DIVERSE SAMPLING IN DIFFUSION MODELS WITH DIVERGENCE-FREE PARTICLE GUIDANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Modern generative models can produce high-quality samples, but independent samples under the same condition often yields highly similar outputs. Particle-based methods increase diversity by letting the samples in a batch interact, typically through a repulsive force. These forces, however, also push each sample away from the data distribution and produce visible artifacts. We introduce EDDY, a training-free particle guidance method designed to avoid this failure mode. Instead of a repulsive gradient, EDDY couples particles through anti-symmetric matrix fields passed through a Stein operator, a family of drift perturbations that leaves the Fokker–Planck equation invariant. Unlike repulsive guidance, these interactions do not alter a particle's distribution while the batch is independent. For the same reason they cannot create diversity on their own, so EDDY pairs them with a negatively correlated initialization that keeps each particle's prior exact. To use EDDY with perceptual kernels such as DINOv2, we approximate its second-order terms with finite differences and Hutchinson estimates. Across FLUX.1-dev, FLUX.2-klein and SDXL, EDDY achieves higher image quality and prompt alignment than existing particle guidance methods at matched diversity.

1 INTRODUCTION

Diffusion and flow matching models (Albergo & Vanden-Eijnden, 2023; Ho et al., 2020; Lipman et al., 2023; Liu et al., 2023; Sohl-Dickstein et al., 2015; Song et al., 2021b) have become a prominent class of generative models, enabling high-quality sampling across multiple modalities, such as images Dhariwal & Nichol (2021); Lipman et al. (2023); Rombach et al. (2022); Song et al. (2021a), molecules Corso et al. (2023); Hoogeboom et al. (2022); Xu et al. (2022); Jing et al. (2022), and texts Austin et al. (2021); Li et al. (2022); Lou et al. (2024); Sahoo et al. (2024). These models can be viewed as transporting a simple prior distribution to a complex target distribution through time-dependent stochastic or deterministic dynamics. One advantage of this formulation is that generation can be modified at inference time through guidance Ho & Salimans (2021); Rombach et al. (2022): the drift can be adjusted to favor a condition, satisfy a constraint, or steer samples toward desired properties. However, when multiple samples are drawn for the same condition, the resulting set may contain similar outputs.

This issue has motivated a growing body of work on diverse sampling Kynkäänniemi et al. (2024); Lu et al. (2024); Sadat et al. (2024). In many applications, one does not want a single output, but rather a small set of samples that explores different plausible modes of the conditional distribution. Naively drawing independent samples is often inefficient, partly because guidance improves sample fidelity at the expense of diversity Ho & Salimans (2021). Recent methods addressed this issue by introducing interactions between samples, for example, by optimizing initial noise latents Harrington et al. (2025); Kim et al. (2026) or adding repulsive particle potentials during sampling Corso et al. (2024); Jalali et al. (2026); Morshed & Boddeti (2025). These approaches can improve diversity but can suffer from severe drawbacks. Mainly, we found that adding repulsive terms during sampling creates noticeable artifacts, while optimizing the noise latents had a limited effect on diversity (see examples in Fig. 1).

In this work, we ask why particle-based methods produce artifacts, and whether the interactions can be designed to avoid them. Repulsive methods add the gradient of a similarity potential to the drift. Even when the particles are independent, such a gradient moves probability mass away from

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.



Figure 1: Qualitative comparison on the prompt *"a chef's display of ingredients and pastry creations"* PG (top-left) and DF (top-middle): Increase in variety with visible artifacts. CADS (top-right): High-quality samples but with low prompt adherence (no chef or ingredients in any image, burgers instead of pastries). I.I.D (bottom-left): Low diversity between samples. EDDY (bottom-right): Diverse output that preserves quality and prompt fidelity.

the data distribution (Figure 2), so every sample is pushed off the data manifold whether or not the batch needs to spread out. We look instead for interactions whose effect on a sample's distribution vanishes when the particles start as independent. The Fokker-Planck equation that governs the density evolution provides such a family: drift perturbations obtained by applying a Stein-type operator to anti-symmetric matrix fields change individual trajectories without changing the density evolution. EDDY builds its pairwise interactions from this family. For each pair of particles, it forms an anti-symmetric matrix from the difference of two outer products involving the kernel's repulsive direction, and averages the resulting perturbations over the batch.

However, the property that protects an marginals also keeps it independent: starting from i.i.d. samples, the joint distribution remains a product, so the interaction adds no diversity (Claim 2). EDDY therefore starts from a negatively correlated prior, which keeps each particle's initial distribution exactly Gaussian. Once the particles are coupled, the interaction can generate diversity but also shift the marginal. We then show empirically that our construction allows us to achieve better diversity-distortion tradeoffs relative to other methods.

We further observe that a substantial part of the diversity EDDY produces in practice comes from an unexpected source: the discretization of the sampler (see Figure A.1 in the appendix). For independent particles, the guidance's contribution to the density evolution cancels exactly in continuous time, but a finite sampling step keeps this cancellation only to first order in the step size. The remainder involves pairs of particles, since each particle's guidance depends on its neighbors, so it makes the particles dependent; empirically, this dependence increases diversity. With the few, large steps used in text-to-image sampling, this remainder is not negligible. It can also shifts each particle's marginal, however we observe a very favorable tradeoff.

For practical text-to-image generation, we often wish to define similarity not directly in pixel space, but in a perceptual feature space such as DINO or CLIP embeddings. In this setting, computing the exact kernel can be prohibitively expensive due to second order derivatives. We therefore approximate the expensive elements with finite differences and Hutchinson trace estimation. These approximations add a third source of distortion, on top of the coupling and the discretization; empirically, EDDY still achieves higher sample quality than competing methods at matched diversity.

We make the following contributions: (i) **EDDY**, a training-free particle guidance method that builds pairwise interactions from anti-symmetric Stein perturbations, together with a negatively correlated initialization that preserves each particle's prior. (ii) **A scalable approximation** that uses only kernel evaluations and first-order feature gradients, running with perceptual kernels such as DINOv2

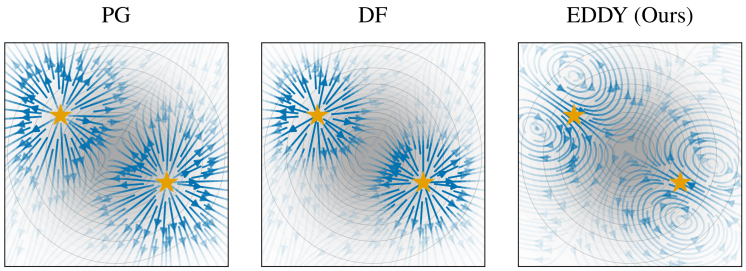

Figure 2: Transport fields induced by PG (left), DF (middle) and EDDY (right) for a standard Gaussian target (gray contours): how two particles (stars) affect a third test particle at each location. PG and DF applies the mean negative RBF kernel gradient, a purely repulsive field that distorts the sampling distribution even between independent particles. EDDY applies a divergence-free field.

at 10–15% overhead on FLUX. (iii) **An empirical evaluation** on three text-to-image backbones showing a better diversity–quality trade-off than PG and DiverseFlow at matched diversity, plus a synthetic study where the target is known. (iv) **An analysis that motivates the design**: anti-symmetric interactions cannot correlate an independent batch, which is why EDDY needs a coupled initialization; and unlike repulsive guidance, their distortion of the marginal arises only through the coupling.

2 Background

Throughout the paper, state variables such as $x_t \in \mathbb{R}^d$ and functions such as μ_t and p_t carry implicit argument x_t unless stated otherwise; and ∇ (div) denotes the gradient (resp. divergence) with respect to x_t.

2.1 Diffusion Models and Flow Matching

Modern generative modeling transports a simple prior $p_0 = \mathcal{N}(0, I_d)$ to a target p_1 over $t \in [0, 1]$ through continuous-time dynamics. DDPM Ho et al. (2020); Song et al. (2021b) fixes a Gaussian noising process with noise schedule β_t and samples $x_1 \sim p_1$ by integrating the time-reversal SDE Anderson (1982),

$$dx_t = \left[\tfrac{1}{2}\beta_t\, x_t + \beta_t\, \nabla \log p_t(x_t)\right] dt + \sqrt{\beta_t}\, dW_t, \tag{1}$$

forward in t, where W_t is the Wiener process. Optimal-transport flow matching (OT-FM) Albergo & Vanden-Eijnden (2023); Lipman et al. (2023); Liu et al. (2023) instead integrates the deterministic ODE

$$\frac{\partial}{\partial t} x_t = \mathbb{E}[x_1 - x_0 \mid x_t]. \tag{2}$$

Both frameworks therefore fall under a common Itô SDE

$$dx_t = \mu_t(x_t)\, dt + \sigma_t\, dW_t, \tag{3}$$

with drift $\mu : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ and volatility $\sigma : [0, 1] \to \mathbb{R}_{\geq 0}$: diffusion takes $\sigma_t > 0$ and absorbs the score into the drift as in equation 1, while OT-FM has $\sigma \equiv 0$ and learns μ_t directly. We assume throughout that any condition (e.g. a prompt) is already incorporated into the drift. The score $\nabla \log p_t(x_t)$ is available in both: diffusion models output it up to a known scaling factor, and in flow matching it follows from $\mathbb{E}[x_1 - x_0 \mid x_t]$ by Tweedie's formula Efron (2011),

$$\nabla \log p_t(x_t) = \frac{t\mathbb{E}[x_1 - x_0 \mid x_t] - x_t}{1 - t}. \tag{4}$$

see Appendix C for the derivation.

2.2 Fokker–Planck Equation and its Symmetries

The state x_t in equation 3 is a random variable whose marginal density p_t on $\mathbb{R}^d$ evolves deterministically under the Fokker-Planck equation (FPE),

$$\frac{\partial}{\partial t} p_t \;=\; -\operatorname{div}(p_t\, \mu_t) \;+\; \frac{\sigma_t^2}{2}\, \Delta p_t, \tag{5}$$

so each pair (μ, σ) induces a path of marginals $\{p_t\}_{t\in[0,1]}$ from p_0 to p_1. This mapping is not one-to-one: many (μ, σ) produce the identical evolution of p_t. These are the *symmetries* of the FPE: changes that leave the density evolution, and hence every marginal, invariant while reshaping how individual particles move.

Definition 1 (Matrix Stein operator)**.** For a matrix field $F : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, the Stein operator (Stein, 1972) over F is defined as

$$\mathcal{A}_p(F) \;=\; \frac{1}{p}\operatorname{div}(pF) \;=\; \operatorname{div} F \;+\; F\,\nabla \log p, \tag{6}$$

where $\operatorname{div} F$ is the row-wise divergence: $(\operatorname{div} F)_i = \sum_j \frac{\partial}{\partial x_j} F_{ij}$.

Claim 1 (Symmetry of the FPE)**.** *Let $A : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be any smooth anti-symmetric matrix field (i.e., $A^\top = -A$), that can depend on t. Then the guidance $\psi_t = \mathcal{A}_{p_t}(A)$ yields the same Fokker–Planck equation 5 and hence the same marginal p_t for all t. (Hua et al., 2025; Richter-Powell et al., 2022)*

Proof. Substituting the guided drift $\tilde{\mu}_t = \mu_t + \psi_t$ into equation 5, the additional contribution is $\operatorname{div}(p_t\,\mathcal{A}_p(A)) = \operatorname{div}(p_t \nabla(p_t A)/p_t) = \operatorname{div}(\operatorname{div}(A\,p_t))$ where the inner divergence acts on a matrix and the outer divergence acts on a vector. For any anti-symmetric A, this double divergence vanishes identically by the symmetry of mixed partial derivatives ($\frac{\partial^2}{\partial x_i \partial x_j}(A_{ij}p_t) + \frac{\partial^2}{\partial x_j \partial x_i}(A_{ji}p_t) = 0$ since $A_{ji} = -A_{ij}$). □

The operator $\mathcal{A}_p$ in equation 6 is the classical Stein operator $f \mapsto \operatorname{div}(f) + \langle f, \nabla \log p\rangle$, but acting on each row of the matrix F independently.

3 Related Work

Diverse Conditional Sampling in Diffusion Models. Guidance is a powerful mechanism to gain sampling control in diffusion and flow models. Early methods in this area of study are classifier guidance Dhariwal & Nichol (2021) and classifier-free guidance (CFG) Ho & Salimans (2021), both trade-off diversity for fidelity. Several studies improve diversity while maintaining high fidelity in CFG. Inspired by the forward process in diffusion models, CADS (Sadat et al., 2024) injects random noise into the conditioning signal. This increases sample diversity while maintaining strong image quality; however, the noisy conditioning leads to a noticeable decline in prompt alignment. Kynkäänniemi et al. (2024) limit guidance to a middle interval of sampling steps. Koulischer et al. (2026) adapt the guidance scale from the posterior likelihood of the current time step and the condition prior. Lu et al. (2024) push samples away from a reference set based on approximate denoised values.

Diverse Sampling in Diffusion Models using Particles. Several approaches achieve diverse sampling in diffusion and flow models through interactions between particles. Kim et al. (2026) and Harrington et al. (2025) both optimize the initial noise samples: the former with an InfoNCE loss Oord et al. (2018) on approximate denoised values, the latter with an objective combining quality, diversity and noise regularization. Parmar et al. (2026) sub-sample a subset of particles at each flow step using quadratic integer programming that promotes both quality and diversity. Kirchhof et al. (2025) shift particles whose approximate denoised values fall inside balls centered at the elements of some reference set. More related to our approach are PG (Corso et al., 2024), DiverseFlow (Morshed & Boddeti, 2025), and SPARKE Jalali et al. (2026), all of which modify the SDE during sampling using kernels. PG adds a fixed potential function based on similarity kernels, DiverseFlow uses determinantal point processes (DPP) as a repulsion force on the approximate denoised values, and SPARKE, designed to operate on semantically similar prompts simultaneously, adds a repulsion term based on the RKE score Jalali et al. (2023). Unlike EDDY, all of these approaches modify the marginal distribution of particles, hence sacrificing fidelity or quality for diversity. We note that PG discussed the possibility of preserving the marginal distribution of particles by learning the potential function, but they did not follow this path in their experimental evaluations. Lastly, Liu et al. (2025) use importance weighting to correct the bias introduced by a repulsion term; however, this requires learning an additional model to track the weights, unlike EDDY which is training-free.

4 THE COST OF DIVERGENCE-FREE GUIDANCE

In this section and the next, time t is implicit unless stated otherwise. That includes $p = p_t$, $x^i = x_t^i$, and derived quantities that depend on t. The *marginal* $p_t(x_t^i)$ describes the distribution of a single particle at time t. When sampling a batch of n particles jointly, their collective behavior is governed by a *joint* distribution $\boldsymbol{p}_t(\boldsymbol{x}_t)$ where $\boldsymbol{x}_t = (x_t^1, \dots, x_t^n)$. We would like a sampler that (i) preserves p_t for every particle individually, and (ii) couples the particles non-trivially and thereby increases diversity.

Claim 1 is stated for *external* fields: the matrix field A depends solely on the particle's own position, so guided particles never interact and stay independent. Both claims below concern *interacting* fields, in which $A_t^{(i)}$ sees the whole batch, and together they analyze such guidance. Claim 2 shows it is free but inert on an i.i.d. batch: the joint stays a product, so the marginals are exact and no diversity is gained. Claim 3 shows exactly how the marginal is perturbed when initialized from a non-i.i.d. initialization.

Setup. Let $\boldsymbol{x}_t = (x_t^i)_{i=1}^n$ be n particles in $\mathbb{R}^d$, $d \geq 2$, each individually evolving by drift μ_t under a probability path p_t which is always nonzero, and let each particle i receive the additional drift $\psi_t^i(\boldsymbol{x}_t) = \mathcal{A}_{p_t}(A_t^{(i)}(\boldsymbol{x}_t))$ for an anti-symmetric matrix field $A_t^{(i)}$ which is Lipschitz in $\boldsymbol{x}_t$. Write $\boldsymbol{p}_t$ for the joint law of $\boldsymbol{x}_t$ and $\tilde{p}_t$ for the marginal law of x_t^i.

Claim 2 (Independence propagates)**.** *If $\boldsymbol{p}_0 = p_0^{\otimes n}$ then $\boldsymbol{p}_t = p_t^{\otimes n}$ for every $t \geq 0$.*

Proof sketch. The *joint* density $\boldsymbol{p}_t$ on $\mathbb{R}^{nd}$ obeys the FPE

$$\frac{\partial}{\partial t}\boldsymbol{p}_t \;=\; -\sum_i \operatorname{div}_i\big(\boldsymbol{p}_t\,(\mu_t + \psi_t^i)\big) \;+\; \frac{\sigma_t^2}{2}\sum_i \Delta_i \boldsymbol{p}_t, \tag{7}$$

where div_i and Δ_i act on the x_t^i block. Substitute the product ansatz $\boldsymbol{p}_t = p_t^{\otimes n}$. Each guidance term contributes

$$\operatorname{div}_i\big(\boldsymbol{p}_t\,\psi_t^i\big) \;=\; \Big(\prod_{j\neq i} p_t(x_t^j)\Big)\,\operatorname{div}_i \operatorname{div}_i\big(p_t(x_t^i)\,A_t^{(i)}(\boldsymbol{x}_t)\big) \;=\; 0, \tag{8}$$

because the factors $j \neq i$ do not depend on x_t^i and pass through div_i, and the remaining double divergence vanishes by the mixed-partials cancellation of Claim 1 applied in the x_t^i block. The surviving terms are n copies of the single-particle FPE equation 5, which $p_t^{\otimes n}$ satisfies by construction. Note that equation 7 is *linear* in $\boldsymbol{p}_t$ and admits a unique solution, so the product law is that solution for every t. □

Every marginal is therefore exactly p_t and nothing else happens. The batch is distributed exactly as n independent samples, which generates no additional diversity. We can generate diversity from a coupled initialization, and the next theorem says what that costs.

Claim 3 (The marginal price of coupling)**.** *At any time at which $\tilde{p}_t = p_t$, the marginal picks up the additive term $\operatorname{div}(p_t\, b_t^{(i)})$ per unit time,*

$$\frac{\partial}{\partial t}\tilde{p}_t \;=\; \frac{\partial}{\partial t}p_t \;+\; \operatorname{div}\big(p_t\, b_t^{(i)}\big), \tag{9}$$

where the marginal bias

$$b_t^{(i)}(x) \;=\; \mathbb{E}\Big[\, A_t^{(i)}\big(x, x_t^{-i}\big)\,\nabla_x \log \boldsymbol{p}_t\big(x_t^{-i} \mid x_t^i = x\big) \;\Big|\; x_t^i = x \Big] \tag{10}$$

vanishes identically if the particles are independent at time t.

Proof sketch. Write $x := x_t^i$ and $y := x_t^{-i}$, and integrate equation 7 over y. Every term with div_j or Δ_j, $j \neq i$, is a divergence in a variable the integral covers and drops out, leaving the marginal equation

$$\frac{\partial}{\partial t}\tilde{p}_t \;=\; -\operatorname{div}_x\Big[\mu_t\,\tilde{p}_t \;+\; \int \boldsymbol{p}_t(x,y)\,\psi_t^i(x,y)\,dy\Big] \;+\; \frac{\sigma_t^2}{2}\,\Delta_x \tilde{p}_t. \tag{11}$$

Factoring $\boldsymbol{p}_t(x, y) = \tilde{p}_t(x)\, c_t(y \mid x)$ and moving the conditional inside the divergence via the product rule $\mathrm{div}_x(p_t A_t^{(i)} c_t) = c_t\ \mathrm{div}_x(p_t A_t^{(i)}) + p_t A_t^{(i)} \nabla_x c_t$ splits the guidance integral in two,

$$\int \boldsymbol{p}_t\, \psi_t^i\, dy \;=\; \tilde{p}_t\big[\, \mathcal{A}_{p_t}(\bar{A}_t) \;-\; b_t^{(i)}\,\big], \qquad \bar{A}_t(x) \;:=\; \mathbb{E}\big[\, A_t^{(i)}(x, x_t^{-i}) \;\big|\; x_t^i = x\,\big]. \tag{12}$$

Both enter equation 11 inside div_x, so at a time where $\tilde{p}_t = p_t$ what decides whether the marginal moves is their divergence. Claim 1 kills the first, $\bar{A}_t$ being an average of anti-symmetric matrices and depending on x alone:

$$\mathrm{div}_x\big(p_t\, \mathcal{A}_{p_t}(\bar{A}_t)\big) \;=\; \mathrm{div}_x\, \mathrm{div}_x\big(p_t \bar{A}_t\big) \;=\; 0. \tag{13}$$

It kills $A_t^{(i)}(\cdot, y)$ for each *fixed* y too, so the second would vanish as well if div_x could pass inside the y-average; it cannot, because that average is weighted by $c_t(y \mid x)$ and the product rule charges for the weight's own x-dependence:

$$\mathrm{div}_x\!\left(\int \boldsymbol{p}_t\, \psi_t^i\, dy\right) \;=\; \int \Big[\underbrace{\mathrm{div}_x\big(p_t\, \psi_t^i(\cdot, y)\big)}_{=\,0}\, c_t \;+\; p_t\, \psi_t^i \cdot \nabla_x c_t\Big]\, dy \;=\; -\,\mathrm{div}_x\big(p_t\, b_t^{(i)}\big). \tag{14}$$

What survives is a *correlation* between the guidance and the neighbors' conditional score $\nabla_x \log c_t$, which anti-symmetry does not constrain. Substituting it back into equation 11 leaves the marginal evolving with drift $\mu_t - b_t^{(i)}$, which is equation 9 once the target's own evolution equation 5 is subtracted. Independent particles have $c_t(y \mid x) = c_t(y)$, so $\nabla_x \log c_t \equiv 0$ and $b_t^{(i)} \equiv 0$. □

The full proofs are in Appendix A.1.

Remark 1. *Claim 2 is a continuous-time statement. An explicit Euler sampler may couples even an i.i.d. batch by treating time as discrete.*

5 METHOD

5.1 CONSTRUCTING EDDY-RBF

We first instantiate the principle into a sampling algorithm with negatively correlated priors and RBF kernel. We will, in the next subsection, generalize the construction to arbitrary similarity kernels, but this requires approximations to scale to high dimensions.

First, we couple the prior. Claim 2 shows that the drift alone cannot introduce dependence, thus it is required. We negatively correlate the priors while preserving their marginals by: drawing n particles independently from p_0, subtracting their empirical mean and rescaling by $\sqrt{n/(n-1)}$. For more details, see Claim 4 (Appendix A.2.1).

Second, we perturb the drift. For each particle x^i and each neighbor $x^j \neq x^i$, we form the anti-symmetric matrix $A^{(ij)}$ as follows:

$$A^{(ij)} \;=\; v^{(j)} \otimes r^{(ij)} \;-\; r^{(ij)} \otimes v^{(j)}, \tag{15}$$

where $\otimes$ is the outer product of vectors, $v^{(j)}$ is a vector that depends on neighbor j, and $r^{(ij)} = -\nabla k(x^i, x^j)$ is the repulsive direction induced by a similarity kernel k. Intuitively, $r^{(ij)} \otimes v^{(j)}$ is a rank-one matrix whose range is the desired direction $r^{(ij)}$, and subtracting its transpose makes it anti-symmetric. We let $v^{(j)}$ depend on neighbour j to avoid additional costly terms in the divergence and to incorporate information about neighbour j. In our experiments, we found that $v^{(j)} = \mu^{(j)}$ (the drift) works well for OT-FM as well as the synthetic experiment, while $v^{(j)} = \sigma\, \nabla \log p(x^j)$ works well for SDXL where $\sigma = \sigma_t$ is the current noise in the schedule.

To account for the contribution of all particles on x^i we average the contribution over all neighbors, which yields the per-particle guidance: $\psi^i_{\text{EDDY}} \;=\; \frac{1}{n-1} \sum_{j \neq i} \mathcal{A}_p(A^{(ij)})$.. Each $A^{(ij)}$ is anti-symmetric, so this quantity disturbs every particle's marginal only through the marginal bias equation 10.

By Definition 1, $\mathcal{A}_p(A^{(ij)}) = \operatorname{div} A^{(ij)} + A^{(ij)}\, s^{(i)}$ where $s^{(i)} = \nabla \log p(x^i)$. The score $s^{(i)}$ is available either directly for diffusion or via Tweedie's formula using Eq. 4 for flow matching, reducing the problem to evaluating the divergence:

$$\operatorname{div} A^{(ij)} \;=\; (\operatorname{div} r^{(ij)})\, v^{(j)} \;+\; \cancel{(\nabla v^{(j)})\, r^{(ij)}} \;-\; \cancel{(\operatorname{div} v^{(j)})\, r^{(ij)}} \;-\; (\nabla r^{(ij)})\, v^{(j)}. \tag{16}$$

Where, the terms in Eq. 16 disappear because $v^{(j)}$ does not depend on x^i, so its gradient and divergence with respect to x^i are zero. The remaining terms are pure kernel quantities, $\operatorname{div} r^{(ij)} = -\Delta k(x^i, x^j)$ and $(\nabla r^{(ij)})\, v^{(j)} = -\nabla^2 k(x^i, x^j)\, v^{(j)}$, both **do not** require backpropagation through the generative model. These combine into $\operatorname{div} A^{(ij)} = K^{(ij)}\, v^{(j)}$, where $K^{(ij)} = \nabla^2 k(x^i, x^j) - \Delta k(x^i, x^j)\, I_d$ is the divergence-free matrix kernel (Narcowich & Ward, 1994; Ni et al., 2025), classically used to represent flow and electromagnetic fields. Eq. 5.1 thus simplifies to:

$$\psi^i_{\text{EDDY}} \;=\; \frac{1}{n-1} \sum_{j \neq i} \big(A^{(ij)}\, s^{(i)} \;+\; K^{(ij)}\, v^{(j)} \big). \tag{17}$$

For the RBF kernel $k(x^i, x^j) = \exp\big(-\|x^i - x^j\|^2/\gamma\big)$ with bandwidth γ, all the kernel derivatives admit closed forms, resulting in an easy to calculate formula (see Appendix A.3) that we name EDDY-RBF.

5.2 Constructing EDDY

EDDY-RBF assumes the kernel derivatives are given in closed form. In our text-to-image experiments (Section 6.2) we use an RBF kernel applied to DINOv2 features, so the second-order derivatives in Eq. 17 are no longer known analytically. Moreover, to exactly compute the Laplacian element of EDDY would require $\mathcal{O}(d)$ backpropagations, which is impractical for high-dimensional problems.

We therefore drop the closed-form assumption and approximate both second-order quantities using only fast kernel evaluations and first-order kernel gradients (as available from CLIP or DINO). This adds a third source of distortion, on top of the marginal bias of Claim 3, yet EDDY still outperforms competing approaches in our experiments.

For the Hessian–vector product, we use a central difference with two backward passes,

$$\nabla^2 k(x^i, x^j)\, v^{(j)} \;\approx\; \frac{\nabla k(x^i_t + \epsilon v^{(j)}, x^j_t) - \nabla k(x^i_t - \epsilon v^{(j)}, x^j_t)}{2\epsilon}. \tag{18}$$

For the Laplacian, we use a forward-only Hutchinson trace estimator with $m = 5$ Rademacher probes $r_1, \ldots, r_m \overset{\text{iid}}{\sim} \operatorname{Rad}(\{\pm 1\}^d)$, requiring $2m + 1$ forward passes,

$$\Delta k(x^i, x^j) \;\approx\; \frac{1}{m} \sum_{\ell=1}^{m} \frac{k(x^i_t + \epsilon r_\ell, x^j_t) - 2k(x^i, x^j) + k(x^i_t - \epsilon r_\ell, x^j_t)}{\epsilon^2}. \tag{19}$$

Empirically, the guidance has its largest effect in the early, high-noise portion of the trajectory, so we apply ψ^i_{EDDY} only for the first fraction $r_{\text{stop}} = 0.2$ of sampling steps, cutting the overhead fivefold with negligible loss in output quality. Algorithm 1 in the appendix summarizes the full procedure, and Table A.6 reports average sampling times for EDDY, I.I.D, PG and DF under the same 0.2 fraction.

6 Experiments

We first evaluate on a synthetic low-dimensional model where standard statistical tests can resolve how far the marginal distribution has moved, and then on modern text-to-image generation models.

6.1 2D Gaussian Mixture Model

Experimental Setting. We evaluate EDDY-RBF on a 2-dimensional data model for which we can compute the process analytically. Specifically, we use OT-FM to transport a batch of 5 particles from

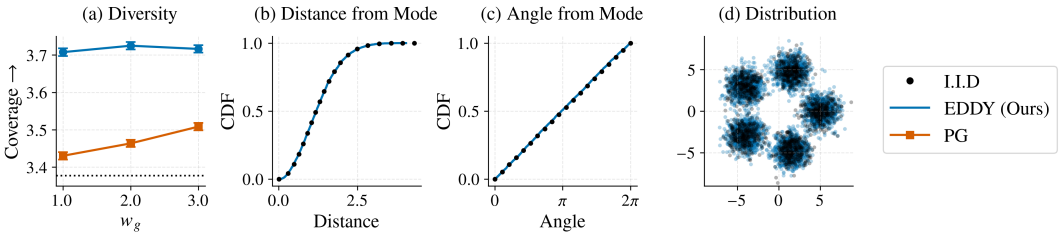

Figure 3: (a) Mean mode coverage of EDDY and PG vs guidance weight w_g. I.I.D baseline (dotted line) denoting the expected coverage for independent particles. (b) Empirical CDF of nearest-mode distances for EDDY ($w_g = 2.0$) versus I.I.D. (c) Empirical CDF of nearest-mode angle for EDDY ($w_g = 2.0$) versus I.I.D. (d) Samples from EDDY and I.I.D.

the prior $p_0 = \mathcal{N}(0, I_2)$ to target p_1 an equal mixture of five Gaussians with unit variance, where the ℓth Gaussian is centered at $5 \cdot \big(\sin(2\pi\ell/5), \cos(2\pi\ell/5)\big)$. We simulate 5,000 batches with 10,000 uniform Euler steps. Appendix B.1 repeats the protocol for a Langevin-corrected SDE sampler.

Evaluation metrics. We measure diversity by the expected mode coverage: how many unique modes are a nearest neighbor for some particle. The baseline coverage for independent (non-interacting) particles can be calculated directly as $5\big[1-(1-\frac{1}{5})^5\big] \approx 3.36$. We test fidelity with two test statistics: the distance of each sample to the nearest Gaussian center and its angle around it. We run Kolmogorov–Smirnov, Mann–Whitney and Welch two-sample tests between EDDY-RBF, PG and I.I.D samples, taking one particle per batch to obtain independent marginal samples.

Experimental results. Figure 3(a) shows that mode coverage increases with EDDY's guidance coefficient w_g outperforming PG. At all operating points EDDY-RBF passes all six tests and its distance and angle CDFs stay indistinguishable from I.I.D (Fig. 3(b,c)).

6.2 Text-to-Image Generation

Experimental Setting. We compare EDDY against PG (Corso et al., 2024) and DF (Morshed & Boddeti, 2025) on both a flow matching model and a latent diffusion model. We additionally evaluate CADS (Sadat et al., 2024), but because it induces diversity by perturbing the conditioning signal rather than by coupling particles, it trades prompt adherence for diversity along a different axis than the methods above; we therefore report it separately in Appendix B.3. We also attempted to compare against CNO (Kim et al., 2026), but official code was unavailable at the time of submission and, despite substantial hyperparameter tuning, our own implementation never reached a configuration that improved diversity while staying competitive in image quality. We therefore exclude it from the main comparison to avoid drawing potentially misleading conclusions. We additionally report that using solely the coupled prior does not change our metrics.

We test on FLUX.1-dev (Black Forest Labs, 2024) and FLUX.2-klein-base-9B (Black Forest Labs, 2026) for flow matching and Stable Diffusion XL (Podell et al., 2024) (SDXL) for diffusion on 2048 randomly sampled validation prompts from MS-COCO (Lin et al., 2014), generating 4 images per caption. During inference, we use EDDY, PG and DF with a RBF kernel on features extracted by composing a fast latent decoder with small DINOv2 (Darcet et al., 2024), which we found to work better than the CLIP Radford et al. (2021) similarity measure. For a fair comparison, we tune each method's hyperparameters to a comparable diversity regime; the chosen values are listed in Table A.3 and the backbone settings in Table A.2 (Appendix B.2). Despite extensive tuning, we could not bring DF into a comparable diversity regime on SDXL, and therefore omit it from that backbone.

Evaluation metrics. We measure diversity by using the DINO pairwise similarity score between images as was done in Corso et al. (2024) with large DINOv2. Quality can be divided into two components: prompt alignment and image quality. We measure prompt alignment via the CLIP-Score (Hessel et al., 2021), and we measure image quality using CMMD (Jayasumana et al., 2024), Aesthetic score LAION-AI Team (2022), and FID Heusel et al. (2017). CMMD and FID are computed by independent samples from the marginal distributions, meaning we only use one particle per batch. To set a FID score for the baseline samples, we compute FID between two non-overlapping sets of samples from the base I.I.D model.

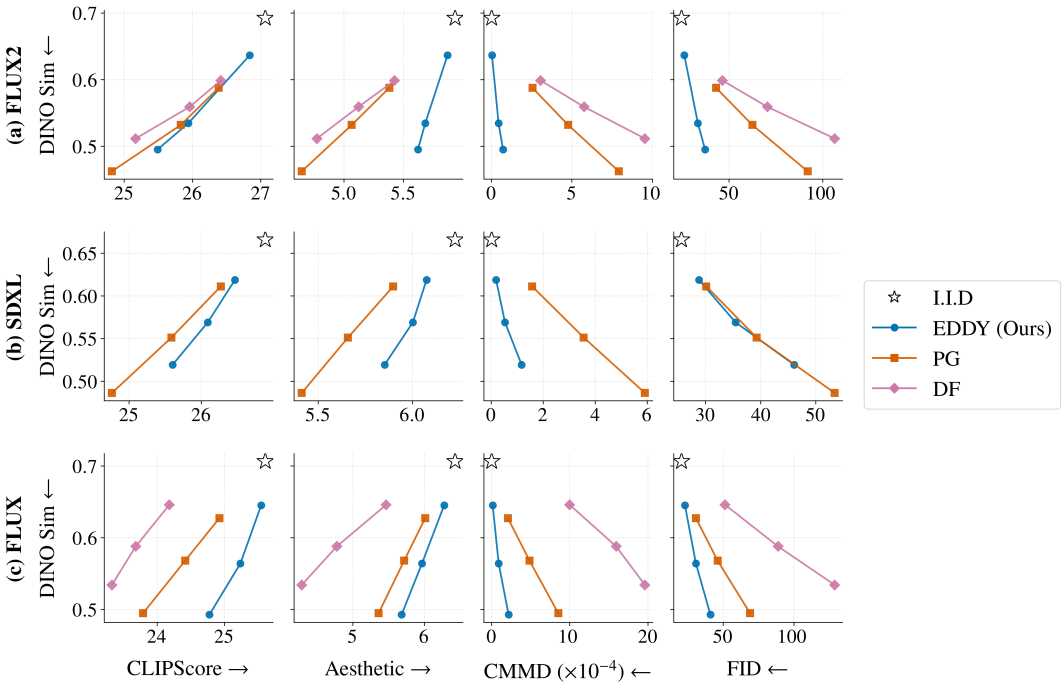

Figure 4: Diversity-quality trade-off curves for EDDY, Particle Guidance (PG), and DiverseFlow (DF) on MS-COCO, for (a) FLUX.2-klein-base-9B, (b) SDXL, and (c) FLUX.1-dev. Each point corresponds to a different guidance strength; the white star marks the I.I.D baseline. The y-axis measures pairwise DINO similarity (← lower = more diverse).

Experimental results. Quantitative results for FLUX.1-dev, SDXL, and FLUX.2-klein-base-9B are presented in Fig. 4, with exact values reported in Tables A.4c, A.4b, and A.4a in the appendix. As shown, EDDY consistently achieves higher image quality than PG and DF at matched diversity levels, across all four quality metrics and all three backbones. Qualitatively, PG and DF introduce pronounced visual artifacts, whereas EDDY at worst exhibits mild blurring, resulting in less perceptual degradation overall. Representative samples are provided in the supplementary material.

7 Limitations

Our approach has several limitations, and the first is theoretical. EDDY does not preserve the marginal exactly, and Claim 2 shows no method in its family could; Section 6.1 measures the marginal bias, but for text-to-image generation the marginal is inaccessible. Second, part of the reported diversity is solver-dependent (Appendix B.1) and need not survive a finer solver. Third, EDDY adds runtime overhead: roughly 14% on FLUX.1-dev, 10% on FLUX.2-klein-base-9B and 6 seconds on SDXL (Table A.6). Finally, we observe occasional blurriness with FLUX.1-dev, a known issue for that model which may reflect a mode of diversity consistent with the data; we do not observe it on SDXL, whose image quality is lower.

8 Conclusion

We asked whether particle guidance can promote diversity without disturbing each sample's own distribution, and proved that it cannot: independently initialized particles are rigid, while coupled particles may disturb the marginal. The closed form is what makes the negative result constructive: it yields EDDY, which improves the diversity–fidelity trade-off at matched diversity.

AI use statement

In this work, we used generative AI tools for proofreading, copyediting, and improving the phrasing of the manuscript. We have not used generative AI tools for generating primary research ideas,

analyzing experimental data, or writing the literature review, and tasks related to generating synthetic datasets are not applicable to this work. Additionally, we used generative AI tools for drafting boilerplate code for data visualization. We have reviewed all AI-assisted work. Specifically, all AI-suggested text edits were manually reviewed by the lead author to ensure original intent was preserved, and all AI-generated code was executed, tested, and verified for accuracy against our raw dataset. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Brian D. O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, May 1982. ISSN 0304-4149. doi: 10.1016/0304-4149(82)90051-5. URL https://www.sciencedirect.com/science/article/pii/0304414982900515.

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in neural information processing systems*, 34:17981–17993, 2021.

Black Forest Labs. FLUX.1 [dev], 2024. URL https://huggingface.co/black-forest-labs/FLUX.1-dev.

Black Forest Labs. FLUX.2, 2026. URL https://huggingface.co/black-forest-labs/FLUX.2-klein-base-9B.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi S. Jaakkola. DiffDock: Diffusion steps, twists, and turns for molecular docking. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=kKF8_K-mBbS.

Gabriele Corso, Yilun Xu, Valentin De Bortoli, Regina Barzilay, and Tommi S. Jaakkola. Particle guidance: non-I.I.D. diverse sampling with diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=KqbCvIFBY7.

Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=2dnO3LLiJ1.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Bradley Efron. Tweedie's Formula and Selection Bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. ISSN 0162-1459. URL https://www.jstor.org/stable/23239562.

Anne Harrington, A Koepke, Shyamgopal Karthik, Trevor Darrell, and Alexei A Efros. It's never too late: Noise optimization for collapse recovery in trained diffusion models. *arXiv preprint arXiv:2601.00090*, 2025.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. CLIPScore: A Reference-free Evaluation Metric for Image Captioning. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 7514–7528, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.595. URL https://aclanthology.org/2021.emnlp-main.595/.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/hash/8a1d694707eb0fefe65871369074926d-Abstract.html.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. URL https://openreview.net/forum?id=qw8AKxfYbI.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Mengjian Hua, Eric Vanden-Eijnden, and Ricky TQ Chen. Simulation-free differential dynamics through neural conservation laws. In *Conference on Uncertainty in Artificial Intelligence*, pp. 1730–1744. PMLR, 2025.

Mohammad Jalali, Cheuk Ting Li, and Farzan Farnia. An information-theoretic evaluation of generative models in learning multi-modal distributions. *Advances in Neural Information Processing Systems*, 36:9931–9943, 2023.

Mohammad Jalali, Haoyu LEI, Amin Gohari, and Farzan Farnia. SPARKE: Scalable prompt-aware diversity and novelty guidance in diffusion models via RKE score. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=1YLpf8nUIq.

Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Rethinking FID: Towards a better evaluation metric for image generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9307–9315, 2024.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in neural information processing systems*, 35:24240–24253, 2022.

Byungjun Kim, Soobin Um, and Jong Chul Ye. Diverse text-to-image generation via contrastive noise optimization. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=EVRMnAREc3.

Michael Kirchhof, James Thornton, Louis Béthune, Pierre Ablin, Eugene Ndiaye, and Marco Cuturi. Shielded diffusion: Generating novel and diverse images using sparse repellency. In *International Conference on Machine Learning*, pp. 30911–30942. PMLR, 2025.

Felix Koulischer, Florian Handke, Johannes Deleu, Thomas Demeester, and Luca Ambrogioni. Feedback guidance of diffusion models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=8ySOcf7UpM.

Tuomas Kynkäänniemi, Miika Aittala, Tero Karras, Samuli Laine, Timo Aila, and Jaakko Lehtinen. Applying guidance in a limited interval improves sample and distribution quality in diffusion models. *Advances in Neural Information Processing Systems*, 37:122458–122483, 2024.

LAION-AI Team. Laion-aesthetics predictor v2, 2022. URL https://github.com/christophschuhmann/improved-aesthetic-predictor.

Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. Diffusion-LM improves controllable text generation. *Advances in neural information processing systems*, 35:4328–4343, 2022.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European conference on computer vision*, pp. 740–755. Springer, 2014.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Xinshuang Liu, Runfa Blark Li, Shaoxiu Wei, and Truong Nguyen. Importance-weighted non-IID sampling for flow matching models. *arXiv preprint arXiv:2511.17812*, 2025.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=CNicRIVIPA.

Jack Lu, Ryan Teehan, and Mengye Ren. ProCreate, don't reproduce! propulsive energy diffusion for creative generation. In *European Conference on Computer Vision*, pp. 397–414. Springer, 2024.

Mashrur M Morshed and Vishnu Boddeti. DiverseFlow: Sample-efficient diverse mode coverage in flows. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 23303–23312, 2025.

Francis J. Narcowich and Joseph D. Ward. Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. *Mathematics of Computation*, 63(208):661–661, January 1994. ISSN 0025-5718. doi: 10.1090/S0025-5718-1994-1254147-6. URL https://www.ams.org/mcom/1994-63-208/S0025-5718-1994-1254147-6/.

Xingyu Ni, Jingrui Xing, Xingqiao Li, Bin Wang, and Baoquan Chen. Representing flow fields with divergence-free kernels for reconstruction. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 8(4):1–21, 2025.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Gaurav Parmar, Or Patashnik, Daniil Ostashev, Kuan-Chieh Jackson Wang, Kfir Aberman, Srinivasa Narasimhan, and Jun-Yan Zhu. Scaling group inference for diverse and high-quality generation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=IyTNxjTuWT.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=di52zR8xgf.

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.

Jack Richter-Powell, Yaron Lipman, and Ricky T. Q. Chen. Neural Conservation Laws: A Divergence-Free Perspective. *Advances in Neural Information Processing Systems*, 35:38075–38088, December 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/f8d39584f87944e5dbe46ec76f19e20a-Abstract-Conference.html.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann M. Weber. CADS: Unleashing the diversity of diffusion models through condition-annealed sampling. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=zMoNrajk2X.

Subham S Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. *Advances in Neural Information Processing Systems*, 37:130136–130184, 2024.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *9th International Conference on Learning Representations ICLR*. OpenReview.net, 2021b.

Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, volume 6.2, pp. 583–603. University of California Press, January 1972.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. GeoDiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=PzcvxEMzvQC.

Appendix of "Diverse Sampling in Diffusion Models with Divergence-Free Particle Guidance"

Table of Contents

A EDDY Theory **15**

A.1 Proofs of Claims 2 and 3 15

A.2 EDDY Algorithm 17

A.2.1 Coupled Prior 17

A.2.2 Complexity 18

A.2.3 Full Procedure 18

A.3 EDDY-RBF Closed-Form Formula 18

B Experiments **20**

B.1 Synthetic ODE Results 20

B.2 Text-to-Image Experiment Details and Results 20

B.3 Comparison to CADS 22

B.4 Text-to-Image Ablation Study 23

C Score Derivation via Tweedie's formula in Flow Matching **24**

A EDDY THEORY

A.1 PROOFS OF CLAIMS 2 AND 3

SETUP

Consider the n-particle system of Section 4,

$$dx_t^i = \big[\mu_t(x_t^i) + \psi_t^i(x_t^i, x_t^{-i})\big]\,dt + \sigma_t\,dW_t^i, \qquad i = 1, \dots, n, \tag{20}$$

with independent Brownian motions $W^1, \dots, W^n$ and $\psi_t^i(\boldsymbol{x}_t) := \mathcal{A}_{p_t}\big(A_t^{(i)}(\boldsymbol{x}_t)\big)$, the Stein operator taken with respect to the single-particle marginal p_t. We assume, as in Section 4:

1. each $A_t^{(i)}(\boldsymbol{x}_t)$ is a Lipschitz anti-symmetric matrix field;
2. μ_t and ψ_t^i are Lipschitz in all arguments, uniformly in t, and have at most linear growth;
3. σ_t is bounded, and $p_t > 0$ everywhere.
4. for each $t > 0$ the joint density is sub-Gaussian together with its first two derivatives: $\boldsymbol{p}_t + \|\nabla \boldsymbol{p}_t\| + \|\nabla^2 \boldsymbol{p}_t\| \le C_t\, e^{-c_t\|\boldsymbol{x}_t\|^2}$ for some $C_t, c_t > 0$.

Assumption (2) and (4) allow us to assume the flux in both claims vanish at infinity. Write $\boldsymbol{p}_t$ for the joint density of $\boldsymbol{x}_t = (x_t^1, \dots, x_t^n)$ on $\mathbb{R}^{nd}$, and let div_i, ∇_i, Δ_i denote the divergence, gradient and Laplacian acting on the x^i block only. Both claims of Section 4 share these assumptions, and the proof of Claim 3 below uses them unchanged.

PROOF OF CLAIM 2

Proof. **Step 1: The joint Fokker–Planck equation.** The tuple $\boldsymbol{x}_t$ evolves under a single SDE on $\mathbb{R}^{nd}$, with velocity field $V_t(\boldsymbol{x}_t) = \big(\mu_t(x_t^i) + \psi_t^i(x_t^i, x_t^{-i})\big)_{i=1}^n$ and Brownian motion $W_t = (W_t^1, \dots, W_t^n)$,

$$d\boldsymbol{x}_t \;=\; V_t(\boldsymbol{x}_t)\,dt \;+\; \sigma_t\,dW_t, \tag{21}$$

so its Fokker–Planck equation is

$$\frac{\partial}{\partial t}\boldsymbol{p}_t \;=\; -\,\mathrm{div}(\boldsymbol{p}_t V_t) \;+\; \frac{\sigma_t^2}{2}\Delta \boldsymbol{p}_t. \tag{22}$$

Step 2: The product law is a solution. Substitute the ansatz $\boldsymbol{p}_t = p_t^{\otimes n}$, with p_t the solution of the single-particle equation 5. Block by block,

$$\mathrm{div}_i\big(\boldsymbol{p}_t\, V_t\big) \;=\; \prod_{j\neq i} p_t(x_t^j)\;\cdot\;\Big(\mathrm{div}_i(p_t\mu_t) \;+\; \cancel{\mathrm{div}_i(p_t\psi_t^i)}\Big)\,, \tag{23}$$

because the factors $j \neq i$ do not depend on x_t^i and pass through div_i, while $p_t\psi_t^i = \mathrm{div}_i(p_t A_t^{(i)})$ makes the second term a double divergence, which vanishes by the mixed-partials cancellation of Claim 1 applied in the x_t^i block. Summing over i and using the linearity of the Laplacian leaves n copies of the single-particle equation, which p_t satisfies by construction.

Step 3: Conclusion. The joint equation inherits its regularity from assumptions (1)–(3), so it admits a unique solution — and the product law is that solution for every t. □

The marginals are preserved — and nothing else happens: the batch is distributed exactly as n independent samples, which is the sampler we were trying to improve on. This holds for every admissible field, so no choice of $A_t^{(i)}$ buys diversity from an i.i.d. start; only the initialization can. Claim 2 leaves the coupled case open, and this is where Claim 3 takes over: we keep the system equation 20 and the assumptions listed with it, drop the product initialization, and locate exactly what obstructs $\tilde{p}_t = p_t$.

MARGINAL NOTATION

Write $x := x^i$ and $y := x^{-i}$ for the i-th particle and its neighbors, $\boldsymbol{p}_t(x, y)$ for the joint density under the modified dynamics and $\tilde{p}_t(x)$ for the marginal of particle i, and factor $\boldsymbol{p}_t(x, y) = \tilde{p}_t(x)\, c_t(y \mid x)$ with c_t the conditional law of the neighbors.

PROOF OF CLAIM 3

Step 1: Marginal evolution and the bias term. The joint density obeys the Fokker–Planck equation 7, which in the notation of this appendix reads

$$\partial_t \boldsymbol{p}_t \;=\; -\sum_{j=1}^{n} \operatorname{div}_j\big(\boldsymbol{p}_t\,(\mu_t + \psi_t^j)\big) \;+\; \frac{\sigma_t^2}{2} \sum_{j=1}^{n} \Delta_j \boldsymbol{p}_t, \tag{24}$$

and we marginalize it over y, that is, integrate it over every block x^j with $j \neq i$. Every such block drops out: writing $\Delta_j = \operatorname{div}_j \nabla_j$, both of its terms are divergences in a variable the integral covers, so

$$\int \operatorname{div}_j\big(\boldsymbol{p}_t\,(\mu_t + \psi_t^j)\big)\, dy \;=\; -\int \boldsymbol{p}_t\,(\mu_t + \psi_t^j)\cdot \nabla_j 1\, dy \;=\; 0,$$
$$\int \Delta_j \boldsymbol{p}_t\, dy \;=\; \int \operatorname{div}_j\big(\nabla_j \boldsymbol{p}_t\big)\, dy \;=\; -\int \nabla_j \boldsymbol{p}_t \cdot \nabla_j 1\, dy \;=\; 0.$$

The $j = i$ terms survive instead, div_i and Δ_i acting on x, which the y-integral treats as a parameter. The base drift factors out of the one integral that is left, leaving

$$\partial_t \tilde{p}_t \;=\; -\operatorname{div}_x\!\left[\mu_t\, \tilde{p}_t \;+\; \int \boldsymbol{p}_t(x, y)\, \psi_t^i(x, y)\, dy\right] \;+\; \tfrac{\sigma_t^2}{2}\, \Delta_x \tilde{p}_t. \tag{25}$$

We now extract the marginal bias from that remaining integral. Substitute $\psi_t^i = \frac{1}{p_t} \operatorname{div}_x(p_t A_t^{(i)})$ and $\boldsymbol{p}_t = \tilde{p}_t\, c_t$, and write

$$\bar{A}_t(x) \;:=\; \int A_t^{(i)}(x, y)\, c_t(y \mid x)\, dy \;=\; \mathbb{E}\big[A_t^{(i)}(x, x_t^{-i}) \mid x_t^i = x\big] \tag{26}$$

for the conditionally averaged field. Then

$$\begin{aligned}
\int \boldsymbol{p}_t\, \psi_t^i\, dy \;&=\; \frac{\tilde{p}_t}{p_t} \int c_t\, \operatorname{div}_x\big(p_t A_t^{(i)}\big) dy \\
&=\; \frac{\tilde{p}_t}{p_t} \left[\int \operatorname{div}_x\big(p_t A_t^{(i)} c_t\big) dy \;-\; \int p_t A_t^{(i)}\, \nabla_x c_t\, dy\right] \\
&=\; \frac{\tilde{p}_t}{p_t} \left[\operatorname{div}_x\big(p_t \bar{A}_t\big) \;-\; p_t \int A_t^{(i)}\, \nabla_x \log c_t\; c_t\, dy\right] \\
&=\; \tilde{p}_t\, \big[\, \mathcal{A}_{p_t}(\bar{A}_t) \;-\; b_t^{(i)}\, \big].
\end{aligned} \tag{27}$$

Line by line: the first pulls $\tilde{p}_t/p_t$ out of the integral, neither factor depending on y; the second applies the product rule $\operatorname{div}_x(p_t A_t^{(i)} c_t) = c_t\, \operatorname{div}_x(p_t A_t^{(i)}) + p_t A_t^{(i)} \nabla_x c_t$; the third moves div_x outside the y-integral, legal because the domain of integration does not depend on x, and substitutes $\nabla_x c_t = c_t \nabla_x \log c_t$; the fourth reads off $\mathcal{A}_{p_t}(\bar{A}_t) = \frac{1}{p_t} \operatorname{div}_x(p_t \bar{A}_t)$ and the marginal bias $b_t^{(i)}$ of equation 10.

Step 2: Only the coupling term survives. Both terms of equation 27 enter the marginal equation 25 inside $\operatorname{div}_x[\,\cdot\,]$, so what decides whether the marginal moves is not the terms themselves but their divergence. Take a time at which $\tilde{p}_t = p_t$, so that $\boldsymbol{p}_t = p_t\, c_t$, and take the two terms in turn.

The first term is inert. $\bar{A}_t$ is an average of anti-symmetric matrices, hence anti-symmetric, and it depends on x alone, so Claim 1 applies to it verbatim:

$$\operatorname{div}_x\big(p_t\, \mathcal{A}_{p_t}(\bar{A}_t)\big) \;=\; \operatorname{div}_x \operatorname{div}_x\big(p_t \bar{A}_t\big) \;=\; 0. \tag{28}$$

The second term is not, and this is where the price is charged. Claim 1 applies just as well to $A_t^{(i)}(\cdot, y)$ for each *fixed* y, which is anti-symmetric in x, so the guidance is pointwise inert:

$$\operatorname{div}_x\big(p_t\,\psi_t^i(\cdot,y)\big) \;=\; \operatorname{div}_x \operatorname{div}_x\big(p_t A_t^{(i)}(\cdot,y)\big) \;=\; 0 \qquad \text{for every fixed } y. \tag{29}$$

The guidance would therefore leave the marginal alone if div_x could be taken inside the y-average. It cannot: that average is weighted by $c_t(y \mid x)$, and the product rule charges for the weight's own x-dependence,

$$\begin{aligned}
\operatorname{div}_x\!\left(\int \boldsymbol{p_t}\,\psi_t^i\,dy\right) \;&=\; \int \operatorname{div}_x\big(p_t\,\psi_t^i\,c_t\big)\,dy \\
&=\; \int \Big[\underbrace{\operatorname{div}_x\big(p_t\,\psi_t^i\big)}_{=0}\,c_t \;+\; p_t\,\psi_t^i\cdot\nabla_x c_t\Big]\,dy \\
&=\; \int p_t\,\psi_t^i\cdot\nabla_x c_t\,dy.
\end{aligned} \tag{30}$$

Comparing equation 30 with equation 27 and equation 28 names what is left:

$$\operatorname{div}_x\big(p_t\,b_t^{(i)}\big) \;=\; -\,p_t\,\mathbb{E}\big[\,\psi_t^i\cdot\nabla_x \log c_t(y\mid x)\;\big|\;x_t^i = x\,\big]. \tag{31}$$

Substituting equation 27 into equation 25, with equation 28 removing the averaged field, gives

$$\partial_t \tilde{p}_t \;=\; -\operatorname{div}\big(p_t(\mu_t - b_t^{(i)})\big) + \tfrac{\sigma_t^2}{2}\Delta p_t \;=\; \partial_t p_t \;+\; \operatorname{div}\big(p_t\,b_t^{(i)}\big), \tag{32}$$

whose right-hand equality subtracts the target's own evolution 5 and is the additive term 9 of Claim 3. The identity is exact and holds at the instant t alone: the marginal agrees with p_t there and is displaced from it at rate $\operatorname{div}(p_t b_t^{(i)})$.

Step 3: The bias is charged on the coupling alone. Every factor of $b_t^{(i)}$ that is not the field itself is the neighbors' conditional score $\nabla_x \log c_t(y \mid x)$, which measures how much the neighbors' law moves when particle i moves. If the particles are independent at time t then $c_t(y \mid x) = c_t(y)$, so that score is identically zero and $b_t^{(i)} \equiv 0$: the displacement in equation 32 vanishes and the marginal is exactly p_t. Claim 2 is the special case in which this holds at every time at once. □

A.2 EDDY ALGORITHM

A.2.1 COUPLED PRIOR

We need an initialization whose joint is not a product, as Claim 2 rules out everything else, but whose every marginal is still exactly $p_0 = \mathcal{N}(0, I_d)$, so that the marginal starts on the exact side of Claim 3 and departs only by accumulating the bias equation 10. The following construction draws independent particles and then pushes them apart by removing their common component.

Claim 4 (Coupled prior)**.** *Let* $x_0^1, \ldots, x_0^n \overset{\text{iid}}{\sim} \mathcal{N}(0, I_d)$ *and define*

$$\tilde{x}_0^i \;=\; \sqrt{\frac{n}{n-1}}\,\big(x_0^i - \bar{x}_0\big) \quad \textit{where} \quad \bar{x}_0 = \frac{1}{n}\sum_i x_0^i. \tag{33}$$

Then each pair $i \neq j$ *is negatively correlated with* $\operatorname{Cov}\big(\tilde{x}_0^i, \tilde{x}_0^j\big) = -\frac{1}{n-1} I_d$ *while each particle's marginal is preserved.*

Proof. Since $\operatorname{Cov}(x_0^i, \bar{x}_0) = \operatorname{Cov}(\bar{x}_0, \bar{x}_0) = \frac{1}{n} I_d$, bilinearity gives $\operatorname{Cov}\big(x_0^i - \bar{x}_0,\, x_0^j - \bar{x}_0\big) = \big(\delta_{ij} - \frac{1}{n}\big) I_d$, which the factor $\frac{n}{n-1}$ rescales to I_d for $i = j$ and to $-\frac{1}{n-1} I_d$ for $i \neq j$. Each $\tilde{x}_0^i$ is a linear image of the jointly Gaussian $(x_0^1, \ldots, x_0^n)$ and has mean zero, so its marginal is exactly $\mathcal{N}(0, I_d) = p_0$. □

Note that for $n = 2$ it reduces to $\tilde{x}_0^2 = -\tilde{x}_0^1$, the classical antithetic-sampling pair.

Because each $\tilde{x}_0^i$ is exactly p_0-distributed, the marginals are exact at $t = 0$ and depart from p_t only through the accumulated marginal bias. The per-particle evaluation protocols of Section 6.2 therefore measure that departure rather than assume it away; Section 6.1 reports its size where the marginal is analytically accessible.

A.2.2 COMPLEXITY

Assume a kernel call (forward or backward) takes $\mathcal{O}(1)$, and that a forward and a backward pass through the feature extractor take F and B respectively[1]. Then EDDY (Algorithm 1) costs $\mathcal{O}\big(n(mF + B) + n^2\big)$ per iteration for the entire batch.

To see this, note that both estimators perturb only x^i: the Hutchinson probes $r_1, \dots, r_m$ of equation 19 are drawn independently of j, so the $2m + 1$ perturbed feature evaluations of particle i are shared across all $n - 1$ of its neighbours, and the kernel gradients required by equation 18 are obtained from two backward passes, giving $\mathcal{O}(mF + B)$ feature-extractor work per particle. Every remaining quantity in equation 17 is an $\mathcal{O}(1)$ kernel call or the equivalent vector algebra, evaluated once per ordered pair, which contributes $\mathcal{O}(n^2)$. Note that the quadratic term involves no network evaluation, so at the particle counts we use the cost is dominated by the $n(mF + B)$ passes and is effectively linear in n.

A.2.3 FULL PROCEDURE

Algorithm 1 EDDY Sampling

Input: drift μ_t, volatility σ_t, kernel k; steps T with $dt = 1/T$; FD step ϵ, Hutchinson probes m; stop ratio sr; particle count n; scaling of EDDY's correction term w_g
Output: n approximate samples from p_1

Sample $z^1, \dots, z^n \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, I_d)$
Set $x_0^i \leftarrow \sqrt{\frac{n}{n-1}}\left(z^i - \frac{1}{n}\sum_{j=1}^n z^j\right)$ for every i
for $t \leftarrow 0$ **to** $1 - dt$ **by** dt **do**
  for $i = 1$ **to** n **do**
    if $t < sr$ **then**
      Compute $\psi^i_{\text{EDDY}} = \frac{1}{n-1}\sum_{j\neq i}\big(A^{(ij)}\, s^{(i)} + K^{(ij)}\, v^{(j)}\big)$; use the FD estimators in Eq. 18 and Eq. 19 if necessary;
    else
      $\psi^i_{\text{EDDY}} \leftarrow \mathbf{0}$
    end
    $x^i_{t+dt} \leftarrow$ SDE step (Eq. 3) with drift $\mu^{(i)} + w_g \cdot \psi^i_{\text{EDDY}}$
  end
end
return $x_1^1, \dots, x_1^n$

A.3 EDDY-RBF CLOSED-FORM FORMULA

Define $\delta^{(ij)} = x^i - x^j$. For the RBF kernel $k(x^i, x^j) = \exp\big(-\|\delta^{(ij)}\|^2/\gamma\big)$ with bandwidth γ, the kernel quantities admit closed forms,

$$r^{(ij)} = -\nabla k(x^i, x^j) = \frac{2}{\gamma}\, k(x^i, x^j)\, \delta^{(ij)}, \tag{34}$$

$$\nabla r^{(ij)} = -\nabla^2 k(x^i, x^j) = \frac{2}{\gamma}\, k(x^i, x^j)\left(I_d - \frac{2}{\gamma}\, \delta^{(ij)}\delta^{(ij)\top}\right), \tag{35}$$

$$\operatorname{div} r^{(ij)} = -\Delta k(x^i, x^j) = \frac{2}{\gamma}\, k(x^i, x^j)\left(d - \frac{2}{\gamma}\|\delta^{(ij)}\|^2\right), \tag{36}$$

substituted into equation 17, give the closed-form variant we call EDDY-RBF,

$$\psi^i_{\text{EDDY}} = \frac{2}{\gamma(n-1)}\sum_{j\neq i} k(x^i, x^j)\big(C^{\delta}_{ij}\, \delta^{(ij)} + C^{v}_{ij}\, v^{(j)}\big), \tag{37}$$

[1] These assumptions were taken with the RBF kernel and DINO in mind.

with coefficients

$$C_{ij}^{\delta} = \frac{2}{\gamma}\langle \delta^{(ij)}, v^{(j)} \rangle - \langle v^{(j)}, s^{(i)} \rangle, \tag{38}$$

$$C_{ij}^{v} = d - 1 - \frac{2}{\gamma}\|\delta^{(ij)}\|^2 + \langle \delta^{(ij)}, s^{(i)} \rangle. \tag{39}$$

B EXPERIMENTS

B.1 SYNTHETIC ODE RESULTS

In the synthetic experiments, we transport a unit Gaussian $\mathcal{N}(0, I_2)$ to an equal mixture of five Gaussians with unit variance where the ℓth Gaussian is centered at $5 \cdot \big(\sin(2\pi\ell/5), \cos(2\pi\ell/5)\big)$. Specifically, we use an optimal transport model:

$$\mu_{\text{ODE}}(x_t) = \mathbb{E}_{x_0 \sim p_0,\, x_1 \sim p_1}[x_1 - x_0 \mid x_t] \tag{40}$$

We simulate 5,000 batches of 5 particles, using 10,000 uniform time steps from $t = 0$ to $t = 1$. We present goodness-of-fit statistical tests in Table A.1

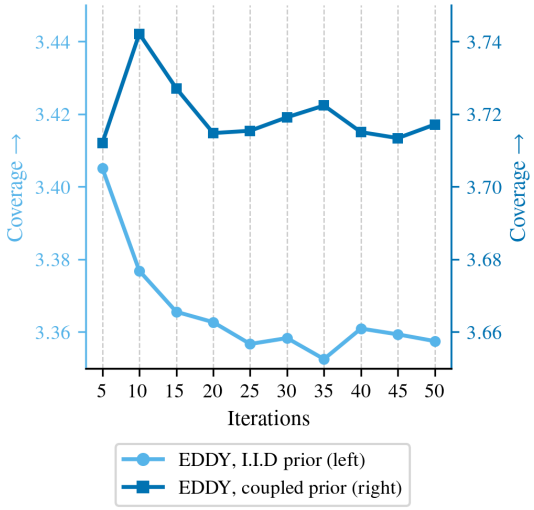

Figure A.1: EDDY with i.i.d and coupled prior over increasing number of iterations. It can be seen that EDDY couples i.i.d initialization in discrete time while theory in continuous time prohibit such behavior.

Table A.1: Goodness-of-fit statistics for the ODE synthetic experiment of Section 6.1 (Fig. 3). We test whether the radial distance and angular position of EDDY and PG samples match I.I.D samples using Kolmogorov–Smirnov (KS), Mann–Whitney (MW), and Welch tests at guidance weights $w_g \in \{1.0, 2.0, 3.0\}$; bold marks $p < 0.05$.

Metric	Method	Test	$w_g = 1.0$		$w_g = 2.0$		$w_g = 3.0$	
			stat	*p*	*stat*	*p*	*stat*	*p*
Distance	EDDY	KS	0.013	0.776	0.013	0.807	0.022	0.194
		MW	0.497	0.663	0.499	0.912	0.491	0.126
		Welch	−0.337	0.736	−0.031	0.975	−1.346	0.178
	PG	KS	0.024	0.123	0.024	0.118	0.019	0.304
		MW	0.492	0.143	0.493	0.225	0.493	0.258
		Welch	−1.188	0.235	−0.953	0.341	−0.656	0.512
Angle	EDDY	KS	0.012	0.837	0.011	0.942	0.015	0.661
		MW	0.502	0.773	0.497	0.648	0.497	0.650
		Welch	0.287	0.774	−0.459	0.647	−0.457	0.648
	PG	KS	0.021	0.239	0.014	0.678	0.020	0.249
		MW	0.506	0.311	0.504	0.509	0.500	0.965
		Welch	1.011	0.312	0.660	0.509	0.016	0.987

B.2 TEXT-TO-IMAGE EXPERIMENT DETAILS AND RESULTS

For the text-to-image experiment, we generate 2048 randomly selected MS-COCO prompts (Lin et al., 2014) using three backbones. Specifically, we use FLUX.1-dev and FLUX.2-klein-base-9B

for OT-FM, and SDXL for LDM. All generated images are of resolution 512×512, where in SDXL case we generate 1024×1024 and downsample to match the its training distribution. Additionally, to implement EDDY's kernel we compose `facebook/dinov2-with-registers-small` over a fast (tiny) decoder. We list the per backbone parameters and fast decoders in Table A.2.

For each guidance algorithm we perturb a single parameter to control the diversity–quality tradeoff and hold the rest fixed, while applying guidance only during the first 20% of the inference iterations. We perturb the guidance weight w_g for EDDY and PG, and the coefficient for DF. The remaining parameters are common to all diversity levels and backbones: EDDY uses $m = 5$ Hutchinson trace samples, $\gamma = 0.05, \epsilon = 0.00046$ for FLUX.2, $\gamma = 0.025, \epsilon = 0.001$ for FLUX.1 and SDXL; PG has no further parameters; and DF uses bandwidth multiplier $h = 0.1$, norm budget $p = 2.0$ and quality clamp $\varepsilon = 0.01$. We could not find a configuration matching the target diversity tiers for DF on SDXL, and therefore omit it from that backbone. Appendix B.3 reports CADS separately, together with its hyperparameters. We list the perturbed parameter for each algorithm, backbone and diversity tier in Table A.3.

The evaluation are reported in Table A.4c, Table A.4a and Table A.4b. FID and CMMD are computed as the mean over ℓ of variation ℓ of the compared algorithm against variation $(\ell+1) \bmod n$ of I.I.D.

Table A.2: Base model generation parameters. The guidance scale refers to CFG in the SDXL case.

Model	**Inference steps**	**Guidance scale**	**Fast Decoder**
FLUX.1-dev	30	3.5	`madebyollin/taef1`
FLUX.2-klein-base-9B	28	3.5	`madebyollin/taef2`
SDXL	20	7.5	`madebyollin/taesdxl`

Table A.3: Chosen hyperparameters per backbone, for each method and diversity level (low, medium and high).

	EDDY (w_g)			PG (w_g)			DF (w_g)		
Model	L	M	H	L	M	H	L	M	H
FLUX.1	45	80	115	4000	5000	6000	1500	2000	2500
FLUX.2	50	130	160	40000	50000	60000	10000	12000	14000
SDXL	80	120	160	1200	1600	2000	–	–	–

Table A.4: Pareto frontiers in (a) FLUX.2 9B Base, (b) SDXL and (c) FLUX.1-dev. DINO Sim represent similarity between pairwise samples (lower = more diverse). FID and CMMD represent quality (lower = higher quality). CLIPScore represent prompt alignment (higher = more aligned).

	Method	DINO Sim ↓	CMMD ($\times 10^{-4}$) ↓	FID ↓	CLIP Score ↑	Aesthetic ↑
	I.I.D	0.693 ± 0.004	0.02 ± 0.00	24.562 ± 0.039	27.066 ± 0.068	5.934 ± 0.010
Low	EDDY (Ours)	0.637 ± 0.004	$\mathbf{0.04} \pm 0.01$	$\mathbf{26.127} \pm 0.160$	$\mathbf{26.840} \pm 0.040$	$\mathbf{5.870} \pm 0.010$
	PG	0.588 ± 0.004	2.56 ± 0.03	43.121 ± 0.760	26.384 ± 0.062	5.381 ± 0.010
	DF	0.599 ± 0.004	3.06 ± 0.02	46.524 ± 0.391	26.415 ± 0.063	5.425 ± 0.010
Medium	EDDY (Ours)	0.535 ± 0.005	$\mathbf{0.45} \pm 0.01$	$\mathbf{33.370} \pm 0.165$	25.941 ± 0.048	$\mathbf{5.683} \pm 0.011$
	PG	0.532 ± 0.004	4.77 ± 0.08	62.468 ± 0.686	25.830 ± 0.060	5.066 ± 0.011
	DF	0.559 ± 0.004	5.76 ± 0.05	70.456 ± 0.492	$\mathbf{25.960} \pm 0.061$	5.123 ± 0.010
High	EDDY (Ours)	0.495 ± 0.005	$\mathbf{0.73} \pm 0.03$	$\mathbf{37.221} \pm 0.703$	$\mathbf{25.489} \pm 0.052$	$\mathbf{5.622} \pm 0.012$
	PG	0.462 ± 0.004	7.92 ± 0.08	92.031 ± 1.787	24.821 ± 0.060	4.646 ± 0.012
	DF	0.512 ± 0.004	9.54 ± 0.05	106.481 ± 1.007	25.169 ± 0.061	4.773 ± 0.010

(a) Pareto frontiers on FLUX.2 9B Base.

	Method	DINO Sim ↓	CMMD ($\times 10^{-4}$) ↓	FID ↓	CLIP Score ↑	Aesthetic ↑
	I.I.D	0.666 ± 0.004	0.02 ± 0.00	25.595 ± 0.073	26.886 ± 0.063	6.227 ± 0.009
Low	EDDY (Ours)	0.619 ± 0.004	$\mathbf{0.19} \pm 0.01$	$\mathbf{28.787} \pm 0.262$	$\mathbf{26.467} \pm 0.036$	$\mathbf{6.075} \pm 0.010$
	PG	0.611 ± 0.004	1.58 ± 0.04	30.131 ± 0.216	26.271 ± 0.061	5.898 ± 0.011
Medium	EDDY (Ours)	0.569 ± 0.004	$\mathbf{0.53} \pm 0.01$	$\mathbf{35.432} \pm 0.302$	$\mathbf{26.091} \pm 0.041$	$\mathbf{6.003} \pm 0.010$
	PG	0.551 ± 0.005	3.56 ± 0.10	39.253 ± 0.563	25.587 ± 0.063	5.658 ± 0.012
High	EDDY (Ours)	0.519 ± 0.004	$\mathbf{1.17} \pm 0.06$	$\mathbf{46.066} \pm 1.373$	$\mathbf{25.603} \pm 0.047$	$\mathbf{5.853} \pm 0.011$
	PG	0.486 ± 0.004	5.92 ± 0.06	53.471 ± 0.351	24.762 ± 0.066	5.413 ± 0.014

(b) Pareto frontiers on SDXL.

	Method	DINO Sim ↓	CMMD ($\times 10^{-4}$) ↓	FID ↓	CLIP Score ↑	Aesthetic ↑
	I.I.D	0.707 ± 0.004	0.01 ± 0.00	20.530 ± 0.105	25.610 ± 0.071	6.430 ± 0.009
Low	EDDY (Ours)	0.645 ± 0.004	$\mathbf{0.16} \pm 0.01$	$\mathbf{23.049} \pm 0.300$	$\mathbf{25.549} \pm 0.038$	$\mathbf{6.278} \pm 0.010$
	PG	0.627 ± 0.004	2.11 ± 0.04	30.822 ± 0.425	24.927 ± 0.066	6.013 ± 0.010
	DF	0.646 ± 0.004	10.00 ± 0.14	51.423 ± 0.624	24.182 ± 0.065	5.467 ± 0.010
Medium	EDDY (Ours)	0.564 ± 0.005	$\mathbf{0.92} \pm 0.03$	$\mathbf{30.809} \pm 0.748$	$\mathbf{25.240} \pm 0.042$	$\mathbf{5.968} \pm 0.013$
	PG	0.568 ± 0.004	4.86 ± 0.08	46.133 ± 0.659	24.418 ± 0.064	5.719 ± 0.011
	DF	0.588 ± 0.004	15.94 ± 0.10	88.871 ± 0.747	23.683 ± 0.061	4.777 ± 0.010
High	EDDY (Ours)	0.493 ± 0.004	$\mathbf{2.19} \pm 0.06$	$\mathbf{41.036} \pm 0.270$	$\mathbf{24.779} \pm 0.044$	$\mathbf{5.683} \pm 0.014$
	PG	0.495 ± 0.004	8.54 ± 0.23	69.051 ± 1.328	23.792 ± 0.065	5.362 ± 0.012
	DF	0.534 ± 0.003	19.62 ± 0.08	128.945 ± 0.360	23.327 ± 0.062	4.287 ± 0.009

(c) Pareto frontiers on FLUX.1-dev.

B.3 Comparison to CADS

CADS (Sadat et al., 2024) induces diversity by a different mechanism than the methods compared in Section 6.2. Rather than coupling the particles of a batch, it anneals noise into the conditioning signal itself, so the samples remain mutually independent and diversity arises from each one being generated under a slightly different condition. This makes the diversity–quality trade-off it traces qualitatively different: the quantity it spends is not image fidelity but prompt adherence. We therefore report it here rather than alongside PG and DF.

We tune the noise scale to the same three diversity tiers used in the main comparison, holding $\tau_1 = 0.7$, $\tau_2 = 0.8$, rescale $=$ False and guidance scale $g = 5.0$ fixed; the chosen noise scales on FLUX.1-dev are 0.05, 0.07 and 0.08. Table A.5 compares CADS against EDDY at matched diversity on FLUX.1-dev.

CADS attains strong image-quality scores — at medium and high diversity it beats EDDY on CMMD, FID and Aesthetic, while its CLIPScore fall significantly from EDDY and I.I.D. Figure 1 shows this directly: for the prompt *"a chef's display of ingredients and pastry creations"*, the CADS samples are photorealistic but showing no chef, and in some photos no pastries.

On SDXL and FLUX.2-klein-base-9B we could not reproduce even this trade-off. Despite extensive hyperparameter tuning, every configuration we found either yielded high diversity at very low prompt adherence or left diversity essentially unchanged from the baseline, with no usable regime in between. We therefore evaluate CADS on FLUX.1-dev only.

Table A.5: CADS versus EDDY at matched diversity tiers on FLUX.1-dev. Bold marks the better of the two per metric. DINO Sim represents similarity between pairwise samples (lower = more diverse); FID and CMMD represent quality (lower = higher quality); CLIPScore represents prompt alignment (higher = more aligned).

	Method	DINO Sim ↓	CMMD ($\times 10^{-4}$) ↓	FID ↓	CLIP Score ↑	Aesthetic ↑
	I.I.D	0.707 ± 0.004	0.01 ± 0.00	20.530 ± 0.105	25.610 ± 0.071	6.430 ± 0.009
Low	EDDY	0.645 ± 0.004	$\mathbf{0.16} \pm 0.01$	$\mathbf{23.049} \pm 0.300$	$\mathbf{25.549} \pm 0.038$	6.278 ± 0.010
	CADS	0.618 ± 0.005	0.31 ± 0.01	25.032 ± 0.107	23.164 ± 0.068	$\mathbf{6.563} \pm 0.010$
Medium	EDDY	0.564 ± 0.005	0.92 ± 0.03	30.809 ± 0.748	$\mathbf{25.240} \pm 0.042$	5.968 ± 0.013
	CADS	0.575 ± 0.005	$\mathbf{0.57} \pm 0.01$	$\mathbf{27.876} \pm 0.271$	22.530 ± 0.074	$\mathbf{6.609} \pm 0.010$
High	EDDY	0.493 ± 0.004	2.19 ± 0.06	41.036 ± 0.270	$\mathbf{24.779} \pm 0.044$	5.683 ± 0.014
	CADS	0.522 ± 0.005	$\mathbf{0.80} \pm 0.02$	$\mathbf{31.420} \pm 0.108$	21.584 ± 0.080	$\mathbf{6.615} \pm 0.010$

B.4 Text-to-Image Ablation Study

Table A.6: Wall-clock overhead (seconds per generation) for each guidance configuration on a NVIDIA A100 40GB PCIe GPU.

Method	SDXL	Flux.1	Flux.2
I.I.D	8.0 ± 0.0s	25.0 ± 0.5s	40.1 ± 0.4s
PG	9.0 ± 0.0s	25.9 ± 0.6s	41.1 ± 0.3s
DF	–	25.1 ± 0.6s	40.9 ± 0.4s
EDDY	14.0 ± 0.0s	28.7 ± 0.7s	44.0 ± 0.0s

Table A.7: Wall-clock overhead (seconds per generation) for EDDY for varied number of particles on a NVIDIA A100 40GB PCIe GPU.

n	SDXL	Flux.1	Flux.2
3	9.9 ± 0.4s	23.2 ± 0.4s	34.1 ± 0.3s
4	14.0 ± 0.0s	28.7 ± 0.7s	44.0 ± 0.0s
5	19.9 ± 0.3s	35.3 ± 0.7s	53.0 ± 0.0s
6	25.0 ± 0.0s	41.1 ± 0.5s	63.0 ± 0.0s

Table A.8: Change in each metric relative to I.I.D when prior is coupled but no guidance applied, with 512 MS-COCO prompts tested. Bold: $p < 0.05$, Welch z-test.

Backbone	DINO pairwise sim	Aesthetic score	CLIPScore
SDXL	-0.0041	$\mathbf{-0.1338}$	$+0.0126$
FLUX.1	-0.0001	-0.0140	$+0.1684$
FLUX.2	-0.0005	-0.0101	-0.0046

C SCORE DERIVATION VIA TWEEDIE'S FORMULA IN FLOW MATCHING

First, let us state Tweedie's formula Efron (2011): given clean x and estimation $\hat{x} = x + \epsilon$ with independent noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I_d)$ it follows that $\mathbb{E}[x \mid \hat{x}] = \hat{x} + \sigma^2 \nabla \log p(\hat{x})$.

In training, a flow matching model is trained to predict $x_1 - x_0$ given $x_t = tx_1 + (1-t)x_0$ for noise $x_0 \sim p_0 = \mathcal{N}(0, I_d)$ and data point $x_1 \sim p_1$. Thus, given a state x_t the model will estimate $\mathbb{E}[x_1 - x_0 \mid x_t]$ for some unknown x_1.

From one hand, using Tweedie's formula over "clean" tx_1 and estimation x_t with independent noise $(1-t)x_0 \sim \mathcal{N}\big(0, (1-t)^2 I_d\big)$, we get

$$\mathbb{E}[tx_1 \mid x_t] = x_t + (1-t)^2 \nabla \log p_t(x_t) \tag{41}$$

From the other hand, it easily follows from the x_t's definition that $tx_1 = t(1-t)(x_1 - x_0) + tx_t$, so by taking expectation conditioned on x_t we get

$$\mathbb{E}[tx_1 \mid x_t] = t(1-t)\,\mathbb{E}[x_1 - x_0 \mid x_t] + tx_t \tag{42}$$

Putting it all together

$$\nabla \log p_t(x_t) = \frac{t\mathbb{E}[x_1 - x_0 \mid x_t] - x_t}{1-t} \tag{43}$$