FLINT: Coupling Proximal Initialization and Bounded Stochasticity for Flow-Matching Inverse Problems

Junseo Bang[1] **Daewon Choi**[2] **Dong Ju Mun**[1] **Se Young Chun**[1,2,3†]
[1]Dept. of ECE [2]IPAI [3]INMC & AIIS, Seoul National University, Republic of Korea
{qkdwnstj10, eodnjschl, dongju1126, sychun}@snu.ac.kr

Abstract

Flow matching models have emerged as a dominant paradigm for generative posterior sampling in imaging inverse problems. Existing solvers follow two distinct lines of work: trajectory-space methods, which refine sampling dynamics from random initial noise, and source-space methods, which optimize the initial noise itself. While most prior work treats these as independent design choices, we show that they are inherently coupled. We derive a unified stability bound that links initialization proximity and cumulative sampling stochasticity through a multiplicative envelope, complemented by two empirical observations. First, we identify a sufficiency regime in which a moderately optimized proximal source initialization is enough to attain the performance plateau, rendering exhaustive source-space optimization unnecessary. Second, this initial advantage is fragile: it is eroded by excessive stochasticity accumulation during guided sampling, necessitating a mechanism that bounds the cumulative noise injection. These findings motivate FLINT, a two-stage framework that operationalizes this coupling. FLINT establishes a high-fidelity starting point via a manifold-constrained proximal objective in only a few iterations, then performs guided sampling under a bounded stochasticity schedule that helps prevent erasure of the initialization advantage. Across diverse inverse problems, FLINT delivers state of the art perceptual quality and competitive distortion metrics while maintaining superior computational efficiency.

1 Introduction

Inverse problems arise throughout computational imaging and computer vision [5, 22, 37, 34, 67], where the goal is to recover a clean image $\mathbf{x}_0 \in \mathbb{R}^d$ from a degraded measurement $\boldsymbol{y} = \mathcal{A}\boldsymbol{x}_0 + \mathbf{n}$, with $\mathcal{A}$ denoting the forward operator and $\mathbf{n} \sim \mathcal{N}(0, \sigma_y^2\mathbf{I})$ additive Gaussian noise. Since $\mathcal{A}$ is typically many-to-one, the problem is ill-posed and requires a prior over natural images [55, 48, 57, 56]. From a Bayesian perspective, this reduces to sampling from the posterior $p(\mathbf{x}_0 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x}_0) \cdot p(\mathbf{x}_0)$, where the likelihood $p(\mathbf{y} \mid \mathbf{x}_0)$ encodes the forward model and the prior $p(\mathbf{x}_0)$ regularizes solutions toward the natural image manifold. Deep generative models have emerged as powerful priors with diffusion models [21, 54] and flow matching models [33, 32]. Large-scale rectified flow models [18] achieve photorealistic generation quality, rendering them highly attractive for posterior sampling.

Existing inverse problem solvers leveraging these priors typically fall into two categories (Sec. 2.2). *Trajectory-space methods* [11, 52, 25, 51, 66, 27, 17, 42, 29, 40, 3] steer the sampling path toward the measurement-consistent manifold at each step but rely on a fixed, random starting point $\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, leaving reconstruction quality inherently sensitive to the specific realization of the initial noise (Fig. 1(a) and (b)). *Source-space methods* [8, 60, 4, 44, 23, 62, 65, 39] perform inference directly on the initial noise $\boldsymbol{x}_1$, keeping the subsequent generative trajectory unmodified (Fig. 1(c) and (d)). Most prior work has treated initialization and sampling dynamics as decoupled modules, an implicit

† corresponding author.

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

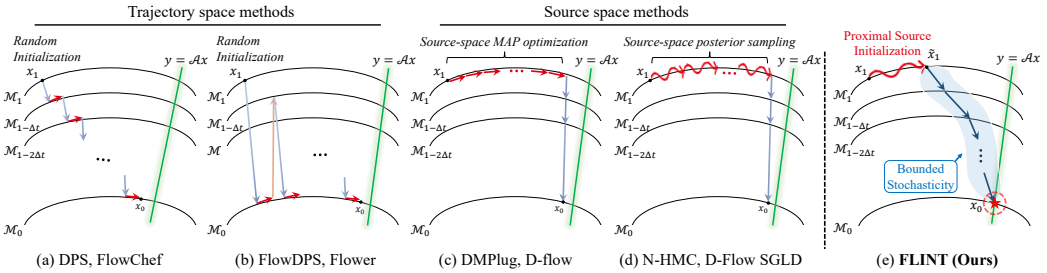

Figure 1: **Illustration of generative paradigms for inverse problems**. (a) and (b) depict trajectory-space methods, while panels (c) and (d) depict source-space optimization methods. (a) Likelihood score corrections are applied directly at each noisy intermediate state $\mathbf{x}_t$. (b) Likelihood score corrections are applied in clean image space via the Tweedie estimate $\hat{\mathbf{x}}_{0|t}$, followed by renoising to the generative manifold at the next timestep. (c) Source-space MAP methods reparameterize the clean image via the deterministic generative map Φ as $\mathbf{x}_0 = \Phi(\mathbf{x}_1)$ and optimize over the initial noise $\mathbf{x}_1$ to minimize data fidelity. (d) Source-space posterior sampling draws samples from $p(\mathbf{x}_1 \mid \mathbf{y})$ via Hamiltonian Monte Carlo or Langevin dynamics. (e) By strategically coupling proximal source initialization with bounded stochasticity during guided posterior sampling, FLINT bounds the sampling trajectory, preserving the initialization advantage and yielding high-fidelity convergence.

assumption that overlooks the coupling between initialization proximity and sampling stochasticity, leaving their joint impact on reconstruction quality largely unexplored.

In this work, we reveal that these two design choices are not independent but inherently coupled, and that exploiting their synergy is beneficial for high-fidelity posterior sampling. Combining a stability bound that decomposes into a source-space initialization term and a cumulative stochasticity term with empirical observations, we identify two complementary findings. First, a proximal source initialization that anchors the starting noise within a sufficient proximity of a source-space solution can attain high-quality reconstruction without exhaustive optimization. Second, the benefits of this proximal start are fragile. They are eroded by the excessive accumulation of stochasticity during guided posterior sampling, necessitating a mechanism to bound the cumulative injected noise. Together, these findings lead to *FLINT* (Flow-based posterior sampling with proximal source initialization and bounded stochasticity), a two-stage framework that operationalizes this coupling. Stage 1 implements a well-informed starting point by optimizing a manifold-constrained proximal objective for only a few iterations. Stage 2 then employs a bounded stochasticity schedule to steer the trajectory while preventing the erasure of the informed initialization (Fig. 1(e)). Our main contributions are:

- **Theoretical Coupling:** We derive a unified stability bound for flow-based posterior sampling, showing that the terminal deviation from a deterministic reference is controlled by initialization proximity and cumulative sampling stochasticity. We empirically identify a sufficiency regime, where moderate source-space optimization is enough to reach the performance plateau.
- **FLINT Framework:** We introduce FLINT, a training-free two-stage sampler that bridges source-space and trajectory-space methods. It utilizes a manifold-constrained proximal objective for informed initialization and a bounded stochasticity schedule that helps preserve the initialization advantage during guided posterior sampling.
- **Experimental Results:** Extensive experiments across imaging inverse problems show that FLINT achieves state-of-the-art perceptual quality and competitive or superior distortion metrics, with a markedly favorable efficiency-performance profile.

2 Background

2.1 Rectified Flow

Rectified Flow [33] transports between the data distribution π_0 and the noise distribution $\pi_1 = \mathcal{N}(\mathbf{0}, \mathbf{I})$ via an ODE, where $t = 1$ corresponds to pure noise and $t = 0$ to the data space. Training pairs $(\boldsymbol{x}_0, \boldsymbol{x}_1) \sim \pi_0 \times \pi_1$ are linearly interpolated as $\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$, and a velocity network

$\boldsymbol{v}_\theta$ is trained to match the straight-line direction $(\boldsymbol{x}_1 - \boldsymbol{x}_0)$ as follows.

$$\min_\theta \mathbb{E}_{t\sim\mathcal{U}[0,1],\, \boldsymbol{x}_0\sim\pi_0,\, \boldsymbol{x}_1\sim\pi_1} \left[\|(\boldsymbol{x}_1 - \boldsymbol{x}_0) - \boldsymbol{v}_\theta(\boldsymbol{x}_t, t)\|_2^2 \right]. \tag{1}$$

Samples are generated by integrating the ODE from $t = 1$ to $t = 0$ over a fixed schedule $\{\tau_i^{(N)}\}_{i=1}^N$ with $N \in [28, 50]$ steps in practice. We denote the resulting deterministic generative map as $\Phi : \mathbb{R}^d \to \mathbb{R}^d$, so that $\boldsymbol{x}_0 = \Phi(\boldsymbol{x}_1)$. Rearranging the linear interpolation at any intermediate state $\boldsymbol{x}_t$ yields a flow-based Tweedie's formula $\hat{\boldsymbol{x}}_{0|t} := \mathbb{E}[\boldsymbol{x}_0 \mid \boldsymbol{x}_t] \approx \boldsymbol{x}_t - t\,\boldsymbol{v}_\theta(\boldsymbol{x}_t, t)$, which provides a single step clean image estimate from any intermediate state [26].

2.2 Posterior Sampling for Inverse Problems

Given the measurement model $\boldsymbol{y} = \mathcal{A}\boldsymbol{x}_0 + \mathbf{n}$, posterior sampling aims to draw samples from $p(\boldsymbol{x}_0 \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{x}_0)\, p(\boldsymbol{x}_0)$ [54, 53, 11, 52, 12]. Existing solvers built on diffusion or flow matching priors divide into two families. *Trajectory-space methods* steer intermediate states toward measurement consistency via the Tweedie estimate $\hat{\boldsymbol{x}}_{0|t} := \mathbb{E}[\boldsymbol{x}_0 \mid \boldsymbol{x}_t]$, obtaining the posterior mean either via Tweedie's formula or by integrating a probability flow ODE from $\boldsymbol{x}_t$ toward $t = 0$. Within this family, one subfamily enforces data consistency directly in the noisy latent space by approximating $\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y} \mid \boldsymbol{x}_t) \approx \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y} \mid \hat{\boldsymbol{x}}_{0|t})$ [11, 52, 25, 10, 47, 46, 41, 30], while the other enforces consistency in clean image space via $\hat{\boldsymbol{x}}_{0|t}$ and then renoises back to the $(t+\Delta t)$-manifold of the generative prior [51, 66, 28, 27, 17, 42, 40, 29, 3]. Both subfamilies initialize from $\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, leaving the quality of the starting point entirely uncontrolled. *Source-space methods* instead confine to the initial noise $\boldsymbol{x}_1$ via the deterministic generative map $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ defined by the full reverse ODE, leaving the generative trajectory unmodified. One line of work casts source-space inference as MAP estimation [8, 60, 4], reparameterizing $\boldsymbol{x}_0 = \Phi(\boldsymbol{x}_1)$ and optimizing $\boldsymbol{x}_1$ to minimize data fidelity. Another targets $p(\boldsymbol{x}_1 \mid \boldsymbol{y})$ directly via Langevin dynamics, Hamiltonian Monte Carlo, or related MCMC updates [44, 23, 62, 65, 39], enabling broader posterior exploration. Both subfamilies nonetheless share a substantial computational burden: evaluating $\nabla_{\boldsymbol{x}_1} \mathcal{J}(\Phi(\boldsymbol{x}_1))$ requires backpropagating through the full N-step reverse ODE.

2.3 Reducing the Computational Cost of Source-Space Optimization

Source-space methods optimize $\boldsymbol{x}_1$ by minimizing a data consistency objective $\mathcal{J}(\boldsymbol{x}_0)$. Computing $\nabla_{\boldsymbol{x}_1} \mathcal{J}(\boldsymbol{x}_0)$ incurs two compounding costs: obtaining $\boldsymbol{x}_0$ requires an N-step reverse ODE, and backpropagating through this chain requires evaluating the Jacobian product $\prod_i \frac{\partial \boldsymbol{x}_{i-1}}{\partial \boldsymbol{x}_i}$.

Truncated sampling. To reduce the forward pass cost, we adopt truncated sampling with $K \ll N$ steps [49, 31]. The flow ODE is integrated for the first $K-1$ steps along the N-step time schedule $\{\tau_i^{(N)}\}_{i=1}^N$, after which a clean image estimate is formed via the flow-based Tweedie's formula:

$$\boldsymbol{x}_{t_{\text{stop}}} = \boldsymbol{x}_1 - \textstyle\sum_{i=1}^{K-1} (\tau_i - \tau_{i+1})\, \boldsymbol{v}_\theta(\boldsymbol{x}_{\tau_i},\, \tau_i) \quad \hat{\boldsymbol{x}}_0 = \boldsymbol{x}_{t_{\text{stop}}} - t_{\text{stop}} \cdot \boldsymbol{v}_\theta(\boldsymbol{x}_{t_{\text{stop}}},\, t_{\text{stop}}), \tag{2}$$

where $\tau_i = \tau_i^{(N)}$ and $t_{\text{stop}} = \tau_{K-1}$. Denoising to t_{stop} before applying Tweedie's formula yields an efficient surrogate $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_1)$.

Efficient gradient computation. The backpropagation cost has been mitigated by strategies such as adjoint methods and gradient checkpointing [38, 58, 7, 43, 14], though these often trade memory for computation. Shortcut Diffusion Optimization (SDO) [16] reinterprets sampling as Picard fixed-point iteration and shows that $\nabla_{\mathbf{x}_1} \mathcal{J}(\mathbf{x}_0)$ is well-approximated by $\frac{\partial \mathcal{J}}{\partial \mathbf{x}_0} \frac{\partial \mathbf{x}_{N-1}}{\partial \mathbf{x}_1}$, retaining the computational graph only at the first denoising step with bounded error. Our method combines truncated sampling and the SDO shortcut, as detailed in Algorithm 2, to substantially reduce per-iteration overhead.

3 Analysis: Coupling of Proximal Initialization and Sampling Stochasticity

This section analyzes how source-space proximity and sampling stochasticity jointly affect an upper bound on the mean-square terminal deviation from a reference trajectory during guided sampling. An initial noise placed near the minimizer of the posterior energy $\mathcal{E}(\boldsymbol{x}_1) \coloneqq -\log p(\boldsymbol{x}_1 \mid \boldsymbol{y})$ admits an

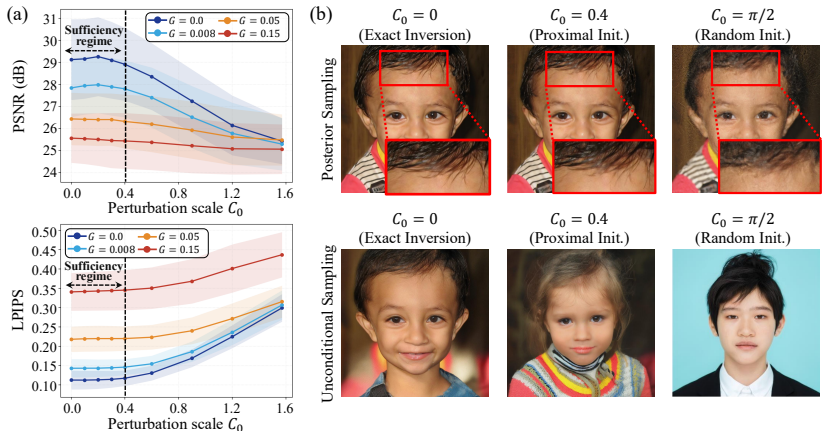

Figure 2: **Interplay of proximal initialization and sampling stochasticity** on Gaussian deblurring. (a) PSNR and LPIPS as functions of the perturbation scale C_0 across varying the cumulative stochasticity G, demonstrating the existence of a sufficiency regime ($C_0 \leq 0.4$ in this setup) where proximal initialization performs on par with exact inversion. (b) Visual comparisons for $C_0 \in \{0, 0.4, \pi/2\}$. While a representative proximal initialization within this regime ($C_0 = 0.4$) matches the high fidelity of exact inversion under posterior sampling ($G = 0.008$), it preserves measurement-consistent structures even in unconditional sampling, unlike the unrelated output from random initialization.

energy upper bound controlled by its distance to the minimizer, whereas a random Gaussian draw typically lies far from this minimizer in high dimensions and thus yields a looser distance-based upper bound. This argument is formalized in Appendix A.1. Whether such an initialization advantage is retained during sampling depends on how the proximity gap propagates through the sampling trajectory, which we address by analyzing the reverse-time posterior flow SDE [2, 54]:

$$d\boldsymbol{x}_t \;=\; \left[v_t(\boldsymbol{x}_t \mid \boldsymbol{y}) - \frac{g_t^2}{2}\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t \mid \boldsymbol{y})\right] dt \;+\; g_t\, d\bar{\boldsymbol{W}}_t, \tag{3}$$

where $v_t(\boldsymbol{x}_t \mid \boldsymbol{y})$ is the posterior velocity, $\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t \mid \boldsymbol{y})$ is the posterior score, and g_t is the diffusion coefficient controlling stochastic departure.

Proposition 3.1 (Terminal stability bound)**.** *Let $\boldsymbol{x}_1^* \in \mathbb{R}^d$ be any reference point in the d-dimensional source space, $\boldsymbol{x}_0^*$ the terminal output of the drift ODE of* (3) *with only the Brownian term removed which is initialized at $\boldsymbol{x}_1^*$, and $\Phi_g(\boldsymbol{x}_1)$ the terminal output of* (3) *initialized at $\boldsymbol{x}_1$. Let $G := \int_0^1 g_t^2\, dt$ denote the cumulative stochasticity. Suppose the posterior velocity satisfies a one-sided Lipschitz condition with parameter $m_v(t)$ (Assumption A.4) and the posterior score satisfies a one-sided Lipschitz condition with parameter $m_s(t)$ (Assumption A.5). Then any $\tilde{\boldsymbol{x}}_1 \in \mathbb{R}^d$ satisfying $\|\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1^*\|_2 \leq \delta$ obeys:*

$$\mathbb{E}\big[\|\Phi_g(\tilde{\boldsymbol{x}}_1) - \boldsymbol{x}_0^*\|_2^2\big] \leq 2\, e^{2M_+}\left(\underbrace{\delta^2}_{\substack{\text{initialization}\\ \text{error term}}} + \underbrace{d \cdot G}_{\substack{\text{stochastic}\\ \text{departure term}}}\right),\; M_+ := \max\Big(0,\; \sup_{s\in[0,1]} \int_0^s m(u)\, du\Big), \tag{4}$$

where $m(t) := m_v(t) + \frac{g_t^2}{2}\, m_s(t)$ is the effective contraction rate of the posterior drift. (Proof in Appendix A.2.)

Remark 1 (Canonical reference: prior-typical and measurement-consistent initialization)**.** A canonical instantiation of Proposition 3.1 selects $\boldsymbol{x}_1^*$ as a reference point that is simultaneously prior typical and measurement consistent, satisfying $A(\Phi(\boldsymbol{x}_1^*)) \approx \boldsymbol{y}$ under the unconditional generative map Φ. Proposition 3.1 neither bounds reconstruction error relative to the ground truth nor establishes convergence to the exact posterior.

Theoretical insights. Proposition 3.1 places the initialization error and the stochastic departure in a single additive bound under a common prefactor e^{2M_+}, yielding two key insights independent of the choice of reference. *(i) Sufficiency of proximal initialization.* When $\delta^2 \ll d \cdot G$, the stochastic departure dominates the bound, so further reducing δ has only a small relative effect on the upper

bound. Achieving sufficient proximity thus controls the initialization term adequately without exhaustive optimization. *(ii) Attenuation by excessive stochasticity.* Increasing G inflates the stochastic departure term $d \cdot G$, so that, for a fixed M_+, the bound grows with G regardless of how well δ has been reduced. Bounding G throughout sampling therefore helps preserve the benefit of any principled starting point.

We explore the two insights empirically under the setting of Remark 1. Using FlowDPS [26] as the posterior sampler on Gaussian deblurring (FFHQ [24], 100 samples), we instantiate $\boldsymbol{x}_1^*$ as the unconditional ODE inversion of the ground truth image and construct perturbed initializations $\tilde{\boldsymbol{x}}_1$ via spherical linear interpolation (SLERP) between $\boldsymbol{x}_1^*$ and a Gaussian noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$:

$$\tilde{\boldsymbol{x}}_1 = \frac{\sin(\theta - C_0)}{\sin\theta}\, \boldsymbol{x}_1^* + \frac{\sin C_0}{\sin\theta}\, \boldsymbol{\epsilon}, \qquad \theta = \cos^{-1}\left(\frac{\boldsymbol{x}_1^* \cdot \boldsymbol{\epsilon}}{\|\boldsymbol{x}_1^*\|_2 \|\boldsymbol{\epsilon}\|_2}\right), \tag{5}$$

where $C_0 \in [0, \pi/2]$ controls the perturbation scale. This construction approximately preserves the Gaussian typical norm $\|\tilde{\boldsymbol{x}}_1\|_2 \approx \sqrt{d}$, providing a clean empirical counterpart to the initialization error term in (4). Figure 2(a) visualizes the perception-distortion landscape as a joint function of the perturbation scale C_0, reflecting δ^2 along the horizontal axis, and the cumulative stochasticity G across the curve family.

Sufficiency of proximal source initialization. Across all G curves in Figure 2(a), reconstruction quality remains nearly flat within an initial proximity range before it begins to degrade. The exact extent of this flat region (here, up to $C_0 \approx 0.4$) depends on the specific setup. We refer to this empirically observed region as the "sufficiency regime". Within this regime, further reducing δ provides negligible gain. Figure 2(b) qualitatively corroborates this. Under posterior sampling ($G = 0.008$), a moderately proximal initialization from this regime (e.g., $C_0 = 0.4$) achieves reconstructions visually comparable to exact inversion ($C_0 = 0$), whereas random initialization ($C_0 = \pi/2$) incurs a substantial quality drop. Unconditional sampling further confirms that the proximal noise encodes measurement-consistent structure, while the random noise produces an unrelated image.

Attenuation by excessive stochasticity. Turning to the curve family in Figure 2(a), increasing G degrades reconstruction quality. This is qualitatively consistent with Proposition 3.1: as G grows, the stochastic departure term $d \cdot G$ rises directly, so the bound increases regardless of how well δ has been reduced. Consequently, a well chosen initialization alone is insufficient. Bounding G throughout guided sampling helps preserve the advantage conferred by proximal noise initialization. Additional empirical validation is provided in Appendix B.

4 Method: The FLINT Framework

The analysis in Sec. 3 motivates the two-stage design of FLINT. Since exhaustive source space optimization is redundant under the sufficiency regime (Sec. 3), Stage 1 (Sec. 4.1) efficiently anchors the initial noise via a manifold constrained proximal objective with spectrally weighted data consistency. Stage 2 (Sec. 4.2) then employs a bounded stochasticity schedule during guided posterior sampling, preventing excessive stochasticity from eroding the fidelity attained by this initialization (Sec. 3).

4.1 Stage 1: Proximal Source Initialization

Manifold-constrained proximal objective. Unlike source space MAP methods that incur massive computational costs for a deterministic exact solution, or MCMC approaches that exhaustively characterize the source posterior distribution, Stage 1 focuses solely on rapidly attaining sufficient proximity, a measurement-informed initialization statistically consistent with the generative prior. To derive a principled objective, we employ a variational relaxation, minimizing the KL divergence from a surrogate distribution $q_\phi(\boldsymbol{x}_1) = \mathcal{N}(\boldsymbol{x}_1 \mid \boldsymbol{\mu}_\phi, \rho^2 \mathbf{I})$ to the intractable source posterior $p(\boldsymbol{x}_1 \mid \boldsymbol{y})$:

$$\begin{aligned} &\min_{\phi}\ \mathrm{KL}(q_\phi(\boldsymbol{x}_1) \parallel p(\boldsymbol{x}_1 \mid \boldsymbol{y})) \\ &= \underbrace{-\mathbb{E}_{q_\phi}[\log p(\boldsymbol{y} \mid \boldsymbol{x}_1)]}_{\text{data fidelity}} + \underbrace{\mathrm{KL}(q_\phi(\boldsymbol{x}_1) \parallel p(\boldsymbol{x}_1))}_{\text{prior regularization}} + \log p(\boldsymbol{y}) \end{aligned} \tag{6}$$

Algorithm 1 The FLINT solver for flow-based inverse problems

Stage 1: Source-Space Proximal Initialization

Require: $\boldsymbol{y}$, $\mathcal{A}$, v_θ, β_{KL}, λ_{cum}

Initialize $\boldsymbol{\mu}_\phi \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $s = 1, \ldots, S$ **do**
  $\boldsymbol{x}_1 \leftarrow \boldsymbol{\mu}_\phi + \rho^{(s)}\boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\hat{\boldsymbol{x}}_0 \leftarrow \widehat{\Phi}_{\mathrm{tr}}(\boldsymbol{x}_1)$ ▷ Algorithm 2
  $\mathcal{J} \leftarrow \ell_{\mathrm{IR}}(\boldsymbol{y}, \hat{\boldsymbol{x}}_0) + \beta_{\mathrm{KL}}\,\mathrm{KL}(q_\phi(\boldsymbol{x}_1)\|p(\boldsymbol{x}_1)) + \lambda_{\mathrm{cum}}\,\mathcal{R}(\boldsymbol{\mu}_\phi)$ ▷ ℓ_{IR}: Eq. (8)
  $\boldsymbol{\mu}_\phi \leftarrow \mathrm{Adam}(\nabla_{\boldsymbol{\mu}_\phi}\mathcal{J})$
end for
return $\tilde{\boldsymbol{x}}_1 = \boldsymbol{\mu}_\phi$

Stage 2: Proximity-Preserving Posterior Sampling with Bounded Stochasticity

Require: $\tilde{\boldsymbol{x}}_1$, $\boldsymbol{y}$, $\mathcal{A}$, v_θ, λ, c, $c_\emptyset$, $\eta(t)$ ▷ $\eta(t)$: Eq. (10)

$\boldsymbol{x}_1 \leftarrow \tilde{\boldsymbol{x}}_1$
for $t : 1 \to 0$ **do**
  $v_t \leftarrow v_\theta(\boldsymbol{x}_t, c_\emptyset) + \lambda\left(v_\theta(\boldsymbol{x}_t, c) - v_\theta(\boldsymbol{x}_t, c_\emptyset)\right)$
  $\hat{\boldsymbol{x}}_{0|t} \leftarrow \boldsymbol{x}_t - \sigma_t v_t$, $\hat{\boldsymbol{x}}_{1|t} \leftarrow \boldsymbol{x}_t + (1-\sigma_t)v_t$
  $\tilde{\boldsymbol{x}}_{0|t} \leftarrow \arg\min_{\boldsymbol{x}} \mathcal{L}\big(\mathcal{A}(\boldsymbol{x}),\, \boldsymbol{y}\big)$ ▷ init. at $\hat{\boldsymbol{x}}_{0|t}$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\boldsymbol{x}_{t+\Delta t} \leftarrow (1-\sigma_{t+\Delta t})\tilde{\boldsymbol{x}}_{0|t} + \sigma_{t+\Delta t}\hat{\boldsymbol{x}}_{1|t} + \eta(t)\boldsymbol{\epsilon}$
end for
return $\boldsymbol{x}_0$

where $\log p(\boldsymbol{y})$ is constant. Crucially, the standard KL term only constrains the global norm of $\boldsymbol{\mu}_\phi$ but cannot preclude higher-order departures from Gaussianity in source space. To encourage the optimized noise adheres to the prior manifold, we augment the objective with a higher order cumulant regularizer $\mathcal{R}(\boldsymbol{\mu}_\phi)$ that penalizes non-Gaussian drift during optimization (Appendix C.1).

Combining this regularizer with the computationally efficient truncated sampling $\widehat{\Phi}_{\mathrm{tr}}$ (Sec. 2.3) and the data consistency loss (Sec. 4.1), we formulate the manifold-constrained proximal objective:

$$\mathcal{J}(\phi) = \mathbb{E}_{\boldsymbol{x}_1 \sim q_\phi}\Big[\ell_{\mathrm{IR}}(\boldsymbol{y},\, \widehat{\Phi}_{\mathrm{tr}}(\boldsymbol{x}_1))\Big] + \beta_{\mathrm{KL}} \cdot \mathrm{KL}(q_\phi(\boldsymbol{x}_1) \,\|\, p(\boldsymbol{x}_1)) + \lambda_{\mathrm{cum}} \cdot \mathcal{R}(\boldsymbol{\mu}_\phi), \tag{7}$$

where the expectation is approximated via a single sample Monte Carlo estimator $\boldsymbol{x}_1 = \boldsymbol{\mu}_\phi + \rho\boldsymbol{\epsilon}$ with a fixed annealing schedule for ρ. Gradients of $\mathcal{J}(\phi)$ are computed via the SDO shortcut combined with $\widehat{\Phi}_{\mathrm{tr}}$, yielding a low per-iteration cost. Within just a few iterations, this optimization yields a proximity-initialized noise, a starting point residing in a low energy posterior region.

Spectrally weighted data consistency. The truncated ODE estimate $\hat{\boldsymbol{x}}_0 = \widehat{\Phi}_{\mathrm{tr}}(\boldsymbol{x}_1)$ is spectrally biased: with few reverse sampling steps, high-frequency components remain largely unresolved [9], making data consistency gradients unreliable over the full spectrum and destabilizing the proximal optimization. We address this by estimating frequency dependent reliability with the Wiener filter framework [63]. For the flow-matching interpolant $\boldsymbol{x}_{t_{\mathrm{stop}}} = (1 - t_{\mathrm{stop}})\boldsymbol{x}_0 + t_{\mathrm{stop}}\boldsymbol{\epsilon}$, the signal and noise powers at frequency ω are $(1 - t_{\mathrm{stop}})^2 S_{x_0}(\omega)$ and t_{stop}^2, respectively, where $S_{x_0}(\omega) \propto \|\omega\|^{-\alpha}$ follows the well-established pink noise model for natural images [50]. Applying the Wiener filter formula directly yields the following data consistency loss:

$$\ell_{\mathrm{IR}}(\boldsymbol{y}, \hat{\boldsymbol{x}}_0) = \|W(t_{\mathrm{stop}}) \odot \mathcal{F}[A^*(A\hat{\boldsymbol{x}}_0 - \boldsymbol{y})]\|_2^2,\ W(\omega, t_{\mathrm{stop}}) = \frac{(1 - t_{\mathrm{stop}})^2 S_{x_0}(\omega)}{(1 - t_{\mathrm{stop}})^2 S_{x_0}(\omega) + t_{\mathrm{stop}}^2}, \tag{8}$$

Here, $\mathcal{F}$ and A^* denotes the 2D Fourier transform, and the adjoint of the forward operator, respectively. $A^*(A\hat{\boldsymbol{x}}_0 - \boldsymbol{y})$ back-projects the measurement residual to the image domain before spectral filtering. The weight $W \in [0, 1]$ emphasizes reliable low frequencies and decays toward 0 at high frequencies. By suppressing gradients from unreliable frequency components, Eq. (8) provides stable low-frequency supervision, yielding faster convergence than full-spectrum data consistency.

4.2 Stage 2: Proximity-Preserving Posterior Sampling with Bounded Stochasticity

Stage 2 starts from the proximity-initialized source $\tilde{\boldsymbol{x}}_1 = \boldsymbol{\mu}_\phi$ produced by Stage 1, which provides a measurement-informed starting point. Its role is not to rediscover a measurement-consistent low-frequency structure from scratch, but rather to perform the proximity-preserving stochastic refinement along the sampling trajectory. We adopt the framework of the second sub-family of trajectory-space methods described in Sec. 2.2. At each reverse step, this approach obtains a clean image estimate via the flow-based Tweedie's formula, applies a data consistency correction in the clean image space, and renoises the corrected estimate onto the generative prior manifold at the next timestep. Concretely, the next state is computed as:

$$\boldsymbol{x}_{t+\Delta t} = (1 - \sigma_{t+\Delta t})\,\tilde{\boldsymbol{x}}_{0|t}(\boldsymbol{y}) + \sigma_{t+\Delta t}\,\hat{\boldsymbol{x}}_{1|t} + \eta(t)\,\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{9}$$

where σ_t denotes the noise level, and the update reinterpolates the corrected clean estimate $\tilde{\boldsymbol{x}}_{0|t}(\boldsymbol{y})$ and noise estimate $\hat{\boldsymbol{x}}_{1|t}$ back onto the $(t + \Delta t)$-manifold while injecting stochasticity scale $\eta(t)$.

Table 1: **Quantitative comparison on a flow matching model** averaged over 1,000 FFHQ and 800 DIV2K samples using 50 NFEs under Gaussian noise ($\sigma_y = 0.03$).

	Super-Resolution $\times 8$				Super-Resolution $\times 12$				Motion Deblurring				Gaussian Deblurring			
Method	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓
FFHQ (768 × 768)																
FlowChef	27.94	0.773	104.19	0.438	26.21	0.725	170.39	0.554	25.42	0.733	88.78	0.478	26.62	0.748	103.13	0.488
FlowDPS	27.08	0.711	28.91	0.379	26.66	0.736	30.55	0.354	25.02	0.688	45.54	0.427	25.80	0.699	47.21	0.420
FLAIR	27.58	0.697	29.73	0.285	26.77	0.709	24.75	0.307	29.39	0.755	**13.83**	0.259	28.68	0.744	15.24	0.259
Flower	28.36	0.793	35.97	0.351	27.15	0.762	76.44	0.444	26.38	0.756	71.93	0.4410	26.96	0.759	86.03	0.462
FlowLPS	26.67	0.691	30.41	0.386	25.84	0.685	62.16	0.429	25.80	0.715	21.25	0.365	26.49	0.718	23.46	0.364
Ours	**29.63**	**0.795**	**21.60**	**0.217**	**27.74**	**0.763**	**24.26**	**0.268**	**30.77**	**0.799**	14.95	**0.186**	**29.88**	**0.783**	**15.14**	**0.196**
DIV2K (768 × 768)																
FlowChef	22.58	0.586	89.97	0.578	21.14	0.516	141.43	0.693	20.31	0.510	106.78	0.647	21.08	0.517	133.41	0.675
FlowDPS	21.60	0.511	39.23	0.503	20.84	0.483	47.37	0.554	19.79	0.453	55.84	0.599	20.35	0.456	61.20	0.614
FLAIR	21.11	0.484	**30.70**	0.385	19.92	0.439	43.12	0.461	24.16	0.644	17.33	0.303	22.74	0.572	27.73	0.378
Flower	22.86	0.595	41.19	0.479	**21.69**	**0.546**	71.51	0.592	21.16	0.544	80.38	0.581	21.56	0.538	91.03	0.622
FlowLPS	21.78	0.516	34.05	0.450	20.82	0.479	58.35	0.507	20.64	0.507	33.12	0.451	20.98	0.491	39.14	0.465
Ours	**23.10**	**0.598**	33.21	**0.356**	20.35	0.496	**35.01**	**0.432**	**26.55**	**0.728**	**15.69**	**0.197**	**24.13**	**0.629**	**25.39**	**0.301**

Bounded stochasticity schedule. Proposition 3.1 motivates limiting stochastic accumulation but does not uniquely determine how the stochasticity budget should be allocated over time. We design the following stochasticity schedule:

$$\eta(t) = \lambda_s \cdot \sqrt{1 - \sigma_t} \cdot \sigma_t, \tag{10}$$

where λ_s controls the overall magnitude. The shape is deliberately chosen to endow Eq. (9) with a proximity-preserving role. At early timesteps $t \approx 1$ ($\sigma_t \approx 1$), $\eta(t)$ remains close to zero, suppressing stochastic perturbations during the phase where the coarse image layout is determined in flow-based models. Large perturbations at this stage would erode the low-energy proximal state established by Stage 1. As t decreases to 0, $\eta(t)$ again decays to zero, effectively concentrating stochasticity in the intermediate timesteps where perceptual details are refined without disturbing the global structure. With $\lambda_s < 1$, the bell-shaped profile of $\eta(t)$ keeps $\int_0^1 \eta^2(t)\,\mathrm{d}t$ bounded, aiming to preserve the initialization advantage throughout guided sampling. The overall algorithm is summarized in Algorithm 1.

5 Experiments

5.1 Experimental Setup

Datasets and metrics. Experiments are conducted on FFHQ [24] (1024^2, 1k samples) and DIV2K [1] (2K, 800 samples), with PSNR, SSIM [61], LPIPS [68], and patch-based FID [20] (256×256 patches, InceptionV3) as evaluation metrics. Additionally, for ablation studies conducted on a smaller subset (100 samples), we report patch-based KID [6].

Baselines. We compare against recent flow matching approaches: FlowChef [41], FlowDPS [26], FLAIR [17], Flower [42], and FlowLPS [40], all built upon Stable Diffusion 3.5-Medium (768^2, NFE 50). For FLINT, the same 50 NFE budget is split into 21 NFEs for Stage 1 (K=3 truncated steps $\times$ 7 iterations) and 29 NFEs for Stage 2. Baseline configurations are detailed in Appendix E.

Problem setting. We evaluate on several linear inverse problems: SR$\times$8, $\times$12, motion deblurring with a 61×61 kernel at intensity 0.5, and Gaussian deblurring with a 61×61 kernel at standard deviation 3.0. All tasks incorporate additive Gaussian noise with $\sigma_y = 0.03$ (corresponding to 1.5% of the normalized pixel range). For text conditioning, a fixed prompt ("A high quality photo of a face") is used for FFHQ, while image adaptive captions from DAPE [64] are used for DIV2K.

5.2 Experimental Results

Flow matching model. As shown in Table 1, FLINT consistently outperforms all flow matching baselines, yielding favorable results across both distortion and perceptual quality metrics. Fig. 3 corroborates these results, with structural artifacts of baseline reconstructions visibly suppressed. Extended evaluations on diffusion model backbones (Stable Diffusion 1.5) are provided in Appendix D.

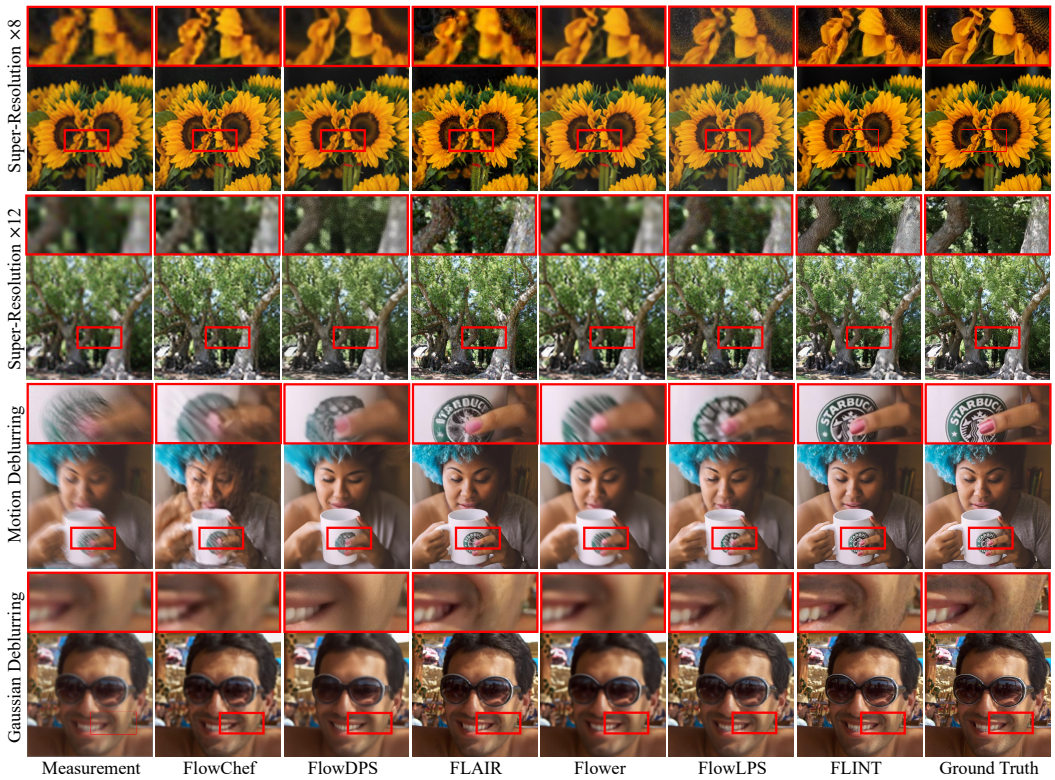

Figure 3: **Qualitative comparison** for FFHQ and DIV2K datasets.

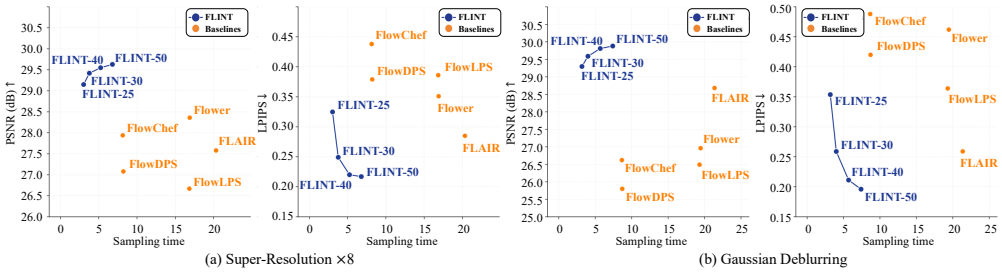

Figure 4: **Computational cost and performance.** PSNR and LPIPS vs. sampling time (sec) for SR×8 and Gaussian deblurring. Numbers after method names denote its NFEs.

Table 2: **Quantitative comparison on a diffusion model** over 1,000 FFHQ samples using 200 NFEs under additive Gaussian noise ($\sigma_y = 0.03$).

	Super-Resolution ×4				Motion Deblurring			
Method	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓
PSLD	26.07	0.712	46.69	0.309	25.06	0.703	59.61	0.367
ReSample	25.96	**0.849**	50.97	0.225	24.18	0.765	92.86	0.397
DDPG	29.73	0.809	84.12	0.305	26.62	0.752	94.41	0.358
TReg	25.22	0.756	37.35	0.244	25.42	0.701	45.38	0.274
Ours	**29.88**	0.776	**22.36**	**0.161**	**30.05**	**0.791**	**33.32**	**0.182**

Computational cost and performance. Fig. 4 evaluates SR ×8 and Gaussian deblurring on 1000 FFHQ samples, varying Stage 2 over $\{4, 9, 19, 29\}$ steps (total budgets $\{25, 30, 40, 50\}$). With as few as 30 NFEs, FLINT surpasses nearly all baselines at the full 50 NFE budget on both PSNR and LPIPS. This advantage stems from the proximal initialization. Stage 1 requires only a small number of optimization iterations at low cost enabled by Sec. 2.3, which reduces the computational burden of Stage 2 by positioning the initial noise in a favorable region of the source space, requiring only a

Table 3: **Ablation studies** on Gaussian deblurring (FFHQ, 100 samples). (a) Stage 1: Components include MCPO and SWDC. (b) Stage 2: Performance of the BSS profile across λ_s.

Conf.	MCPO	SWDC	PSNR↑	SSIM↑	KID↓	LPIPS↓
A	✗	–	28.56	0.641	0.015	0.327
B	✓	✗	29.60	0.744	0.008	0.218
C	✓	✓	30.11	0.788	0.005	0.192

Schedule	λ_s	PSNR↑	SSIM↑	KID↓	LPIPS↓
front-loaded	0.3	29.60	0.769	0.022	0.193
	0.8	29.19	0.727	0.008	0.230
	1.3	28.44	0.639	0.018	0.438
	1.8	28.32	0.629	0.027	0.499
Bell-shaped	0.3	29.59	0.771	0.031	0.208
	0.8	30.11	0.788	0.005	0.192
	1.3	29.88	0.784	0.024	0.305
	1.8	29.62	0.768	0.026	0.328

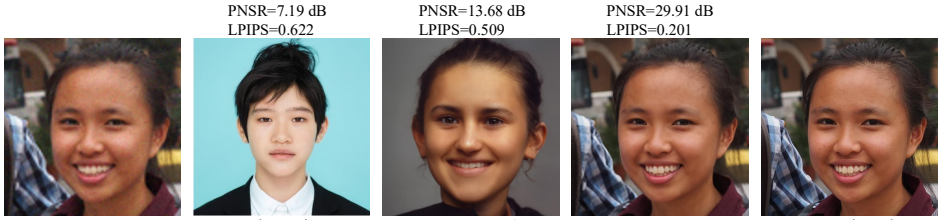

Figure 5: **Initialization quality** on SR×8. Unconditional sampling from the Stage 1 output (third column) demonstrates that FLINT's optimized noise recovers faithful coarse structures, contrasting with the unrelated results from random initialization (second column). Starting from this informed point, Stage 2 achieves high-fidelity reconstruction. Metrics are averaged over 100 FFHQ samples.

short guided trajectory to reach the subspace consistent with the measurement. Extended runtime comparisons are provided in Appendix F.2.

Diffusion model. To assess applicability beyond flow matching, we further evaluate FLINT on Stable Diffusion 1.5 (512^2, 200 NFEs) against PSLD [46], ReSample [51], DDPG [19], and TReg [28]. As shown in Table 2, FLINT achieves the best PSNR, FID, and LPIPS on SR×4 and motion deblurring, as well as the best SSIM on the latter task. Detailed settings and qualitative comparisons are provided in Appendix D, with baseline configurations in Appendix E.2.

5.3 Ablation Studies

Effectiveness of stage 1: MCPO and SWDC. Table 3a isolates the contribution of each component within Stage 1, the Manifold-Constrained Proximal Objective (MCPO) and Spectrally Weighted Data Consistency (SWDC). Comparing Conf. A (random) against Conf. C (full proximal) confirms that the posterior sampling started from proximal-initialized noise yields substantial gains across all metrics. Figure 5 visually corroborates this: unconditional sampling from the proximal noise recovers a coarse structure consistent with the measurement, whereas the same procedure from a random Gaussian draw produces an entirely unrelated image. Incorporating the SWDC (B to C) improves reconstruction quality by suppressing gradients from unreliable high-frequency components during the optimization, ensuring stable convergence.

Sensitivity to the bounded stochasticity schedule (BSS). Table 3b ablates the stochasticity schedule along two axes: the magnitude λ_s and the temporal shape of $\eta(t)$, comparing our bell-shaped schedule (Eq. 10) against a front-loaded alternative $\eta_{\text{mono}}(t) = \lambda_s \sqrt{1-(1-\sigma_t)^2}\,\sigma_t$ that decreases monotonically over $t : 1 \to 0$ and peaks at $t = 1$. Two key trends emerge. (i) Magnitude: While Sec. 3 warns against excessive stochasticity (large $\lambda_s \geq 1.3$) due to the amplified stochastic departure, suppressing it ($\lambda_s \to 0.3$) limits the FLINT's ability to perform stochastic exploration concentrated at middle timesteps, resulting in suboptimal performance. The moderate setting ($\lambda_s = 0.8$) achieves the optimal balance, bounded to preserve the low-energy proximal state yet active enough to achieve a high-fidelity reconstruction. (ii) Shape: at matched λ_s, the bell-shaped schedule generally yields superior performance, particularly as λ_s increases, as front loading disrupts the proximal noise precisely when coarse structure is being established.

5.4 Discussion

The theoretical and empirical findings in Sec. 3 predict that proximal initialization reduces the initialization gap, while bounded stochasticity preserves this advantage throughout guided sampling. Fig. 6 corroborates this prediction by evaluating the reconstruction quality of Tweedie's estimate along the sampling trajectory (visualizations in Appendix F.3). FLINT enters the measurement-consistent subspace earlier than baselines and sustains higher reconstruction quality through convergence. The Appendix further provides extensive analyses on the effect of the cumulant regularizer (Appendix F.4), comparisons with alternative initialization strategies including forward noising initialization [36] and flow inversion in both deterministic [15] and stochastic variants [45] (Appendix F.5), the effect of prolonged source space optimization (Appendix F.6), and a posterior variance analysis confirming the absence of mode collapse (Appendix F.7).

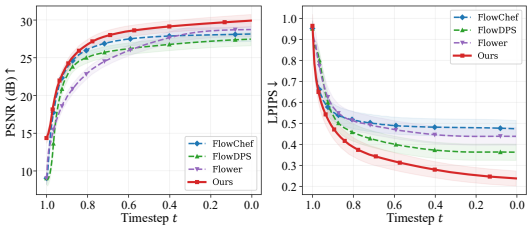

Figure 6: **Quantitative comparison of the sampling process** via Tweedie's estimates.

6 Conclusion

We introduce FLINT, a two-stage framework that bridges source-space and trajectory-space methods by exploiting their coupling, offering a new design principle for flow-based inverse solvers. Through a unified stability analysis, we motivate coupling proximal source initialization with bounded stochasticity to preserve the initialization advantage during posterior sampling. FLINT employs a manifold-constrained proximal objective followed by a strategic stochasticity schedule, consistently improving reconstruction quality across diverse inverse problems at a competitive sampling cost.

Limitation. FLINT inherits the limitations of the underlying pretrained generative models, including training data biases and reduced capacity for out-of-distribution images. Reconstruction quality may degrade on scenes with high structural complexity ($e.g.$, multiple objects or instances), where the increased ill-posedness makes it difficult to simultaneously satisfy data fidelity and prior consistency.

Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)], the National Research Foundation of Korea(NRF) grants funded by the Korea government(MSIT) (No. RS-2026-25585456, No. RS-2025-02263628), and Samsung Electronics Co., Ltd. (IO251217-14748-01).

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, 2017.

[2] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

[3] Junseo Bang, Dong Ju Mun, Hoigi Seo, Seongmin Hong, and Se Young Chun. Triadic dynamics aware diffusion posterior sampling for inverse problems: Optimizing guidance and stochasticity schedules. In *ICML*, 2026.

[4] Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. In *ICML*, 2024.

[5] Mario Bertero, Patrizia Boccacci, and Christine De Mol. *Introduction to inverse problems in imaging*. CRC press, 2021.

[6] Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. In *International Conference on Learning Representations*, 2018.

[7] Tianqi Chen, Bing Xu, Chiyuan Zhang, and Carlos Guestrin. Training deep nets with sublinear memory cost. *arXiv preprint arXiv:1604.06174*, 2016.

[8] Hamadi Chihaoui, Abdelhak Lemkhenter, and Paolo Favaro. Blind image restoration via fast diffusion inversion. *NeurIPS*, 2024.

[9] Jooyoung Choi, Jungbeom Lee, Chaehun Shin, Sungwon Kim, Hyunwoo Kim, and Sungroh Yoon. Perception prioritized training of diffusion models. In *CVPR*, 2022.

[10] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *CVPR*, 2023.

[11] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *ICLR*, 2023.

[12] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *NeurIPS*, 2022.

[13] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *CVPR*, 2022.

[14] Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *ICLR*, 2024.

[15] Yingying Deng, Xiangyu He, Changwang Mei, Peisong Wang, and Fan Tang. Fireflow: Fast inversion of rectified flow for image semantic editing. In *ICML*, 2025.

[16] Hongkun Dou, Zeyu Li, Xingyu Jiang, Hongjue Li, Lijun Yang, Wen Yao, and Yue Deng. You only look one step: Accelerating backpropagation in diffusion sampling with gradient shortcuts. *IEEE TPAMI*, 2025.

[17] Julius Erbach, Dominik Narnhofer, Andreas Robert Dombos, Bernt Schiele, Jan Eric Lenssen, and Konrad Schindler. Solving inverse problems with FLAIR. In *NeurIPS*, 2025.

[18] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *ICML*, 2024.

[19] Tomer Garber and Tom Tirer. Image restoration by denoising diffusion models with iteratively preconditioned guidance. In *CVPR*, 2024.

[20] 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. *NeurIPS*, 2017.

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 2020.

[22] Jari P Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*. Springer, 2005.

[23] Adhithyan Kalaivanan, Zheng Zhao, Jens Sjölund, and Fredrik Lindsten. Ess-flow: Training-free guidance of flow-based models as inference in source space. *arXiv preprint arXiv:2510.05849*, 2025.

[24] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.

[25] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *NeurIPS*, 35, 2022.

[26] Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. Flowdps : Flow-driven posterior sampling for inverse problems. In *ICCV*, 2025.

[27] Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. Flowdps: Flow-driven posterior sampling for inverse problems. In *ICCV*, 2025.

[28] Jeongsol Kim, Geon Yeong Park, Hyungjin Chung, and Jong Chul Ye. Regularization by texts for latent diffusion inverse solvers. In *ICLR*, 2025.

[29] Minwoo Kim, Seunghyeok Shin, and Hongki Lim. FAST-DIPS: Adjoint-free analytic steps and hard-constrained likelihood correction for diffusion-prior inverse problems. In *ICLR*, 2026.

[30] Haechang Lee, Dong Ju Mun, Hyunwoo Lee, Kyung Chul Lee, Seongmin Hong, Gwanghyun Kim, Seung Ah Lee, and Se Young Chun. Fidelity-preserving zero-shot diffusion models for highly ill-posed inverse problems in lensless imaging. *Image and Vision Computing*, 165:105786, 2026.

[31] Sangyun Lee, Yilun Xu, Tomas Geffner, Giulia Fanti, Karsten Kreis, Arash Vahdat, and Weili Nie. Truncated consistency models. In *ICLR*, 2025.

[32] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *ICLR*, 2023.

[33] Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *ICLR*, 2023.

[34] Michael T McCann, Kyong Hwan Jin, and Michael Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Process. Mag.*, 2017.

[35] Peter McCullagh. *Tensor methods in statistics: Monographs on statistics and applied probability*. Chapman and Hall/CRC, 2018.

[36] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *ICLR*, 2022.

[37] Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. Deep learning techniques for inverse problems in imaging. *IEEE JSAIT*, 2020.

[38] Jiachun Pan, Jun Hao Liew, Vincent Tan, Jiashi Feng, and Hanshu Yan. AdjointDPM: Adjoint sensitivity method for gradient backpropagation of diffusion probabilistic models. In *ICLR*, 2024.

[39] Meet Hemant Parikh, Yaqin Chen, and Jian-Xun Wang. D-flow sgld: Source-space posterior sampling for scientific inverse problems with flow matching. *arXiv preprint arXiv:2602.21469*, 2026.

[40] Jonghyun Park and Jong Chul Ye. Flowlps: Langevin-proximal sampling for flow-based inverse problem solvers. *arXiv preprint arXiv:2512.07150*, 2025.

[41] Maitreya Patel, Song Wen, Dimitris N. Metaxas, and Yezhou Yang. Flowchef: Steering of rectified flow models for controlled generations. In *ICCV*, 2025.

[42] Mehrsa Pourya, Bassam El Rawas, and Michael Unser. FLOWER: A flow-matching solver for inverse problems. In *ICLR*, 2026.

[43] Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation. *arXiv preprint arXiv:2310.03739*, 2023.

[44] Vishal Purohit, Matthew Repasky, Jianfeng Lu, Qiang Qiu, Yao Xie, and Xiuyuan Cheng. Posterior sampling via langevin dynamics based on generative priors. *arXiv preprint arXiv:2410.02078*, 2024.

[45] Litu Rout, Yujia Chen, Nataniel Ruiz, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Semantic image inversion and editing using rectified stochastic differential equations. In *ICLR*, 2025.

[46] Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. Solving linear inverse problems provably via posterior sampling with latent diffusion models. *NeurIPS*, 2023.

[47] François Rozet, Gérôme Andry, François Lanusse, and Gilles Louppe. Learning diffusion priors from observations by expectation maximization. *NeurIPS*, 2024.

[48] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 1992.

[49] Youssef Shehata, Benjamin Holzschuh, and Nils Thuerey. Improved sampling of diffusion models in fluid dynamics with tweedie's formula. In *ICLR*, 2025.

[50] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 2001.

[51] Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. Solving inverse problems with latent diffusion models via hard data consistency. In *ICLR*, 2024.

[52] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *ICLR*, 2023.

[53] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *ICLR*, 2022.

[54] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021.

[55] Andrei N Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Sov Dok*, 1963.

[56] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, 2018.

[57] Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *IEEE Global Conference on Signal and Information Processing*, 2013.

[58] Bram Wallace, Akash Gokul, Stefano Ermon, and Nikhil Naik. End-to-end diffusion latent optimization improves classifier guidance. In *ICCV*, 2023.

[59] Wolfgang Walter. *Ordinary Differential Equations*, volume 182 of *Graduate Texts in Mathematics*. Springer, 1st edition, 1998.

[60] Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun. Dmplug: A plug-in method for solving inverse problems with diffusion models. *NeurIPS*, 2024.

[61] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 2004.

[62] Zifan Wang, Alice Harting, Matthieu Barreau, Michael M. Zavlanos, and Karl Henrik Johansson. Source-guided flow matching. In *ICLR*, 2026.

[63] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. The MIT press, 1949.

[64] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *CVPR*, 2024.

[65] Yingzhi Xia, Setthakorn Tanomkiattikun, Liangli Zhen, and ZAIWANG GU. Noise-adaptive diffusion sampling for inverse problems without task-specific tuning. In *ICLR*, 2026.

[66] Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *CVPR*, 2025.

[67] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 2017.

[68] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

Appendix

A Theoretical Analysis: Assumptions, Propositions, and Proofs

A.1 Posterior Energy Bound of Principled Initialization

This appendix formalizes the interpretation introduced in Remark 1 (Section 3): when the reference point $\boldsymbol{x}_1^*$ in Proposition 3.1 is taken to be the posterior energy minimizer, source-space proximity controls a likelihood energy upper bound, while a random Gaussian draw typically lies farther away in source space. A principled initialization refers to an initial noise placed deliberately close to the source-space solution, as opposed to a random draw from the prior. We define the posterior energy as the negative log-posterior up to a constant:

$$E(\boldsymbol{x}_1) \;:=\; -\log p(\boldsymbol{x}_1 \mid \boldsymbol{y}) + C \;=\; \underbrace{\frac{1}{2\sigma_y^2}\|\boldsymbol{y} - \mathcal{A}(\Phi(\boldsymbol{x}_1))\|_2^2}_{E_{\text{like}}(\boldsymbol{x}_1)} \;+\; \underbrace{\frac{1}{2}\|\boldsymbol{x}_1\|_2^2}_{E_{\text{prior}}(\boldsymbol{x}_1)}, \tag{S1}$$

and choose a source-space solution $\boldsymbol{x}_1^* \in \arg\min_{\boldsymbol{x}_1} E(\boldsymbol{x}_1)$. Throughout the Section A.1, $\boldsymbol{n}$ denotes the reference residual, defined by $\boldsymbol{n} \;:=\; \boldsymbol{y} - \mathcal{A}(\Phi(\boldsymbol{x}_1^*))$.

We state the assumptions underlying Proposition A.1.

Assumption A.1 (Lipschitz generative map)**.** *The generative map $\Phi\colon \mathbb{R}^d \to \mathbb{R}^d$ is L-Lipschitz continuous: for all $\boldsymbol{x}_1, \boldsymbol{x}_1' \in \mathbb{R}^d$,*

$$\|\Phi(\boldsymbol{x}_1) - \Phi(\boldsymbol{x}_1')\|_2 \le L\,\|\boldsymbol{x}_1 - \boldsymbol{x}_1'\|_2. \tag{S2}$$

Assumption A.2 (Lipschitz forward operator)**.** *The forward operator $\mathcal{A}$ is L_A-Lipschitz continuous in a neighbourhood $\mathcal{U}$ of $\Phi(\boldsymbol{x}_1^*)$: for all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{U}$,*

$$\|\mathcal{A}(\boldsymbol{u}) - \mathcal{A}(\boldsymbol{v})\|_2 \le L_A\,\|\boldsymbol{u} - \boldsymbol{v}\|_2. \tag{S3}$$

When $\mathcal{A}$ is linear, $L_A = \|\mathcal{A}\|_{\text{op}}$ (the largest singular value of $\mathcal{A}$). For nonlinear operators, a finite local Lipschitz constant can be chosen on a sufficiently small neighbourhood whenever $\mathcal{A}$ is continuously differentiable.

Assumption A.3 (Typical-set norm of a source-space solution)**.** *The source-space solution $\boldsymbol{x}_1^*$ satisfies*

$$\|\boldsymbol{x}_1^*\|_2 = \Theta(\sqrt{d}). \tag{S4}$$

This holds when $\boldsymbol{x}_1^* \approx \Phi^{-1}(\boldsymbol{x}_0^*)$ and the pretrained flow model transports $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ to the data distribution, since Gaussian concentration implies $\|\boldsymbol{z}\|_2 \approx \sqrt{d}$ with high probability for $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$.

Proposition A.1 (Posterior energy bound)**.** *Let Assumptions A.1–A.3 hold. Let $\tilde{\boldsymbol{x}}_1 \in \mathbb{R}^d$ satisfy $\|\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1^*\|_2 \le \delta$. Then:*

(i) ***Likelihood energy upper bound:***

$$E_{\text{like}}(\tilde{\boldsymbol{x}}_1) \;\le\; \frac{\big(\|\boldsymbol{n}\|_2 + L_A \cdot L \cdot \delta\big)^2}{2\sigma_y^2}. \tag{S5}$$

(ii) ***Prior energy near-equality:*** *Both $\tilde{\boldsymbol{x}}_1$ (constructed via SLERP with $\|\tilde{\boldsymbol{x}}_1\|_2 \approx \sqrt{d}$) and $\boldsymbol{x}_1^{\text{rand}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ satisfy*

$$E_{\text{prior}}(\tilde{\boldsymbol{x}}_1) \;\approx\; E_{\text{prior}}(\boldsymbol{x}_1^{\text{rand}}) \;\approx\; \frac{d}{2}. \tag{S6}$$

(iii) ***Concentration of random initialization:*** *For $\boldsymbol{x}_1^{\text{rand}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$,*

$$\|\boldsymbol{x}_1^{\text{rand}} - \boldsymbol{x}_1^*\|_2 = \Theta(\sqrt{d}) \quad \textit{with high probability.} \tag{S7}$$

Since the posterior energy is the sum of the likelihood and prior terms, Claims (i) and (ii) together give a proximity-dependent upper bound on the posteirior energy.

Proof. We now prove the three claims of Proposition A.1.

Part (i): Upper bound on $E_{\text{like}}(\tilde{\boldsymbol{x}}_1)$. Since $\boldsymbol{y} = A(\boldsymbol{x}_0^*) + \boldsymbol{n}$ with $\boldsymbol{x}_0^* = \Phi(\boldsymbol{x}_1^*)$, we have

$$\begin{aligned}
\left\|\boldsymbol{y} - A(\Phi(\tilde{\boldsymbol{x}}_1))\right\|_2 &= \left\|A(\Phi(\boldsymbol{x}_1^*)) + \boldsymbol{n} - A(\Phi(\tilde{\boldsymbol{x}}_1))\right\|_2 \\
&\le \|\boldsymbol{n}\|_2 + \left\|A(\Phi(\boldsymbol{x}_1^*)) - A(\Phi(\tilde{\boldsymbol{x}}_1))\right\|_2 && \text{(triangle inequality)} \\
&\le \|\boldsymbol{n}\|_2 + L_A\left\|\Phi(\boldsymbol{x}_1^*) - \Phi(\tilde{\boldsymbol{x}}_1)\right\|_2 && \text{(Assumption A.2)} \\
&\le \|\boldsymbol{n}\|_2 + L_A \cdot L\,\|\boldsymbol{x}_1^* - \tilde{\boldsymbol{x}}_1\|_2 && \text{(Assumption A.1)} \\
&\le \|\boldsymbol{n}\|_2 + L_A \cdot L \cdot \delta. && (\|\tilde{\boldsymbol{x}}_1 - \boldsymbol{x}_1^*\|_2 \le \delta)
\end{aligned}$$

Squaring and dividing by $2\sigma_y^2$ gives the stated bound.

Part (ii): Prior energy near-equality. For $\boldsymbol{x}_1^{\text{rand}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$, the strong law of large numbers gives $\|\boldsymbol{x}_1^{\text{rand}}\|_2^2 = d(1 + o(1))$ with high probability, so $E_{\text{prior}}(\boldsymbol{x}_1^{\text{rand}}) \approx d/2$. For $\tilde{\boldsymbol{x}}_1$ constructed via SLERP (Eq. 5), the interpolation preserves $\|\tilde{\boldsymbol{x}}_1\|_2 \approx \sqrt{d}$ by construction, giving $E_{\text{prior}}(\tilde{\boldsymbol{x}}_1) \approx d/2$.

Part (iii): Concentration of $\boldsymbol{x}_1^{\text{rand}}$. Expand the squared distance:

$$\|\boldsymbol{x}_1^{\text{rand}} - \boldsymbol{x}_1^*\|_2^2 = \|\boldsymbol{x}_1^{\text{rand}}\|_2^2 - 2\langle \boldsymbol{x}_1^{\text{rand}},\, \boldsymbol{x}_1^* \rangle + \|\boldsymbol{x}_1^*\|_2^2. \tag{S8}$$

Term 1. By the strong law of large numbers, $\|\boldsymbol{x}_1^{\text{rand}}\|_2^2 = d(1 + o(1))$ with high probability.

Term 2. Since $\boldsymbol{x}_1^{\text{rand}}$ is independent of $\boldsymbol{x}_1^*$, $\mathbb{E}[\langle \boldsymbol{x}_1^{\text{rand}}, \boldsymbol{x}_1^* \rangle^2] = \|\boldsymbol{x}_1^*\|_2^2 = \Theta(d)$, so $|\langle \boldsymbol{x}_1^{\text{rand}}, \boldsymbol{x}_1^* \rangle|/d \to 0$ in L^2, and the cross term is $o(d)$.

Term 3. By Assumption A.3, $\|\boldsymbol{x}_1^*\|_2^2 = \Theta(d)$.

Combining all three terms, $\|\boldsymbol{x}_1^{\text{rand}} - \boldsymbol{x}_1^*\|_2^2 = \Theta(d)$, and therefore $\|\boldsymbol{x}_1^{\text{rand}} - \boldsymbol{x}_1^*\|_2 = \Theta(\sqrt{d})$ with high probability. □

Remark 1 (Total energy gap)**.** Since Claim (ii) establishes that E_{prior} is approximately equal for both initializations, the total energy gap is driven entirely by the likelihood term:

$$E(\tilde{\boldsymbol{x}}_1) - E(\boldsymbol{x}_1^{\text{rand}}) \;\approx\; E_{\text{like}}(\tilde{\boldsymbol{x}}_1) - E_{\text{like}}(\boldsymbol{x}_1^{\text{rand}}). \tag{S9}$$

Claim (i) bounds $E_{\text{like}}(\tilde{\boldsymbol{x}}_1)$ by a quantity controlled by δ, while Claim (iii) implies that $\|\boldsymbol{x}_1^{\text{rand}} - \boldsymbol{x}_1^*\|_2 = \Theta(\sqrt{d}) \gg \delta$, incurring a substantially larger likelihood energy upper bound. Consequently, when $\delta \ll \sqrt{d}$ and the same Lipschitz bound applies to both initializations, principled initialization admits a tighter distance-based posterior energy upper bound than a random draw.

A.2 Proof of Proposition 3.1

In the reverse-time convention [2, 54], t decreases from 1 (noise) to 0 (data), so $dt < 0$. Set $\tau = 1 - t$ and $\mathbf{X}_\tau := \mathbf{x}_{1-\tau}$. Define the sampling velocity $\bar{v}_t(\cdot|\mathbf{y}) := -v_t(\cdot|\mathbf{y})$, retaining the original noise-level index t; its value at sampling time τ is $\bar{v}_{1-\tau}$. Then (3) becomes

$$d\mathbf{X}_\tau = \left[\bar{v}_{1-\tau}(\mathbf{X}_\tau|\mathbf{y}) + \tfrac{g_{1-\tau}^2}{2}\nabla_{\mathbf{X}_\tau}\log p_{1-\tau}(\mathbf{X}_\tau|\mathbf{y})\right] d\tau + g_{1-\tau}\, d\mathbf{W}_\tau, \tag{S10}$$

where $\mathbf{W}_\tau$ is a standard Wiener process in increasing sampling time.

Assumption A.4 (One-sided Lipschitz posterior velocity)**.** *The posterior velocity $\bar{v}_t(\cdot|\mathbf{y})$ satisfies a one-sided Lipschitz condition with a measurable, integrable parameter $m_v : [0,1] \to \mathbb{R}$: for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ and $t \in [0,1]$,*

$$\left\langle \bar{v}_t(\mathbf{x}|\mathbf{y}) - \bar{v}_t(\mathbf{x}'|\mathbf{y}),\; \mathbf{x} - \mathbf{x}' \right\rangle \;\le\; m_v(t)\,\|\mathbf{x} - \mathbf{x}'\|_2^2.$$

The parameter $m_v(t)$ may take any sign; negative values indicate contraction in the sampling direction.

Assumption A.5 (One-sided Lipschitz posterior score)**.** *The posterior score $\nabla_{\mathbf{x}}\log p_t(\mathbf{x}|\mathbf{y})$ satisfies a one-sided Lipschitz condition with a measurable, integrable parameter $m_s : [0,1] \to \mathbb{R}$: for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ and $t \in [0,1]$,*

$$\left\langle \nabla\log p_t(\mathbf{x}|\mathbf{y}) - \nabla\log p_t(\mathbf{x}'|\mathbf{y}),\; \mathbf{x} - \mathbf{x}' \right\rangle \;\le\; m_s(t)\,\|\mathbf{x} - \mathbf{x}'\|_2^2.$$

Remark 2 (Comparison with two-sided Lipschitz)**.** A standard two-sided Lipschitz constant bounds the operator norm of the full Jacobian, which is always nonnegative. The one-sided constant above bounds to the maximum eigenvalue of the symmetric part of the Jacobian, which can be *negative*. In particular, log-concavity of $p_t(\cdot|\mathbf{y})$ allows a nonpositive choice of $m_s(t)$, in which case the score term contributes a nonpositive term to the effective contraction rate. This can reduce the prefactor in our bound by offsetting expansion from the posterior velocity.

Setup and notation. We work with (S10) for $\tau \in [0,1]$, with $\mathbf{y}$ and the deterministic schedule g fixed. The total drift is

$$b(\mathbf{x},\tau) := \bar{v}_{1-\tau}(\mathbf{x}|\mathbf{y}) + \frac{g_{1-\tau}^2}{2}\nabla_{\mathbf{x}}\log p_{1-\tau}(\mathbf{x}|\mathbf{y}).$$

Under Assumptions A.4–A.5, b inherits a one-sided Lipschitz property with a *time-varying* parameter:

$$\big\langle b(\mathbf{x},\tau) - b(\mathbf{x}',\tau),\ \mathbf{x}-\mathbf{x}'\big\rangle \ \le\ m(1-\tau)\,\|\mathbf{x}-\mathbf{x}'\|_2^2, \qquad m(t) := m_v(t) + \tfrac{g_t^2}{2}\,m_s(t). \quad \text{(S11)}$$

Its time integral defines the contraction parameter:

$$M \ :=\ \int_0^1 m(t)\,dt \ =\ \int_0^1 m_v(t)\,dt \ +\ \tfrac{1}{2}\int_0^1 g_t^2\,m_s(t)\,dt.$$

Note that M may be negative, zero, or positive.

Triangle inequality decomposition. We decompose the error via two intermediate reference trajectories. Couple all processes with the same Wiener process $\bar{\mathbf{W}}_t$:

$$\|\Phi_g(\tilde{\mathbf{x}}_1) - \mathbf{x}_0^\star\|_2 \ \le\ \underbrace{\|\Phi_g(\tilde{\mathbf{x}}_1) - \Phi_g(\mathbf{x}_1^\star)\|_2}_{\text{Term A}} + \underbrace{\|\Phi_g(\mathbf{x}_1^\star) - \mathbf{x}_0^\star\|_2}_{\text{Term B}}.$$

Squaring and applying $(a+b)^2 \le 2a^2+2b^2$ gives

$$\mathbb{E}\big[\|\Phi_g(\tilde{\mathbf{x}}_1) - \mathbf{x}_0^\star\|_2^2\big] \ \le\ 2\,\mathbb{E}\big[\|\Phi_g(\tilde{\mathbf{x}}_1) - \Phi_g(\mathbf{x}_1^\star)\|_2^2\big] + 2\,\mathbb{E}\big[\|\Phi_g(\mathbf{x}_1^\star) - \mathbf{x}_0^\star\|_2^2\big]. \quad \text{(S12)}$$

Bounding Term A (initialization sensitivity). Drive both trajectories $\Phi_g(\tilde{\mathbf{x}}_1)$ and $\Phi_g(\mathbf{x}_1^\star)$ with the same Wiener process. Define the pathwise error $\mathbf{e}_\tau := \mathbf{X}_\tau^{(\tilde{\mathbf{x}}_1)} - \mathbf{X}_\tau^{(\mathbf{x}_1^\star)}$. Subtracting the two SDEs, the diffusion terms cancel exactly:

$$d\mathbf{e}_\tau = \big[b(\mathbf{X}_\tau^{(\tilde{\mathbf{x}}_1)},\tau) - b(\mathbf{X}_\tau^{(\mathbf{x}_1^\star)},\tau)\big]\,d\tau,$$

so $\mathbf{e}_\tau$ satisfies a deterministic ODE. Differentiating the squared norm and applying (S11) gives, for almost every τ, almost surely,

$$\frac{d}{d\tau}\|\mathbf{e}_\tau\|_2^2 \ =\ 2\big\langle \mathbf{e}_\tau,\ b(\mathbf{X}_\tau^{(\tilde{\mathbf{x}}_1)},\tau) - b(\mathbf{X}_\tau^{(\mathbf{x}_1^\star)},\tau)\big\rangle \ \le\ 2m(1-\tau)\,\|\mathbf{e}_\tau\|_2^2.$$

The initial condition is $\|\mathbf{e}_0\|_2^2 = \|\tilde{\mathbf{x}}_1 - \mathbf{x}_1^\star\|_2^2 \le \delta^2$. Gronwall's inequality [59] gives the pathwise bound

$$\|\mathbf{e}_1\|_2^2 \ \le\ \exp\Big(\int_0^1 2m(1-\tau)\,d\tau\Big)\delta^2 \ =\ e^{2M}\,\delta^2.$$

Taking expectations yields

$$\mathbb{E}\big[\|\Phi_g(\tilde{\mathbf{x}}_1) - \Phi_g(\mathbf{x}_1^\star)\|_2^2\big] \ \le\ e^{2M}\,\delta^2. \quad \text{(S13)}$$

Bounding Term B (stochastic departure accumulation). Define $\boldsymbol{\xi}_\tau := \mathbf{X}_\tau^{(\eta)} - \mathbf{X}_\tau^{(0)}$, the difference between

- $\mathbf{X}_\tau^{(\eta)}$: trajectory of (S10) initialized at $\mathbf{X}_0^{(\eta)} = \mathbf{x}_1^\star$, and
- $\mathbf{X}_\tau^{(0)}$: drift ODE trajectory, $d\mathbf{X}_\tau^{(0)} = b(\mathbf{X}_\tau^{(0)},\tau)\,d\tau$, initialized at $\mathbf{X}_0^{(0)} = \mathbf{x}_1^\star$, so that $\mathbf{X}_1^{(0)} = \mathbf{x}_0^\star$.

Here (0) denotes removal of only the Brownian term, with the same schedule g retained in b. Both trajectories share the same total drift b, so

$$d\boldsymbol{\xi}_\tau = \big[b(\mathbf{X}_\tau^{(\eta)}, \tau) - b(\mathbf{X}_\tau^{(0)}, \tau)\big]d\tau + g_{1-\tau}\, d\mathbf{W}_\tau, \qquad \boldsymbol{\xi}_0 = \mathbf{0}.$$

Applying Itô's lemma to $\|\boldsymbol{\xi}_\tau\|_2^2$ gives

$$d\|\boldsymbol{\xi}_\tau\|_2^2 = 2\langle \boldsymbol{\xi}_\tau, d\boldsymbol{\xi}_\tau\rangle + d\, g_{1-\tau}^2\, d\tau.$$

Using the one-sided bound (S11), a standard localization and integrating-factor argument, followed by Fatou's lemma, yields Gronwall's estimate in the original time variable:

$$\mathbb{E}\|\boldsymbol{\xi}_1\|_2^2 \;\le\; d\int_0^1 g_t^2\, \exp\Big(2\int_0^t m(u)\, du\Big)\, dt \;\le\; e^{2M_+}\, d\cdot G, \tag{S14}$$

where $M_+ := \max\big(0,\ \sup_{s\in[0,1]} \int_0^s m(u)\, du\big)$ controls the worst-case partial accumulation along the trajectory. Therefore,

$$\mathbb{E}\big[\|\Phi_g(\mathbf{x}_1^\star) - \mathbf{x}_0^\star\|_2^2\big] \;\le\; e^{2M_+}\, d\cdot G. \tag{S15}$$

Combining the bounds. Since $M \le M_+$, substituting (S13) and (S15) into (S12) gives

$$\mathbb{E}\big[\|\Phi_g(\tilde{\mathbf{x}}_1) - \mathbf{x}_0^\star\|_2^2\big] \;\le\; 2\, e^{2M}\, \delta^2 + 2\, e^{2M_+}\, d\cdot G \;\le\; 2\, e^{2M_+}\big(\delta^2 + d\cdot G\big).$$

This completes the proof.

B Additional Empirical Validation: Erosion by Excessive Stochasticity

This section provides additional empirical validation for the erosion phenomenon identified in Section 3, complementing the joint analysis of Figure 2(a) with a controlled experiment that isolates the effect of exessive cumulative stochasticity from the initialization gap. We fix the initialization at either proximal ($C_0 = 0.4$) or random ($C_0 = \pi/2$) and vary the cumulative stochasticity G. Under this design, Figure S1(a) shows that at low G, proximal initialization substantially outperforms random initialization across both PSNR and LPIPS, confirming that a well-chosen starting point confers a meaningful reconstruction advantage when cumulative stochasticity is bounded. As G grows, the performance gap narrows and eventually vanishes. The direction of this trend agrees with the monotone dependence of the bound in Proposition 3.1 on G, although the bound itself does not predict that the gap closes. Figure S1(b) corroborates this qualitatively. At low stochasticity ($G = 0.008$), the proximal initialization produces reconstructions that are perceptually sharper and more faithful to the ground truth than the random initialization, demonstrating that the initialization advantage translates into visible quality differences. At high stochasticity ($G = 0.12$), both initialization types yield reconstructions of comparable and degraded quality, visually confirming that excessive cumulative noise injection overrides any benefit from a principled starting point. Together, these results suggest that bounding G plays a key role in preserving the initialization advantage throughout guided sampling.

C Further Details on FLINT Algorithm

C.1 Stage 1: Proximal Source Initialization

We cast source-space initialization as variational optimization by introducing a variational distribution $q_\phi(\boldsymbol{x}_1) = \mathcal{N}(\boldsymbol{x}_1 \mid \boldsymbol{\mu}_\phi, \rho^2\mathbf{I}_d)$ over the initial noise and minimizing the objective in Eq. (7) of the main paper. The expectation in Eq. (7) is approximated by a single reparameterized sample per iteration, i.e., $\boldsymbol{x}_1 = \boldsymbol{\mu}_\phi + \rho\,\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. We find this sufficient in practice given the smoothness of ℓ_{IR} over the spectrally filtered subspace 4.1. Gradients $\nabla_\phi \mathcal{J}(\phi)$ are computed via the SDO shortcut combined with the truncated ODE forward pass.

Closed-form KL term. Since both q_ϕ and the prior $p(\boldsymbol{x}_1) = \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ are Gaussian, the KL divergence admits a closed form. For two d-dimensional Gaussians with diagonal covariance,

$$\mathrm{KL}\big(\mathcal{N}(\boldsymbol{\mu}, \rho^2\mathbf{I}_d) \,\|\, \mathcal{N}(\mathbf{0}, \mathbf{I}_d)\big) = \frac{1}{2}\Big[\mathrm{tr}(\rho^2\mathbf{I}_d) + \boldsymbol{\mu}^\top \mathbf{I}_d^{-1}\boldsymbol{\mu} - d + \log\frac{\det(\mathbf{I}_d)}{\det(\rho^2\mathbf{I}_d)}\Big], \tag{S16}$$

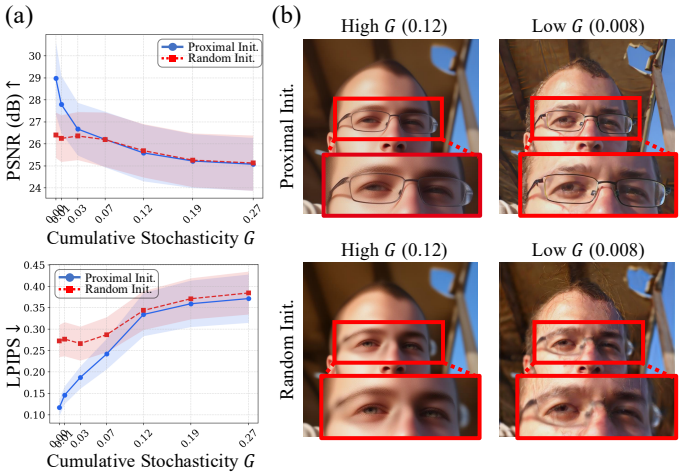

Figure S1: Additional empirical validation for erosion by excessive stochasticity. (a) PSNR and LPIPS as a function of the cumulative stochasticity G for proximal initialization ($C_0 = 0.4$) and random initialization ($C_0 = \pi/2$). Proximal initialization outperforms random initialization at low G, while the performance gap narrows as G increases. (b) Qualitative comparison across initialization types (proximal vs. random, rows) and stochasticity levels (high $G = 0.12$ vs. low $G = 0.008$, columns).

which, upon substituting $\det(\rho^2 \mathbf{I}_d) = \rho^{2d}$ and $\log\det(\mathbf{I}_d) = 0$, simplifies to

$$\mathrm{KL}\big(q_\phi \,\|\, p\big) = \frac{1}{2}\Big(\|\boldsymbol{\mu}_\phi\|_2^2 + d\rho^2 - d - d\log\rho^2\Big). \tag{S17}$$

This term simultaneously penalizes the magnitude of $\boldsymbol{\mu}_\phi$ (through $\|\boldsymbol{\mu}_\phi\|_2^2$) and constrains the variance ρ^2 toward unity, regularizing q_ϕ toward $p(\boldsymbol{x}_1)$ in second-order statistics.

Closed-form cumulant regularizer. The KL term alone cannot preclude higher-order departures from the Gaussian prior, for instance, a heavy-tailed or skewed coordinate distribution that shares the first two moments of $\mathcal{N}(0,1)$. To address this, we include the cumulant regularizer $\mathcal{R}(\boldsymbol{\mu}_\phi)$, which penalizes non-zero standardized skewness and excess kurtosis of the per-channel coordinate distribution of $\boldsymbol{\mu}_\phi$. Treating the HW spatial coordinates of channel c as an empirical sample, define the per-channel empirical mean and variance:

$$\bar{\mu}_c = \frac{1}{HW}\sum_{i=1}^{HW}(\mu_\phi)_{c,i}, \qquad \hat{s}_c^2 = \frac{1}{HW}\sum_{i=1}^{HW}\big[(\mu_\phi)_{c,i} - \bar{\mu}_c\big]^2. \tag{S18}$$

The per-channel standardized skewness $\gamma_{1,c}$ and excess kurtosis $\gamma_{2,c}$ are then

$$\gamma_{1,c}(\boldsymbol{\mu}_\phi) = \frac{1}{HW}\sum_{i=1}^{HW}\left(\frac{(\mu_\phi)_{c,i} - \bar{\mu}_c}{\hat{s}_c}\right)^3, \tag{S19}$$

$$\gamma_{2,c}(\boldsymbol{\mu}_\phi) = \frac{1}{HW}\sum_{i=1}^{HW}\left(\frac{(\mu_\phi)_{c,i} - \bar{\mu}_c}{\hat{s}_c}\right)^4 - 3, \tag{S20}$$

where both quantities vanish for a perfectly Gaussian coordinate distribution. The regularizer averages these penalties over the C channels:

$$\mathcal{R}(\boldsymbol{\mu}_\phi) = \frac{1}{C}\sum_{c=1}^{C}\left[\frac{1}{12}\gamma_{1,c}^2(\boldsymbol{\mu}_\phi) + \frac{1}{48}\gamma_{2,c}^2(\boldsymbol{\mu}_\phi)\right]. \tag{S21}$$

The coefficients $\frac{1}{12}$ and $\frac{1}{48}$ are not chosen ad hoc but follow from the Edgeworth expansion of $\log p$ around $\mathcal{N}(0,1)$: at third and fourth cumulant order, the expansion yields contributions of $\frac{\kappa_3^2}{12}$ and

$\frac{\kappa_4^2}{48}$ to $\mathrm{KL}(p\|\mathcal{N}(0,1))$ [35]. Consequently, $\mathcal{R}$ constitutes a second-order approximation to the KL divergence between the empirical marginal coordinate distribution of each channel and $\mathcal{N}(0,1)$. Together, the KL term controls the norm of $\boldsymbol{\mu}_\phi$, while $\mathcal{R}$ suppresses skewness and excess kurtosis, jointly reducing the risk of the optimized noise drifting outside the training support of the pretrained flow model during optimization.

Noise annealing schedule. The variance ρ is linearly annealed over K optimization iterations:

$$\rho^{(s)} = \rho_{\max} - \frac{s}{S}(\rho_{\max} - \rho_{\min}), \quad s = 0, 1, \ldots, S, \tag{S22}$$

with $\rho_{\min} = 0.02$. Large ρ in early iterations smooths the optimization landscape and prevents premature convergence to suboptimal $\boldsymbol{\mu}_\phi$. As ρ decays toward $\rho_{\min}$, gradient estimates concentrate around $\boldsymbol{\mu}_\phi$, enabling fine-grained refinement within the identified low-energy region.

C.2 Stage 2: Proximity-Preserving Posterior Sampling with Bounded Stochasticity

At the start of Stage 2, the sample $\tilde{\boldsymbol{x}}_1 = \boldsymbol{\mu}_\phi$ produced by Stage 1 replaces the standard random initialization, positioning the sampler within a low posterior energy region from the outset. Each reverse step proceeds through four sequential operations.

Text-conditional velocity field. At each timestep t, the pretrained flow model is queried under the text condition c with a fixed classifier-free guidance scale λ:

$$v_t(\boldsymbol{x}_t) = v_\theta(\boldsymbol{x}_t, c_\emptyset) + \lambda \left(v_\theta(\boldsymbol{x}_t, c) - v_\theta(\boldsymbol{x}_t, c_\emptyset)\right). \tag{S23}$$

Decomposition via Tweedie's formula. The velocity $v_t(\boldsymbol{x}_t)$ simultaneously yields a clean image estimate $\hat{\boldsymbol{x}}_{0|t}$ and a source noise estimate $\hat{\boldsymbol{x}}_{1|t}$ via the flow-based Tweedie's formula:

$$\hat{\boldsymbol{x}}_{0|t} = \boldsymbol{x}_t - \sigma_t\, v_t(\boldsymbol{x}_t), \qquad \hat{\boldsymbol{x}}_{1|t} = \boldsymbol{x}_t + (1-\sigma_t)\, v_t(\boldsymbol{x}_t), \tag{S24}$$

where σ_t denotes the noise schedule.

Measurement-consistency correction.

$$\tilde{\boldsymbol{x}}_{0|t}(\boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \mathcal{L}\big(\mathcal{A}(\boldsymbol{x}),\, \boldsymbol{y}\big) = \hat{\boldsymbol{x}}_{0|t} - \beta(t)\, \nabla_{\hat{\boldsymbol{x}}_{0|t}} \mathcal{L}\big(\mathcal{A}(\hat{\boldsymbol{x}}_{0|t}),\, \boldsymbol{y}\big), \tag{S25}$$

where $\mathcal{L}$ measures the discrepancy between the forward projection of the current estimate and the observed measurement, and $\beta(t)$ is a time-varying data consistency guidance scale. The loss $\mathcal{L}$ in Eq. (S25) is instantiated as a weighted interpolation between a Back-Projection (BP) objective and a Least-Squares (LS) objective [19]:

$$\mathcal{L}\big(\mathcal{A}(\hat{\boldsymbol{x}}_{0|t}),\, \boldsymbol{y}\big) = (1-\omega_t) \left\|\mathcal{A}^\dagger(\boldsymbol{y}) - \mathcal{A}^\dagger\big(\mathcal{A}(\hat{\boldsymbol{x}}_{0|t})\big)\right\|_2 + \omega_t \left\|\boldsymbol{y} - \mathcal{A}(\hat{\boldsymbol{x}}_{0|t})\right\|_2, \tag{S26}$$

where $\mathcal{A}^\dagger$ is the pseudo-inverse of $\mathcal{A}$. The interpolation weight $\omega_t = (1-\sigma_\omega(t))^\phi$ with $\phi = 0.8$ transitions from a BP dominated correction at high noise levels to an LS dominated correction as $t \to 0$. For flow matching models, $\sigma_\omega(t) = \sigma_t$. For diffusion models, $\sigma_\omega(t) = 1 - \bar{\alpha}_t$.

Euler update with bounded stochasticity. To preserve the initialization advantage from Stage 1, the next state is obtained by renoising the corrected clean estimate $\tilde{\boldsymbol{x}}_{0|t}(\boldsymbol{y})$ to the $(t+\Delta t)$-manifold with stochasticity injected into the source noise component under the bounded schedule $\eta(t)$:

$$\boldsymbol{x}_{t+\Delta t} = (1-\sigma_{t+\Delta t})\, \tilde{\boldsymbol{x}}_{0|t}(\boldsymbol{y}) + \sigma_{t+\Delta t}\, \hat{\boldsymbol{x}}_{1|t} + \eta(t)\, \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{S27}$$

The complete inference procedure is given in Algorithm 1.

D FLINT in Diffusion Model

D.1 Datasets and Metrics

Quantitative evaluation is conducted on 1,000 high resolution images from the FFHQ dataset [24]. Distortion is measured via PSNR and SSIM [61], and perceptual quality via LPIPS [68] and a patch based FID [20] computed on 128×128 patches using InceptionV3 features.

Algorithm 2 Truncated Sampling with SDO Shortcut

Require: $\boldsymbol{x}_1$, $\boldsymbol{v}_\theta$, $(c, c_\varnothing)$, λ, K

$\boldsymbol{x} \leftarrow \boldsymbol{x}_1$
for $i = 1, \ldots, K-1$ **do**
  $\boldsymbol{v} \leftarrow \boldsymbol{v}_\theta(\boldsymbol{x}, c_\varnothing) + \lambda\left(\boldsymbol{v}_\theta(\boldsymbol{x}, c) - \boldsymbol{v}_\theta(\boldsymbol{x}, c_\varnothing)\right)$
  if $i \neq 1$ **then**
    $\boldsymbol{v} \leftarrow \mathrm{stop_grad}(\boldsymbol{v})$ ▷ SDO shortcut: retain graph only at $i = 1$
  end if
  $\boldsymbol{x} \leftarrow \boldsymbol{x} + (\tau_{i+1} - \tau_i)\,\boldsymbol{v}$ ▷ Euler update; $\tau_i = \tau_i^{(N)}$
end for
$t_{\mathrm{stop}} \leftarrow \tau_{K-1}$
$\boldsymbol{v}_{t_{\mathrm{stop}}} \leftarrow \mathrm{stop_grad}(\boldsymbol{v}_\theta(\boldsymbol{x},\, c_\varnothing) + \lambda\left(\boldsymbol{v}_\theta(\boldsymbol{x},\, c) - \boldsymbol{v}_\theta(\boldsymbol{x},\, c_\varnothing)\right))$
$\hat{\boldsymbol{x}}_0 \leftarrow \boldsymbol{x} - t_{\mathrm{stop}} \cdot \boldsymbol{v}_{t_{\mathrm{stop}}}$ ▷ Tweedie's formula (Eq. (2))
return $\hat{\boldsymbol{x}}_0$

D.2 Baselines

We evaluate our approach against diffusion based baselines implemented on Stable Diffusion 1.5 (512^2, NFE 200), encompassing PSLD [46], ReSample [51], DDPG [19], and TReg [28]. Hyperparameters for each baseline are provided in E.2.

D.3 Problem Setting

Diffusion based experiments cover two inverse problems: super-resolution $4\times$ and motion deblurring with a kernel size of 61 and intensity of 0.5. Across all settings, Gaussian noise with standard deviation of 0.03 (1.5% of the pixel range) is added to the measurements.

D.4 Experimental Results for Diffusion Model

To evaluate the generalizability of FLINT beyond flow matching frameworks, we extend the proposed method to diffusion models and benchmark it against diffusion based baselines on super resolution $\times 4$ and motion deblurring. As summarized in Table 2, FLINT achieves the best distortion perception trade-off across both tasks, demonstrating that proximal initialization via variational framework and guided posterior sampling with a bounded stochasticity schedule generalize effectively to the diffusion model setting. The qualitative comparisons in Fig. S2 corroborate these findings, with FLINT producing reconstructions of higher visual fidelity and structural coherence than competing baselines. Collectively, these results confirm that the core principles of FLINT are not specific to rectified flow priors but transfer robustly across distinct generative mechanisms.

E Implementation Details

E.1 Implementation Details for FLINT

E.1.1 Stage 1: Proximal Source Initialization

Optimization setup. The optimization parameters $\boldsymbol{\mu}_\phi$ are optimized using the Adam optimizer. The regularization coefficients are set to $\beta_{\mathrm{KL}} = 0.1$ and $\lambda_{\mathrm{cum}} = 10$ for FFHQ under the flow matching framework, with the learning rate (α_{lr}) set to 0.06. The noise annealing schedule uses $\sigma_{\max} = 0.5$ and $\sigma_{\min} = 0.02$ throughout all flow matching experiments. Hyperparameters for DIV2K and diffusion model experiments are summarized in Table S1.

Truncated sampling configuration. For flow matching experiments, we use $K = 3$ truncated ODE steps per optimization iteration, with 7 optimization iterations in total, yielding $\mathrm{NFE}_{\mathrm{Stage1}} = K \times 7 = 21$ function evaluations. The remaining 29 NFEs are allocated to Stage 2 guided posterior sampling, for a total budget of 50 NFEs. For diffusion model experiments, we use $K = 10$ truncated ODE steps per optimization iteration with 10 optimization iterations, yielding $\mathrm{NFE}_{\mathrm{Stage1}} = K \times 10 = 100$ function evaluations, with the remaining 100 NFEs allocated to Stage 2, for a total of 200 NFEs.

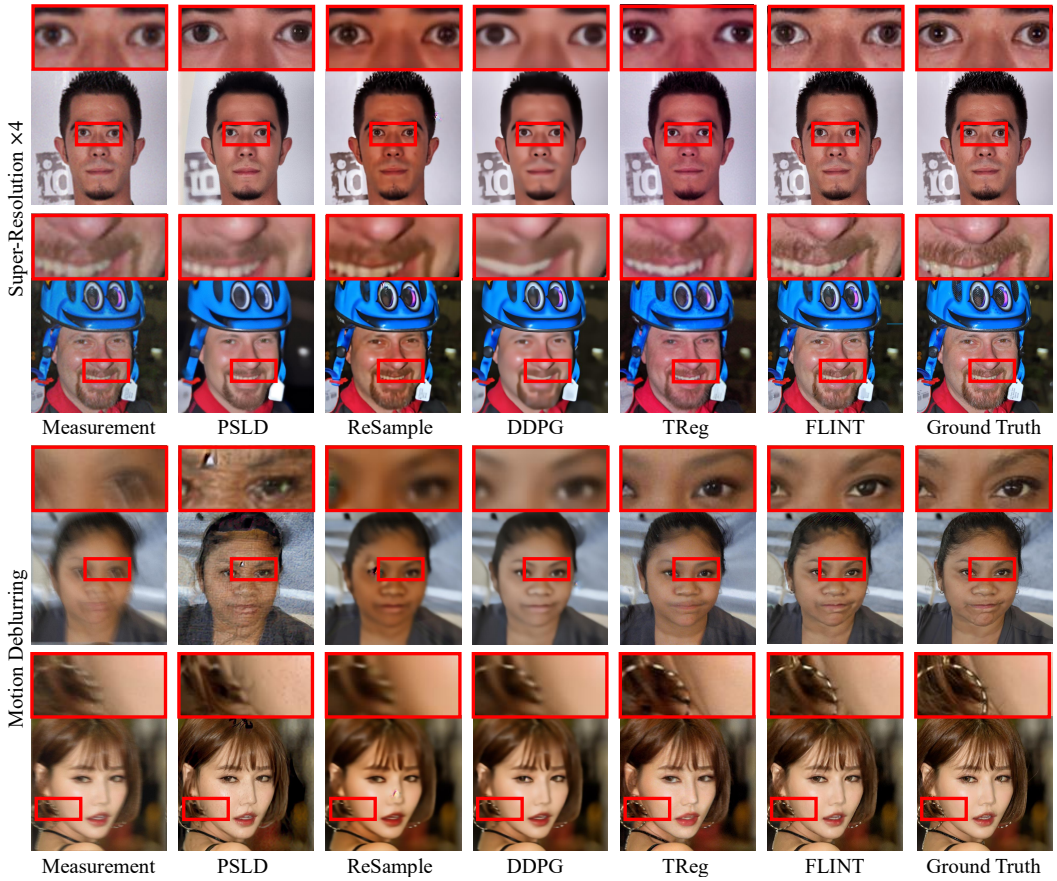

Figure S2: Qualitative comparison for FFHQ dataset on super-resolution 4× and motion deblurring based on diffusion model (SD 1.5).

Table S1: Stage 1 hyperparameters across datasets and generative frameworks.

Framework	Dataset	β_{KL}	λ_{cum}	$\sigma_{\max}$	α_{lr}
Flow Matching (SD3.5-M)	FFHQ	0.1	10.0	0.5	0.006
	DIV2K	0.01	1.0	0.5	0.01
Diffusion (SD1.5)	FFHQ	0.01	10.0	0.02	0.02

Calibration for spectrally weighted data consistency. The Wiener filter weight $W(\omega, t_{\text{stop}})$ in Eq. 8 assumes a pink noise prior $S_{x_0}(\omega) \propto \|\omega\|^{-\alpha}$ for natural images [50], consistent with the $1/f^2$ power spectrum, with $\alpha = 2.0$ and $t_{\text{stop}} = 0.976$ fixed across all tasks and datasets. While α determines the spectral shape of the prior, it leaves the proportionality constant undetermined, causing the transition frequency between near unity and near zero mask values to shift in an uncontrolled manner. Calibration resolves this ambiguity by rescaling the raw mask to attain a prescribed value $W_{\text{at_cutoff}}$ at a designated reference frequency ω_c, thereby anchoring the transition between reliable and unreliable frequency components as a half power point. We uniformly adopt $\omega_c = 0.05$, $W_{\text{at_cutoff}} = 0.5$, and $\alpha = 2.0$ across all tasks and datasets without task specific tuning, indicating that the calibration scheme is robust to variations in degradation type and image domain.

E.1.2 Stage 2: Proximity-Preserving Posterior Sampling with Bounded Stochasticity

Stochasticity magnitude. The overall magnitude λ_s of the bounded stochasticity schedule $\eta(t)$ in Eq. 10 governs the total amount of noise injected throughout the guided sampling trajectory and is likewise task-dependent. A complete specification of the λ_s values used across all datasets and inverse problems is provided in Table S2.

Data consistency guidance scale. The time-varying guidance scale $\beta(t)$ in Eq. S25 controls the step size of the measurement correction at each reverse timestep and is task-dependent. Specifically, $\beta(t)$ follows a monotonically decreasing schedule interpolating between a minimum value $\beta_{\min}$ and a maximum value $\beta_{\max}$ over the sampling trajectory. The complete set of $(\beta_{\min}, \beta_{\max})$ values used across all datasets and inverse problems is provided in Table S3.

Table S2: Stochasticity magnitude λ_s across tasks and datasets.

Framework	Dataset	SR $\times 8$	SR $\times 12$	Motion Deblurring	Gaussian Deblurring
Flow Matching (SD3.5-M)	FFHQ	0.9	0.9	0.6	0.8
	DIV2K	0.9	0.8	0.9	0.9
Framework	Dataset	SR $\times 4$	Motion Deblur		
Diffusion (SD1.5)	FFHQ	0.3	0.3		

Table S3: Data consistency guidance scale $(\beta_{\min}, \beta_{\max})$ configurations across tasks and datasets.

Framework	Dataset	SR $\times 8$	SR $\times 12$	Motion Deblurring	Gaussian Deblurring
Flow Matching (SD3.5-M)	FFHQ	(70, 130)	(30, 130)	(110, 130)	(60, 140)
	DIV2K	(125, 50)	(120, 0)	(450, 250)	(160, 100)
Framework	Dataset	SR $\times 4$	Motion Deblur		
Diffusion (SD1.5)	FFHQ	(10, 100)	(50, 20)		

E.2 Implementation Details for Baseline Hyperparameter Settings

Flow Matching Model Baselines For all flow matching evaluations, we utilize Stable Diffusion 3.5-Medium as the generative backbone with the time scheduler shift factor set to 4.0 and the number of function evaluations fixed at 50. All experiments are conducted on a single NVIDIA A100 GPU. The specific configurations for each method are as follows:

- **FlowChef** The number of gradient descent steps and the step size serving as the data consistency guidance scale are set to 3 and 1, respectively, via grid search on the calibration set. The CFG scale is fixed at 2.0.
- **FlowDPS** FlowDPS is evaluated in its standard configuration with 50 NFE and a CFG scale of 2, using a step size of 15 with 3 gradient descent steps for the inpainting task and a step size of 10 with 3 gradient descent steps for all remaining tasks.
- **FLAIR** We employ 15 gradient descent steps for data consistency optimization with a CFG scale of 2.0 in its standard configuration. The learning rate is set to 0.1 for deblurring tasks, consistent with the original paper. For SR $\times 8$ and SR $\times 12$ tasks, the optimal learning rates are determined via grid search on the calibration set: 2 for both tasks on FFHQ, 3 for SR $\times 8$ on DIV2K, and 6 for SR $\times 12$ on DIV2K.
- **Flower** We employ 15 steps of gradient descent with a step size of 15 for data consistency optimization via a grid search on the calibration set. Consistent with the original authors, the optional uncertainty term is found to destabilize restoration, and thus γ is set to 0 with the CFG scale fixed at 2.0.
- **FlowLPS** We employ 6 steps of gradient descent with a step size of 20 for data consistency optimization via a grid search on the calibration set. Consistent with the original paper, we use a fixed number of Langevin-Proximal update steps, $N_p = 15$, and a CFG scale of 2.0. A constant step size $\zeta_t = 10^{-4}$ is used for Langevin dynamics across all timesteps. For the proximal optimization, we utilized the SGD optimizer. We use an initial learning rate of 0.5 for super-resolution and 0.1 for other tasks. To ensure stability during optimization, we apply a learning rate decay schedule: for super-resolution, the learning rate is decayed by a factor of 0.85 every 5 iterations, and for inpainting, by 0.65 every 10 iterations.
- **FLINT** We employ 6 steps of gradient descent for the data consistency update by default, increasing to 9 for DIV2K motion deblurring and to 12 for DIV2K SR$\times$12, with the CFG scale fixed at 1.0. Remaining implementation details follow the setup described in Sec. E.1.

Diffusion Model Baselines For all diffusion-based evaluations, we utilize Stable Diffusion 1.5 as the generative backbone with the NFE fixed at 100. To ensure a competitive comparison, hyperparameters for all baselines are determined via an extensive grid search on a calibration set. All experiments are conducted on a single NVIDIA A100 GPU. Specific configurations for each method are as follows:

- **PSLD** We use the paper's default loss weights $\omega = 1.0$ (measurement fidelity) and $\gamma = 0.1$ (gluing), applied identically to both super-resolution and motion deblurring. Following the official implementation, the gluing gradient is added to the DDIM update with no additional step size, and $\mathcal{A}^\top$ approximates the pseudo-inverse for all tasks. We disable classifier-free guidance, following PSLD's official setup with an empty null prompt.
- **Resample** We adopt the resampling hyperparameter $\gamma \left(\frac{1-\bar{\alpha}_{t-1}}{\bar{\alpha}_t}\right)\left(1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}\right)$ with $\gamma = 40$ as proposed in the original paper. To enforce hard data consistency, the skip step size is set to 1, accompanied by 10 steps of gradient descent for SR$\times$4 task and 20 steps of gradient descent for motion deblurring task with a step size of 30. The CFG scale is fixed at 1.0.
- **DDPG** We set hyperparameters for SR$\times$4 and motion deblurring set to $\{\gamma, \zeta, \tilde{\eta}\} = \{10.0, 0.8, 0.3\}$ and $\{5.0, 0.6, 0.6\}$, respectively, with the step size μ_t set to the theoretically safe value μ_t^*. Here, γ controls the guidance transition from back-projection to least-squares, ζ balances stochastic noise injection against reconstruction accuracy, and $\tilde{\eta}$ regularizes the back-projection operator to prevent noise amplification.
- **TReg** We follow TReg's algorithmic pipeline (CG-based data-consistency every third sampling step within the first 170 of 200 DDIM steps, corresponding to $t \gtrsim 145$ on the standard schedule; adaptive negation of the null prompt; no DPS branch). The numerical hyperparameters $\{\text{CFG scale}, \lambda_{\text{cg}}, \text{lr}_{\text{null}}\} = \{2.0,\ 0.1,\ 10^{-5}\}$ are calibrated for our setting (FFHQ $\times$ SD1.5 $\times$ $\sigma_y = 0.03$, SR$\times$4 and motion deblurring).
- **FLINT** We employ 8 steps of gradient descent for the data consistency update for super-resolution $\times$4 and 6 steps for motion deblurring, with the CFG scale fixed at 1.0. Remaining implementation details follow the setup described in Sec. E.1.

F Additional Experiments

F.1 Additional Task: Inpainting

Beyond the four inverse problems in Sec. 5.2, we evaluate FLINT on the inpainting task using a flow matching model. As shown in Table S4 and Fig. S3, FLINT attains state of the art performance on box inpainting across all metrics. Proximal initialization via variational framework combined with guided posterior sampling under a bounded stochasticity schedule suppresses both boundary artifacts and blurry textures prevalent in baseline posterior sampling approaches. Moreover, FLINT achieves a favorable position on the perception–distortion trade-off, attaining competitive reconstruction fidelity (PSNR, SSIM) while improving perceptual quality (LPIPS, FID).

Table S4: Quantitative comparison on inpainting based on flow matching model. All metrics are averaged over 1,000 FFHQ and 800 DIV2K samples, with each method allocated 50 NFEs under additive Gaussian noise ($\sigma_y = 0.03$).

	FFHQ (768 $\times$ 768)				**DIV2K (768 $\times$ 768)**			
Method	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓
FlowChef	22.82	0.785	57.23	0.375	20.01	0.606	70.08	0.463
FlowDPS	21.74	0.783	33.56	0.345	20.36	0.618	29.97	0.415
FLAIR	**28.34**	<u>0.867</u>	11.86	<u>0.141</u>	**23.74**	**0.826**	12.09	<u>0.167</u>
Flower	<u>27.12</u>	0.811	37.94	0.334	22.66	0.687	36.36	0.359
FlowLPS	26.65	0.829	**9.59**	0.249	22.28	0.739	**8.17**	0.257
Ours	26.52	**0.872**	<u>10.63</u>	**0.113**	<u>22.91</u>	<u>0.825</u>	<u>8.78</u>	**0.123**

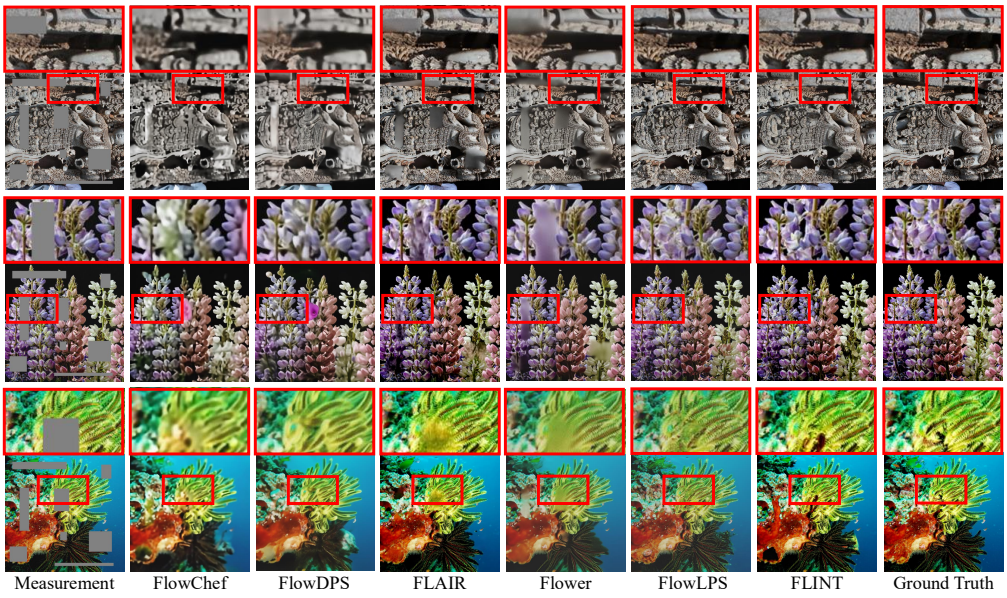

Figure S3: Qualitative comparison for DIV2K datasets on inpainting task.

F.2 Additional Runtime Comparison

We evaluate the inference efficiency of FLINT by comparing its mean runtime with baselines based on flow matching and diffusion models. As shown in Table S5, we report the per image sampling cost (s/image) averaged over 10 samples.

Table S5: Runtime comparison of methods based on flow matching and diffusion models. Sampling time per image, in seconds, is averaged over 10 samples.

Flow Matching Model (SD3.5-M)				
Method	SR $\times 8$ Runtime(s)↓	SR $\times 12$ Runtime(s)↓	Motion Deblur Runtime(s)↓	Gaussian Deblur Runtime(s)↓
FlowChef	8.14	8.14	9.28	8.66
FlowDPS	8.22	8.18	9.47	8.71
FLAIR	20.33	20.28	22.94	21.37
Flower	16.89	16.83	23.36	19.46
FlowLPS	16.84	16.78	22.82	19.31
Ours (FFHQ)	6.79	6.81	6.71	7.47
Ours (DIV2K)	6.79	9.26	7.92	7.41

Diffusion Model (SD1.5)		
Method	SR $\times 4$ Runtime(s)↓	Motion Deblur Runtime(s)↓
PSLD	65.65	65.49
ReSample	161.29	289.24
DDPG	56.98	98.04
TReg	18.61	18.74
FLINT **(Ours)**	64.25	49.56

F.3 Additional Discussion 1: Qualitative Comparison of Sampling Process

Figure S4 visualizes the Tweedie's estimate at representative intermediate timesteps for FLINT and the competing baselines, complementing the quantitative trajectory analysis in Figure 6. At early timesteps, the baseline Tweedie's estimates remain dominated by noise and exhibit only weak alignment with the underlying measurement, whereas FLINT's estimate already reflects the coarse structure of the target reconstruction. As sampling progresses, FLINT's intermediate estimates progressively refine perceptual detail without disturbing the coarse layout established at initialization, whereas the baselines continue to form the coarse layout itself through much of the sampling trajectory. These patterns visualize, in trajectory form, the two complementary mechanisms identified in Section 3: proximal initialization places the sampler within a low-energy posterior region from the outset, and the bounded stochasticity schedule preserves this advantage as detail-level refinement proceeds. Together, they allow FLINT to converge along a noticeably shorter path through the generative manifold than the baselines.



Figure S4: Qualitative comparison of the Tweedie estimate's along the sampling trajectory for FLINT and competing baselines.

Table S6: Ablation on cumulant regularizer $\mathcal{R}(\mu_\phi)$ coefficient λ_{cum} for Motion Deblurring, averaged over 100 FFHQ samples under Gaussian noise ($\sigma_y = 0.03$).

Method	PSNR↑	SSIM↑	KID↓	LPIPS↓
$\lambda_{\text{cum}} = 0$ (OFF)	30.32	**0.802**	0.015	0.263
$\lambda_{\text{cum}} = 10.0$ (ON)	**30.39**	0.800	**0.008**	**0.224**

F.4 Additional Discussion 2: Effect of Cumulant Regularization

The cumulant regularizer $\mathcal{R}(\mu_\phi)$ incorporated into the manifold-constrained proximal objective of stage 1 serves to enforce strict adherence of the optimized noise to the statistical characteristics of the Gaussian prior manifold. While the standard KL divergence primarily constrains lower-order statistics such as mean and variance, this regularization term suppresses non-Gaussian drift by penalizing departures in higher-order moments, including skewness and excess kurtosis. Table S6 shows that on motion deblurring (FFHQ, 100 samples), incorporating this term improves perceptual quality metrics, suggesting that controlling higher order statistics keeps the proximal initialization within the valid input support of the pretrained generative model, enabling full utilization of proximity-preserving posterior sampling of stage 2.

F.5 Additional Discussion 3: Comparison with Pseudo-Inverse-Based Initialization Strategies

We compare FLINT against three initialization strategies that derive the starting noise from the pseudo-inverse measurement, with Stage 2 fixed: forward-noising initialization [36, 13], which perturbs the pseudo-inverse via a single linear interpolation; deterministic inversion [15], which integrates the reverse generative ODE; and stochastic inversion [45], which integrates a controlled SDE biased toward typical Gaussian noise. None of these approaches jointly satisfies the two proximal conditions: proximity to a measurement-consistent region of source space and consistency with the prior typical set.

The forward-noising path does not coincide with the generative trajectory and satisfies neither condition. Deterministic inversion faithfully reproduces the pseudo-inverse but propagates the spectral distortions of the degradation operator into source space, yielding atypical noise. Stochastic inversion enforces typicality through its controller, yet the measurement operator enters neither the control objective nor the dynamics, causing the optimized noise to drift toward a generic typical sample without measurement awareness. FLINT's variational formulation enforces both conditions jointly through a closed-form prior-matching term and a data-fidelity term that explicitly involves the measurement operator. Table S7 confirms this analysis, with FLINT attaining the best score across every metric.

Table S7: Quantitative comparison of alternative initialization strategies that derive the initial noise from the pseudo-inverse measurement, with Stage 2 fixed. Gaussian deblurring (FFHQ, 100 samples).

	Gaussian Deblurring			
Method	PSNR↑	SSIM↑	KID↓	LPIPS↓
Forward-Noising	28.64	0.616	0.020	0.449
Deterministic-Inv.	29.97	0.788	0.021	0.231
Stochastic-Inv.	29.81	0.751	0.010	0.209
Ours	**30.11**	**0.788**	**0.005**	**0.192**

F.6 Additional Discussion 4: Excessive Source-Space Optimization

This section examines the effect of prolonged source-space optimization on reconstruction quality, corroborating the proximal-level optimization budget adopted in Stage 1. Excessive source-space optimization pushes the initial noise out-of-distribution (OOD), reducing reconstruction quality. To quantify this, we track the horizontal lag-1 spatial correlation of the optimized noise across optimization steps. As shown in Figure S5(c), the correlation remains stable within the proximal regime and gradually increases beyond it, eventually exceeding the 95% acceptance threshold for i.i.d. $\mathcal{N}(0, I)$, indicating that the optimized noise progressively drifts toward the boundary of the prior's typical set. This OOD drift manifests directly in reconstruction quality under guided posterior sampling (Figure S5(a)). While prolonged optimization yields marginal fidelity gains (0.36 dB PSNR from step 10 to 40), perceptual quality degrades substantially, with LPIPS increasing by over 35% (from 0.202 to 0.274), as the OOD noise can no longer be rendered into natural image structures by the pretrained generative model. These results corroborate the proximal design choice of Stage 1. Prolonged source-space optimization beyond a proximal level yields limited fidelity gains while reducing perceptual quality. A short proximal initialization is therefore preferable to exhaustive source-space optimization.

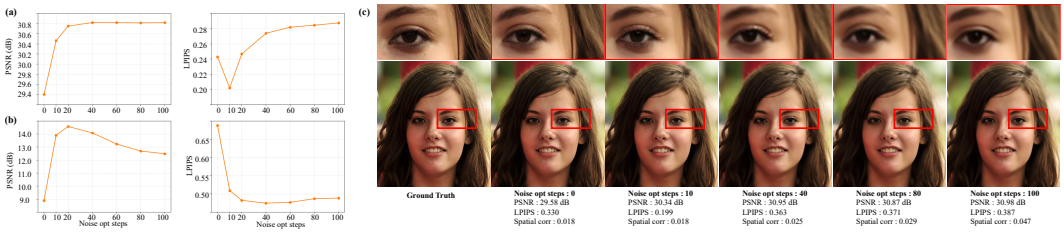

Figure S5: Effect of excessive source-space optimization on Gaussian deblurring (FFHQ, 100 samples). (a) PSNR and LPIPS of FLINT reconstructions across noise optimization steps. Beyond the proximal regime (10 steps), PSNR plateaus while LPIPS degrades by over 35%, as the OOD noise can no longer be rendered into natural image structures by the pretrained generative model. (b) PSNR and LPIPS of unconditional sampling rendered directly from the optimized noise (NFE = 29, matched to Stage 2). (c) PSNR, LPIPS, and horizontal lag-1 spatial correlation of the optimized noise on a representative sample. The correlation remains stable within the proximal regime and trends upward beyond it, indicating a gradual drift toward the boundary of the prior's typical set.

F.7 Additional Discussion 5: Posterior Variance

To verify that FLINT does not suffer from mode collapse, we assess the posterior variance $\mathrm{Var}[\mathbf{x}|\mathbf{y}]$ for the super-resolution $\times 12$ task by drawing 50 samples for a fixed observation $\mathbf{y}$ and computing their pixel-wise standard deviation. Fig. S6 presents this analysis for FLINT alongside FLAIR, Flower, FlowDPS, and FlowLPS. FLINT maintains meaningful posterior coverage with variance concentrated in high-frequency texture regions, precisely the regions left underdetermined by the likelihood term in the highly ill-posed task, confirming that FLINT does not degenerate into a single-mode solution. This diversity arises from two complementary mechanisms. In Stage 1, the spectrally weighted

data consistency loss anchors the low-frequency structure of the proximity-initialized noise to the measurement while suppressing gradients from high-frequency components, leaving uncertainty in fine-grained details. Combined with annealed noise in the variational optimization, this encourages the optimizer to explore a range of measurement-consistent proximal points that share the same coarse layout but differ in their high-frequency realization. In Stage 2, the bounded bell-shaped stochasticity schedule concentrates noise injection in intermediate timesteps where perceptual detail is refined, actively generating varying high-frequency components across samples while keeping the low-frequency structure stable. Together, these results confirm that FLINT reliably explores the posterior rather than collapsing to a single reconstruction.



Figure S6: Posterior variance analysis for super-resolution $\times 12$. For each method, 50 samples are drawn from the posterior $p(\mathbf{x}|\mathbf{y})$ for a fixed observation $\mathbf{y}$, with four representative samples, the posterior mean $\mathbb{E}[\mathbf{x}|\mathbf{y}]$, and the pixel-wise standard deviation Std$[\mathbf{x}|\mathbf{y}]$ displayed. FLINT produces diverse posterior samples with meaningful variance in perceptually significant regions, confirming the absence of mode collapse.

G Additional Qualitative Results



Figure S7: Additional qualitative comparison for the super resolution tasks (super-resolution $\times 8$ and super-resolution $\times 12$) on the FFHQ dataset.

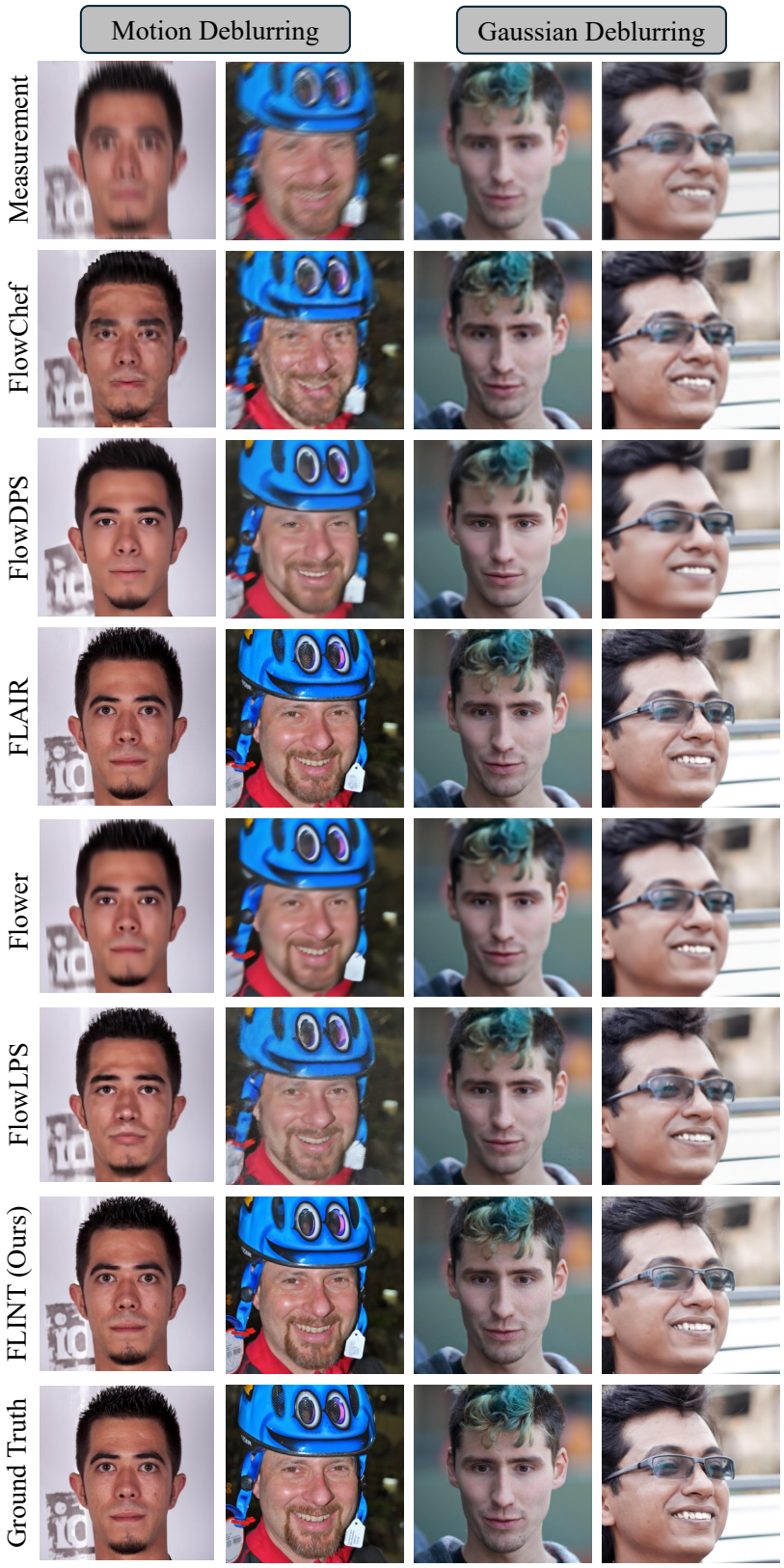

Figure S8: Additional qualitative comparison for the deblurring tasks (motion deblurring and gaussian deblurring) on the FFHQ dataset.



Figure S9: Additional qualitative comparison for the super resolution tasks (super-resolution $\times 8$ and super-resolution $\times 12$) on the DIV2K dataset.



Figure S10: Additional qualitative comparison for the deblurring tasks (motion deblurring and gaussian deblurring) on the DIV2K dataset.



Figure S11: Qualitative failure cases of FLINT on highly complex scenes (DIV2K dataset). Increased ill-posedness in dense, multi-object images disrupts the balance between measurement fidelity and the generative prior. This results in either hallucinated structures, where the model generates spurious details to satisfy the prior, or blurry textures, where conflicting updates fail to resolve high-frequency information. These cases highlight the inherited limitations of the pretrained backbone's capacity for complex out-of-distribution scenes.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The Abstract and Introduction (Sec. 1) clearly reflect the paper's contributions and scope.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: This paper discusses the limitations in Sec. 6.

   Guidelines:

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

3. **Theory assumptions and proofs**

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

   Answer: [Yes]

Justification: This paper presents a theoretical analysis comprising **Proposition 1** in Sec. 3 of the main paper, with their corresponding mathematical proofs and assumptions provided in Appendix Subsec. A.1, Subsec. A.2.

Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: This paper provides sufficient information that can reproduce the results of the main experiment in Sec. 5, Sec. E, and additional experiments in Sec. D, Sec. F.

   Guidelines:

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

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

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

Answer: [Yes]

Justification: We will release code and sample data to reproduce the experimental results in Sec. 5 upon acceptance of the paper.

Guidelines:

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

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: This paper provides the details of the experiments sufficiently in Sec. 5, Sec. E, and Sec. D.

   Guidelines:

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

7. **Experiment statistical significance**

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

   Answer: [Yes]

   Justification: This paper offers several tables for quantitative comparison and figures for qualitative comparison in the main paper as well as the Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: This paper provides the details of computing resources and time of execution in Fig. 4 and Tab. S5.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: This paper fully complies with the NeurIPS code of ethics.

   Guidelines:

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

10. **Broader impacts**

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

    Answer: [Yes]

    Justification: In Sec. 1 and Sec. 6 of the main paper, we emphasize that the our work opens a new perspective on the design of flow-based inverse problem solvers, that combines source space warmup initialization via variational framework with guided sampling under a bounded stochasticity schedule.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.

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

11. **Safeguards**

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

    Answer: [N/A]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: We have fully cited all methods, datasets, and pre-trained foundation models utilized in this paper.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

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

    Answer: [Yes]

    Justification: The assets introduced in this paper have been documented, and we publicly plan to release the code in the future.

    Guidelines:

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

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

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

    Answer: [N/A]

    Justification: Our study does not involve crowdsourcing or research with human subjects.

    Guidelines:

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

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

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

    Answer: [N/A]

    Justification: No new data collection or direct interaction with human subjects was conducted in this study.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

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

    Answer: [N/A]

    Justification: The core method on this paper does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

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