\theta) - \inf_{\vartheta \in \Theta} \mathcal{L}(\vartheta)}_{:=\epsilon_{\mathrm{opt}},\ \text{population training suboptimality}} + \underbrace{\inf_{\vartheta \in \Theta} \mathcal{R}(\vartheta)}_{:=\epsilon_{\mathrm{app}},\ \text{model-class approximation error}} := \delta. \tag{14}$$

See App. A.2 for proof. The decomposition (14) shows that the marginal risk $\mathcal{R}(\theta)$ is controlled by population suboptimality of the training objective $\mathcal{L}$ used in practice and model-class approximation error. By the MeanFlow identity (5), the exact instantaneous and interval-average velocity fields jointly achieve zero marginal risk (Geng et al., 2026b), so a sufficiently expressive model class can in principle make ϵ_{app} small. We therefore seek to bound posterior error in terms of $\mathcal{R}(\theta)$.

We also note that, for original MeanFlow (Geng et al., 2026a), the parameter-dependent gap $\mathcal{L}_{\mathrm{MF}}(\theta) - \mathcal{R}_{\mathrm{MF}}(\theta)$ prevents the decomposition (14) in general. Small training suboptimality and approximation error therefore need not imply small marginal risk.

We establish posterior stability by first controlling the error of the implemented iMF transport, building on the flow-map error analysis of (Boffi et al., 2024). For the analysis, we use independent coupling $X \sim \pi_0, Z \sim \gamma$ and time $R, T \overset{\text{iid}}{\sim} \mathrm{Unif}[0,1]$[2]. Also, let $V = Z - X$, $W_t = (1-t)X + tZ$, $v(w,t) = \mathbb{E}[V | W_t = w]$, and $p_t = \mathrm{Law}(W_t)$. Thus, $p_0 = \pi_0$ and $p_1 = \gamma$.

Assumption 3.1 (Prior and marginal flow)**.** *The prior $\pi_0 \in \mathcal{P}_2(\mathbb{R}^d)$. For each $t \in [0,1]$ and p_t-a.e. w, the ODE $\frac{d}{ds}\xi_s^{t,w} = v(\xi_s^{t,w}, s)$, $\xi_t^{t,w} = w$, admits absolutely continuous solutions on $[0,t]$ whose solution maps $\varphi_{s,t}(w) := \xi_s^{t,w}$ form a measurable flow satisfying $(\varphi_{s,t})_{\#} p_t = p_s, 0 \le s \le t \le 1$.*

Under Assumption 3.1, the map $S := \varphi_{0,1}$ is an exact prior transport: $S_{\#}\gamma = \pi_0$.

Assumption 3.2 (Regularity of the learned average velocity)**.** *The field u_θ is continuously differentiable on $\mathbb{R}^d \times (0,1)^2$, and both u_θ and its first derivatives extend continuously to $\mathbb{R}^d \times [0,1]^2$.*

For $r, t \in [0,1]$, define the marginal iMF regression residual

$$\bar{q}_\theta(w,r,t) := u_\theta(w,r,t) - v(w,t) + (t-r)\left[\partial_t u_\theta(w,r,t) + \partial_w u_\theta(w,r,t)\, v_\theta(w,t)\right]$$
$$G_\theta(r) := \int_r^1 \|\bar{q}_\theta(\cdot,r,t)\|^2_{L^2(p_t)}\, dt.$$

Thus, the iMF risk in (9) is rewritten as $\mathcal{R}_{\mathrm{iMF}}(\theta) = \mathbb{E}\|\bar{q}_\theta(W_T, R, T)\|^2$. We impose the following.

Assumption 3.3 (Regularity across starting times)**.** *$G_\theta : [0,1] \to [0,\infty)$ is L_G-Lipschitz, i.e.,*

$$|G_\theta(r) - G_\theta(r')| \le L_G |r - r'|, \qquad \forall r, r' \in [0,1].$$

Remark 3.4. Assumption 3.3 holds if $\bar{q}_\theta$ is continuously differentiable in r and both $\bar{q}_\theta$ and $\partial_r \bar{q}_\theta$ are uniformly bounded in $L^2(p_t)$ over $0 \le r \le t \le 1$. Under the finite-second-moment assumption above, these conditions hold for finite affine-SiLU networks with direct spatial and time inputs; the resulting constants may depend on the network parameters. See App. A.3 for details and proof.

Fix a generation schedule $1 = t_K > t_{K-1} > \cdots > t_0 = 0$, put $h_k = t_k - t_{k-1}$.

Assumption 3.5.

$$\sum_{k=1}^{K} \int_{t_{k-1}}^{t_k} (t - t_{k-1})^2 \left\| (\partial_w u_\theta)(\cdot, t_{k-1}, t)(v - v_\theta)(\cdot, t) \right\|^2_{L^2(p_t)} dt \le B_u^2 \mathcal{R}_{\mathrm{FM}}(\theta). \tag{15}$$

where $\mathcal{R}_{\mathrm{FM}}$ is the marginal flow-matching risk (3) of the auxiliary velocity head.

Remark 3.6. A sufficient condition for (15) is $\sup_{w,\, 0 \le r < t \le 1} (t-r) \|\partial_w u_\theta(w,r,t)\|_{\mathrm{op}} \le B_u$. In particular, if u_θ is L_u-Lipschitz in w, uniformly over (r,t) (e.g., affine–SiLU networks), then this condition holds with $B_u = L_u$, since $t - r \le 1$. The weighted bound in Assumption 3.5 is weaker: it only controls the Jacobian acting on $v - v_\theta$ in the $L^2(p_t)$ sense. See Appendix A.4 for more details.

[2] A more general schedule is possible, but we use a simpler one for clarity of presentation.

Theorem 3.7. *Let $F_{k,\theta}(w) = w - h_k u_\theta(w, t_{k-1}, t_k)$ be L_k-Lipschitz, and let $S_\theta = F_{1,\theta} \circ \cdots \circ F_{K,\theta}$ be the few-step iMF map (13). Define $\Lambda_k := \prod_{j<k} L_j$, $A_K := \sum_{k=1}^K \Lambda_k \sqrt{h_k}$, $B_K := \left(\sum_{k=1}^K \Lambda_k^2 h_k\right)^{1/2}$, $\Lambda_1 = 1$. Under Assumptions 3.1, 3.2, 3.3 and 3.5, if $\delta \le L_G/2$,*

$$\mathbb{E}_\gamma \|S_\theta(Z) - S(Z)\|^2 \le \left[A_K (2L_G\delta)^{1/4} + \frac{B_u B_K}{\sqrt{\lambda}} \delta^{1/2}\right]^2. \tag{16}$$

If u_θ is uniformly L_u-Lipschitz in space, then $\Lambda_k \le e^{L_u}$, $A_K \le e^{L_u}\sqrt{K}$ and $B_K \le e^{L_u}$. Proof of Theorem 3.7 is given in Appendix A.5. The bound in (16) links transport error to population training suboptimality and model-class approximation error through δ in (14). The bound vanishes when $\delta = 0$, *without* requiring the step sizes h_k to tend to zero. This shows the few-step nature of the result. In particular, $K = 1$ gives $A_1 = B_1 = 1$ and a transport guarantee for the one-step map (12), with no Lipschitz condition on S_θ needed in (16). We support Theorem 3.7 with an experiment in Appendix B.1, showing that decreasing δ is accompanied by decreasing transport-map error $\mathbb{E}_\gamma \|S_\theta(Z) - S(Z)\|^2$ for both one- and six-step generation. Define terms in Theorem 3.7:

$$\mathcal{E} := \mathbb{E}_{Z\sim\gamma} \|S_\theta(Z) - S(Z)\|^2, \qquad b_K(\delta) := A_K (2L_G\delta)^{1/4} + \frac{B_u B_K}{\sqrt{\lambda}} \delta^{1/2}. \tag{17}$$

We now connect the posterior error to the transport error. Fix y, we make the following assumption.

Assumption 3.8 (Weight regularity)**.** *There exist finite constants $M_y > 0$ and $L_y \ge 0$ such that*

$$0 \le \ell_y(x) \le M_y, \qquad |\ell_y(x) - \ell_y(x')| \le L_y \|x - x'\|, \qquad \forall x, x' \in \mathbb{R}^d.$$

Moreover, $Z_y := \int \ell_y(x)\, \pi_0(dx) > 0$.

With the notation of Section 3.1, the normalized source posteriors are

$$\nu^y(dz) = \frac{\ell_y(S(z))}{Z_y}\gamma(dz), \qquad \nu^y_\theta(dz) = \frac{\ell_y(S_\theta(z))}{Z_{\theta,y}}\gamma(dz), \quad Z_{\theta,y} := \int \ell_y(S_\theta(z))\, \gamma(dz),$$

where ν^y_θ is defined whenever $Z_{\theta,y} > 0$. Here $S_\# \nu^y = \pi^y$. When $\ell_y > 0$, write $\Phi_y = -\log \ell_y$.

Theorem 3.9. *Under Assumption 3.8, if $L_y \mathcal{E}^{1/2} \le Z_y/2$, then $Z_{\theta,y} \ge Z_y/2 > 0$ and*

$$W_2(\nu^y_\theta, \nu^y) \le C_y(d)^{1/2} \mathcal{E}^{1/4}, \qquad C_y(d) := \frac{4L_y}{Z_y}\left[\sqrt{d^2 + 4d + 1} + \frac{M_y(1+d)}{Z_y}\right]. \tag{18}$$

If, additionally, S_θ is L_{S_θ}-Lipschitz, then $W_2\big((S_\theta)_\# \nu^y_\theta, \pi^y\big) \le L_{S_\theta} C_y(d)^{1/2} \mathcal{E}^{1/4} + \sqrt{\frac{M_y}{Z_y}} \mathcal{E}^{1/2}$.

Proof is in Appendix A.6. Combining these results, we derive the following.

Corollary 3.10. *Suppose the hypotheses of Theorem 3.7 and Assumption 3.8 hold. Let L_{S_θ} be a Lipschitz constant of S_θ, and let $C_y(d)$ be as in Theorem 3.9. If $\delta \le L_G/2, L_y b_K(\delta) \le Z_y/2$, then $Z_{\theta,y} \ge Z_y/2 > 0$ and*

$$W_2(\nu^y_\theta, \nu^y) \le C_y(d)^{1/2} b_K(\delta)^{1/2}, \tag{19}$$

$$W_2\big((S_\theta)_\# \nu^y_\theta, \pi^y\big) \le L_{S_\theta} C_y(d)^{1/2} b_K(\delta)^{1/2} + \sqrt{\frac{M_y}{Z_y}} b_K(\delta). \tag{20}$$

3.3 Source parallel tempering with improved MeanFlow

Given the learned iMF, we aim to sample from ν^y_θ. Although the source reference is Gaussian, the pulled-back likelihood can induce multimodality, and the source retains the data dimension. We use source parallel tempering (Wang et al., 2026a) to improve mixing between modes, together with local kernels like pCN and Gaussian-reference split HMC to address high dimensionality. Both admit formulations directly on infinite-dimensional Hilbert spaces, motivating their use for high-dimensional Gaussian-reference targets (Beskos et al., 2008; Cotter et al., 2013; Beskos et al., 2011).

Parallel tempering (Geyer, 1991) runs replicas at inverse temperatures $0 = \beta_0 < \cdots < \beta_J = 1$, targeting $\nu^y_{\theta,\beta}(\mathrm{d}z) \propto \ell_y(S_\theta(z))^\beta \gamma(\mathrm{d}z)$. Local MCMC kernels explore each target, while Metropolis swaps between adjacent replicas promote transitions between modes. Each sampling sweep locally updates every replica once, then attempts one Metropolis swap per adjacent temperature pair. The coldest replica ($\beta = 1$) targets ν^y_θ. The two variants below differ in their local MCMC kernels.

iMF+SPT+pCN. We use pCN at every replica as in (Wang et al., 2026a). Writing $\Phi_y = -\log \ell_y$ and $\Psi_\theta = \Phi_y \circ S_\theta$, the proposal is $z' = \cos(h_\beta)z + \sin(h_\beta)\xi$, where $\xi \sim \gamma$ and $h_\beta \in (0, \pi/2]$, with acceptance probability $1 \wedge \exp\{-\beta[\Psi_\theta(z') - \Psi_\theta(z)]\}$. At $\beta = 0$, every proposal is accepted; and choosing $h_0 = \pi/2$ gives an independent Gaussian draw $z' = \xi \sim \gamma$.

iMF+SPT+hybrid. When Ψ_θ is differentiable, we replace the *coldest* pCN update ($\beta = 1$) with split HMC (Beskos et al., 2011) to exploit local posterior geometry, while retaining pCN at hotter replicas to limit computational cost. Refresh $p \sim \gamma$ and define the Hamiltonian $H(z,p) = \frac{1}{2}\|z\|^2 + \frac{1}{2}\|p\|^2 + \Psi_\theta(z)$. Starting from $(z_0, p_0) = (z, p)$, apply L steps of Strang's splitting:

$$\begin{aligned} p_{\ell+1/2} &= p_\ell - \tfrac{h}{2}\nabla\Psi_\theta(z_\ell), \quad \triangleright \text{ half momentum kick} \\ \begin{pmatrix} z_{\ell+1} \\ \widetilde{p}_{\ell+1} \end{pmatrix} &= \begin{pmatrix} \cos(h)I & \sin(h)I \\ -\sin(h)I & \cos(h)I \end{pmatrix} \begin{pmatrix} z_\ell \\ p_{\ell+1/2} \end{pmatrix}, \quad \triangleright \text{ exact harmonic rotation} \\ p_{\ell+1} &= \widetilde{p}_{\ell+1} - \tfrac{h}{2}\nabla\Psi_\theta(z_{\ell+1}), \quad \triangleright \text{ half momentum kick} \end{aligned} \tag{21}$$

Accept $(z', p') = (z_L, -p_L)$ with probability $1 \wedge \exp\{H(z,p) - H(z',p')\}$; otherwise retain z.

Sampler geometric convergence. Wang et al. (2026a) establish geometric W_2 convergence to the learned posterior for SPT with all pCN updates. We extend this guarantee to SPT+hybrid. Let $\nu_n = \mathrm{Law}(z_n)$, where z_n is the coldest replica's state after n sampling sweeps. Under the conditions in Appendix A.7, including a source weight bounded above and away from zero, the cold marginal of either sampler satisfies $W_2(\nu_n, \nu_\theta^y) \le C\rho^n$ for some $C < \infty$ and $\rho \in (0,1)$. Together with posterior stability (Theorem 3.9, Corollary 3.10), this yields an end-to-end error bound with a geometrically decaying sampling term and a transport term that vanishes with δ; see Corollary A.3.

4 Experiments

4.1 Synthetic experiments

Problem setting. We consider two priors on $x \in \mathbb{R}^{32}$, each formed from 16 independent two-dimensional blocks. For $u_j, v_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, the priors are

$$x_{2j-1} = u_j, \qquad x_{2j} = \begin{cases} 0.62(u_j^2 - 1) + 0.30v_j, & \text{Banana}, \\ 1.05\sin(1.65u_j) + 0.15v_j, & \text{Sine}, \end{cases} \qquad j = 1, \dots, 16.$$

Both problems use linear observations with independent Gaussian noise, giving the likelihood

$$p(y \mid x) = \prod_{j=1}^{16} \mathcal{N}\big(y_j;\, x_{2j} + 0.35x_{2j-1},\, 0.2^2\big)\,.$$

We condition on the fixed observations $y_1 = y_2 = 2$ and $y_j = -1$ for $j = 3, \dots, 16$.

Prior learning and sampling. We evaluate iMF+SPT+pCN and iMF+SPT+hybrid against two baselines: FM+SPT+pCN and diffusion posterior sampling (DPS) (Chung et al., 2023). The two iMF methods share a six-step transport. The FM baseline uses a 100-step RK4 transport. For each problem, the iMF, FM, and diffusion prior models are trained on the same fixed dataset of 20,000 prior samples. Each method returns 6,000 samples. Figure 4 compares samples from the learned priors with those from the true banana and sine priors. See Appendix B.2 for more details.

Results. The blockwise structure of the problems enables accurate numerical posterior references for assessing approximation quality. Table 1 summarizes the results, and Figures 5 and 6 compare posterior sample histograms from each method with numerical references for the true banana and sine. Using the same learned iMF transport, the hybrid sampler increases mean coordinate ESS (effective sample size) by 3.25× on banana and 4.46× on sine, and coordinate ESS per second by 2.39× and 3.09×, respectively, compared with iMF+SPT+pCN. It also achieves the lowest mean sliced Wasserstein and binned TV errors, although the differences in sliced Wasserstein error are small relative to the variability across seeds. With its configured 100-step RK4 transport,

Table 1: Synthetic posterior sampling. See Appendix B.2 for more details.

Method	$SW_2 \downarrow$	Bin TV $\downarrow$	Coord. ESS $\uparrow$	Coord. ESS/s $\uparrow$	Time (s) $\downarrow$
Banana					
DPS	0.462 ± 0.071	0.311 ± 0.002	—	—	$\mathbf{40.8 \pm 2.9}$
FM+SPT+pCN	0.219 ± 0.081	0.106 ± 0.008	373 ± 14	0.0505 ± 0.0025	7388 ± 214
iMF+SPT+pCN	0.154 ± 0.012	0.109 ± 0.006	345 ± 5	2.24 ± 0.25	155.4 ± 16.8
iMF+SPT+hybrid	$\mathbf{0.142 \pm 0.031}$	$\mathbf{0.085 \pm 0.005}$	$\mathbf{1123 \pm 51}$	$\mathbf{5.37 \pm 0.26}$	209.2 ± 1.0
Sine					
DPS	0.931 ± 0.025	0.286 ± 0.001	—	—	$\mathbf{39.1 \pm 0.1}$
FM+SPT+pCN	0.104 ± 0.021	0.129 ± 0.014	394 ± 14	0.0554 ± 0.0020	7099 ± 47
iMF+SPT+pCN	0.093 ± 0.016	0.131 ± 0.010	376 ± 36	2.54 ± 0.32	148.8 ± 4.7
iMF+SPT+hybrid	$\mathbf{0.088 \pm 0.011}$	$\mathbf{0.127 \pm 0.004}$	$\mathbf{1678 \pm 47}$	$\mathbf{7.83 \pm 0.29}$	214.3 ± 2.9

FM+SPT+pCN requires $35.32\times$ and $33.13\times$ the sampling time of the iMF hybrid on banana and sine, respectively. DPS exhibits substantially larger distributional errors, with systematic deviations from the reference posterior visible in the histograms.

4.2 CLIP-guided ImageNet generation

We evaluate CLIP-guided 256×256 class-conditional ImageNet generation: for a fixed ImageNet class (e.g., bald eagle) and text prompt, the goal is to sample images from that class that favor high image-text alignment. We use the cosine similarity from a frozen pretrained CLIP model (Radford et al., 2021) as the guidance reward. Let $z \sim \mathcal{N}(0, I)$ and let $G_c(z)$ denote a frozen pretrained transport from the Gaussian source distribution to images in class c. For a text prompt p, the learned source posterior is $\nu_\theta^p(z) \propto \exp\big(-\frac{1}{2}\|z\|^2 + \tau \mathrm{CLIP}(G_c(z), p)\big)$ with $\tau = 200$.

Comparison We use the pretrained iMF for class-conditional ImageNet generation (Geng et al., 2026b). We compare iMF+SPT+pCN and iMF+SPT+hybrid against their multi-step counterpart SiT+SPT+pCN, where SiT (Ma et al., 2024) is a pretrained class-conditional flow-matching prior integrated using 125 Heun steps. These methods use eight temperature replicas and eight independent ladders. For iMF-based methods, we run 250 adaptation sweeps, 250 burn-in sweeps, and retain 1,000 post-burn-in samples per ladder, yielding 8,000 retained samples per method. SiT+SPT+pCN is extremely slow in this experiment, we evaluate it under a matched per-prompt sampling-time budget equal to the measured hybrid sampling time (more than 10 hours), with an additional one-hour allowance for initialization, scoring, and image export. We also include best-of-K (Karthik et al., 2023), $K = 12$, as a reward-based selection baseline, retaining the image with the highest CLIP reward among K learned-prior independent samples. We produce $8,000$ best-of-K images.

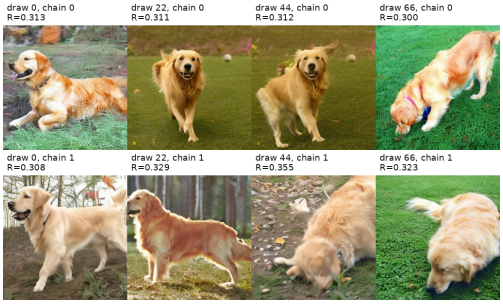
Results. Table 2 reports CLIP rewards and sampling costs for two class-prompt pairs, while Figures 1 and 2 show samples (R is the reward) from two ladders per method. The two iMF methods achieve comparable rewards and both outperform SiT+SPT+pCN. Their samples more faithfully depict a *running* golden retriever amid *fallen autumn leaves*, whereas SiT+SPT+pCN often misses the autumn setting within its sampling budget of 10-11 hours. For the bald eagle prompt, all three methods depict eagles with *outstretched wings*, but the iMF methods more consistently include *snow-capped mountains* whereas SiT+SPT+pCN depicts the eagle against the plain blue sky. In these examples, the hybrid sampler offers no clear improvement in reward or sample quality over iMF+SPT+pCN. Finally, best-of-K achieves the lowest rewards, suggesting guidance is more effective than independent prior draws. The full results are given in Appendix B.3.2.

5 Conclusion

We develop few-step iMF source-space inference, linking population training suboptimality and approximation error to posterior stability and end-to-end error bounds under stated assumptions. Hybrid SPT improves efficiency on synthetic benchmarks, while one-step iMF reduces CLIP-guided ImageNet sampling costs relative to multi-step SiT. Quantifying how model-class approximation error ϵ_{app} decreases with capacity while controlling the regularity constants remains open.

Table 2: CLIP reward and measured sampling cost.

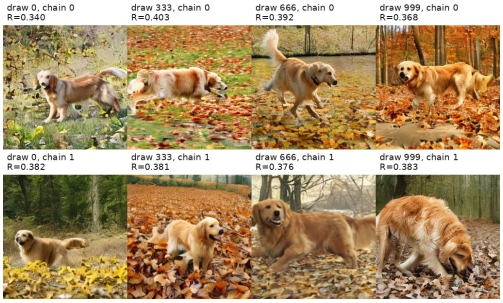
Method	Mean CLIP reward ↑	Sampling time (h)	Mean sweep time (s)
Golden retriever: *"A photograph of a golden retriever running through fallen autumn leaves."*			
Best-of-K	0.2864	2.49	–
SiT+SPT+pCN	0.3122	10.51	370.89
iMF+SPT+pCN	0.3809	2.40	5.75
iMF+SPT+hybrid	**0.3824**	10.95	26.28
Bald eagle: *"A photograph of a bald eagle soaring above snow-capped mountains, wings spread wide."*			
Best-of-K	0.3278	2.49	–
SiT+SPT+pCN	0.3481	10.70	370.27
iMF+SPT+pCN	**0.3936**	2.44	5.85
iMF+SPT+hybrid	0.3914	11.15	26.76



(a) SiT + SPT + pCN.

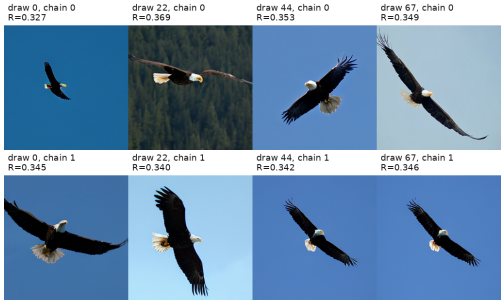


(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

Figure 1: Class: golden retriever, prompt: *"A photograph of a golden retriever running through fallen autumn leaves."*



(a) SiT + SPT + pCN.



(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

Figure 2: Class: bald eagle, prompt: *"A photograph of a bald eagle soaring above snow-capped mountains, wings spread wide."*

AI USE STATEMENT

We used generative AI tools to assist with language editing; literature search and paper summarization; algorithm implementation, comparative experiments, and code testing; preliminary (unus-

able) drafts and templates; and developing and checking mathematical arguments; and reviewing the manuscript. We did not use generative AI to generate synthetic datasets. The remaining disclosure categories are not applicable to this work.

The authors reviewed the AI-assisted material, including text, code, and mathematical arguments. We take full responsibility for the final manuscript, its claims, and all accompanying artifacts, including those produced with generative AI assistance.

Reproducibility statement

We provide code for the iMF methods and SiT+SPT+pCN as supplementary material on OpenReview and will make it publicly available. All theoretical results are stated with explicit assumptions, with complete proofs provided in the appendix.

References

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. *arXiv preprint arXiv:2402.14017*, 2024.

Alexandros Beskos, Gareth Roberts, Andrew Stuart, and Jochen Voss. Mcmc methods for diffusion bridges. *Stochastics and Dynamics*, 8(03):319–350, 2008.

Alexandros Beskos, Frank J Pinski, Jesús Marıa Sanz-Serna, and Andrew M Stuart. Hybrid monte carlo on hilbert spaces. *Stochastic Processes and their Applications*, 121(10):2201–2230, 2011.

MJ Betancourt, Simon Byrne, and Mark Girolami. Optimizing the integrator step size for hamiltonian monte carlo. *arXiv preprint arXiv:1411.6669*, 2014.

Pier Giovanni Bissiri, Chris C Holmes, and Stephen G Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78 (5):1103–1130, 2016.

Nicholas M Boffi, Michael S Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. *arXiv preprint arXiv:2406.07507*, 2024.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Yuxin Chen, David Keyes, Kody JH Law, and Hatem Ltaief. Accelerated dimension-independent adaptive metropolis. *SIAM Journal on Scientific Computing*, 38(5):S539–S565, 2016.

Zilan Cheng, Li-Lian Wang, and Zhongjian Wang. Preconditioned one-step generative modeling for bayesian inverse problems in function spaces. *arXiv preprint arXiv:2603.14798*, 2026.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

Simon L Cotter, Gareth O Roberts, Andrew M Stuart, and David White. Mcmc methods for functions: modifying old algorithms to make them faster. *Statistical Science*, pp. 424–446, 2013.

David J Earl and Michael W Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. *Advances in Neural Information Processing Systems*, 38:75460–75482, 2026a.

Zhengyang Geng, Yiyang Lu, Zongze Wu, Eli Shechtman, J Zico Kolter, and Kaiming He. Improved mean flows: On the challenges of fastforward generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30467–30476, 2026b.

Charles J Geyer. Markov chain monte carlo maximum likelihood. 1991.

David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In *International conference on machine learning*, pp. 2404–2414. PMLR, 2019.

Ernst Hairer, Gerhard Wanner, and Syvert P Nørsett. *Solving ordinary differential equations I: Nonstiff problems*. Springer, 1993.

Matthew Hoffman, Pavel Sountsov, Joshua V Dillon, Ian Langmore, Dustin Tran, and Srinivas Vasudevan. Neutra-lizing bad geometry in hamiltonian monte carlo using neural transport. *arXiv preprint arXiv:1903.03704*, 2019.

Matthew D Hoffman, Andrew Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

Shyamgopal Karthik, Karsten Roth, Massimiliano Mancini, and Zeynep Akata. If at first you don't succeed, try, try again: Faithful diffusion-based text-to-image generation by selection. *arXiv preprint arXiv:2305.13308*, 2023.

Tomasz Korbak, Ethan Perez, and Christopher Buckley. Rl with kl penalties is better viewed as bayesian inference. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 1083–1091, 2022.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

Juanwu Lu and Ziran Wang. On variance reduction in learning mean flows. *arXiv preprint arXiv:2605.09235*, 2026.

Sebastian Lunz, Ozan Öktem, and Carola-Bibiane Schönlieb. Adversarial regularizers in inverse problems. *Advances in neural information processing systems*, 31, 2018.

Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, pp. 23–40. Springer, 2024.

Felix Otto and Cédric Villani. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

Matthew D Parno and Youssef M Marzouk. Transport map accelerated markov chain monte carlo. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):645–682, 2018.

Dhruv Patel and Assad A Oberai. Bayesian inference with generative adversarial network priors. *arXiv preprint arXiv:1907.09987*, 2019.

Dhruv V Patel, Deep Ray, and Assad A Oberai. Solution of physics-based bayesian inverse problems with deep generative priors. *Computer Methods in Applied Mechanics and Engineering*, 400: 115428, 2022.

Vishal Purohit, Matthew Repasky, Jianfeng Lu, Qiang Qiu, Yao Xie, and Xiuyuan Cheng. Consistency posterior sampling for diverse image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28327–28336, June 2025.

Stefan T Radev, Marvin Schmitt, Lukas Schumacher, Lasse Elsemüller, Valentin Pratz, Yannik Schälte, Ullrich Köthe, and Paul-Christian Bürkner. Bayesflow: Amortized bayesian workflows with neural networks. *arXiv preprint arXiv:2306.16015*, 2023.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Zaccharie Ramzi, Benjamin Remy, Francois Lanusse, Jean-Luc Starck, and Philippe Ciuciu. Denoising score-matching for uncertainty quantification in inverse problems. *arXiv preprint arXiv:2011.08698*, 2020.

Andrew M Stuart. Inverse problems: a bayesian perspective. *Acta numerica*, 19:451–559, 2010.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

Shih-Hsin Wang, Joel Keller, Taos Transue, Drake Benjamin Brown, Thomas Strohmer, and Bao Wang. Test-time guidance for flow-based generative models via parallel tempering on source distributions. In *International Conference on Machine Learning*, 2026a.

Zifan Wang, Alice Harting, Matthieu Barreau, Michael Zavlanos, and Karl H Johansson. Source-guided flow matching. In *International Conference on Learning Representations*, volume 2026, pp. 106484–106514, 2026b.

Benjamin J. Zhang, Youssef M. Marzouk, and Konstantinos Spiliopoulos. Transport map unadjusted langevin algorithms: Learning and discretizing perturbed samplers. *Foundations of Data Science*, 7(3):705–736, 2025.

A Proofs and clarifications

A.1 Equivalence between $\mathcal{L}_{\mathrm{iMF}}$ and $\mathcal{R}_{\mathrm{iMF}}$

Write $\mathcal{L}_{\mathrm{iMF}}(\theta) = \mathbb{E}\|A_\theta - V\|_2^2$, where the predictor $A_\theta := u_\theta(W_T, R, T) + (T - R)\,[\partial_t u_\theta(W_T, R, T) + (\partial_w u_\theta)(W_T, R, T)\, v_\theta(W_T, T)]$ is measurable with respect to (W_T, R, T). Since (R, T) is independent of (X, Z),

$$\mathbb{E}[V \mid W_T, R, T] = v(W_T, T).$$

Hence

$$\mathbb{E}\langle A_\theta - v(W_T, T),\, V - v(W_T, T)\rangle = 0.$$

Expanding the squared norm therefore gives

$$\mathcal{L}_{\mathrm{iMF}}(\theta) = \mathcal{R}_{\mathrm{iMF}}(\theta) + \underbrace{\mathbb{E}\|V - v(W_T, T)\|_2^2}_{C},$$

where C is independent of θ.

A.2 Proof of relation (14)

Since $\mathcal{R}(\theta) = \mathcal{L}(\theta) + C$, it holds $\inf \mathcal{R} = \inf \mathcal{L} + C$. Therefore, $\mathcal{R}(\theta) = \mathcal{L}(\theta) - \inf \mathcal{L} + \inf \mathcal{R}$.

A.3 Details on Remark 3.4

Set

$$A := \sup_{0\le r\le t\le 1} \|\bar{q}_\theta(\cdot, r, t)\|_{L^2(p_t)}, \qquad B := \sup_{0\le r\le t\le 1} \|\partial_r \bar{q}_\theta(\cdot, r, t)\|_{L^2(p_t)}.$$

By hypothesis, $A, B < \infty$. For $0 \le r < s \le t \le 1$, the fundamental theorem of calculus and Minkowski's inequality give

$$\|\bar{q}_\theta(\cdot, r, t) - \bar{q}_\theta(\cdot, s, t)\|_{L^2(p_t)} \le \int_r^s \|\partial_r \bar{q}_\theta(\cdot, a, t)\|_{L^2(p_t)}\, da \le B(s - r).$$

Consequently,

$$\left| \|\bar{q}_\theta(\cdot, r, t)\|^2_{L^2(p_t)} - \|\bar{q}_\theta(\cdot, s, t)\|^2_{L^2(p_t)} \right| \le 2AB(s - r).$$

Splitting the integrals defining $G_\theta(r)$ and $G_\theta(s)$ therefore yields

$$\begin{aligned} |G_\theta(r) - G_\theta(s)| &\le \int_r^s \|\bar{q}_\theta(\cdot, r, t)\|^2_{L^2(p_t)}\, dt \\ &\quad + \int_s^1 \left| \|\bar{q}_\theta(\cdot, r, t)\|^2_{L^2(p_t)} - \|\bar{q}_\theta(\cdot, s, t)\|^2_{L^2(p_t)} \right| dt \\ &\le A^2(s - r) + 2AB(1 - s)(s - r) \\ &\le (A^2 + 2AB)|s - r|. \end{aligned}$$

Now suppose that both u_θ and v_θ are finite compositions of affine maps and coordinatewise SiLU activations, with direct spatial and time inputs. The SiLU activation $\rho(a) = a/(1 + e^{-a})$ is smooth and has bounded first and second derivatives. By the chain rule, these networks have bounded first and second input derivatives and at most linear growth. Since the time inputs lie in $[0, 1]$, there is a finite constant C_θ such that

$$\|u_\theta(w, r, t)\| + \|v_\theta(w, t)\| \le C_\theta(1 + \|w\|),$$

and the first and second input derivatives of u_θ are uniformly bounded.

Because v and v_θ do not depend on r,

$$\partial_r \bar{q}_\theta = \partial_r u_\theta - \partial_t u_\theta - (\partial_w u_\theta)v_\theta + (t - r)\left[\partial_{rt} u_\theta + (\partial_{rw} u_\theta)v_\theta\right].$$

The preceding network bounds imply, after enlarging C_θ if necessary,

$$\begin{aligned} \|\bar{q}_\theta(w, r, t)\| &\le C_\theta(1 + \|w\|) + \|v(w, t)\|, \\ \|\partial_r \bar{q}_\theta(w, r, t)\| &\le C_\theta(1 + \|w\|), \end{aligned}$$

uniformly over $0 \le r \le t \le 1$.

Let $M_2 := \mathbb{E}\|X\|^2 < \infty$ where $X \sim \pi_0$ (Assumption 3.1). Independence of X and $Z \sim \mathcal{N}(0, I_d)$ gives

$$\mathbb{E}\|W_t\|^2 = (1 - t)^2 M_2 + t^2 d \le M_2 + d.$$

Moreover, since $v(W_t, t) = \mathbb{E}[Z - X \mid W_t]$, conditional Jensen's inequality gives

$$\|v(\cdot, t)\|^2_{L^2(p_t)} \le \mathbb{E}\|Z - X\|^2 = M_2 + d.$$

These estimates establish $A, B < \infty$, proving the network claim. The resulting constants may depend on the network parameters.

A.4 Details on Remark 3.6

Write $e_\theta(w, t) := v(w, t) - v_\theta(w, t)$. The operator-norm bound implies, for $r < t$,

$$(t - r)\|\partial_w u_\theta(w, r, t)e_\theta(w, t)\| \le B_u\|e_\theta(w, t)\|.$$

Squaring, integrating against p_t, and summing over the schedule intervals gives

$$\sum_{k=1}^{K} \int_{t_{k-1}}^{t_k} (t - t_{k-1})^2 \|\partial_w u_\theta(\cdot, t_{k-1}, t) e_\theta(\cdot, t)\|^2_{L^2(p_t)}\, dt$$
$$\leq B_u^2 \sum_{k=1}^{K} \int_{t_{k-1}}^{t_k} \|e_\theta(\cdot, t)\|^2_{L^2(p_t)}\, dt = B_u^2 \mathcal{R}_{\mathrm{FM}}(\theta),$$

where the last equality uses $T \sim \mathrm{Unif}[0,1]$ and the fact that the schedule intervals partition $[0,1]$.

If u_θ is uniformly L_u-Lipschitz in w, then $\|\partial_w u_\theta(w,r,t)\|_{\mathrm{op}} \leq L_u$. Since $t - r \leq 1$, the sufficient condition follows with $B_u = L_u$.

Finally, (15) controls only the Jacobian acting on e_θ, averaged over space and time; it does not require a uniform operator-norm bound. For example, when $v_\theta = v$, its left-hand side vanishes.

For finite affine-SiLU networks with direct spatial and time inputs, the spatial Jacobian is uniformly bounded (similar to the argument in Appendix A.3):

$$L_u := \sup_{w \in \mathbb{R}^d,\, (r,t) \in [0,1]^2} \|\partial_w u_\theta(w,r,t)\|_{\mathrm{op}} < \infty.$$

A.5 Proof of Theorem 3.7

Population risk control:

By (14) and the definition of the joint marginal risk,

$$\mathcal{R}_{\mathrm{iMF}}(\theta) + \lambda \mathcal{R}_{\mathrm{FM}}(\theta) = \mathcal{R}(\theta) = \delta.$$

Therefore, $\mathcal{R}_{\mathrm{iMF}}(\theta) \leq \delta$ and $\mathcal{R}_{\mathrm{FM}}(\theta) \leq \delta/\lambda$.

Recovering fixed-time slices:

Independent uniform time sampling on $[0,1]^2$ implies

$$\int_0^1 G_\theta(r)\, dr = \int_0^1 \int_r^1 \|\bar{q}_\theta(\cdot, r, t)\|^2_{L^2(p_t)}\, dt\, dr \leq \int_0^1 \int_0^1 \|\bar{q}_\theta(\cdot, r, t)\|^2_{L^2(p_t)}\, dt\, dr = \mathcal{R}_{\mathrm{iMF}}(\theta).$$

Fix $r \in [0,1]$ and $h \in (0,1]$. Choose an interval $I = [a,b] \subset [0,1]$ such that $a \leq r \leq b$ and $b - a = h$. By Assumption 3.3,

$$G_\theta(r) \leq G_\theta(s) + L_G |s - r|, \qquad s \in I.$$

Averaging this inequality over I gives

$$G_\theta(r) \leq \frac{1}{h} \int_I G_\theta(s)\, ds + \frac{L_G}{h} \int_I |s - r|\, ds.$$

The first term is at most δ/h, since $G_\theta \geq 0$ and $\int_0^1 G_\theta(s)\, ds \leq \mathcal{R}_{\mathrm{iMF}}(\theta) \leq \delta$. For the second term, note that

$$\int_a^b |s - r|\, ds = \frac{(r-a)^2 + (b-r)^2}{2} \leq \frac{((r-a) + (b-r))^2}{2} = \frac{h^2}{2}.$$

Consequently,

$$G_\theta(r) \leq \frac{\delta}{h} + \frac{L_G h}{2}. \tag{22}$$

If $0 < \delta \leq L_G/2$, choosing $h = \sqrt{2\delta/L_G} \in (0,1]$ to minimize the bound in (22), we get

$$G_\theta(r) \leq (2 L_G \delta)^{1/2}, \qquad \forall r \in [0,1]. \tag{23}$$

If $\delta = 0$, the same bound follows by letting $h \downarrow 0$ in (22).

Local transport error:

Define

$$q_\theta(w,r,t) := u_\theta(w,r,t) - v(w,t) + (t-r)\left[\partial_t u_\theta(w,r,t) + \partial_w u_\theta(w,r,t)v(w,t)\right]$$
$$= \bar{q}_\theta(w,r,t) + (t-r)\partial_w u_\theta(w,r,t)\big(v(w,t) - v_\theta(w,t)\big).$$

Fix $0 \le r < t \le 1$ and, for p_t-almost every w, set (see Assumption 3.1 for the flow definition)

$$w_s := \varphi_{s,t}(w), \qquad D_s := (s-r)u_\theta(w_s,r,s) - w_s, \qquad s \in [r,t].$$

Assumptions 3.1 and 3.2 ensure that D_s is absolutely continuous. The chain rule gives $\frac{d}{ds}D_s = q_\theta(w_s,r,s)$ for almost every s. Since $D_r = -\varphi_{r,t}(w)$ and $D_t = (t-r)u_\theta(w,r,t) - w$, integration gives

$$(t-r)u_\theta(w,r,t) - w + \varphi_{r,t}(w) = \int_r^t q_\theta(\varphi_{s,t}(w),r,s)\,ds. \tag{24}$$

For the fixed generation schedule, let $F_k := \varphi_{t_{k-1},t_k}$ be the exact step map, and define

$$Q_k := \int_{t_{k-1}}^{t_k} \|\bar{q}_\theta(\cdot,t_{k-1},t)\|^2_{L^2(p_t)}\,dt,$$
$$J_{v,k} := \int_{t_{k-1}}^{t_k} (t-t_{k-1})^2\|\partial_w u_\theta(\cdot,t_{k-1},t)(v-v_\theta)(\cdot,t)\|^2_{L^2(p_t)}\,dt.$$

Equation (24) gives

$$F_{k,\theta}(w) - F_k(w) = -\int_{t_{k-1}}^{t_k} q_\theta(\varphi_{s,t_k}(w),t_{k-1},s)\,ds.$$

Using $(\varphi_{s,t_k})_\# p_{t_k} = p_s$, Minkowski's inequality, and Cauchy-Schwarz in time, we obtain

$$\begin{aligned}
\|F_{k,\theta} - F_k\|_{L^2(p_{t_k})} &\le \int_{t_{k-1}}^{t_k} \|q_\theta(\varphi_{s,t_k}(\cdot),t_{k-1},s)\|_{L^2(p_{t_k})}\,ds \\
&= \int_{t_{k-1}}^{t_k} \|q_\theta(\cdot,t_{k-1},s)\|_{L^2(p_s)}\,ds \\
&\le \int_{t_{k-1}}^{t_k} \|\bar{q}_\theta(\cdot,t_{k-1},s)\|_{L^2(p_s)}\,ds \\
&\quad + \int_{t_{k-1}}^{t_k} (s-t_{k-1})\|\partial_w u_\theta(\cdot,t_{k-1},s)(v-v_\theta)(\cdot,s)\|_{L^2(p_s)}\,ds \\
&\le \sqrt{h_k}\big(\sqrt{Q_k} + \sqrt{J_{v,k}}\big).
\end{aligned} \tag{25}$$

Composing the step maps:

The flow property gives $S = F_1 \circ \cdots \circ F_K = \varphi_{0,1}$. Define the intermediate compositions

$$H_k := F_{1,\theta} \circ \cdots \circ F_{k,\theta} \circ F_{k+1} \circ \cdots \circ F_K, \qquad k = 0,\dots,K,$$

where empty compositions are the identity. Thus $H_0 = S$, $H_K = S_\theta$, and

$$S_\theta(z) - S(z) = \sum_{k=1}^{K}\big(H_k(z) - H_{k-1}(z)\big).$$

The maps H_k and H_{k-1} differ only at the kth step. Their common inner composition is the exact flow $F_{k+1}\circ\cdots\circ F_K = \varphi_{t_k,1}$, while their common outer composition $F_{1,\theta}\circ\cdots\circ F_{k-1,\theta}$ has Lipschitz constant at most $\Lambda_k = \prod_{j<k} L_j$. Hence

$$\|H_k(z) - H_{k-1}(z)\| \le \Lambda_k\|F_{k,\theta}(\varphi_{t_k,1}(z)) - F_k(\varphi_{t_k,1}(z))\|.$$

Since $(\varphi_{t_k,1})_\#\gamma = p_{t_k}$, taking $L^2(\gamma)$ norms and applying the triangle inequality yields

$$\|S_\theta - S\|_{L^2(\gamma)} \le \sum_{k=1}^{K} \Lambda_k\|F_{k,\theta} - F_k\|_{L^2(p_{t_k})}.$$

Consequently,

$$\begin{aligned}\|S_\theta - S\|_{L^2(\gamma)} &\le \sum_{k=1}^{K} \Lambda_k \|F_{k,\theta} - F_k\|_{L^2(p_{t_k})} \\ &\le \sum_{k=1}^{K} \Lambda_k \sqrt{h_k}\big(\sqrt{Q_k} + \sqrt{J_{v,k}}\big) \\ &\le A_K (2L_G\delta)^{1/4} + B_K \left(\sum_{k=1}^{K} J_{v,k}\right)^{1/2} \\ &\le A_K (2L_G\delta)^{1/4} + \frac{B_u B_K}{\sqrt{\lambda}}\delta^{1/2}.\end{aligned}$$

The third line uses Cauchy–Schwarz over k for the $J_{v,k}$ terms. Squaring proves (16). For $K = 1$, we have $h_1 = \Lambda_1 = A_1 = B_1 = 1$, so the same argument gives the one-step transport bound. In this case there is no outer learned map in the composition estimate, and no Lipschitz condition on $F_{1,\theta}$ is needed.

Finally, if u_θ is uniformly L_u-Lipschitz in space, then

$$L_k \le 1 + h_k L_u, \qquad \Lambda_k \le \exp\left(L_u \sum_{j<k} h_j\right) \le e^{L_u}.$$

Since $\sum_k h_k = 1$, Cauchy–Schwarz yields $A_K \le e^{L_u}\sqrt{K}$ and $B_K \le e^{L_u}$. Also $L_{S_\theta} \le \prod_k L_k \le e^{L_u}$, as used in the data-space posterior bounds. All estimates concern the fixed generation schedule.

A.6 Proof of Theorem 3.9

Write $e := \mathcal{E}^{1/2}$, $a(z) := \ell_y(S_\theta(z))$ and $b(z) := \ell_y(S(z))$. The weight assumptions give

$$|Z_{\theta,y} - Z_y| \le \int |a - b|\, d\gamma \le L_y e.$$

Hence $L_y e \le Z_y/2$ ensures $Z_{\theta,y} \ge Z_y/2 > 0$. Let $U_2(z) = 1 + \|z\|^2$ and use the weighted-variation convention

$$\|\mu - \nu\|_{U_2} := \sup_{|f| \le U_2} \left| \int f\, d\mu - \int f\, d\nu \right|.$$

For each such f, decompose the normalized densities as

$$\int f\, d(\nu_\theta^y - \nu^y) = \frac{1}{Z_{\theta,y}} \int f(a - b)\, d\gamma + \frac{Z_y - Z_{\theta,y}}{Z_y Z_{\theta,y}} \int f b\, d\gamma.$$

Since $\mathbb{E}_\gamma U_2 = 1 + d$ and $\mathbb{E}_\gamma U_2^2 = d^2 + 4d + 1$, Cauchy-Schwarz gives

$$\begin{aligned}\|\nu_\theta^y - \nu^y\|_{U_2} &\le \frac{L_y e}{Z_{\theta,y}} \left[\sqrt{d^2 + 4d + 1} + \frac{M_y(1 + d)}{Z_y}\right] \\ &\le \frac{C_y(d)}{2} e.\end{aligned}$$

For completeness, couple the common part of two measures identically and couple their remaining positive parts. The inequality $\|z - z'\|^2 \le 2\|z\|^2 + 2\|z'\|^2$ then gives

$$W_2^2(\mu, \nu) \le 2 \int \|z\|^2\, d|\mu - \nu|(z) \le 2\|\mu - \nu\|_{U_2}.$$

This proves (18). All source second moments are finite because the normalized weights are bounded relative to γ.

For the data-space statement, first note that $S_\#\gamma = \pi_0$ implies $S_\#\nu^y = \pi^y$ by the definition of a pushforward. Insert the intermediate measure $(S_\theta)_\#\nu^y$. Lipschitz propagation and the triangle inequality give

$$W_2\big((S_\theta)_\#\nu_\theta^y, \pi^y\big) \le L_{S_\theta} W_2(\nu_\theta^y, \nu^y) + W_2\big((S_\theta)_\#\nu^y, S_\#\nu^y\big).$$

The common-input coupling bounds the second term by

$$W_2^2\big((S_\theta)_\#\nu^y, S_\#\nu^y\big) \le \mathbb{E}_{\nu^y}\|S_\theta - S\|^2 \le \frac{M_y}{Z_y}\mathcal{E},$$

which proves the data-space stability. Substituting (17) yields (19)–(20).

A.7 Geometric convergence of the source samplers

Fix θ and y, and write

$$\omega(z) := \ell_y(S_\theta(z)), \qquad \mu_j(\mathrm{d}z) := Z_j^{-1}\omega(z)^{\beta_j}\gamma(\mathrm{d}z), \qquad Z_j := \int \omega^{\beta_j}\,\mathrm{d}\gamma, \qquad \Pi := \bigotimes_{j=0}^{J}\mu_j.$$

Thus $\mu_0 = \gamma$ and $\mu_J = \nu_\theta^y$. Let η_n be the joint law after n production sweeps and ν_n its cold marginal. We use $\chi^2(\eta\|\Pi) := \int(\mathrm{d}\eta/\mathrm{d}\Pi - 1)^2\,\mathrm{d}\Pi$.

Assumption A.1 (Source weight and production sampler)**.** *The source weight is measurable and satisfies $0 < m \le \omega(z) \le M < \infty$ for every $z \in \mathbb{R}^d$, where $m < M$. The ladder $0 = \beta_0 < \cdots < \beta_J = 1$, $J \ge 1$, and all local-kernel parameters are fixed during production. Each sweep independently applies the local kernels to all replicas, then attempts each adjacent Metropolis swap once in a prescribed order; this order may vary with the sweep index. Every pCN angle satisfies $h_{\beta_j} \in (0, \pi/2]$. For the hybrid sampler, $\Psi_\theta = -\log\omega$ is continuously differentiable, and the cold update is the Metropolis-corrected split HMC kernel of Section 3.3, with fresh independent Gaussian momentum at each update. The initial joint law satisfies $D_0 := \chi^2(\eta_0\|\Pi) < \infty$.*

The positive lower bound is an additional condition for the convergence result. It holds, for example, for exponentiated bounded rewards. Geometric convergence of SPT+pCN is established by Wang et al. (2026a). The following statement gives explicit bounds for both local-kernel choices and extends the guarantee to the hybrid sampler.

Theorem A.2 (Geometric convergence of SPT+pCN and SPT+hybrid)**.** *Under Assumption A.1, set $\kappa := M/m$. For* iMF+SPT+pCN*, define*

$$g_{\mathrm{pCN}} := \min_{0\le j\le J} \kappa^{-\beta_j}(1 - \cos h_{\beta_j}), \qquad q_{\mathrm{pCN}} := 1 - g_{\mathrm{pCN}} \in (0,1).$$

Then, for every integer $n \ge 0$,

$$\chi^2(\eta_n\|\Pi) \le D_0 q_{\mathrm{pCN}}^{2n}, \qquad W_2(\nu_n, \nu_\theta^y) \le \sqrt{2\kappa D_0}\, q_{\mathrm{pCN}}^n. \tag{26}$$

For iMF+SPT+hybrid*, define*

$$g_{\mathrm{hot}} := \min_{0\le j<J} \kappa^{-\beta_j}(1 - \cos h_{\beta_j}), \qquad a := \kappa^{-(1-\beta_{J-1})}, \qquad q_{\mathrm{hyb}} := 1 - g_{\mathrm{hot}}^2 a \in (0,1).$$

Then

$$\begin{aligned} \chi^2(\eta_n\|\Pi) &\le D_0 q_{\mathrm{hyb}}^{2\lfloor n/2\rfloor}, \\ W_2(\nu_n, \nu_\theta^y) &\le \sqrt{2\kappa D_0}\, q_{\mathrm{hyb}}^{\lfloor n/2\rfloor}. \end{aligned} \tag{27}$$

The hybrid bounds remain valid if the cold update is replaced by any fixed μ_J-invariant Markov kernel. In particular, no separate spectral-gap assumption on split HMC is required. If $\eta_0 = \gamma^{\otimes(J+1)}$, then $D_0 \le \kappa^{\sum_{j=0}^J \beta_j} - 1$.

Proof. All operator norms below are on the centered space $L_0^2(\Pi) := \{f \in L^2(\Pi) : \int f\,\mathrm{d}\Pi = 0\}$, unless another measure is specified. Each local kernel and each Metropolis swap preserves Π, so each is an $L^2(\Pi)$ contraction. Their composition therefore preserves the joint target, whose cold marginal is ν_θ^y.

Contraction of a pCN update. Let Q_j be the Gaussian pCN proposal kernel and A_j its Metropolis correction for μ_j. The Gaussian kernel is self-adjoint and positive on $L^2(\gamma)$, with centered norm $\cos h_{\beta_j}$. Writing $a_j := \mathrm{d}\mu_j/\mathrm{d}\gamma$, its accepted transition measure is

$$\min\{a_j(z), a_j(z')\}\,\gamma(\mathrm{d}z)\mathsf{Q}_j(z, \mathrm{d}z').$$

Hence comparison of Dirichlet forms and variances gives

$$\begin{aligned}\langle f, (I-\mathsf{A}_j)f\rangle_{\mu_j} &\geq (\inf a_j)\langle f, (I-\mathsf{Q}_j)f\rangle_\gamma \\ &\geq \frac{\inf a_j}{\sup a_j}(1-\cos h_{\beta_j})\,\mathrm{Var}_{\mu_j}(f) \\ &\geq \kappa^{-\beta_j}(1-\cos h_{\beta_j})\,\mathrm{Var}_{\mu_j}(f).\end{aligned}$$

The corrected kernel is also positive. Indeed, the identity

$$\min\{a_j(z), a_j(z')\} = \int_0^\infty \mathbf{1}_{\{a_j(z)\geq s\}}\mathbf{1}_{\{a_j(z')\geq s\}}\,\mathrm{d}s$$

expresses the accepted part of its quadratic form as an integral of nonnegative quadratic forms of Q_j; the rejection part is nonnegative as well. Reversibility and positivity therefore imply

$$\|\mathsf{A}_j\|_{L^2_0(\mu_j)\to L^2_0(\mu_j)} \leq 1-\kappa^{-\beta_j}(1-\cos h_{\beta_j}). \tag{28}$$

All-pCN sweeps. When every replica uses pCN, tensorization of (28) bounds the centered norm of the joint local update by q_{pCN}. The subsequent swap sweep is a contraction, so the centered norm after n sweeps is at most q_{pCN}^n. Applying the adjoint operator to $\mathrm{d}\eta_0/\mathrm{d}\Pi - 1$ proves the first bound in (26).

Hybrid sweeps. Let H be the cold invariant kernel and write $\mathsf{L} = (\bigotimes_{j<J}\mathsf{A}_j)\otimes\mathsf{H}$ for the local update. Let E be the orthogonal projection onto functions of the cold coordinate, obtained by integrating all hot coordinates against $\bigotimes_{j<J}\mu_j$. The operators L and E commute, and (28) gives

$$\|\mathsf{LE}\| \leq 1, \qquad \|\mathsf{L}(I-\mathsf{E})\| \leq r, \qquad r := 1-g_{\mathrm{hot}} \in [0,1).$$

For a complete swap sweep S, every swap involving only hot replicas satisfies $\mathsf{ES}_j = \mathsf{S}_j\mathsf{E} = \mathsf{E}$. Since the edge $(J-1, J)$ is attempted once, $\mathsf{ESE} = \mathsf{ES}_{J-1}\mathsf{E}$. On functions of the cold coordinate this is an independence Metropolis kernel with target μ_J and proposal μ_{J-1}. Moreover,

$$\frac{\mathrm{d}\mu_J}{\mathrm{d}\mu_{J-1}}(z) = \frac{Z_{J-1}}{Z_J}\omega(z)^{1-\beta_{J-1}} \leq \kappa^{1-\beta_{J-1}} = a^{-1}.$$

Its transition kernel thus minorizes $a\mu_J$, yielding $\|\mathsf{ESE}\| \leq 1-a$ on $L^2_0(\Pi)$.

Set $\mathsf{D} := \mathsf{E} + r(I-\mathsf{E})$. For $r > 0$, the operator $\mathsf{T} := \mathsf{LE} + r^{-1}\mathsf{L}(I-\mathsf{E})$ is a contraction and satisfies $\mathsf{L} = \mathsf{TD} = \mathsf{DT}$. For $r = 0$ the same factorization holds with $\mathsf{T} = \mathsf{L}$. Consequently,

$$\begin{aligned}\|\mathsf{LSL}\| &\leq \|\mathsf{DSD}\| \\ &\leq r^2 + 2r(1-r) + (1-r)^2(1-a) \\ &= 1 - g_{\mathrm{hot}}^2 a = q_{\mathrm{hyb}}.\end{aligned}$$

Here we expanded $\mathsf{D} = rI + (1-r)\mathsf{E}$ and used that S, ES, and SE are contractions. Each production sweep acts on functions as LS_n, so every two consecutive sweeps contract the centered norm by at most q_{hyb}, regardless of their prescribed swap orders. Applying the adjoint product to the initial density difference proves the χ^2 bound in (27). This argument only uses invariance of H, which proves the assertion for general cold kernels.

Cold-marginal Wasserstein bounds. The density of μ_J relative to γ has upper-to-lower ratio at most κ. The Gaussian logarithmic Sobolev inequality, bounded perturbation, and the implication from logarithmic Sobolev to transport inequalities give (Otto & Villani, 2000)

$$W_2^2(\nu, \mu_J) \leq 2\kappa\,\mathrm{KL}(\nu\|\mu_J) \leq 2\kappa\chi^2(\nu\|\mu_J);$$

see also Wang et al. (2026a, Lemma A.10). Conditional Jensen's inequality gives $\chi^2(\nu_n\|\mu_J) \le \chi^2(\eta_n\|\Pi)$. Combining these inequalities with the joint bounds proves (26) and (27). In particular, the latter has the form $C\rho^n$ with $\rho = \sqrt{q_{\mathrm{hyb}}}$ and $C = \sqrt{2\kappa D_0/q_{\mathrm{hyb}}}$.

Finally, for independent Gaussian initialization,

$$D_0 + 1 = \prod_{j=0}^{J} \left[\left(\int \omega^{\beta_j} \, \mathrm{d}\gamma \right) \left(\int \omega^{-\beta_j} \, \mathrm{d}\gamma \right) \right] \le \kappa^{\sum_{j=0}^{J} \beta_j}.$$

□

Corollary A.3 (End-to-end convergence)**.** *Suppose the assumptions of Theorem 3.7, Proposition 3.9, and Theorem A.2 hold. Assume also that S_θ is L_{S_θ}-Lipschitz, $0 \le \delta \le L_G/2$, and $L_y b_K(\delta) \le Z_y/2$. Define*

$$s_n := \sqrt{2\kappa D_0} \begin{cases} q_{\mathrm{pCN}}^{n}, & \text{for iMF+SPT+pCN,} \\ q_{\mathrm{hyb}}^{\lfloor n/2 \rfloor}, & \text{for iMF+SPT+hybrid.} \end{cases}$$

Then the data-space output law satisfies

$$W_2\big((S_\theta)_\# \nu_n, \pi^y\big) \le L_{S_\theta} s_n + L_{S_\theta} C_y(d)^{1/2} b_K(\delta)^{1/2} + \sqrt{\frac{M_y}{Z_y}} b_K(\delta). \tag{29}$$

For fixed K and $\lambda > 0$, suppose the regularity constants in b_K, L_{S_θ}, $C_y(d)$, M_y/Z_y, κ, and D_0 are uniformly bounded along a sequence of learned transports. If the corresponding rates q_{pCN} or $\sqrt{q_{\mathrm{hyb}}}$ are bounded by a common $\rho_* \in (0,1)$, then

$$W_2\big((S_\theta)_\# \nu_n, \pi^y\big) = O\Big(\rho_*^n + \delta^{1/8}\Big).$$

In particular, the output law converges to the true posterior in W_2 as $n \to \infty$ and $\delta \to 0$.

B Additional Experiments

B.1 Empirical loss-to-map-error relationship

We examine the relationship between the marginal joint risk and transport-map error in Theorem 3.7 using the two-dimensional banana prior

$$X = \big(U,\ 0.62(U^2 - 1) + 0.30V\big), \qquad U, V \overset{\mathrm{iid}}{\sim} \mathcal{N}(0,1).$$

For independent $X \sim \pi_0$ and $Z \sim \mathcal{N}(0, I_2)$, we use the interpolation $W_t = (1-t)X + tZ$ and marginal velocity $v(w,t) = \mathbb{E}[Z - X \mid W_t = w]$. The reference map $S = \varphi_{0,1}$ is the transport induced by this marginal velocity, rather than the explicit triangular map used to generate banana samples.

Numerical reference. Conditioning on the first coordinate of W_t gives a Gaussian conditional distribution for U. Integrating out V analytically reduces evaluation of $v(w,t)$ to one-dimensional Gaussian quadrature, with weights determined by the second coordinate of W_t. We use 128-point Gauss–Hermite quadrature and integrate the marginal ODE from $t = 1$ to $t = 0$ using DOP853 in double precision, with relative and absolute tolerances 10^{-8} and 10^{-10}. Doubling the quadrature order changes the reference velocities by 6.9×10^{-11} RMS over the risk-evaluation inputs. On 128 source samples, doubling the quadrature order and tightening both ODE tolerances tenfold change the reference maps by 8.3×10^{-15} and 8.7×10^{-9} RMS, respectively. These sensitivity checks indicate that numerical reference error is negligible at the observed map-error scale.

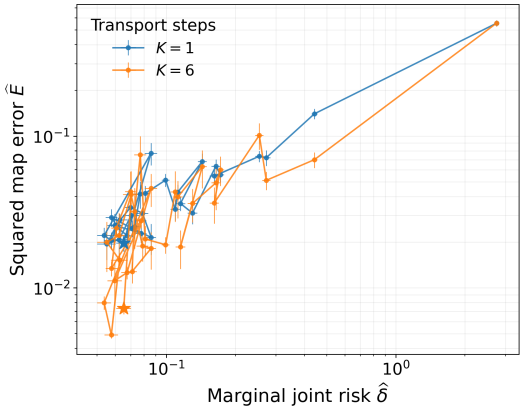


Figure 3: Marginal joint risk versus squared transport-map error on the two-dimensional banana prior for one- and six-step iMF generation. Points represent training checkpoints, connected in training order; stars mark the final checkpoint. Error bars show one Monte Carlo standard error, conditional on the fitted model and numerical reference, and do not represent variability across training runs.

Training and evaluation. We train an iMF model with three hidden SiLU layers of width 128 and separate average-velocity and auxiliary-velocity heads on 20,000 fixed prior samples. Training uses Adam with learning rate 10^{-3}, weight decay 10^{-5}, batch size 512, and 16,000 updates. We use independent data–noise coupling, independent unsorted uniform times $(r, t) \in [0, 1]^2$, and auxiliary-loss weight $\lambda = 1$. The figure reports a single training run with seed 0. At initialization and every 500 updates, we estimate

$$\delta = R_{\mathrm{iMF}} + \lambda R_{\mathrm{FM}}$$

directly from the marginal residuals using 8,192 held-out interpolation samples and the numerical reference velocity. Thus, $\widehat{\delta}$ estimates the marginal risk appearing in Theorem 3.7, rather than the raw training loss containing irreducible regression noise. We also estimate the squared map error

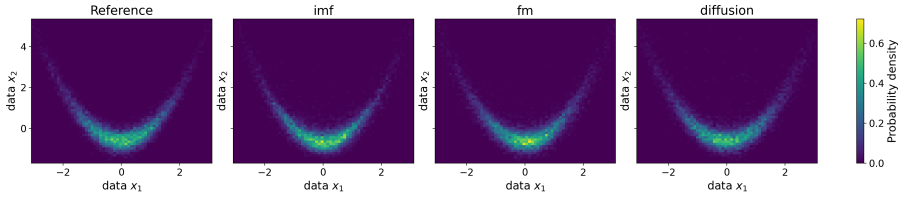
$$\widehat{E}_K = \frac{1}{2048} \sum_{i=1}^{2048} \left\| S_{\theta,K}(z_i) - S(z_i) \right\|_2^2, \qquad z_i \sim \mathcal{N}(0, I_2),$$

for $K \in \{1, 6\}$ uniformly spaced generation steps. The evaluation draws are independent of the training dataset and fixed across checkpoints.

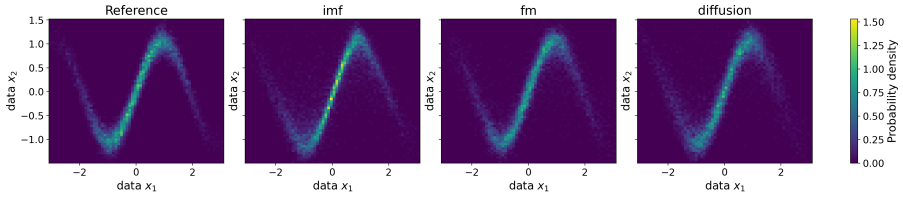
Results. Figure 3 shows that decreasing marginal risk is generally accompanied by decreasing map error for both generation schedules. Over training, $\widehat{\delta}$ decreases from 2.743 to 0.0653, while the squared map error decreases from 0.554 to 0.0198 for one-step generation and from 0.554 to 0.00732 for six-step generation. The relationship is not strictly monotonic across checkpoints. These results provide empirical support for the loss-to-map-error connection in Theorem 3.7, but do not establish its asymptotic rate or uniform control of the regularity constants, which may change during training.

B.2 Synthetic experiments

Prior training. For each scenario, all models are trained on the same fixed dataset of 20,000 prior samples using four-hidden-layer SiLU networks of width 512, with 16,000 optimizer updates at learning rate 10^{-3}. The iMF model uses raw time inputs (r, t) and separate average-velocity and auxiliary-velocity heads; FM and diffusion use four Fourier time harmonics. Both iMF and FM use independent data-noise coupling. For iMF, we sample r, t independently from $\mathrm{Unif}[0, 1]$, and set the auxiliary-loss weight to $\lambda = 1$. We use Adam for iMF with weight decay 10^{-5} and AdamW for FM and diffusion with zero weight decay, both with $(\beta_1, \beta_2) = (0.9, 0.999)$. Gradient clipping at norm 5 is applied only to FM and diffusion. Figure 4 shows that all three models capture the main geometric structure of the banana and sine priors.



(a) Banana prior.



(b) Sine prior.

Figure 4: Learning the priors by iMF, FM, and diffusion model.

Posterior sampling. The two iMF samplers share a six-step transport with uniformly spaced times, while FM+SPT+pCN uses 100 RK4 steps (400 network evaluations per transport). All SPT methods use ten parallel ladders with 24 inverse temperatures $\beta_\ell = (\ell/23)^2$, $\ell = 0, \dots, 23$, initialized from $\mathcal{N}(0, I_{32})$. The pCN variants update every replica with pCN; the hybrid replaces the coldest update with split HMC using six integration steps and initial step size 0.05. During 500 adaptation sweeps, proposal parameters are shared across ladders and updated separately at each temperature using the observed acceptance fraction across the ten ladders. We apply additive updates to the logit pCN noise scale and log HMC step size, proportional to the difference between observed and target acceptance, with respective gains $0.10/\sqrt{1+s/30}$ and $0.055/\sqrt{1+s/30}$ at sweep $s = 0, \dots, 499$. We target acceptance rates of 0.30 for pCN and 0.75 for split HMC. These are practical choices consistent with the pCN tuning range 0.3–0.5 used by Chen et al. (2016) and the general (leapfrog) HMC range 0.6-0.9 discussed by Betancourt et al. (2014). The hottest pCN noise scale remains fixed at 0.999. Each sweep includes odd and even adjacent-swap passes, with their order alternating across sweeps. After adaptation, all proposal parameters are frozen, 500 additional burn-in sweeps are discarded, and 600 consecutive cold states per ladder are retained without thinning, yielding 6,000 samples per method and seed. DPS generates the same number of samples using 1,000 DDPM steps with a linear noise schedule from 10^{-4} to 2×10^{-2}. Scaling DPS guidance by σ_y^{-2} caused numerical instability in our experiments. Following the residual-dependent step-size scheme of Chung et al. (2023, Appendix D.1) and its official implementation, we use guidance $-0.3\nabla_{x_t}\|G(\hat{x}_0(x_t)) - y\|_2$, without explicit noise-variance scaling. Sampling is repeated with seeds $0, 1, 2, 3$; full implementation settings are provided with the released code. Figures 5 and 6 compare sample histograms with numerical posterior references. Both iMF- and FM-based SPT recover the main posterior well, while DPS exhibits systematic deviations from the reference distributions.

Evaluation metrics. Let d be the dimension, C the number of ladders, and N the number of retained samples per ladder. Writing $\rho_{j,c}(\ell)$ for the lag-ℓ autocorrelation of coordinate j in ladder c, the mean coordinate ESS is

$$\text{Coord. ESS} = \frac{1}{d}\sum_{j=1}^{d}\sum_{c=1}^{C}\frac{N}{1+2\sum_{\ell=1}^{\infty}\rho_{j,c}(\ell)}, \qquad \text{Coord. ESS/s} = \frac{\text{Coord. ESS}}{T_{\text{sampling}}}.$$

For distributions μ, ν on $\mathbb{R}^d$, the sliced Wasserstein distance is

$$\text{SW}_2(\mu,\nu) = \left(\int_{\mathbb{S}^{d-1}} W_2^2\big((P_\omega)_\#\mu, (P_\omega)_\#\nu\big)\, d\sigma(\omega)\right)^{1/2}, \qquad P_\omega(x) = \omega^\top x,$$

where σ is the uniform probability measure on the unit sphere. For J two-dimensional blocks and a common partition $\{B_{j,b}\}_{b=1}^{B}$ of each block's first-coordinate marginal, define $p_{j,b} = \mu\{x : x_{j,1} \in$

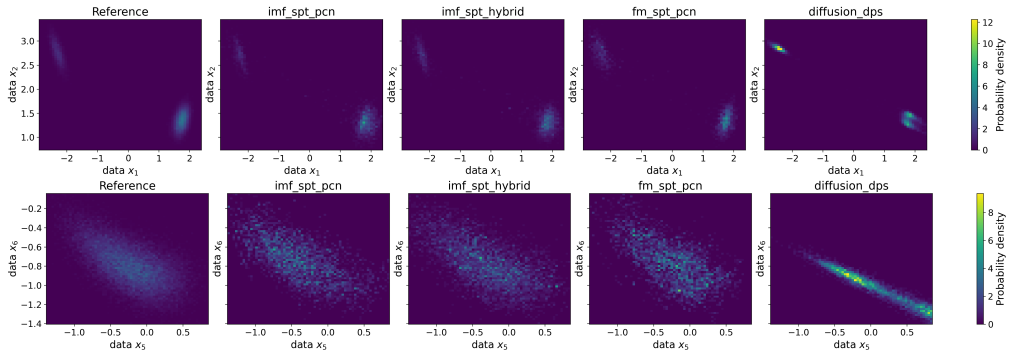


Figure 5: Posterior sample histograms on the banana prior.

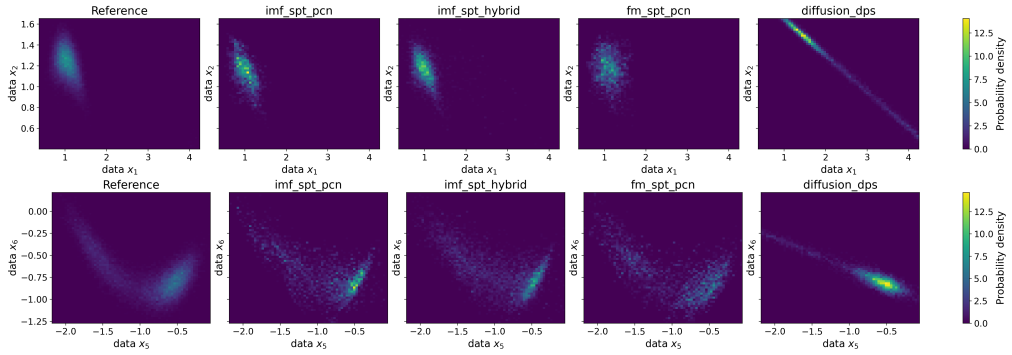


Figure 6: Posterior sample histograms on the sine prior.

$B_{j,b}\}$ and $q_{j,b} = \nu\{x : x_{j,1} \in B_{j,b}\}$. Then

$$\text{Bin TV}(\mu, \nu) = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{2}\sum_{b=1}^{B}|p_{j,b} - q_{j,b}|.$$

B.3 CLIP-GUIDED IMAGENET EXPERIMENTS

B.3.1 EXPERIMENT DETAILS

We use the pretrained iMF-XL/2[3] with one transport step, $\omega = 8$, and guidance interval $[0.4, 0.65]$, and frozen SiT-XL/2[4] with 125 Heun steps (250 velocity evaluations) and classifier-free guidance scale 4.

For each class-prompt pair, we use eight independent ladders of eight replicas with linear inverse temperatures $\beta_k = k/7$, initialized from $\mathcal{N}(0, I)$. The iMF methods run 250 adaptation sweeps, 250 burn-in sweeps, and 1,000 retained sweeps per ladder, yielding 8,000 retained cold states. At each adaptation sweep, we average the local proposal acceptance rates over all eight ladders. We then make progressively smaller updates to the pCN proposal scale to target 0.30 acceptance and, for iMF+SPT+hybrid, to the split-HMC step size at the coldest replica to target 0.75 acceptance. The resulting proposal parameters are fixed before burn-in. iMF+SPT+hybrid uses six-step split HMC at the coldest replica and pCN at all remaining replicas. SiT+SPT+pCN is run after the two iMF methods under a prompt-specific budget: its sampling budget equals the measured iMF+SPT+hybrid sampling time, with an additional one-hour allowance for loading, preflight, scoring, and export. Its available sampling time is allocated as $1/6$, $1/6$, and $2/3$ to adaptation, burn-in, and retained sampling; only complete sweeps are performed.

[3] https://github.com/Lyy-iiis/imeanflow/tree/torch

[4] https://huggingface.co/sairights/sit-xl-2-256x256-sde-cfg/resolve/5a15dcc/pretrained_models/SiT-XL-2-256x256.safetensors

B.3.2 FULL RESULTS

Table 3 reports the complete results, and Figures 7-18 display samples from the first two ladders for each method.

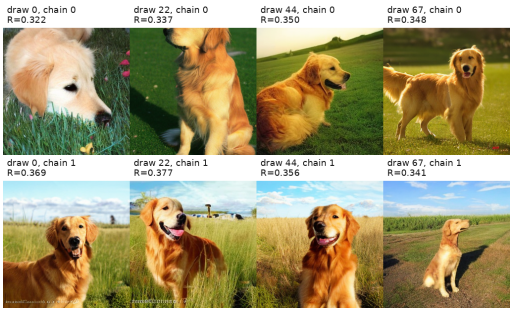
Table 3: CLIP reward over retained samples and sampling cost.

Method	Mean CLIP reward ↑	Sampling time (h)	Mean sweep time (s)
Golden retriever: *"A photograph of a golden retriever running through fallen autumn leaves."*			
Best-of-K	0.2864	2.49	–
SiT+SPT+pCN	0.3122	10.51	370.89
iMF+SPT+pCN	0.3809	2.40	5.75
iMF+SPT+hybrid	**0.3824**	10.95	26.28
Golden retriever: *"A beautiful professional photograph of a golden retriever in a sunlit meadow, sharp focus, natural colors."*			
Best-of-K	0.3073	2.48	–
SiT+SPT+pCN	0.3368	10.79	373.28
iMF+SPT+pCN	0.3741	2.43	5.83
iMF+SPT+hybrid	**0.3754**	11.26	27.02
Golden retriever: *"A colorful painting of a golden retriever standing in falling snow."*			
Best-of-K	0.2885	2.45	–
SiT+SPT+pCN	0.3355	10.96	372.17
iMF+SPT+pCN	0.3945	2.50	6.01
iMF+SPT+hybrid	**0.4011**	11.50	27.58
Golden retriever: *"A photograph of a golden retriever swimming in a clear blue lake."*			
Best-of-K	0.3064	2.45	–
SiT+SPT+pCN	0.3675	10.69	369.97
iMF+SPT+pCN	**0.3847**	2.45	5.88
iMF+SPT+hybrid	0.3845	11.16	26.69
Bald eagle: *"A photograph of a bald eagle perched on a branch surrounded by red and golden autumn leaves."*			
Best-of-K	0.3086	2.50	–
SiT+SPT+pCN	0.3156	10.73	371.46
iMF+SPT+pCN	0.3743	2.46	5.90
iMF+SPT+hybrid	**0.3781**	11.20	26.87
Bald eagle: *"A photograph of a bald eagle catching a fish from a lake, with splashing water."*			
Best-of-K	0.3247	2.59	–
SiT+SPT+pCN	0.3407	10.78	376.60
iMF+SPT+pCN	**0.3625**	2.46	5.90
iMF+SPT+hybrid	0.3596	11.17	26.81
Bald eagle: *"A photograph of a bald eagle soaring above snow-capped mountains, wings spread wide."*			
Best-of-K	0.3278	2.49	–
SiT+SPT+pCN	0.3481	10.70	370.27
iMF+SPT+pCN	**0.3936**	2.44	5.85
iMF+SPT+hybrid	0.3914	11.15	26.76
Bald eagle: *"A colorful painting of a bald eagle perched on a pine branch in falling snow."*			
Best-of-K	0.2984	2.58	–
SiT+SPT+pCN	0.3139	10.74	371.58
iMF+SPT+pCN	**0.3727**	2.41	5.77
iMF+SPT+hybrid	0.3713	11.16	26.78
Mink: *"A photograph of a mink beside a clear forest stream, surrounded by green moss."*			
Best-of-K	0.3629	2.75	–
SiT+SPT+pCN	0.3946	10.72	371.04
iMF+SPT+pCN	**0.4190**	2.45	5.89
iMF+SPT+hybrid	0.4112	11.12	26.70

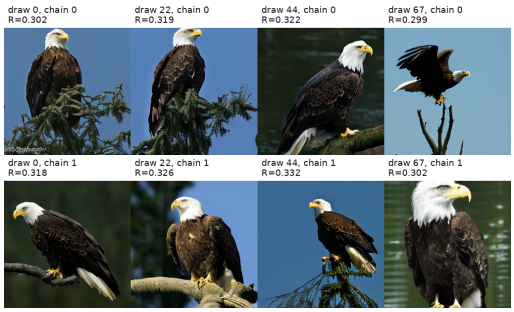
Continued on next page

Table 3 (continued)

Method	Mean CLIP reward ↑	Sampling time (h)	Mean sweep time (s)
Mink: *"A colorful painting of a mink standing in falling snow."*			
Best-of-K	0.3160	2.49	–
SiT+SPT+pCN	0.3487	10.47	369.64
iMF+SPT+pCN	**0.4186**	2.36	5.66
iMF+SPT+hybrid	0.4162	10.99	26.37
Castle: *"A colorful painting of a medieval castle covered in snow."*			
Best-of-K	0.2489	2.48	–
SiT+SPT+pCN	0.2797	10.53	371.82
iMF+SPT+pCN	**0.3618**	2.42	5.79
iMF+SPT+hybrid	0.3569	10.92	26.21
Castle: *"A photograph of a medieval castle on a hill at sunrise, surrounded by green countryside."*			
Best-of-K	0.3023	2.46	–
SiT+SPT+pCN	0.3476	10.96	372.13
iMF+SPT+pCN	0.3820	2.52	6.05
iMF+SPT+hybrid	**0.3849**	11.40	27.35
Aircraft carrier: *"A photograph of an aircraft carrier sailing across the ocean at sunset."*			
Best-of-K	0.2879	2.49	–
SiT+SPT+pCN	0.3286	10.38	373.72
iMF+SPT+pCN	**0.3667**	2.42	5.80
iMF+SPT+hybrid	0.3659	10.90	26.16
Aircraft carrier: *"A watercolor painting of an aircraft carrier on a calm blue sea."*			
Best-of-K	0.2956	2.52	–
SiT+SPT+pCN	0.3055	10.63	367.95
iMF+SPT+pCN	0.3736	2.49	5.97
iMF+SPT+hybrid	**0.3774**	11.07	26.55



(a) SiT + SPT + pCN.



(a) SiT + SPT + pCN.



(b) iMF + SPT + pCN.



(b) iMF + SPT + pCN.



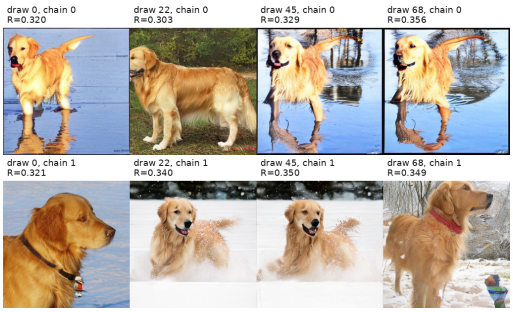
(c) iMF + SPT + hybrid.



(c) iMF + SPT + hybrid.

Figure 7: Class: golden retriever, prompt: *"A beautiful professional photograph of a golden retriever in a sunlit meadow, sharp focus, natural colors."*

Figure 8: Class: bald eagle, prompt: *"A photograph of a bald eagle perched on a branch surrounded by red and golden autumn leaves."*



(a) SiT + SPT + pCN.



(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

Figure 9: Class: golden retriever, prompt: *"A colorful painting of a golden retriever standing in falling snow."*



(a) SiT + SPT + pCN.



(b) iMF + SPT + pCN.

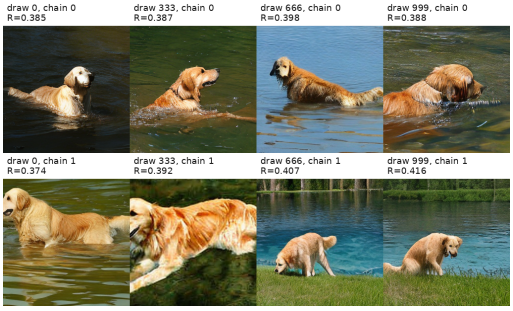


(c) iMF + SPT + hybrid.

Figure 10: Class: bald eagle, prompt: *"A colorful painting of a bald eagle perched on a pine branch in falling snow."*



(a) SiT + SPT + pCN.

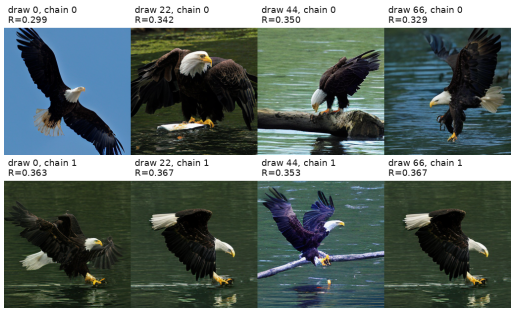


(b) iMF + SPT + pCN.

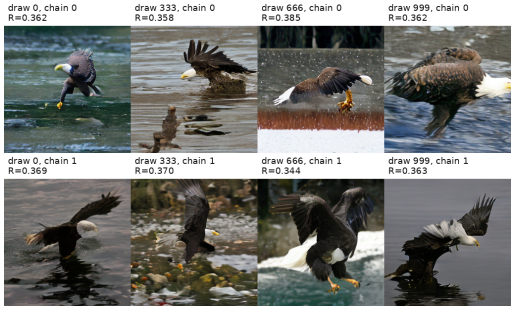


(c) iMF + SPT + hybrid.

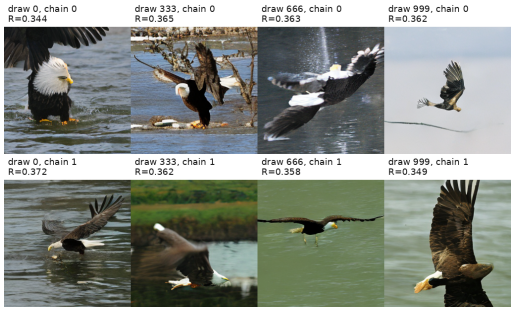
Figure 11: Class: golden retriever, prompt: *"A photograph of a golden retriever swimming in a clear blue lake."*



(a) SiT + SPT + pCN.

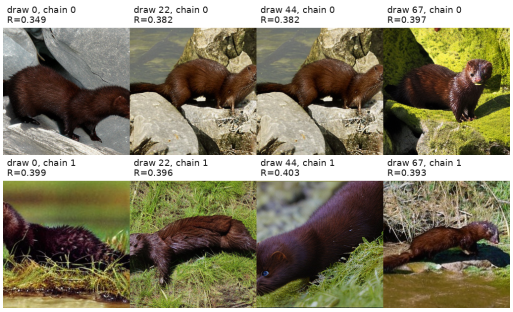


(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

Figure 12: Class: bald eagle, prompt: *"A photograph of a bald eagle catching a fish from a lake, with splashing water."*



(a) SiT + SPT + pCN.



(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

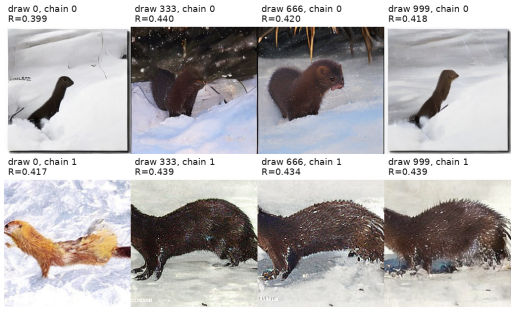
Figure 13: Class: mink, prompt: *"A photograph of a mink beside a clear forest stream, surrounded by green moss."*



(a) SiT + SPT + pCN.



(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

Figure 14: Class: mink, prompt: *"A colorful painting of a mink standing in falling snow."*



(a) SiT + SPT + pCN.

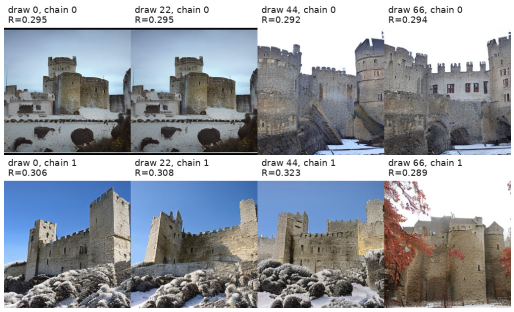


(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

Figure 15: Class: castle, prompt: *"A photograph of a medieval castle on a hill at sunrise, surrounded by green countryside."*



(a) SiT + SPT + pCN.



(b) iMF + SPT + pCN.

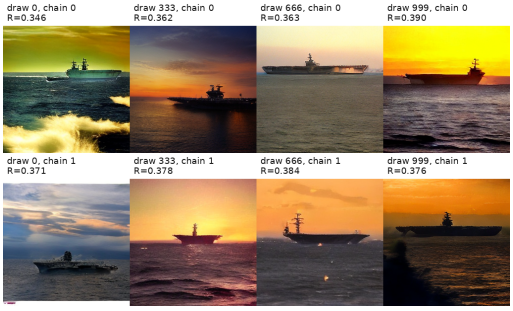


(c) iMF + SPT + hybrid.

Figure 16: Class: castle, prompt: *"A colorful painting of a medieval castle covered in snow."*



(a) SiT + SPT + pCN.

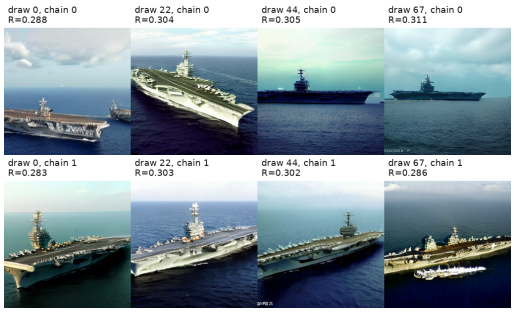


(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

Figure 17: Class: aircraft carrier, prompt: *"A photograph of an aircraft carrier sailing across the ocean at sunset."*



(a) SiT + SPT + pCN.



(b) iMF + SPT + pCN.



(c) iMF + SPT + hybrid.

Figure 18: Class: aircraft carrier, prompt: *"A watercolor painting of an aircraft carrier on a calm blue sea."*