SNaP: One-Step Posterior Sampling for Noisy Inverse Problems

Anonymous authors
Paper under double-blind review

Abstract

Diffusion and flow-matching models can produce high-quality posterior samples for inverse problems, but typically require tens to thousands of network evaluations per draw. MeanFlow enables one-step generation, yet applying it to inverse problems leaves no intermediate steps at which to enforce measurement consistency. We introduce SNaP, a *one-step* MeanFlow posterior sampler for linear inverse problems with Gaussian noise. Its central innovation is a measurement-adapted source: a Gaussian distribution whose mean and anisotropic covariance are determined by the measurement operator, observation, and noise level. The source anchors well-measured directions while preserving variation where the measurements are weak or uninformative. We show that the exact conditional flow transports this source to the true posterior. Across natural-image restoration and multi-coil MRI, SNaP produces diverse, high-quality samples with one network evaluation per draw, 30 to 2250$\times$ faster than iterative samplers.

1 Introduction

Many imaging tasks require recovering an unknown image $\boldsymbol{x}_1$ from noisy linear measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{n}$, where $\boldsymbol{A}$ models the imaging system and $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\mathrm{n}}^2 \boldsymbol{I})$ denotes additive Gaussian noise. This formulation encompasses magnetic resonance imaging (MRI), computed tomography (CT), microscopy, astronomical imaging, and remote sensing (Bertero et al., 2021; McCann et al., 2017; Ongie et al., 2020). Because $\boldsymbol{A}$ is often ill-conditioned or rank-deficient, the measurements do not uniquely determine $\boldsymbol{x}_1$. Therefore, we need prior information on images for a reliable recovery. Classical methods impose handcrafted regularizers such as total variation (Rudin et al., 1992). More recently, deep generative models, including diffusion models (Ho et al., 2020; Song et al., 2021) and flow-matching models (Lipman et al., 2023; Liu et al., 2023b; Albergo et al., 2025), have learned rich image distributions that can serve as priors for inverse problems (Daras et al., 2024).

Most generative inverse solvers use an unconditional generative model as an image prior and impose measurement consistency repeatedly during an iterative sampling trajectory. One family modifies the generative dynamics using data-fidelity gradients or pseudoinverse corrections (Chung et al., 2023; Song et al., 2023a; Kawar et al., 2022; Wang et al., 2023; Pokle et al., 2024; Ben-Hamu et al., 2024). Another alternates generative updates with explicit data-consistency steps, following the plug-and-play paradigm (Zhu et al., 2023; Chung et al., 2024; Zhang et al., 2025; Martin et al., 2025; Pourya et al., 2026). Although these corrections improve agreement with the measurement $\boldsymbol{y}$, they require many sequential network evaluations and therefore make posterior sampling slow. Depending on the method, generating a single sample may require tens to thousands of network evaluations (NFEs). In our CelebA Gaussian-deblurring experiment, for example, DPS (Chung et al., 2023) requires 1000 NFEs and 27 seconds to produce one (128×128) reconstruction (Figure 1).

Consistency models (Song et al., 2023b; Boffi et al., 2025), Shortcut Models (Frans et al., 2025), and MeanFlow(Geng et al., 2025) enable generation with one network evaluation, avoiding the cost of integrating an iterative sampling trajectory. This efficiency is particularly attractive for posterior sampling, where generating a single reconstruction can require tens to thousands of network evaluations. However, collapsing the trajectory into one step also eliminates the intermediate states for enforcing measurement consistency. A direct extension is to condition MeanFlow on the measurement while retaining the standard isotropic source. Recent studies of one-step generation have reported mean-seeking bias in MeanFlow and averaging-induced diversity degradation in class-conditional

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

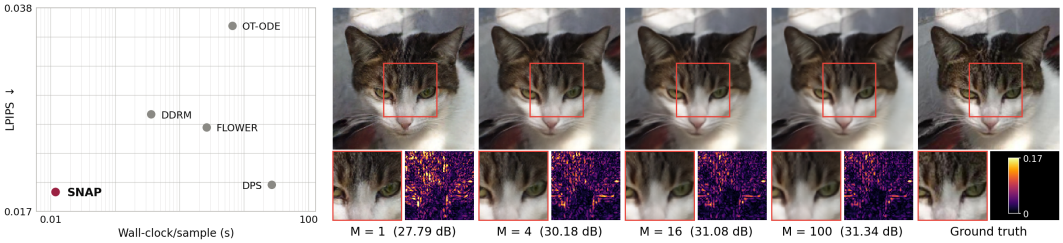

Figure 1: *Left:* LPIPS versus wall-clock time per sample for CelebA Gaussian deblurring. SNaP achieves lowest LPIPS at a fraction of the cost of iterative samplers. *Right:* $4\times$ super-resolution on AFHQ-Cat. Columns show one SNaP draw, averages of $M = 4$, 16, and 100 draws, and the ground truth; zoomed crops, error maps, and PSNR are shown below. Each draw requires one network evaluation. Averaging reduces pixel error and improves PSNR while smoothing fine detail.

flow models (He et al., 2026; Lin et al., 2026). We observe the same failure mode in inverse problems: despite conditioning on the measurement, a MeanFlow with an isotropic source produces nearly identical reconstructions across draws, so averaging offers no improvement (Table 5). This motivates shaping the one-step transport to reflect the geometry of the forward model.

We propose SNaP, a one-step sampler for *noisy* linear inverse problems. Its key innovation is a *measurement-adapted* Gaussian source that builds the forward model and noise level into the starting point of the transport. The source anchors well-measured directions while retaining stochastic variation in weakly measured and null-space directions. Samples from this source can be drawn efficiently, and a single network evaluation maps each draw to a reconstruction.

Our contributions are: **(a)** We extend one-step MeanFlow sampling to noisy linear inverse problems by constructing a measurement-adapted transport aligned with the geometry of the likelihood. **(b)** We derive the source center by minimizing its expected squared displacement to posterior samples among linear estimators. The resulting construction uses a Tikhonov estimate to anchor the transport. **(c)** Across natural-image restoration and multi-coil MRI, SNaP produces diverse, high-quality reconstructions with one network evaluation per draw. Its samples achieve near-unit calibration ratios on the tested CelebA tasks, while sampling 30 to $2250\times$ faster than iterative posterior samplers.

2 Background

Inverse Problems. We consider a linear measurement operator $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ with additive white Gaussian noise,

$$\boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{n}, \qquad \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\mathrm{n}}^2 \boldsymbol{I}). \tag{1}$$

The goal is to infer the unknown image $\boldsymbol{x}_1 \in \mathbb{R}^n$ from the measurement $\boldsymbol{y} \in \mathbb{R}^m$; in posterior sampling, this means drawing samples from $p(\boldsymbol{x}_1 \mid \boldsymbol{y})$. We write the singular value decomposition of $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ and let $\rho := \operatorname{rank}(\boldsymbol{A})$. We order the singular values so that $s_1 \geq \cdots \geq s_\rho > 0$ and set $s_i = 0$ for $\rho < i \leq n$. The corresponding right singular vectors are the columns of $\boldsymbol{V}$. We denote the pseudoinverse by $\boldsymbol{A}^\dagger$ and the orthogonal projection onto $\operatorname{null}(\boldsymbol{A})$ by $\boldsymbol{P} := \boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A}$.

Flow Matching. Flow matching (Lipman et al., 2023; Liu et al., 2023b; Albergo et al., 2025) learns a velocity field that transports a source distribution p_0 to a data distribution p_1. We index time so that $t = 0$ corresponds to the source and $t = 1$ to the data. The flow ψ_t is defined by the ODE $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t(\boldsymbol{x}) = \boldsymbol{v}_t(\psi_t(\boldsymbol{x}))$ for $t \in (0, 1)$, and it maps samples from p_0 to samples from p_1. For the exact marginal velocity field, the terminal map satisfies $(\psi_1)_{\#}p_0 = p_1$. This marginal velocity is generally not available in closed form, so flow matching instead regresses a sample-conditional velocity. Given a coupling of endpoints $(\boldsymbol{x}_0, \boldsymbol{x}_1)$ with $\boldsymbol{x}_0 \sim p_0$ and $\boldsymbol{x}_1 \sim p_1$, the straight-line path

$$\boldsymbol{z}_t = (1-t)\,\boldsymbol{x}_0 + t\,\boldsymbol{x}_1, \qquad \boldsymbol{v} = \boldsymbol{x}_1 - \boldsymbol{x}_0, \tag{2}$$

has constant conditional velocity $\boldsymbol{v}$. A network $\boldsymbol{v}^{\boldsymbol{\theta}}$ is trained by minimizing $\mathbb{E}\,\|\boldsymbol{v}^{\boldsymbol{\theta}}(\boldsymbol{z}_t, t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\|_2^2$. The expected parameter gradient of this conditional regression objective coincides with that of the marginal flow-matching objective (Lipman et al., 2023). The source is commonly chosen as $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ because it is easy to sample; its random initial conditions provide the variability propagated by the deterministic ODE.

Mean Flows. Sampling a flow-matching model requires integrating the velocity $\boldsymbol{v}$, which takes many network evaluations. MeanFlow (Geng et al., 2025; Boffi et al., 2025) avoids this integration by modeling the average velocity $\boldsymbol{u}$ over an interval $[r, t]$, defined as

$$\boldsymbol{u}(\boldsymbol{z}_r, r, t) \;:=\; \frac{1}{t-r}\int_r^t \boldsymbol{v}(\boldsymbol{z}_s, s)\,\mathrm{d}s. \tag{3}$$

Once $\boldsymbol{u}$ is learned, the entire flow reduces to a single evaluation, $\boldsymbol{z}_1 = \boldsymbol{z}_0 + \boldsymbol{u}(\boldsymbol{z}_0, 0, 1)$. Learning $\boldsymbol{u}$ directly from Eq. (3) is intractable, since it requires the integral of $\boldsymbol{v}$. Instead, rewriting Eq. (3) as $(t-r)\,\boldsymbol{u} = \int_r^t \boldsymbol{v}\,\mathrm{d}s$ and differentiating with respect to r, with t fixed, gives the MeanFlow identity

$$\boldsymbol{u}(\boldsymbol{z}_r, r, t) \;=\; \boldsymbol{v}(\boldsymbol{z}_r, r) \;+\; (t-r)\,\frac{\mathrm{d}}{\mathrm{d}r}\boldsymbol{u}(\boldsymbol{z}_r, r, t). \tag{4}$$

This differential identity provides the training relation used by MeanFlow. It also connects MeanFlow to consistency and shortcut models, which learn finite-time flow maps through related consistency relations (Song et al., 2023b; Boffi et al., 2025; Frans et al., 2025).

Measurement-dependent sources. Standard MeanFlow formulations use a fixed, measurement-independent source. Applying MeanFlow to inverse problems therefore requires incorporating the measurement model into the one-step transport. NullFlow (Shi et al., 2026) is a one-step posterior sampler for noiseless linear inverse problems that uses a measurement-dependent source:

$$\boldsymbol{x}_0 \;=\; \boldsymbol{A}^\dagger \boldsymbol{y} + \boldsymbol{P}\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}). \tag{5}$$

For a noiseless measurement, the source satisfies $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{y}$, and restricting the velocity to $\mathrm{null}(\boldsymbol{A})$ keeps the entire trajectory in the measurement-consistent affine set $\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\}$; however, with noisy measurements, the posterior is no longer supported on this set. Moreover, the anchor $\boldsymbol{A}^\dagger\boldsymbol{y}$ contains the row-space noise term $\boldsymbol{A}^\dagger\boldsymbol{n}$, which a velocity restricted to $\mathrm{null}(\boldsymbol{A})$ cannot alter.

3 METHOD

SNaP builds the geometry of the forward model into a one-step posterior transport. Its measurement-dependent source is designed to keep the path to the posterior short while preserving the randomness needed for sampling. It concentrates draws in well-measured directions and retains variation in weakly measured and null-space directions. A conditional MeanFlow then learns to transport these draws toward $p(\boldsymbol{x}_1 \,|\, \boldsymbol{y})$, producing an approximate posterior sample from one source draw and one network evaluation. We now derive the source, characterize the exact transport, and describe how the model is trained and sampled.

Along the straight-line path between a source sample $\boldsymbol{x}_0$ and a posterior sample $\boldsymbol{x}_1$, the sample-conditional velocity is $\boldsymbol{v} = \boldsymbol{x}_1 - \boldsymbol{x}_0$, and therefore $\|\boldsymbol{v}\|_2^2 = \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_2^2$ (Liu et al., 2023b). We keep the path and the conditional-independence structure of the endpoint coupling fixed and use the source distribution as our design variable. This differs from coupling-based methods that modify the endpoint pairing (Pooladian et al., 2023; Tong et al., 2024; Albergo et al., 2024). Unlike approaches that prescribe or learn data-dependent sources (Lee et al., 2022; Mammadov et al., 2026), we derive the source from the measurement model using an explicit design criterion.

Although the posterior is unavailable in closed form, the likelihood provides a tractable geometry for images compatible with the measurement. Under additive Gaussian noise,

$$p(\boldsymbol{y} \,|\, \boldsymbol{x}) \propto \exp\left(-\frac{\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2}{2\sigma_{\mathrm{n}}^2}\right),$$

and is maximized on the least-squares solution set

$$\mathcal{S}_{\boldsymbol{y}} \;:=\; \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2.$$

Its superlevel sets are convex tubes around $\mathcal{S}_{\boldsymbol{y}}$. Along a measured right singular vector $\boldsymbol{v}_i$, their width scales as σ_{n}/s_i: well-measured directions are tightly constrained, weakly measured directions are less constrained, and null-space directions are unconstrained (Fig. 2a). The noise level sets the scale because $\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{y} = -\boldsymbol{n}$. Since these tubes are convex, a straight path between two points in the same tube stays inside it. This geometry motivates a source with little variation in well-measured

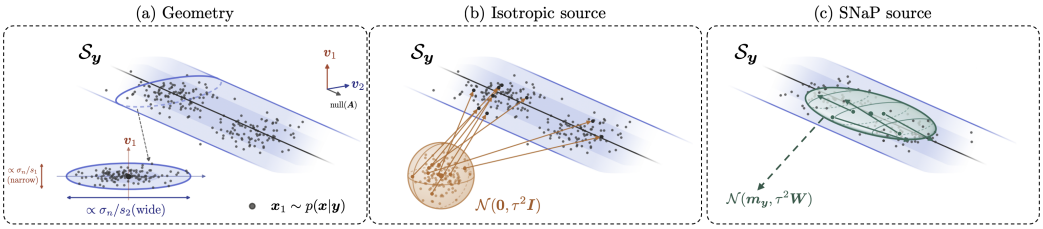

Figure 2: Geometry-aware one-step transport (schematic). **Left:** Likelihood superlevel sets form tubes around $\mathcal{S}_{\boldsymbol{y}}$: narrow in well-measured directions, wider in weakly measured directions, and unbounded along $\mathrm{null}(\boldsymbol{A})$. Gray points represent posterior samples. **Middle:** An isotropic source spreads samples across directions regardless of measurement strength, leading to large source-to-target displacements. **Right:** The SNaP source concentrates samples in well-measured directions while retaining variation in weakly measured and null-space directions, shortening displacements.

directions and greater variation in weakly measured and null-space directions (Fig. 2c), in contrast to an isotropic source (Fig. 2b). We construct this source from the observed $\boldsymbol{y}$, known $\boldsymbol{A}$ and σ_{n}, and auxiliary randomness. We consider sources with an explicit, linear measurement-dependent center:

$$\boldsymbol{x}_0 = \boldsymbol{K}\boldsymbol{y} + \boldsymbol{\xi}, \tag{6}$$

where $\boldsymbol{K}$ is fixed for a given measurement model and $\boldsymbol{\xi}$ has mean $\boldsymbol{0}$ and covariance $\boldsymbol{\Sigma}$, independently of $(\boldsymbol{x}_1, \boldsymbol{n})$. This choice admits an exact analysis and avoids training a separate source model (Mammadov et al., 2026). The family contains the standard isotropic source when $\boldsymbol{K} = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$, and sources centered at the least-squares solution when $\boldsymbol{K} = \boldsymbol{A}^{\dagger}$.

Because the one-step map is deterministic given $(\boldsymbol{x}_0, \boldsymbol{y})$, all sampling randomness originates from the source. For the family in Eq. (6), its expected squared displacement decomposes as

$$\mathbb{E}\big[\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_2^2\big] = \mathrm{tr}(\boldsymbol{R}(\boldsymbol{K})) + \mathrm{tr}(\boldsymbol{\Sigma}), \qquad \boldsymbol{R}(\boldsymbol{K}) := \mathbb{E}\big[(\boldsymbol{x}_1 - \boldsymbol{K}\boldsymbol{y})(\boldsymbol{x}_1 - \boldsymbol{K}\boldsymbol{y})^{\top}\big]. \tag{7}$$

For fixed $\boldsymbol{\Sigma}$, minimizing the displacement reduces to minimizing the center error $\mathrm{tr}(\boldsymbol{R}(\boldsymbol{K}))$ over $\boldsymbol{K}$. Minimizing over $\boldsymbol{\Sigma}$ would instead give $\boldsymbol{\Sigma} = \boldsymbol{0}$ and eliminate source variability, so we optimize the center and choose the spread separately.

Proposition 1. *Let $\boldsymbol{x}_1$ have mean $\boldsymbol{0}$ and covariance $\boldsymbol{C}$, not necessarily Gaussian, and let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{n}$ with $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\mathrm{n}}^2 \boldsymbol{I})$ and $\sigma_{\mathrm{n}} > 0$. For any fixed $\boldsymbol{\Sigma} \succeq \boldsymbol{0}$, the expected squared displacement of the source family in Eq. (6) is uniquely minimized over $\boldsymbol{K}$ by*

$$\boldsymbol{K}_{\boldsymbol{C}} := \boldsymbol{C}\boldsymbol{A}^{\top}\big(\boldsymbol{A}\boldsymbol{C}\boldsymbol{A}^{\top} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}\big)^{-1}. \tag{8}$$

$\boldsymbol{K}_{\boldsymbol{C}}\boldsymbol{y}$ is thus the linear minimum-mean-squared-error estimator of $\boldsymbol{x}_1$ from $\boldsymbol{y}$, with error covariance

$$\boldsymbol{\Sigma}_{\boldsymbol{C}} := \boldsymbol{R}(\boldsymbol{K}_{\boldsymbol{C}}) = \boldsymbol{C} - \boldsymbol{K}_{\boldsymbol{C}}\boldsymbol{A}\boldsymbol{C}.$$

For the isotropic working covariance $\boldsymbol{C} = \tau^2 \boldsymbol{I}$, with $\tau > 0$, the optimal center is the Tikhonov estimate $\boldsymbol{m}_{\boldsymbol{y}} = \boldsymbol{A}^{\top}\big(\boldsymbol{A}\boldsymbol{A}^{\top} + \lambda \boldsymbol{I}\big)^{-1}\boldsymbol{y}$, with $\lambda = \sigma_{\mathrm{n}}^2/\tau^2$, and its error covariance is $\tau^2 \boldsymbol{W}$, where $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{A}^{\top}\big(\boldsymbol{A}\boldsymbol{A}^{\top} + \lambda \boldsymbol{I}\big)^{-1}\boldsymbol{A}$.

The proof is given in App. A.1. Minimizing the displacement over $\boldsymbol{\Sigma}$ would give $\boldsymbol{\Sigma} = \boldsymbol{0}$ and collapse the source. For a chosen working covariance $\boldsymbol{C}$, we instead set $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{C}}$, matching the error covariance of the optimal linear estimator. Under the corresponding Gaussian working model, this is also the posterior covariance.

Because the true image covariance is unknown, we instantiate Proposition 1 with the isotropic working covariance $\boldsymbol{C} = \tau^2 \boldsymbol{I}$, where τ controls the source scale. Under this model, $\boldsymbol{m}_{\boldsymbol{y}}$ is the displacement-minimizing linear center and $\tau^2 \boldsymbol{W}$ is its error covariance. The resulting Tikhonov center regularizes weakly measured directions and avoids amplifying their measurement noise, unlike the least-squares estimate $\boldsymbol{A}^{\dagger}\boldsymbol{y}$. The working model is used only to construct the source; it does not assume that the true image distribution is isotropic or Gaussian. Using this mean and covariance, we define

$$q_{\tau}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{m}_{\boldsymbol{y}}, \tau^2 \boldsymbol{W}). \tag{9}$$

In the right singular basis of $\boldsymbol{A}$,

$$\boldsymbol{W} = \boldsymbol{V}\,\mathrm{diag}(w_i)\boldsymbol{V}^\top, \qquad w_i = \frac{\sigma_\mathrm{n}^2}{\sigma_\mathrm{n}^2 + \tau^2 s_i^2}. \tag{10}$$

The source variance along $\boldsymbol{v}_i$ is therefore $\tau^2 w_i$. It is small in well-measured directions, approaches $\sigma_\mathrm{n}^2/s_i^2$ when $\tau^2 s_i^2 \gg \sigma_\mathrm{n}^2$, and equals τ^2 on $\mathrm{null}(\boldsymbol{A})$. Thus, the source suppresses variation in well-measured directions while retaining it where the measurement is weak or uninformative. As $\sigma_\mathrm{n} \to 0$, the isotropic working source converges to $q_\tau(\cdot \,|\, \boldsymbol{y}) \to \mathcal{N}(\boldsymbol{A}^\dagger \boldsymbol{y}, \tau^2 \boldsymbol{P})$, which recovers the NullFlow source when $\tau = 1$.

Replacing the isotropic source by the working posterior changes the transport, but not its target.

Proposition 2. *Let $\sigma_\mathrm{n} > 0$, and let $\boldsymbol{x}_0 \sim q_\tau(\cdot \,|\, \boldsymbol{y})$ and $\boldsymbol{x}_1 \sim p(\cdot \,|\, \boldsymbol{y})$ be conditionally independent given $\boldsymbol{y}$ and joined by the straight-line path in Eq. (2). Assume that $p(\cdot \,|\, \boldsymbol{y})$ has finite second moments, and that the flow of the marginal velocity field has a terminal limit: writing $\psi_{0\to t}$ for its flow map from time 0 to time t, the limit $\psi_{0\to 1} := \lim_{t\to 1} \psi_{0\to t}$ exists $q_\tau(\cdot \,|\, \boldsymbol{y})$-almost everywhere. Then the one-step map*

$$T_{\boldsymbol{y}}(\boldsymbol{x}) := \boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}, 0, 1 \,|\, \boldsymbol{y}) \tag{11}$$

satisfies

$$(T_{\boldsymbol{y}})_{\#}\, q_\tau(\cdot \,|\, \boldsymbol{y}) = p(\cdot \,|\, \boldsymbol{y}) \qquad \textit{for every } \tau > 0. \tag{12}$$

Here $\boldsymbol{u}$ is the exact average velocity of Eq. (3); the trained network $\boldsymbol{u}^{\boldsymbol{\theta}}$ approximates it, so the guarantee holds for SNaP samples only up to this approximation error. The exact-map result identifies the posterior target of SNaP; the effect of the source choice on the learned one-step approximation is evaluated through the controlled comparison with isotropic sources in Table 5.

3.1 Training and sampling

A direct draw from $q_\tau(\cdot \,|\, \boldsymbol{y})$ would require applying the matrix square root $\boldsymbol{W}^{1/2}$, which is generally impractical for large imaging operators. We instead use perturb-and-solve (Papandreou & Yuille, 2010; Bardsley et al., 2014):

$$\boldsymbol{x}_0 = \boldsymbol{\epsilon} + \boldsymbol{A}^\top \big(\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}\big)^{-1} \big(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\epsilon} - \boldsymbol{n}'\big), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \tau^2 \boldsymbol{I}), \quad \boldsymbol{n}' \sim \mathcal{N}(\boldsymbol{0}, \sigma_\mathrm{n}^2 \boldsymbol{I}). \tag{13}$$

Here, $\boldsymbol{\epsilon}$ and $\boldsymbol{n}'$ are independent of each other and of $(\boldsymbol{x}_1, \boldsymbol{n})$. Lemma 1 shows that $\boldsymbol{x}_0 \,|\, \boldsymbol{y} \sim q_\tau(\cdot \,|\, \boldsymbol{y})$ exactly and that $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ are conditionally independent given $\boldsymbol{y}$. Each source draw requires one solve with $\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}$; operator-specific implementations are given in App. C.2.

Training. The network parameterizes the conditional average velocity

$$\boldsymbol{u}^{\boldsymbol{\theta}}_{r,t} = \boldsymbol{u}^{\boldsymbol{\theta}}(\boldsymbol{z}_r, r, t \,|\, \boldsymbol{y}, \sigma_\mathrm{n}).$$

Because $\boldsymbol{y}$ and σ_n remain fixed along the path, the conditional MeanFlow identity gives the target

$$\boldsymbol{u}_\mathrm{tgt} = \boldsymbol{v} + (t - r)\,\mathrm{sg}\big[\big(\partial_{\boldsymbol{z}} \boldsymbol{u}^{\boldsymbol{\theta}}_{r,t}\big)\boldsymbol{v} + \partial_r \boldsymbol{u}^{\boldsymbol{\theta}}_{r,t}\big], \qquad \boldsymbol{v} = \boldsymbol{x}_1 - \boldsymbol{x}_0. \tag{14}$$

The derivative term is evaluated using one Jacobian–vector product, and $\mathrm{sg}[\cdot]$ prevents gradients from propagating through the target. We minimize

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\mathrm{sg}[\omega(\ell)]\;\ell], \qquad \ell := \frac{1}{n}\big\|\boldsymbol{u}^{\boldsymbol{\theta}}_{r,t} - \boldsymbol{u}_\mathrm{tgt}\big\|_2^2, \tag{15}$$

using the adaptive weight $\omega(\ell) = (\ell + c)^{-p}$ from Geng et al. (2025). Implementation details and Algorithm 1 are given in App. C.3.

Sampling. At test time, SNaP draws $\boldsymbol{x}_0$ using Eq. (13) and returns $\widehat{\boldsymbol{x}}_1 = \boldsymbol{x}_0 + \boldsymbol{u}^{\boldsymbol{\theta}}(\boldsymbol{x}_0, 0, 1 \,|\, \boldsymbol{y}, \sigma_\mathrm{n})$. Each sample requires one source solve and one network evaluation. Redrawing $(\boldsymbol{\epsilon}, \boldsymbol{n}')$ with $\boldsymbol{y}$ fixed produces conditionally independent samples from the learned model.

Table 1: CelebA, 100 test images. We report single draw (M=1) and the average of M draws for SNaP. Note that a single SNaP draw is competitive in LPIPS at one network evaluation, while averaging draws attains the best SSIM on every task and the best PSNR on three of the four. **Best** and **second best** are shown in blue and red.

Method	NFE	Deblurring			Super-resolution			Random inpainting			Box inpainting		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Degraded	–	27.26	0.838	0.214	11.67	0.182	0.859	11.98	0.198	1.070	22.33	0.754	0.217
MSE regressor	1	35.74	0.953	0.029	34.17	0.945	0.033	34.42	0.955	0.022	–	–	–
PnP-GS	23	33.97	0.924	0.041	31.23	0.890	0.065	29.17	0.874	0.066	–	–	–
DDRM	20	35.02	0.946	0.027	32.24	0.927	0.027	32.40	0.942	0.029	–	–	–
DiffPIR	100	34.83	0.938	0.027	31.87	0.892	0.030	32.45	0.924	0.024	30.48	0.918	0.028
OT-ODE	180	32.96	0.920	0.029	31.34	0.903	0.027	28.69	0.871	0.049	29.37	0.919	0.038
PnP-Flow5	500	34.80	0.941	0.046	31.44	0.905	0.056	33.98	0.953	0.021	31.06	0.939	0.043
Flower5-OT	500	35.65	0.954	0.031	33.03	0.931	0.039	33.97	0.952	0.019	31.85	0.952	0.023
DPS	1000	34.27	0.936	0.019	32.23	0.917	0.023	33.43	0.951	0.013	30.26	0.942	0.022
SNaP (M=1)	1	32.90	0.919	0.018	31.27	0.903	0.023	31.97	0.927	0.018	30.73	0.934	0.021
SNaP (M=4)	4	34.98	0.947	0.022	33.26	0.934	0.023	33.97	0.951	0.015	32.78	0.954	0.018
SNaP (M=16)	16	35.68	0.954	0.028	33.95	0.943	0.029	34.64	0.957	0.018	33.43	0.959	0.020
SNaP (M=100)	100	35.91	0.956	0.030	34.16	0.946	0.031	34.84	0.959	0.019	33.69	0.961	0.022

4 Experiments

Datasets. We evaluate on CelebA (128×128), AFHQ-Cat (256×256), and multi-coil fastMRI brain AXT2 (320×320), averaging all metrics over 100 test measurements. Details are in App. C.1.

Forward operators. We adopt the degradation settings of Martin et al. (2025) and Pourya et al. (2026) for natural images, with $\sigma_{\mathrm{n}} = 0.05$ everywhere except random inpainting, where $\sigma_{\mathrm{n}} = 0.01$. Deblurring uses a 61×61 Gaussian kernel with $\sigma_b = 1.0$ on CelebA and 3.0 on AFHQ-Cat; super-resolution uses stride subsampling at 2× on CelebA and 4× on AFHQ-Cat; random inpainting masks 70% of pixels; box inpainting uses a centered 40×40 mask on CelebA and 80×80 on AFHQ-Cat. For MRI, $\boldsymbol{A} = \boldsymbol{MFS}$ with sampling mask $\boldsymbol{M}$, Fourier transform $\boldsymbol{F}$ and per-slice ESPIRiT sensitivities $\boldsymbol{S}$, at ×4 and ×8 Cartesian acceleration, with noise scaled to 20 and 30 dB input SNR. Closed forms of Eq. (13) for each operator are given in App. C.2.

Baselines. For natural images, we compare against a supervised MSE regressor trained with the SNaP backbone, the plug-and-play solvers PnP-GS (Hurault et al., 2022) and PnP-Flow (Martin et al., 2025), and the iterative posterior samplers DDRM (Kawar et al., 2022), DiffPIR (Zhu et al., 2023), OT-ODE (Pokle et al., 2024), Flower (Pourya et al., 2026) and DPS (Chung et al., 2023). Flower and PnP-Flow are run at the recommended setting, where the exact budget of every method is listed in App. C.3. For MRI, we compare against zero-filled, wavelet-ℓ_1 and TV reconstructions, the MSE regressor, and the posterior samplers CSGM (Jalal et al., 2021), DiffPIR, DPS, DAPS (Zhang et al., 2025) and PnP-DM (Wu et al., 2024; Zheng et al., 2025). Variational Flow Maps (VFM) (Mammadov et al., 2026) also place the measurement in the source, but learn it with an adapter rather than constructing it in closed form, and report results for latent-space flow models, whereas SNaP operates in pixel space; we compare the two source constructions in App. B.4. NullFlow (Shi et al., 2026) is not defined for $\sigma_{\mathrm{n}} > 0$ (Sec. 2), so we compare against it from $\sigma_{\mathrm{n}} = 0$ to 0.05 in App. B.5.

Metrics. We report PSNR, SSIM (Wang et al., 2004), LPIPS (Zhang et al., 2018) and FID (Heusel et al., 2017). None of them says whether the spread of the draws is of the right size, so we also report the calibration ratio $\Delta_{\mathrm{cal}} = V/B$, which divides the variance V of the draws by the squared error B of their mean. For an exact posterior sampler the two are equal, so $\Delta_{\mathrm{cal}} = 1$ is ideal: values below 1 indicate draws collapsed toward a point estimate, and values above 1 indicate over-dispersed draws. It is a second-moment summary and does not certify conditional coverage. We give the definition and estimators in App. C.4.

4.1 Main Results

On CelebA (Tab. 1), a single SNaP draw attains the best LPIPS on Gaussian deblurring, and averaging draws attains the best SSIM on all four operators and the best PSNR on three of the four. Results on AFHQ-Cat are reported in App. B.1. SNaP also attains the best FID on two of the three CelebA

Table 2: Distributional and calibration metrics for CelebA. Δ_{cal} is the calibration ratio of App. C.4, for which 1 is ideal. Note that SNaP attains the best or second-best FID among the posterior-sampling baselines in one network evaluation. **Best** and **second best** are shown in blue and red.

Method	**NFE**	**Gaussian deblurring**		**Super-resolution**		**Box inpainting**	
		FID↓	Δ_{cal} ($\to 1$)	FID↓	Δ_{cal} ($\to 1$)	FID↓	Δ_{cal} ($\to 1$)
MSE regressor	1	22.06	–	24.76	–	–	–
DDRM	20	20.26	0.31	18.94	0.19	–	–
Flower1-OT	100	17.86	0.23	23.79	0.24	18.72	0.24
DPS	1000	18.25	0.64	19.94	0.67	15.42	2.15
SNAP (ours)	1	15.80	1.03	16.80	0.99	15.53	1.08

Table 3: fastMRI brain, multi-coil. PSNR and SSIM over 100 test slices, at $\times 4$ and $\times 8$ acceleration and two input SNRs. For SNaP, we report results for a single draw $M = 1$ and average of M draws. Note that among posterior samplers SNaP is very competitive even using one NFE. **Best** and **second best** are shown in blue and red.

		$\times 4$				$\times 8$			
		20 dB		30 dB		20 dB		30 dB	
Method	NFE	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Zero-filled	—	25.35	0.791	25.37	0.793	21.98	0.694	21.99	0.694
Wavelet+ℓ_1	—	26.57	0.672	27.75	0.768	23.02	0.551	23.81	0.643
TV	—	26.19	0.798	26.27	0.799	22.62	0.689	22.66	0.691
MSE regressor	1	33.80	0.909	33.85	0.912	30.71	0.875	30.73	0.878
DiffPIR	100	29.84	0.709	30.01	0.719	26.14	0.642	26.31	0.655
CSGM	1000	27.74	0.734	32.39	0.802	27.74	0.733	28.33	0.754
DPS	1000	28.81	0.646	29.83	0.669	24.89	0.575	25.17	0.574
DAPS	1000	31.55	0.818	32.44	0.782	27.31	0.749	28.01	0.714
PnP-DM	1000	32.16	0.778	33.39	0.786	27.18	0.701	29.00	0.712
SNAP (M=1)	1	32.06	0.889	32.89	0.899	28.10	0.818	29.14	0.845
SNAP (M=4)	4	32.68	0.904	33.80	0.919	29.06	0.843	29.88	0.864
SNAP (M=16)	16	32.85	0.908	34.05	0.924	29.34	0.850	30.09	0.869

operators and the calibration ratio closest to 1 on all three (Tab. 2), so the spread of its draws matches the error they make. On multi-coil fastMRI (Tab. 3) it is competitive with posterior sampler at both accelerations and both noise levels with only **one** NFE. Figures 3 and 4 show the corresponding reconstructions on natural images and on brain MRI.

4.2 Computational cost

Each SNaP sample costs one network evaluation and one source draw by Eq. (13), whose solve with $(\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I})^{-1}$ is diagonal (or Fourier diagonal) for the image operators and reduces to a short conjugate-gradient solve for multi-coil MRI (App. C.2). As reported in Tab. 4, this makes SNaP 30 to $2250\times$ faster per sample than the iterative samplers on CelebA deblurring and 335 to $563\times$ faster on $\times 8$ MRI.

4.3 Ablations

Source. Table 5 isolates the effect of the source by comparing our proposed source to two isotropic sources: the first is the usual flow matching source $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and the other is $\mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I})$ to control for the influence of τ. We hold the other design choices constant and $\boldsymbol{y}$ and σ_{n} are retained in the conditioning. Under the isotropic sources, the calibration ratio falls to 0.00: the draws are nearly identical, yielding a *point estimator* with higher single-draw PSNR and worse LPIPS. The measurement-adapted source keeps the draws dispersed, and its single-draw advantage is perceptual rather than distortion-based. Figures 12 and 13 in App. B.3 show this collapse directly—the draws and their per-pixel spread—and trace both metrics against the averaging budget M.

Additional results. App. B.1 reports the full AFHQ-Cat tables and App. B.2 shows further reconstructions on natural images and brain MRI. App. B.5 compares SNaP with NullFlow across a range of measurement noise levels, and App. B.4 compares the two source constructions on the two-dimensional benchmark of Mammadov et al. (2026). App. B.7 verifies the method against a

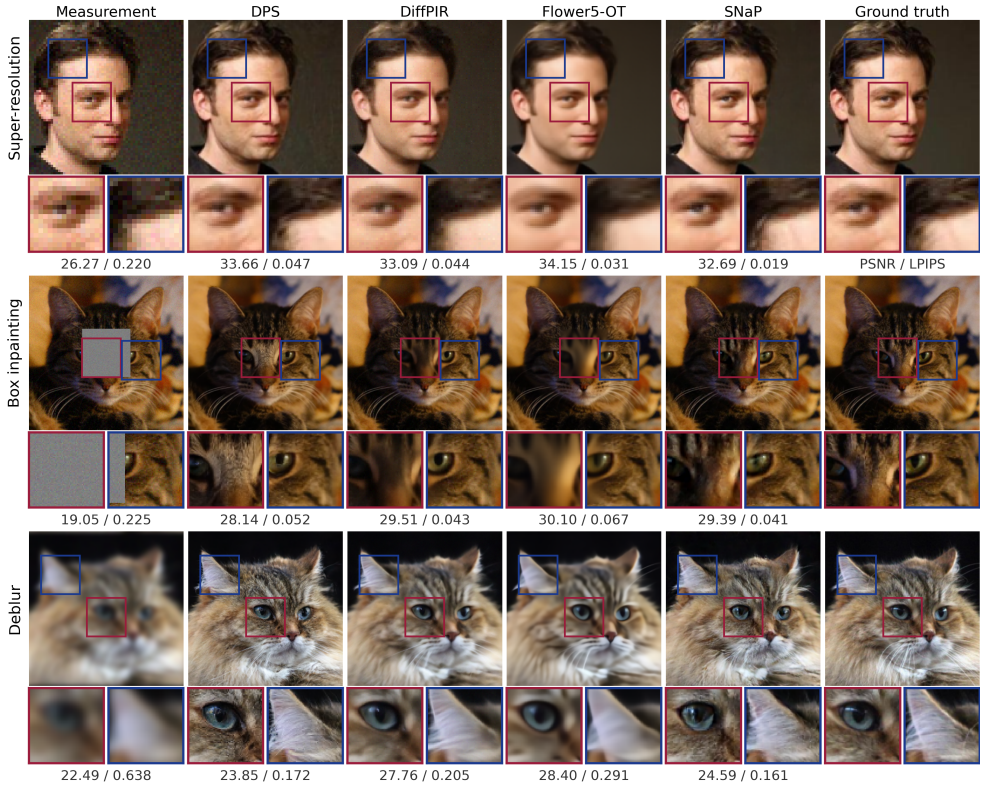

Figure 3: Visual comparison of samples for three imaging tasks, CelebA super-resolution and AFHQ-Cat box inpainting and deblurring, with PSNR and LPIPS for the displayed image. Note that SNaP requires one NFE, while DiffPIR requires 100 and DPS and Flower require 1000. SNaP attains the best LPIPS on all three.

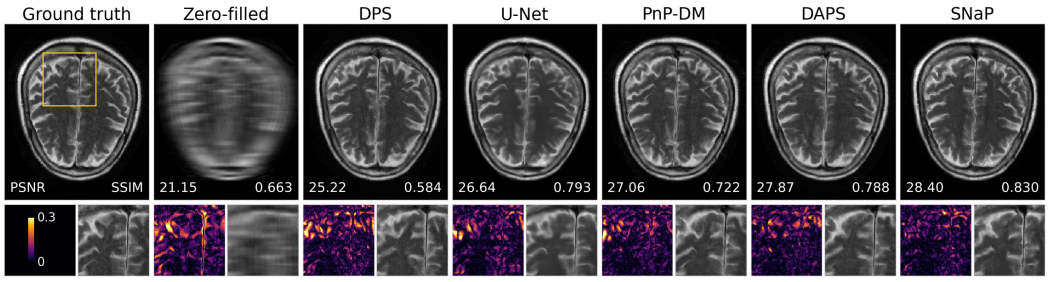

Figure 4: fastMRI brain, multi-coil, $\times 8$ acceleration at 20 dB input SNR. Reconstructions with PSNR and SSIM, and error magnitude and a zoomed in region is displayed. Note that SNaP achieves competitive reconstruction using 1 NFE, while baselines use 1000.

closed-form posterior, and App. B reports the remaining ablations, on the working-prior scale τ, the (r, t) sampling ratio and few-step composition.

5 Related work

Iterative posterior samplers. Most generative solvers enforce consistency with the measurement repeatedly along the sampling trajectory, either by correcting the generative dynamics with a data-fidelity gradient or a pseudoinverse update (Chung et al., 2023; Song et al., 2023a; Kawar et al., 2022; Wang et al., 2023; Pokle et al., 2024; Ben-Hamu et al., 2024; Kim et al., 2025), or by alternating a generative step with an explicit data-consistency step (Zhu et al., 2023; Chung et al., 2024; Zhang et al., 2025; Martin et al., 2025; Pourya et al., 2026). These methods reach high sample quality, but a single posterior sample costs tens to thousands of network evaluations.

Table 4: **Computational cost.** NFE and 1-batch wall-clock time per posterior sample. **Best** in each row is in bold.

CelebA, Gaussian deblurring					
	DPS	OT-ODE	Flower	DDRM	SNAP
NFE ↓	1000	180	100	20	**1**
Time / sample (s) ↓	27.014	6.575	2.614	0.360	**0.012**
Slowdown vs. ours ↓	2251×	548×	218×	30×	–
fastMRI brain, multi-coil ×8					
	CSGM	DAPS	PnP-DM	DPS	SNAP
NFE ↓	1000	1000	1000	1000	**1**
Time / sample (s) ↓	127.78	96.28	92.02	76.07	**0.227**
Slowdown vs. ours ↓	563×	424×	405×	335×	–

Table 5: **Source distribution.** CelebA, Gaussian deblurring at $\sigma_{\mathrm{n}} = 0.05$; only the source differs across rows. M is the number of averaged draws, and Δ_{cal} is the calibration ratio (App. C.4), for which 1 is ideal. An isotropic source collapses onto the posterior mean, where draws are near-identical and averaging brings no gain.

	$M = 1$		$M = 100$		
Source	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	Δ_{cal} ($\to 1$)
$\mathcal{N}(\mathbf{0}, \boldsymbol{I})$	35.85	0.030	35.86	0.030	0.00
$\mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I})$	35.83	0.030	35.84	0.030	0.00
$\mathcal{N}(\boldsymbol{m_y}, \tau^2 \boldsymbol{W})$ (ours)	32.90	**0.018**	**35.91**	0.030	**1.03**

One-step generative models. Consistency models (Song et al., 2023b; Boffi et al., 2025), shortcut models (Frans et al., 2025) and MeanFlow (Geng et al., 2025) eliminate the trajectory and map a source sample to the data in a single network evaluation. Using them for inverse problems requires the measurement to enter through the network conditioning or through the source, since no intermediate state is available at which to enforce data consistency. For real-world super-resolution, MFSR (Wang et al., 2026) distills a text-to-image flow model into a MeanFlow in which the low-resolution image enters only through conditioning, without an explicit forward model or posterior sampling.

Measurement-dependent sources. Bridge models transport from a measurement-dependent source by starting from the degraded observation itself (Liu et al., 2023a; Zhou et al., 2024; Delbracio & Milanfar, 2023), but that source carries no per-direction covariance. NullFlow (Shi et al., 2026) restricts a MeanFlow to $\mathrm{null}(\boldsymbol{A})$ and VFM (Mammadov et al., 2026) learns a measurement-dependent adapter jointly with the flow map, which covers a broader class of problems but yields a source that is trained. MeanFlow has also been applied to PDE-based Bayesian inverse problems with a prior-aligned base measure that does not depend on the measurement (Li et al., 2026).

6 Conclusion

We introduced SNaP, a one-step posterior sampler for linear inverse problems with Gaussian noise. Instead of correcting toward the measurement at every step, SNaP builds it into the transport. The transport starts from a distribution centered at the linear estimate that minimizes the expected transport distance, with a spread that matches the remaining uncertainty. We proved that the exact one-step map carries this distribution to the target posterior. On natural images and multi-coil MRI, SNaP draws a posterior sample in a single network evaluation, 30 to 2250× faster than iterative samplers. On CelebA, averaging its draws gives the best SSIM on all four operators and the best PSNR on three. Its draws also reach the best FID on two of the three operators tested and the calibration ratio closest to 1 on all three.

Limitations. The construction is tied to linear operators, since the working posterior is available in closed form only in that setting, and nonlinear forward models require further work. Sampling the source requires one solve with $\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}$, which is inexpensive for the operators considered here but may be demanding for operators without exploitable structure. Finally, SNaP operates in pixel space, and extending the source construction to latent-space flow models remains open.

Reproducibility Statement

All proofs are provided in the appendix, with the assumptions stated in the corresponding statements. Dataset splits, forward-operator settings and the evaluation protocol are described in App. C.1, the closed forms of the source solve for each operator in App. C.2, the architecture, optimizer and per-operator values of τ in App. C.3, and the metric definitions, including the calibration ratio, in App. C.4. Algorithm 1 states one training step, and the baselines follow the configurations described in §4. Source code is provided.

The Use of Large Language Models

Large language models were used to edit and polish the writing of this paper and to assist with routine code and LaTeX formatting.

References

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic Interpolants: A Unifying Framework for Flows and Diffusions. *J. Mach. Learn. Res.*, 26(209):1–80, 2025.

Michael Samuel Albergo, Mark Goldstein, Nicholas Matthew Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic Interpolants with Data-Dependent Couplings. In *Proc. ICML*, 2024.

Johnathan M Bardsley, Antti Solonen, Heikki Haario, and Marko Laine. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM J. Sci. Comput.*, 36(4):A1895–A1910, 2014.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-Flow: Differentiating through Flows for Controlled Generation. In *Proc. ICML*, 2024.

Mario Bertero, Patrizia Boccacci, and Christine De Mol. *Introduction to inverse problems in imaging*. CRC Press, 2021.

Nicholas Boffi, Michael Albergo, and Eric Vanden-Eijnden. How to Build a Consistency Model: Learning Flow Maps via Self-Distillation. In *Proc. NeurIPS*, 2025.

Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion Posterior Sampling for General Noisy Inverse Problems. In *Proc. ICLR*, 2023.

Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed Diffusion Sampler for Accelerating Large-Scale Inverse Problems. In *Proc. ICLR*, 2024.

Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G Dimakis, and Mauricio Delbracio. A Survey on Diffusion Models for Inverse Problems. *arXiv:2410.00083*, 2024.

Mauricio Delbracio and Peyman Milanfar. Inversion by Direct Iteration: An Alternative to Denoising Diffusion for Image Restoration. *TMLR*, 2023.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One Step Diffusion via Shortcut Models. In *Proc. ICLR*, 2025.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, Zico Kolter, and Kaiming He. Mean Flows for One-step Generative Modeling. In *Proc. NeurIPS*, 2025.

Xiao He, Yang Li, Peizhen Zhang, Songtao Liu, Zhao Zhong, and Nannan Wang. Stabilizing, Scaling & Enhancing MeanFlow for Large-scale Diffusion Distillation. *arXiv preprint arXiv:2605.17834*, 2026.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans Trained by a Two Time-scale Update Rule Converge to a Local Nash Equilibrium. In *Proc. NeurIPS*, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Proc. NeurIPS*, 2020.

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient Step Denoiser for Convergent Plug-and-Play. In *Proc. ICLR*, 2022.

Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust Compressed Sensing MRI with Deep Generative Priors. In *Proc. NeurIPS*, 2021.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising Diffusion Restoration Models. In *Proc. NeurIPS*, 2022.

Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. Flowdps: Flow-driven Posterior Sampling for Inverse Problems. In *Proc. ICCV*, 2025.

Sang-gil Lee, Heeseung Kim, Chaehun Shin, Xu Tan, Chang Liu, Qi Meng, Tao Qin, Wei Chen, Sungroh Yoon, and Tie-Yan Liu. PriorGrad: Improving Conditional Denoising Diffusion Models with Data-Dependent Adaptive Prior. In *Proc. ICLR*, 2022.

Zhiqi Li, Yuchen Sun, Greg Turk, and Bo Zhu. Functional Mean Flow in Hilbert Space. In *Proc. CVPR*, 2026.

Yexiong Lin, Jia Shi, Shanshan Ye, Wanyu Wang, Yu Yao, and Tongliang Liu. SubFlow: Sub-mode Conditioned Flow Matching for Diverse One-Step Generation. *arXiv 2604.12273*, 2026.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. In *Proc. ICLR*, 2023.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I2SB: Image-to-image Schrödinger Bridge. In *Proc. ICML*, 2023a.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *Proc. ICLR*, 2023b.

Abbas Mammadov, So Takao, Bohan Chen, Ricardo Baptista, Morteza Mardani, Yee Whye Teh, and Julius Berner. Variational Flow Maps: Make Some Noise for One-Step Conditional Generation. In *Proc. ICML*, 2026.

Ségolène Tiffany Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. PnP-Flow: Plug-and-Play Image Restoration with Flow Matching. In *Proc. ICLR*, 2025.

Michael T McCann, Kyong Hwan Jin, and Michael Unser. Convolutional Neural Networks for Inverse Problems in Imaging: A Review. *IEEE Signal Process. Mag.*, 34(6):85–95, 2017.

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 J. Sel. Areas Inf. Theory*, 1(1):39–56, 2020.

George Papandreou and Alan L Yuille. Gaussian Sampling by Local Perturbations. In *Proc. NeurIPS*, 2010.

Ashwini Pokle, Matthew J. Muckley, Ricky T. Q. Chen, and Brian Karrer. Training-free Linear Image Inverses via Flows. *TMLR*, 2024.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample Flow Matching: Straightening Flows with Minibatch Couplings. In *Proc. ICML*, 2023.

Mehrsa Pourya, Bassam El Rawas, and Michael Unser. Flower: A Flow-Matching Solver for Inverse Problems. In *Proc. ICLR*, 2026.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear t=Total Variation Based Noise Removal Algorithms. *Physica D*, 60(1-4):259–268, 1992.

Xiao Shi, Edward P Chandler, Chicago Y Park, Shirin Shoushtari, and Ulugbek S Kamilov. NullFlow: One-Step Generative Reconstruction. *arXiv:2606.22696*, 2026.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-Guided Diffusion Models for Inverse Problems. In *Proc. ICLR*, 2023a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *Proc. ICLR*, 2021.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency Models. In *Proc. ICML*, 2023b.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *TMLR*, pp. 1–34, 2024.

Ruiqing Wang, Kai Zhang, Yuanzhi Zhu, Hanshu Yan, Shilin Lu, and Jian Yang. MFSR: MeanFlow Distillation for One Step Real-World Image Super Resolution. *arXiv:2603.20690*, 2026.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model. In *Proc. ICLR*, 2023.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Proc.*, 13(4):600–612, 2004.

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine Bouman. Principled Probabilistic Imaging using Diffusion Models as Plug-and-Play Priors. In *Proc. NeurIPS*, 2024.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving Diffusion Inverse Problem Solving with Decoupled Noise Annealing. In *Proc. CVPR*, 2025.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proc. CVPR*, 2018.

Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy Feng, Caifeng Zou, Yu Sun, Nikola Borislavov Kovachki, Zachary E. Ross, Katherine Bouman, and Yisong Yue. InverseBench: Benchmarking Plug-and-Play Diffusion Priors for Inverse Problems in Physical Sciences. In *Proc. ICLR*, 2025.

Linqi Zhou, Aaron Lou, Samar Khanna, and Stefano Ermon. Denoising Diffusion Bridge Models. In *Proc. ICLR*, 2024.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising Diffusion Models for Plug-and-Play Image Restoration. In *CVPR Workshops*, pp. 1219–1229, 2023.

A PROOFS

A.1 PROOF OF PROPOSITION 1

Proof. Let $\boldsymbol{x}_0 = \boldsymbol{K}\boldsymbol{y} + \boldsymbol{\xi}$ as in Eq. (6).

Since $\boldsymbol{x}_1 - \boldsymbol{x}_0 = (\boldsymbol{x}_1 - \boldsymbol{K}\boldsymbol{y}) - \boldsymbol{\xi}$ and $\boldsymbol{\xi}$ has mean $\mathbf{0}$ and is independent of $\boldsymbol{x}_1 - \boldsymbol{K}\boldsymbol{y}$, for every unit vector $\boldsymbol{e}$,

$$\mathbb{E}\big[(\boldsymbol{e}^\top(\boldsymbol{x}_1 - \boldsymbol{x}_0))^2\big] = \boldsymbol{e}^\top \boldsymbol{R}(\boldsymbol{K})\,\boldsymbol{e} + \boldsymbol{e}^\top \boldsymbol{\Sigma}\,\boldsymbol{e}, \qquad \mathbb{E}[\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_2^2] = \operatorname{tr} \boldsymbol{R}(\boldsymbol{K}) + \operatorname{tr} \boldsymbol{\Sigma}.$$

For fixed $\boldsymbol{\Sigma}$, only $\boldsymbol{R}(\boldsymbol{K})$ depends on $\boldsymbol{K}$.

Let $\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}} := \mathbb{E}[\boldsymbol{x}_1\boldsymbol{y}^\top] = \boldsymbol{C}\boldsymbol{A}^\top$ and $\boldsymbol{C}_{\boldsymbol{y}} := \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^\top] = \boldsymbol{A}\boldsymbol{C}\boldsymbol{A}^\top + \sigma_{\mathrm{n}}^2\boldsymbol{I}$, which is positive definite since $\sigma_{\mathrm{n}} > 0$, so that $\boldsymbol{K}_{\boldsymbol{C}} = \boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}$. Expanding $\boldsymbol{R}(\boldsymbol{K}) = \boldsymbol{C} - \boldsymbol{K}\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}}^\top - \boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}}\boldsymbol{K}^\top + \boldsymbol{K}\boldsymbol{C}_{\boldsymbol{y}}\boldsymbol{K}^\top$ and using $\boldsymbol{K}_{\boldsymbol{C}}\boldsymbol{C}_{\boldsymbol{y}} = \boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}}$,

$$\boldsymbol{R}(\boldsymbol{K}) = \boldsymbol{R}(\boldsymbol{K}_{\boldsymbol{C}}) + (\boldsymbol{K} - \boldsymbol{K}_{\boldsymbol{C}})\,\boldsymbol{C}_{\boldsymbol{y}}\,(\boldsymbol{K} - \boldsymbol{K}_{\boldsymbol{C}})^\top, \qquad \boldsymbol{R}(\boldsymbol{K}_{\boldsymbol{C}}) = \boldsymbol{C} - \boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{y}}^\top = \boldsymbol{C} - \boldsymbol{K}_{\boldsymbol{C}}\boldsymbol{A}\boldsymbol{C}. \tag{16}$$

By Eq. (16), for every unit vector $\boldsymbol{e}$,

$$\boldsymbol{e}^\top \boldsymbol{R}(\boldsymbol{K})\,\boldsymbol{e} = \boldsymbol{e}^\top \boldsymbol{R}(\boldsymbol{K}_{\boldsymbol{C}})\,\boldsymbol{e} + \big\|\boldsymbol{C}_{\boldsymbol{y}}^{1/2}(\boldsymbol{K} - \boldsymbol{K}_{\boldsymbol{C}})^\top \boldsymbol{e}\big\|_2^2 \;\geq\; \boldsymbol{e}^\top \boldsymbol{R}(\boldsymbol{K}_{\boldsymbol{C}})\,\boldsymbol{e},$$

so $\boldsymbol{K}_{\boldsymbol{C}}$ minimizes the displacement along every direction. Taking the trace,

$$\operatorname{tr} \boldsymbol{R}(\boldsymbol{K}) = \operatorname{tr} \boldsymbol{R}(\boldsymbol{K}_{\boldsymbol{C}}) + \big\|\boldsymbol{C}_{\boldsymbol{y}}^{1/2}(\boldsymbol{K} - \boldsymbol{K}_{\boldsymbol{C}})^\top\big\|_F^2,$$

and since $\boldsymbol{C}_{\boldsymbol{y}}^{1/2}$ is invertible, the second term vanishes if and only if $\boldsymbol{K} = \boldsymbol{K}_{\boldsymbol{C}}$. The total displacement is therefore minimized if and only if $\boldsymbol{K} = \boldsymbol{K}_{\boldsymbol{C}}$, for every $\boldsymbol{\Sigma}$.

Isotropic case. For $\boldsymbol{C} = \tau^2\boldsymbol{I}$ and $\lambda = \sigma_{\mathrm{n}}^2/\tau^2$, dividing by τ^2 gives $\boldsymbol{K}_{\boldsymbol{C}} = \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top + \lambda\boldsymbol{I})^{-1}$, so $\boldsymbol{K}_{\boldsymbol{C}}\boldsymbol{y} = \boldsymbol{m}_{\boldsymbol{y}}$, and $\boldsymbol{R}(\boldsymbol{K}_{\boldsymbol{C}}) = \tau^2\big(\boldsymbol{I} - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}\big) = \tau^2\boldsymbol{W}$. □

Remarks. (i) The proof uses only the first two moments of $\boldsymbol{x}_1$, $\boldsymbol{n}$ and $\boldsymbol{\xi}$; neither the images, the noise, nor the perturbation need to be Gaussian. (ii) If $\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C})$, Gaussian conditioning gives $\boldsymbol{x}_1 \,|\, \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K}_{\boldsymbol{C}}\boldsymbol{y}, \boldsymbol{\Sigma}_{\boldsymbol{C}})$, so the source with center $\boldsymbol{K}_{\boldsymbol{C}}\boldsymbol{y}$ and Gaussian perturbation of covariance $\boldsymbol{\Sigma}_{\boldsymbol{C}}$ is the exact posterior; for $\boldsymbol{C} = \tau^2\boldsymbol{I}$ it is the working posterior Eq. (9). (iii) For $\boldsymbol{C} = \tau^2\boldsymbol{I}$, as $\sigma_{\mathrm{n}} \to 0$ with τ fixed, $\lambda \to 0$, and the standard limit of Tikhonov regularization gives $\boldsymbol{K}_{\boldsymbol{C}} \to \boldsymbol{A}^\dagger$ and $\boldsymbol{W} \to \boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A} = \boldsymbol{P}$, so the source tends to $\mathcal{N}(\boldsymbol{A}^\dagger\boldsymbol{y}, \tau^2\boldsymbol{P})$.

A.2 PROOF OF PROPOSITION 2

Proof. Let $p_t(\cdot \,|\, \boldsymbol{y})$ denote the distribution of $\boldsymbol{z}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$, so that $p_0(\cdot \,|\, \boldsymbol{y}) = q_\tau(\cdot \,|\, \boldsymbol{y})$ and $p_1(\cdot \,|\, \boldsymbol{y}) = p(\cdot \,|\, \boldsymbol{y})$. The marginal velocity

$$\boldsymbol{v}_t(\boldsymbol{z}, \boldsymbol{y}) = \mathbb{E}[\boldsymbol{x}_1 - \boldsymbol{x}_0 \,|\, \boldsymbol{z}_t = \boldsymbol{z}, \boldsymbol{y}] \tag{17}$$

generates $\{p_t(\cdot \,|\, \boldsymbol{y})\}_{t\in[0,1]}$ through the continuity equation (Lipman et al., 2023; Albergo et al., 2025).

Since $\sigma_{\mathrm{n}} > 0$, we have $\lambda > 0$ and $w_i \in (0, 1]$ for all i, so $\boldsymbol{W}$ is positive definite. For $t < 1$, the interpolant $\boldsymbol{z}_t$ is therefore the sum of $t\boldsymbol{x}_1$ and an independent Gaussian with nondegenerate covariance $(1-t)^2\tau^2\boldsymbol{W}$. Its density $p_t(\cdot \,|\, \boldsymbol{y})$ is thus smooth and strictly positive. Since $\boldsymbol{x}_1$ has finite second moments, standard arguments show that $\boldsymbol{v}_t(\cdot, \boldsymbol{y})$ is locally Lipschitz, uniformly for $t \in [0, T]$ and every $T < 1$ (Albergo et al., 2025). The ODE $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{z}_t = \boldsymbol{v}_t(\boldsymbol{z}_t, \boldsymbol{y})$ then has a unique solution on $[0, 1)$, and its flow map satisfies $(\psi_{0\to t})_{\#}\, q_\tau(\cdot \,|\, \boldsymbol{y}) = p_t(\cdot \,|\, \boldsymbol{y})$ for every $t < 1$.

Since $\mathbb{E}[\|\boldsymbol{z}_t - \boldsymbol{x}_1\|_2^2] = (1-t)^2\,\mathbb{E}[\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_2^2] \to 0$, the distribution $p_t(\cdot \,|\, \boldsymbol{y})$ converges to $p(\cdot \,|\, \boldsymbol{y})$ as $t \to 1$. By the terminal-limit assumption of Proposition 2, $\psi_{0\to 1}(\boldsymbol{x}) := \lim_{t\to 1} \psi_{0\to t}(\boldsymbol{x})$ exists for q_τ-almost every $\boldsymbol{x}$. Almost-sure convergence implies convergence in distribution, so

$$(\psi_{0\to 1})_{\#}\, q_\tau(\cdot \,|\, \boldsymbol{y}) = p(\cdot \,|\, \boldsymbol{y}). \tag{18}$$

Integrating the ODE from 0 to t and letting $t \to 1$ gives

$$\psi_{0\to1}(\boldsymbol{x}_0) - \boldsymbol{x}_0 = \int_0^1 \boldsymbol{v}_s\big(\psi_{0\to s}(\boldsymbol{x}_0), \boldsymbol{y}\big)\,\mathrm{d}s = \boldsymbol{u}(\boldsymbol{x}_0, 0, 1 \,|\, \boldsymbol{y}), \tag{19}$$

where the last equality is the average velocity in Eq. (3). Hence $\psi_{0\to1} = T_{\boldsymbol{y}}$, and Eq. (18) proves Eq. (12). The target $p(\cdot \,|\, \boldsymbol{y})$ does not depend on τ, so the conclusion holds for every $\tau > 0$. □

A.3 Perturb-and-solve sampling

Direct sampling would require a square root of $\boldsymbol{W}$. The perturb-and-solve sampler avoids this square root while preserving the same Gaussian law.

Lemma 1. *Let $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \tau^2\boldsymbol{I})$ and $\boldsymbol{n}' \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\mathrm{n}}^2\boldsymbol{I})$ be independent of each other and of $(\boldsymbol{x}_1, \boldsymbol{n})$, and let $\boldsymbol{x}_0$ be given by Eq. (13). Then $\boldsymbol{x}_0 \,|\, \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{m}_{\boldsymbol{y}}, \tau^2\boldsymbol{W})$ for every $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and every $\sigma_{\mathrm{n}} > 0$.*

Proof. Write $\boldsymbol{K} := \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top + \lambda\boldsymbol{I})^{-1}$, so that $\boldsymbol{m}_{\boldsymbol{y}} = \boldsymbol{K}\boldsymbol{y}$ and $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{K}\boldsymbol{A}$ by Proposition 1. Substituting these into Eq. (13) gives

$$\boldsymbol{x}_0 = \boldsymbol{m}_{\boldsymbol{y}} + \boldsymbol{W}\boldsymbol{\epsilon} - \boldsymbol{K}\boldsymbol{n}'. \tag{20}$$

Given $\boldsymbol{y}$, this is an affine function of two independent Gaussian vectors, so $\boldsymbol{x}_0$ is Gaussian with mean $\boldsymbol{m}_{\boldsymbol{y}}$ and covariance $\tau^2\boldsymbol{W}^2 + \sigma_{\mathrm{n}}^2\boldsymbol{K}\boldsymbol{K}^\top$. Both $\boldsymbol{W}$ and $\boldsymbol{K}\boldsymbol{K}^\top$ are diagonal in the right singular basis of $\boldsymbol{A}$, with entries w_i and $s_i^2/(s_i^2+\lambda)^2$. Using $\sigma_{\mathrm{n}}^2 = \tau^2\lambda$, the i-th diagonal entry of the covariance is

$$\tau^2 w_i^2 + \tau^2\lambda\frac{s_i^2}{(s_i^2+\lambda)^2} = \tau^2\frac{\lambda^2 + \lambda s_i^2}{(s_i^2+\lambda)^2} = \tau^2\frac{\lambda}{s_i^2+\lambda} = \tau^2 w_i,$$

so the covariance equals $\tau^2\boldsymbol{W}$. □

Each source draw requires applications of $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ and one solve with $\boldsymbol{A}\boldsymbol{A}^\top + \lambda\boldsymbol{I}$. No square root of $\boldsymbol{W}$ is required. Operator-specific implementations are given in App. C.2.

B Additional Results

B.1 Full results on AFHQ-Cat

Table 6: AFHQ-Cat, 100 test images. We report single draw (M=1) and the average of M draws for SNaP. Note that a single SNaP draw is competitive in LPIPS at one network evaluation, and averaging a few draws recovers distortion at a cost still far below the iterative samplers. On AFHQ-Cat the number of time steps used by PnP-Flow and Flower depends on the operator, so their NFE does too; the per-operator budgets are listed in App. C.3. The MSE regressor is omitted on box inpainting, where its single output is the blurred conditional mean of the masked region and is not comparable with the other methods. **Best** and **second best** are shown in blue and red.

Method	NFE	Deblurring			Super-resolution			Random inpainting			Box inpainting		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Degraded	–	24.39	0.532	0.543	11.98	0.212	0.900	13.25	0.214	1.091	21.57	0.736	0.214
MSE regressor	1	29.82	0.806	0.289	29.16	0.817	0.201	33.23	0.915	0.066	–	–	–
PnP-GS	23	28.39	0.787	0.387	24.44	0.639	0.411	29.89	0.841	0.123	–	–	–
PnP-Flow5	500–2500	29.01	0.785	0.312	28.01	0.791	0.167	34.47	0.933	0.044	27.14	0.900	0.127
DDRM	20	29.01	0.784	0.194	27.09	0.781	0.186	32.20	0.907	0.064	–	–	–
DiffPIR	100	28.89	0.773	0.187	23.16	0.635	0.291	31.76	0.881	0.053	27.71	0.880	0.059
OT-ODE	180	27.82	0.735	0.126	26.71	0.737	0.108	29.98	0.849	0.086	24.47	0.873	0.095
Flower5-OT	500–2500	29.73	0.801	0.264	27.09	0.767	0.272	34.41	0.932	0.047	27.32	0.922	0.067
DPS	1000	25.90	0.661	0.148	25.98	0.694	0.144	32.06	0.896	0.036	26.78	0.900	0.054
SNaP (M=1)	1	26.25	0.674	0.159	26.07	0.703	0.130	30.48	0.853	0.066	26.46	0.891	0.054
SNaP (M=4)	4	28.33	0.750	0.157	28.07	0.777	0.120	32.38	0.896	0.052	28.49	0.914	0.047
SNaP (M=16)	16	29.09	0.779	0.231	28.77	0.802	0.157	33.04	0.909	0.056	29.32	0.922	0.056
SNaP (M=100)	100	29.32	0.788	0.320	28.99	0.810	0.187	33.25	0.913	0.058	29.53	0.925	0.069

Table 6 reports the same protocol as Tab. 1 on AFHQ-Cat.

B.2 Additional visual results

Figure 5 shows box inpainting on CelebA.

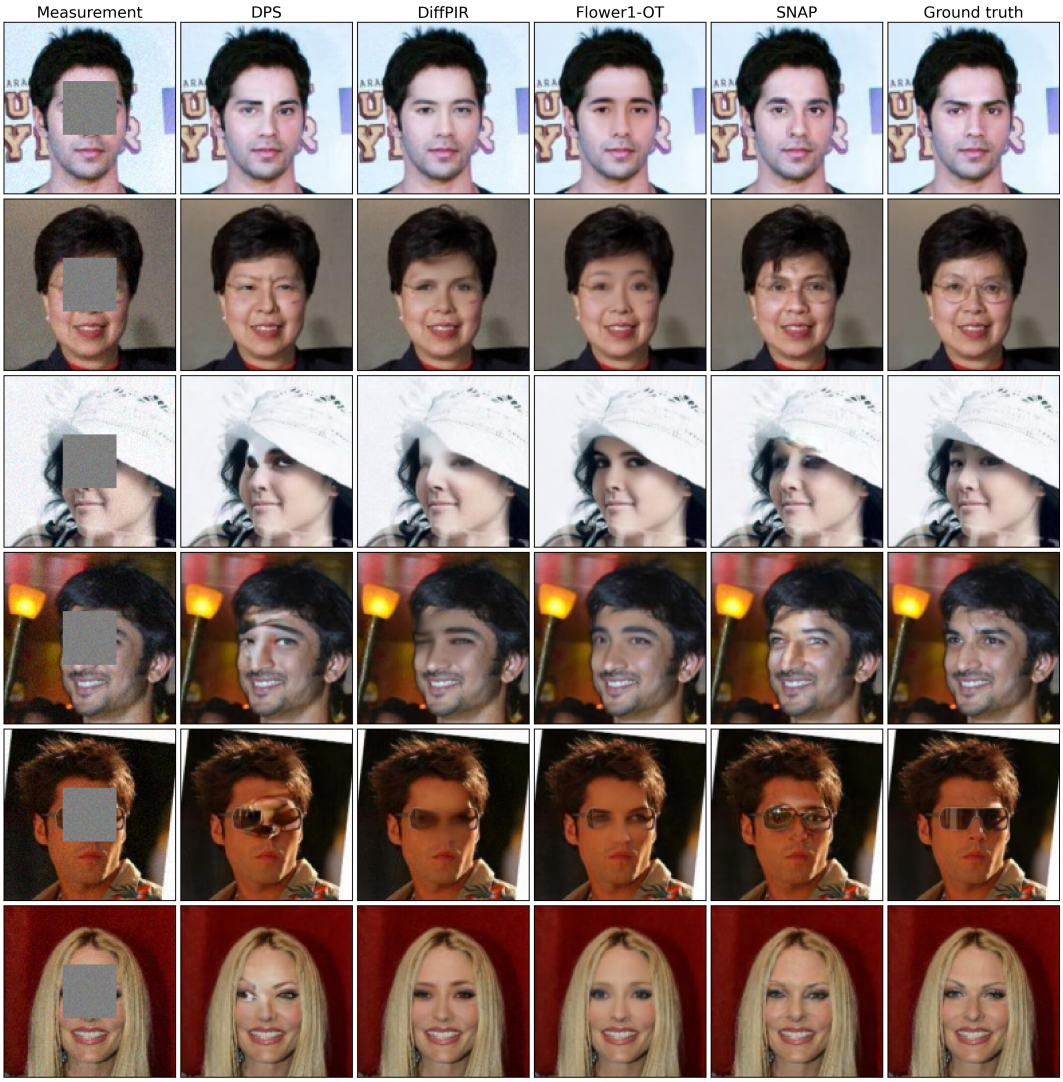

Figure 5: Box inpainting on CelebA.

Figure 6 shows box inpainting on AFHQ-Cat.

Figure 7 shows multi-coil fastMRI at $\times 8$ acceleration and 30 dB input SNR.

Figure 8 shows Gaussian deblurring on CelebA.

Figure 9 shows several SNaP draws for box inpainting.

Figure 10 shows further SNaP draws for box inpainting.

Figure 11 shows further SNaP draws for box inpainting.

B.3 Ablations

Noise misspecification. Since σ_{n} enters the source in closed form, a wrong value at test time changes how the source allocates uncertainty rather than breaking the method: by Eq. (10), over-estimating σ_{n} loosens the source toward the working prior, while under-estimating it shrinks the

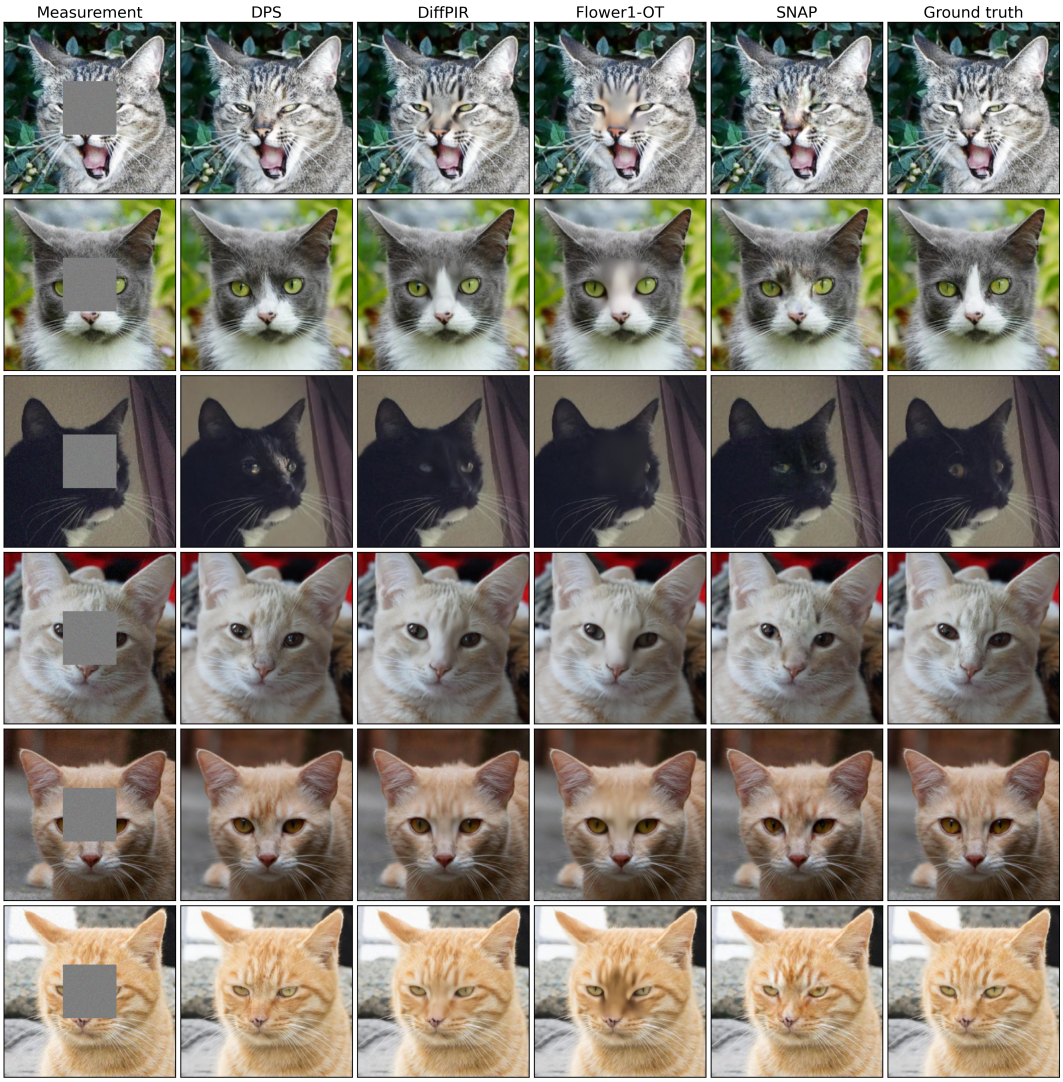

Figure 6: Box inpainting on AFHQ-Cat.

Table 7: CelebA Gaussian deblurring trained at $\sigma_{\mathrm{n}} = 0.05$ and evaluated on y with different measurement noise level. † marks the matched case, where the test noise equals the training noise.

Measurement noise σ_{n}	PSNR↑	SSIM↑	LPIPS↓	input PSNR	Gain
0.01	34.29	0.940	0.020	29.96	+4.33
0.02	34.15	0.939	0.019	29.62	+4.53
0.03	33.89	0.936	0.018	29.12	+4.77
0.04	33.50	0.930	0.017	28.52	+4.98
0.05†	32.90	0.919	0.018	27.81	+5.09

source toward the least-squares anchor. Table 7 tests the first case: the model and its source keep $\sigma_{\mathrm{n}} = 0.05$, while the test measurements carry noise between 0.01 and 0.05, an over-estimate of up to a factor of 5. Reconstruction degrades gracefully. PSNR still rises as the measurements get cleaner, from 32.90 to 34.29 dB, and LPIPS stays between 0.017 and 0.020, although the gain over the input shrinks from 5.09 to 4.33 dB as the mismatch grows. Over this range the method therefore does not require σ_{n} to be known exactly; under-estimation is not tested here.

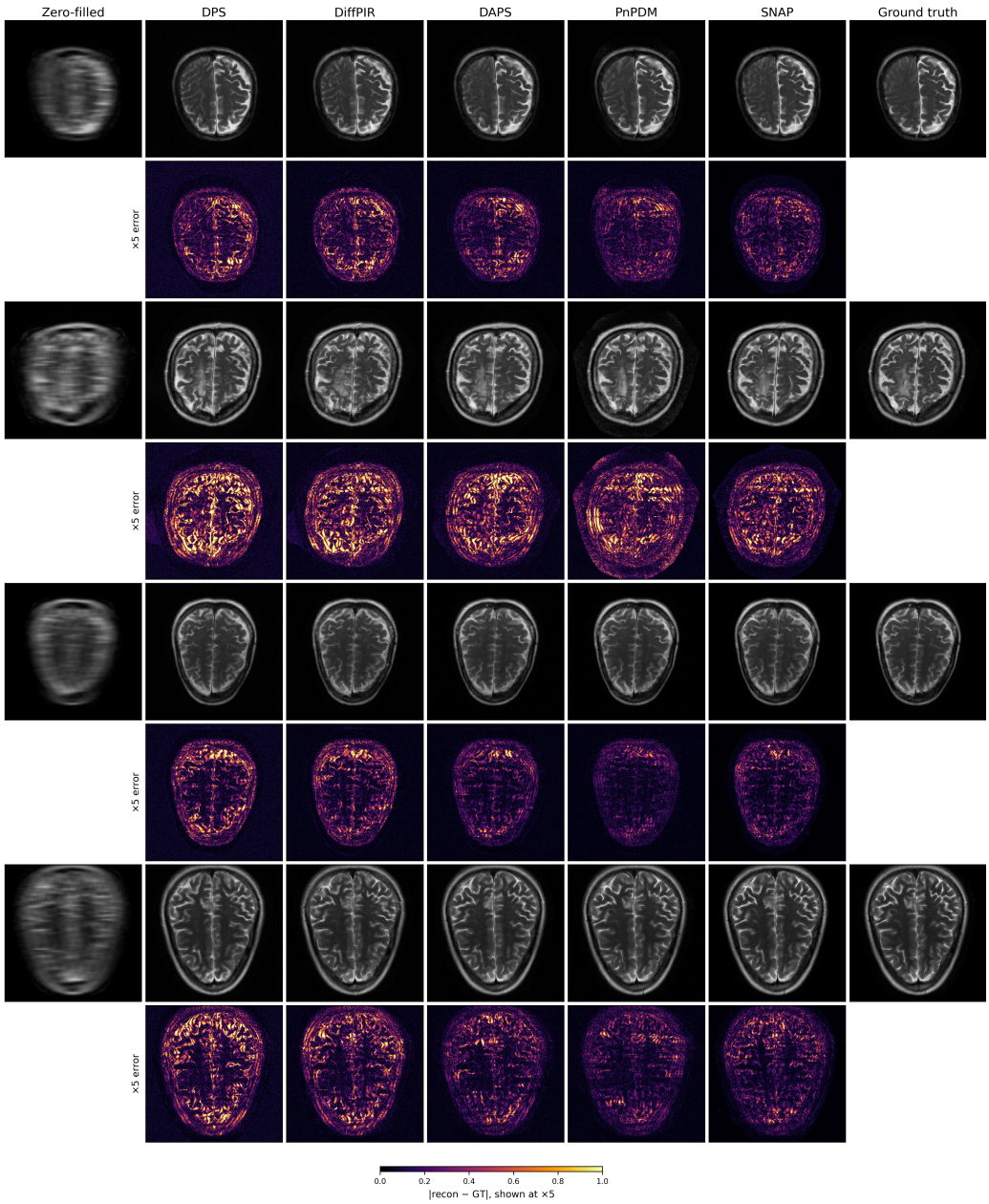

Figure 7: fastMRI brain, multi-coil, $\times 8$ acceleration at 30 dB input SNR.

Working-prior scale τ. Proposition 2 states that τ shapes the source and the transport but not the target, so it is a design parameter rather than a modelling choice. Table 8 reports CelebA Gaussian deblurring for $\tau \in \{0.15, 0.5, 1.0\}$. Distortion is indistinguishable between $\tau = 0.15$ and $\tau = 0.5$ at every M and falls by about 0.4 dB at $\tau = 1.0$, while LPIPS is best at $\tau = 0.5$. The relative spread of the draws is 5.49%, 5.48% and 5.72% for the three settings, essentially unchanged, as the invariance predicts. The method therefore does not require τ to be tuned precisely, and we use the per-operator values listed in App. C.3.

Sampling ratio (r, t). The fraction $p_{\text{ratio}} := \Pr[r{=}t]$ of training samples drawn with $r = t$ controls how often the objective reduces to plain conditional flow matching. Table 9 reports CelebA Gaussian deblurring at $\tau = 0.15$ for $p_{\text{ratio}} \in \{0.25, 0.5, 0.75\}$. Setting $p_{\text{ratio}} = 0.25$ is worse at every M, by about 0.3 dB PSNR at a single draw, while 0.5 and 0.75 are indistinguishable within

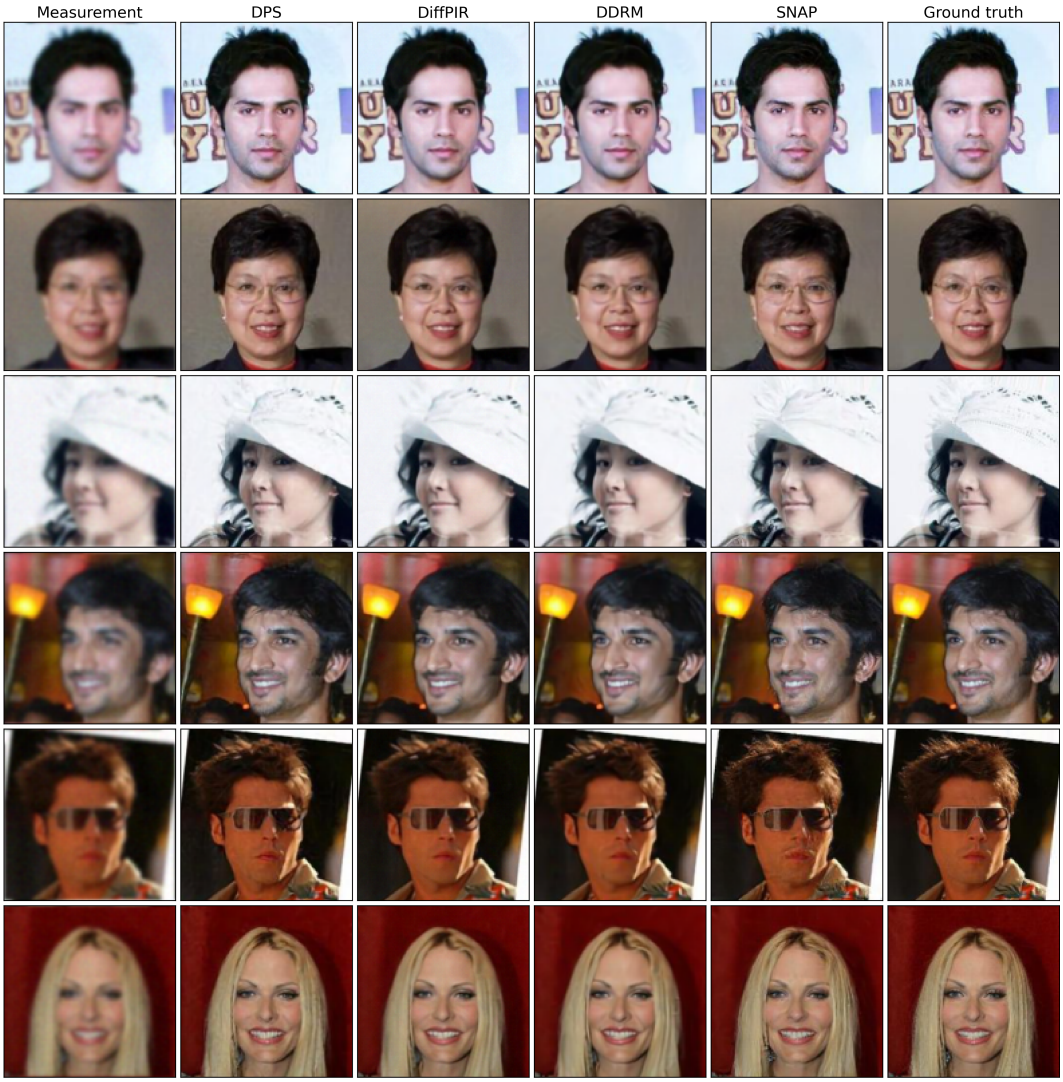

Figure 8: Gaussian deblurring on CelebA.

Table 8: **Working-prior scale τ.** CelebA Gaussian deblurring, 100 test measurements, input PSNR 27.24 dB. M is the number of averaged draws

	$\tau = 0.15$			$\tau = 0.5$			$\tau = 1.0$		
M	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
1	**32.941**	**0.9198**	0.0190	32.922	0.9187	**0.0186**	32.512	0.9060	0.0244
4	**34.966**	**0.9463**	0.0209	34.948	0.9457	**0.0181**	34.588	0.9385	0.0197
16	**35.688**	**0.9537**	0.0275	35.663	0.9532	**0.0234**	35.323	0.9475	0.0249
100	**35.913**	**0.9558**	0.0300	35.885	0.9553	**0.0256**	35.552	0.9500	0.0273

the precision reported. We therefore use $p_{\text{ratio}} = 0.5$ throughout. The trend in M is the same for all three settings, with PSNR and SSIM improving as draws are averaged and LPIPS degrading, as the average moves toward the posterior mean.

Few-step composition. Composing $\boldsymbol{u}_{r,t}$ over a uniform partition of $[0, 1]$ gives k-step sampling with no retraining, and Table 10 reports $k \in \{1, 2, 4, 8\}$ on CelebA Gaussian deblurring. More steps are not monotonically better. All metrics peak at $k = 2$ and decline afterwards, and $k = 8$

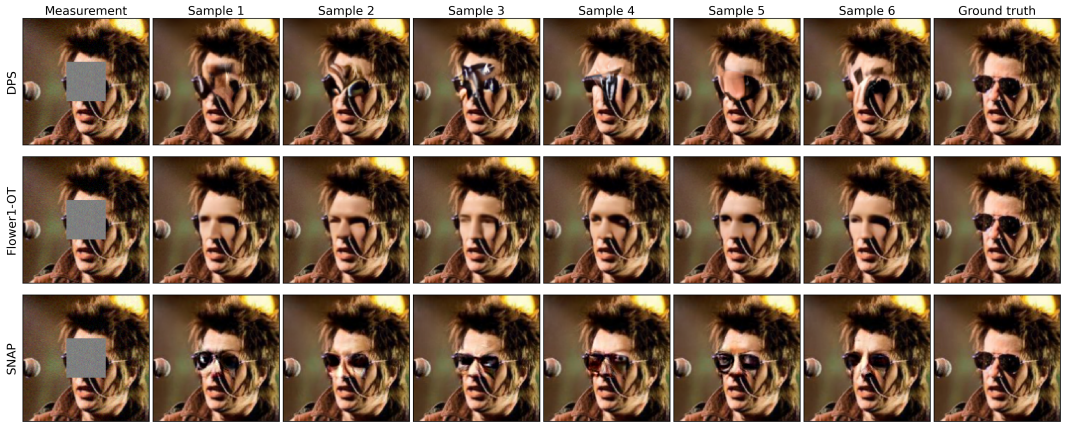

Figure 9: Independent SNaP draws for box inpainting.

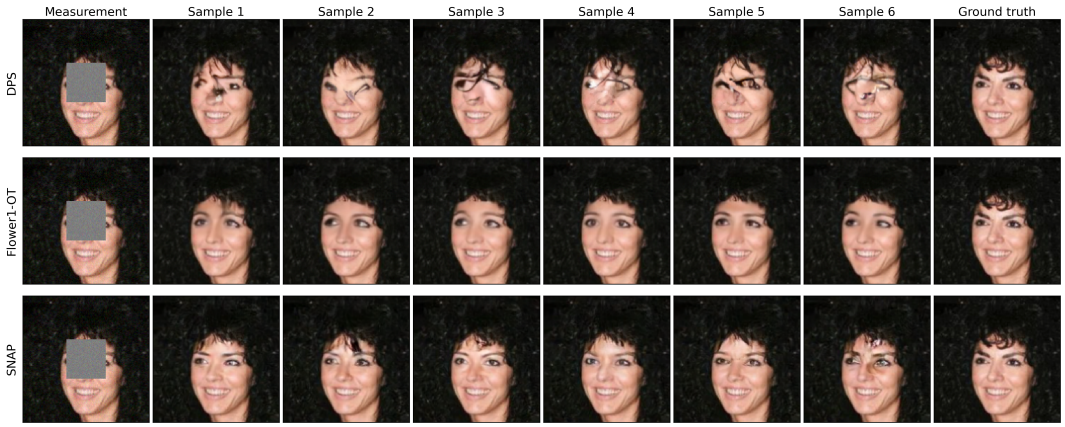

Figure 10: Further independent SNaP draws for box inpainting.

is worse than a single step for $M \geq 16$. The model is trained to jump from the source to the data, so discretizing the trajectory evaluates $\boldsymbol{u}^{\boldsymbol{\theta}}$ at intermediate (r, t) pairs at which its error compounds across steps rather than cancelling. The gain from $k = 2$ is small, so we report $k = 1$ throughout.

Source distribution. Table 5 isolates the source quantitatively; Figures 12 and 13 show the same comparison sample by sample and as a function of the averaging budget M. All three settings share the network, the conditioning, and the training schedule, differing only in the distribution the one-step map starts from: our measurement-adapted posterior source at $\tau = 0.15$ versus isotropic Gaussian sources at $\tau = 0.15$ and $\tau = 1.0$.

Figure 12 shows six draws per source for one measurement. Under the proposed posterior source the draws differ visibly in the hair, the skin texture and the region occluded by the sunglasses, and the standard deviation over draws concentrates on exactly the edges and high-frequency structures that the blur has suppressed, leaving the smooth background, which the measurement already determines, near zero. Under either isotropic Gaussian source the six draws are visually indistinguishable and the standard-deviation map is essentially black at the same $\times 20$ gain: the map has become a point estimator that happens to take a random input. Raising the Gaussian scale from $\tau = 0.15$ to $\tau = 1.0$ does not restore the spread, so the shape of the source is what dictates the spread rather than its size. Figure 13 makes the consequence quantitative.

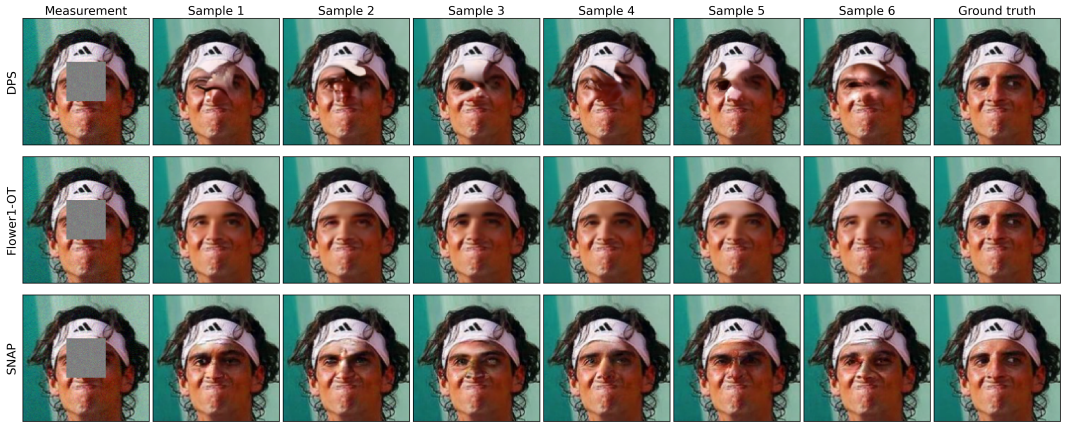

Figure 11: Further independent SNaP draws for box inpainting.

Table 9: (r, t) **sampling ratio.** CelebA Gaussian deblurring, $\tau = 0.15$, 100 test measurements, input PSNR 27.24 dB. M is the number of averaged draws.

	$p_{\text{ratio}} = 0.25$			$p_{\text{ratio}} = 0.5$			$p_{\text{ratio}} = 0.75$		
M	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
1	32.567	0.9130	0.0216	32.941	0.9198	0.0190	32.976	0.9210	0.0189
4	34.679	0.9429	0.0197	34.966	0.9463	0.0209	34.976	0.9465	0.0206
16	35.438	0.9513	0.0266	35.688	0.9537	0.0275	35.682	0.9536	0.0268
100	35.676	0.9537	0.0296	35.913	0.9558	0.0300	35.902	0.9557	0.0292

B.4 Comparison with VFM

VFM (Mammadov et al., 2026) also places the measurement in the source, but learns it: an adapter $q_\phi(\boldsymbol{z} \mid \boldsymbol{y})$ is trained jointly with an unconditional flow map. Its results are reported in the latent space of a pretrained autoencoder, whereas SNaP operates in pixel space, so the published numbers are not directly comparable. We therefore compare the two constructions on the 2D benchmark of Mammadov et al. (2026). The prior is a 4×4 checkerboard on $[-2, 2]^2$ with 20k training samples, observed through $\boldsymbol{y} = \boldsymbol{a}^\top \boldsymbol{x} + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ and $\sigma = 0.1$. We use two forward operators that differ only in the direction of $\boldsymbol{a}$, an aligned operator $\boldsymbol{a} = (1, 0)$, which is the one used in the original paper, and a rotated operator $\boldsymbol{a} = (\cos \frac{\pi}{5}, \sin \frac{\pi}{5})$. All methods share one backbone, a 6-layer width-512 SiLU MLP in $\boldsymbol{x}_1$-prediction form with random-Fourier embeddings of r, t and σ, and one budget, AdamW with learning rate 2×10^{-4}, batch 2048, 50k steps and EMA 0.999. The only thing that varies is where the measurement enters the generative process. The learned-adapter baseline follows VFM. The flow map never sees $\boldsymbol{y}$, and all measurement information is carried by a diagonal Gaussian adapter $q_\phi(\boldsymbol{z} \mid \boldsymbol{y}) = \mathcal{N}(\boldsymbol{\mu}_\phi(\boldsymbol{y}), \operatorname{diag} \boldsymbol{\sigma}^2_\phi(\boldsymbol{y}))$ trained jointly with the flow map under the variational objective, with an observation term, a KL term and an EMA copy of the weights for stability. We also report a frozen-$\boldsymbol{\theta}$ ablation in which the pretrained flow map is held fixed and only q_ϕ is trained. SNaP instead keeps the flow map conditioned on $\boldsymbol{y}$ and draws its source in closed form by Eq. (13), with no adapter and no auxiliary losses.

The source covariance $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^\top/(\|\boldsymbol{a}\|_2^2 + \lambda)$ is diagonal in the basis $\{\boldsymbol{a}, \boldsymbol{a}^\perp\}$, and in the coordinate basis only when the two coincide. For $\boldsymbol{a} = (1, 0)$ they coincide, so a diagonal adapter is exactly expressive and the aligned operator is a parity check. Rotating $\boldsymbol{a}$ breaks the coincidence while holding the prior, the noise level, the architecture and the training budget fixed. The resulting gap can be quantified before any training: the KL divergence from $\mathcal{N}(\boldsymbol{m}_{\boldsymbol{y}}, \tau^2 \boldsymbol{W})$ to its best diagonal approximation is 0 nats for the aligned operator at every τ, and 1.576 nats for the rotated at $\tau = 1$.

On the aligned operator, where the diagonal adapter can represent the exact source, all three methods are close, with SNaP already the most accurate (Table 11). On the rotated operator the two constructions separate. The jointly trained adapter degrades by roughly a factor of seven and the frozen variant by more than an order of magnitude, while SNaP changes far less, since its source

Table 10: **Trajectory steps** k**.** CelebA Gaussian deblurring, $\tau = 0.15$, 100 test measurements, input PSNR 27.24 dB. M is the number of averaged draws. **Best** in each row and metric is in bold.

	$k=1$			$k=2$			$k=4$			$k=8$		
M	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
1	32.941	0.9198	0.0190	**35.084**	**0.9211**	**0.0172**	33.037	0.9205	**0.0172**	32.992	0.9200	0.0173
4	34.966	0.9463	0.0209	**35.044**	**0.9469**	**0.0206**	35.001	0.9465	**0.0206**	34.942	0.9458	0.0210
16	35.688	0.9537	**0.0275**	**35.729**	**0.9540**	**0.0275**	35.688	0.9536	0.0276	35.621	0.9529	0.0281
100	35.913	0.9558	**0.0300**	**35.943**	**0.9561**	0.0302	35.902	0.9557	0.0303	35.833	0.9550	0.0308

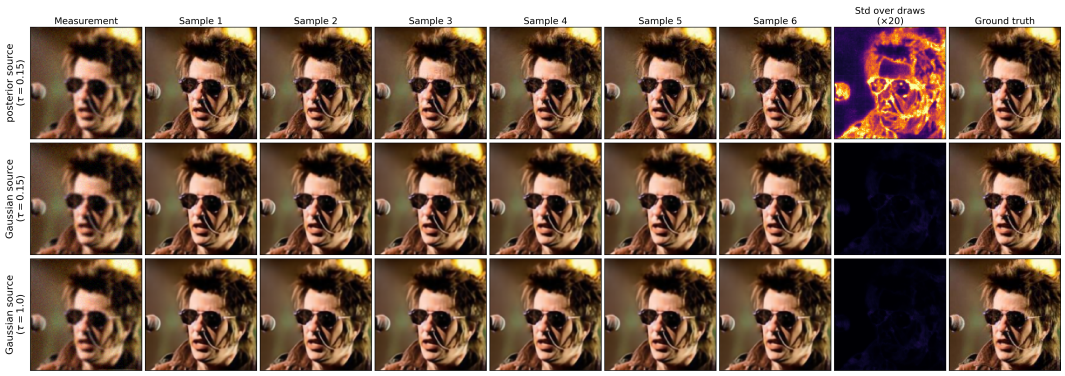

Figure 12: **Source distribution, qualitative.** Six independent draws from a fixed CelebA Gaussian-deblurring measurement under the three sources, with the per-pixel standard deviation over draws on the same $\times 20$ scale in every row. The posterior source varies where the measurement is uninformative; both Gaussian sources collapse onto a single reconstruction.

rotates with the operator by construction. The gaps are significant under a Welch test against the jointly trained adapter, with $p = 0.0005$ on the aligned operator and $p = 0.038$ on the rotated one. Figure 14 shows the mechanism, and the effect is also visible in the support of the samples, with a support accuracy of 0.97 ± 0.004 for SNaP against 0.81 ± 0.03 for the jointly trained adapter on the rotated operator.

B.5 Comparison with NullFlow

NullFlow is defined only for noiseless measurements, so it cannot enter the benchmarks of Sec. 4. We instead sweep the test noise on CelebA box inpainting from $\sigma_n = 0$, where NullFlow is exact, to the $\sigma_n = 0.05$ of our main experiments (Table 12). NullFlow is trained noiselessly with an LPIPS and feature loss, and SNaP at $\sigma_n = 0.05$ with the standard MeanFlow loss, so LPIPS favors NullFlow.

Table 12 reports both methods on CelebA box inpainting, using a single draw and λ frozen at each model's training value. At $\sigma_{\mathrm{n}} = 0$, NullFlow is the stronger method, which is expected, since its source is exact in the noiseless limit and every state satisfies $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$. Its accuracy falls quickly as noise is added, and by $\sigma_{\mathrm{n}} = 0.05$ its SSIM has dropped from 0.977 to 0.855 and its LPIPS has more than quadrupled. This is the failure mode identified in Sec. 2, where the least-squares anchor carries $\boldsymbol{A}^{\dagger}\boldsymbol{n}$ that a null-space flow cannot remove. SNaP changes little over the same range and is the better method at $\sigma_{\mathrm{n}} = 0.05$ on all three metrics. The two methods therefore occupy different regimes, with NullFlow suited to noiseless measurements and SNaP to noisy ones.

B.6 Irreducible regression error: SNaP vs. isotropic sources

Proposition 2 fixes the target of the exact one-step map, but not how hard that map is to learn. We therefore compare sources through the regression problem that training solves. At $r = t$ the MeanFlow target Eq. (14) reduces to the conditional velocity $\boldsymbol{x}_1 - \boldsymbol{x}_0$, which the network regresses on $(\boldsymbol{z}_t, \boldsymbol{y})$ with $\boldsymbol{z}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$. The error left by the best predictor,

$$\mathcal{E}_t(\boldsymbol{y}) := \mathbb{E}[\operatorname{tr}\operatorname{Cov}(\boldsymbol{x}_1 - \boldsymbol{x}_0 \,|\, \boldsymbol{z}_t, \boldsymbol{y}) \,|\, \boldsymbol{y}], \tag{21}$$

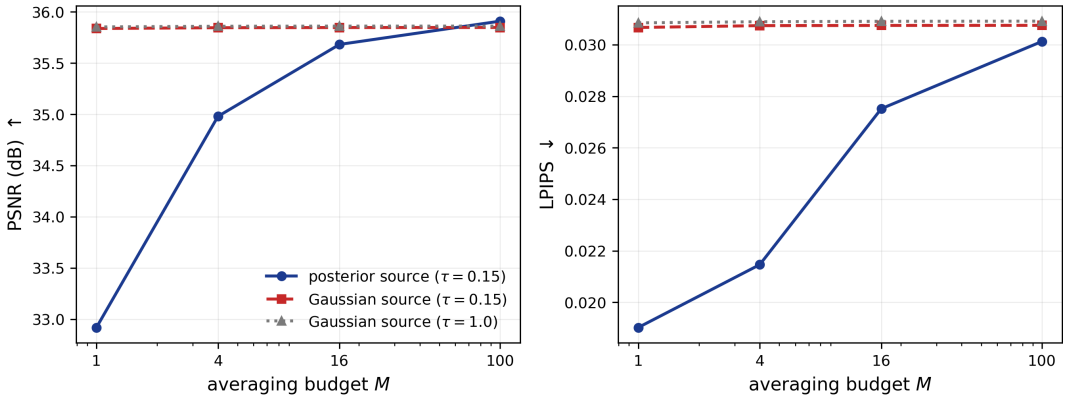

Figure 13: **Source distribution, averaging budget.** PSNR (left, higher is better) and LPIPS (right, lower is better) against the number of averaged draws M on CelebA Gaussian deblurring. Only the posterior source responds to M; the Gaussian sources are flat because their draws carry almost no variance to average away.

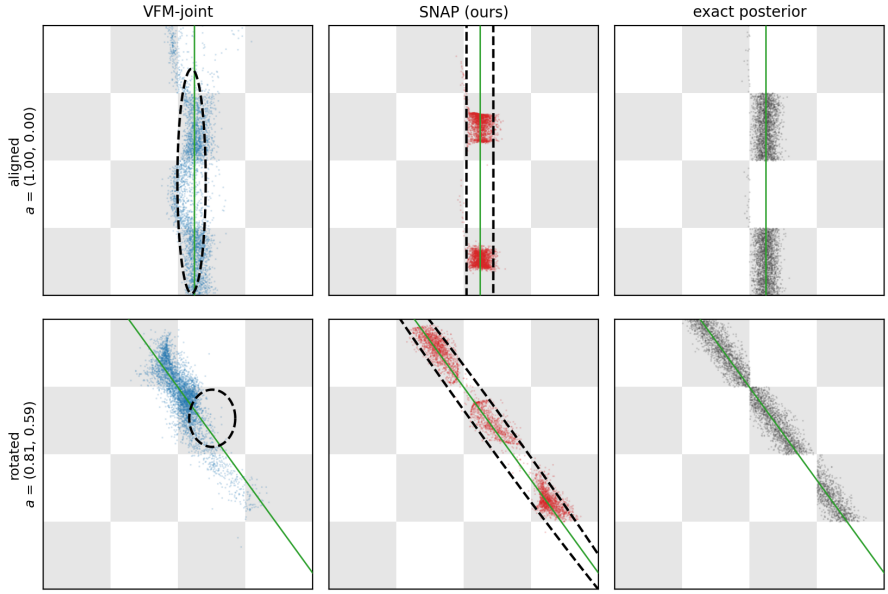

Figure 14: One-step ($K{=}1$) posterior samples on the $4{\times}4$ checkerboard. Rows are the two forward operators and columns are the learned adapter of VFM, the SNaP source, and the exact rejection-sampled posterior. The green line is the measurement constraint $\boldsymbol{a}^\top \boldsymbol{x} = \boldsymbol{y}$, and the dashed ellipse on each method panel is that method's 2σ source covariance, the learned diagonal $q_\phi(\boldsymbol{z} \mid \boldsymbol{y})$ for VFM and the closed-form $\tau^2 \boldsymbol{W}$ for SNaP. In the top row the measurement direction is axis-aligned, both sources align with the constraint, and both recover the posterior. In the bottom row the constraint tilts. The SNaP source tilts with it, whereas the diagonal adapter cannot and degenerates to a near-isotropic blob, so its samples concentrate on one mode and spill off the checkerboard support.

cannot be removed by any network, so a source with a smaller $\mathcal{E}_t$ poses an easier regression problem at time t.

Lemma 2. *Let $\sigma_{\mathrm{n}} > 0$, fix $\boldsymbol{y}$ and $t \in (0,1)$, and let $\boldsymbol{x}_1 \sim p(\cdot \mid \boldsymbol{y})$ have finite second moments. Let $\boldsymbol{x}_0 \sim q_\tau(\cdot \mid \boldsymbol{y})$, and let $\widetilde{\boldsymbol{x}}_0 = \widetilde{\boldsymbol{K}}\boldsymbol{y} + \widetilde{\boldsymbol{\xi}}$ be a source of the form Eq. (6) with $\widetilde{\boldsymbol{\xi}} \sim \mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}})$ and $\widetilde{\boldsymbol{\Sigma}} \succeq \tau^2 \boldsymbol{W}$, each conditionally independent of $\boldsymbol{x}_1$ given $\boldsymbol{y}$. Let $\mathcal{E}_t(\boldsymbol{y})$ and $\widetilde{\mathcal{E}}_t(\boldsymbol{y})$ be the errors Eq. (21) for the interpolants $\boldsymbol{z}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$ and $\widetilde{\boldsymbol{z}}_t = (1-t)\widetilde{\boldsymbol{x}}_0 + t\boldsymbol{x}_1$. Then $\mathcal{E}_t(\boldsymbol{y}) \le \widetilde{\mathcal{E}}_t(\boldsymbol{y})$. In particular, this holds for the isotropic source $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ for every $\sigma \ge \tau$, since $\tau^2 \boldsymbol{W} \preceq \tau^2 \boldsymbol{I} \preceq \sigma^2 \boldsymbol{I}$.*

Table 11: Posterior MMD^2 (↓) on the two-dimensional checkerboard benchmark at $K{=}1$, mean $\pm$ standard deviation over three seeds, each at its validation-selected setting. The aligned operator is a parity check, where a diagonal adapter is exactly expressive, and the rotated operator differs only in the direction of $\boldsymbol{a}$.

Method	Aligned (parity)	Rotated (discriminating)
SNaP	0.0027 ± 0.0003	0.0171 ± 0.0012
VFM (joint)	0.0109 ± 0.0007	0.0737 ± 0.0199
VFM (frozen $\boldsymbol{\theta}$)	0.0102 ± 0.0003	0.1900 ± 0.0015

Table 12: NullFlow vs. SNaP on CelebA box inpainting ($M{=}1$) as test noise grows.

	NullFlow			SNaP		
Test σ_{n}	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
0.00	33.50	0.977	0.012	31.37	0.947	0.025
0.01	33.27	0.970	0.012	31.35	0.947	0.024
0.02	32.65	0.952	0.016	31.30	0.946	0.023
0.05	30.11	0.855	0.055	30.74	0.934	0.021

The lemma requires nothing of the posterior beyond finite second moments, and the center of the source does not enter it: given $\boldsymbol{y}$, the center only shifts $\boldsymbol{z}_t$ by a constant, which conditioning removes. The center is what Proposition 1 chooses to shorten the displacement; the lemma shows that the covariance $\tau^2\boldsymbol{W}$ is what lowers the regression error. Both isotropic sources of Table 5 satisfy its assumption, since $\tau = 0.15 \le 1$ there.

Proof. Given $\boldsymbol{y}$, write $\boldsymbol{x}_0 = \boldsymbol{m}_{\boldsymbol{y}} + \boldsymbol{\eta}$ and $\widetilde{\boldsymbol{x}}_0 = \boldsymbol{c} + \widetilde{\boldsymbol{\xi}}$ with $\boldsymbol{c} := \widetilde{\boldsymbol{K}}\boldsymbol{y}$, where $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \tau^2\boldsymbol{W})$ and $\widetilde{\boldsymbol{\xi}}$ are independent of $\boldsymbol{x}_1$. Since $\boldsymbol{m}_{\boldsymbol{y}}$ and $\boldsymbol{c}$ are constants given $\boldsymbol{y}$, the maps $\boldsymbol{z} \mapsto (\boldsymbol{z} - (1-t)\boldsymbol{m}_{\boldsymbol{y}})/t$ and $\boldsymbol{z} \mapsto (\boldsymbol{z} - (1-t)\boldsymbol{c})/t$ are invertible, so conditioning on $\boldsymbol{z}_t$ and on $\widetilde{\boldsymbol{z}}_t$ is the same as conditioning on

$$\bar{\boldsymbol{z}}_t := \boldsymbol{x}_1 + \tfrac{1-t}{t}\,\boldsymbol{\eta} \qquad \text{and} \qquad \bar{\widetilde{\boldsymbol{z}}}_t := \boldsymbol{x}_1 + \tfrac{1-t}{t}\,\widetilde{\boldsymbol{\xi}},$$

respectively. Each is $\boldsymbol{x}_1$ observed through additive Gaussian noise, and the two differ only in that noise.

Since $\widetilde{\boldsymbol{\Sigma}} - \tau^2\boldsymbol{W} \succeq \boldsymbol{0}$, let $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}} - \tau^2\boldsymbol{W})$ be independent of $(\boldsymbol{x}_1, \boldsymbol{\eta})$ given $\boldsymbol{y}$. Then $\boldsymbol{\eta} + \boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}})$, and because $\widetilde{\mathcal{E}}_t(\boldsymbol{y})$ depends only on the joint law of $\boldsymbol{x}_1$ and $\widetilde{\boldsymbol{z}}_t$ given $\boldsymbol{y}$, we may take $\widetilde{\boldsymbol{\xi}} = \boldsymbol{\eta} + \boldsymbol{\zeta}$. This gives $\bar{\widetilde{\boldsymbol{z}}}_t = \bar{\boldsymbol{z}}_t + \frac{1-t}{t}\,\boldsymbol{\zeta}$, so $\boldsymbol{x}_1 \to \bar{\boldsymbol{z}}_t \to \bar{\widetilde{\boldsymbol{z}}}_t$ is a Markov chain given $\boldsymbol{y}$.

Let $\boldsymbol{\mu}_t := \mathbb{E}[\boldsymbol{x}_1 \mid \bar{\boldsymbol{z}}_t, \boldsymbol{y}]$ and $\widetilde{\boldsymbol{\mu}}_t := \mathbb{E}[\boldsymbol{x}_1 \mid \bar{\widetilde{\boldsymbol{z}}}_t, \boldsymbol{y}]$. The Markov property gives $\mathbb{E}[\boldsymbol{x}_1 \mid \bar{\boldsymbol{z}}_t, \bar{\widetilde{\boldsymbol{z}}}_t, \boldsymbol{y}] = \boldsymbol{\mu}_t$, and the tower rule gives $\widetilde{\boldsymbol{\mu}}_t = \mathbb{E}[\boldsymbol{\mu}_t \mid \bar{\widetilde{\boldsymbol{z}}}_t, \boldsymbol{y}]$. Hence $\boldsymbol{x}_1 - \boldsymbol{\mu}_t$ is orthogonal to $\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}}_t$, and

$$\mathbb{E}\Big[\|\boldsymbol{x}_1 - \widetilde{\boldsymbol{\mu}}_t\|_2^2 \mid \boldsymbol{y}\Big] = \mathbb{E}\Big[\|\boldsymbol{x}_1 - \boldsymbol{\mu}_t\|_2^2 \mid \boldsymbol{y}\Big] + \mathbb{E}\Big[\|\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}}_t\|_2^2 \mid \boldsymbol{y}\Big]. \tag{22}$$

Finally, $\boldsymbol{x}_1 - \boldsymbol{x}_0 = (\boldsymbol{x}_1 - \boldsymbol{z}_t)/(1-t)$ and $\boldsymbol{z}_t$ is fixed once it is conditioned on, so $\mathrm{Cov}(\boldsymbol{x}_1 - \boldsymbol{x}_0 \mid \boldsymbol{z}_t, \boldsymbol{y}) = (1-t)^{-2}\,\mathrm{Cov}(\boldsymbol{x}_1 \mid \boldsymbol{z}_t, \boldsymbol{y})$ and $\mathcal{E}_t(\boldsymbol{y}) = (1-t)^{-2}\,\mathbb{E}[\|\boldsymbol{x}_1 - \boldsymbol{\mu}_t\|_2^2 \mid \boldsymbol{y}]$; likewise $\widetilde{\mathcal{E}}_t(\boldsymbol{y}) = (1-t)^{-2}\,\mathbb{E}[\|\boldsymbol{x}_1 - \widetilde{\boldsymbol{\mu}}_t\|_2^2 \mid \boldsymbol{y}]$. Dividing Eq. (22) by $(1-t)^2$ gives

$$\widetilde{\mathcal{E}}_t(\boldsymbol{y}) - \mathcal{E}_t(\boldsymbol{y}) = \frac{1}{(1-t)^2}\,\mathbb{E}\Big[\|\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}}_t\|_2^2 \mid \boldsymbol{y}\Big] \;\ge\; 0.$$

□

Displacement. The same source also shortens the path relative to the isotropic source. Let $\widetilde{\boldsymbol{x}}_0 \sim \mathcal{N}(\boldsymbol{0}, \tau^2\boldsymbol{I})$ be independent of $(\boldsymbol{x}_1, \boldsymbol{n})$, and let $E_i := \mathbb{E}[(\boldsymbol{v}_i^\top \boldsymbol{x}_1)^2]$ for a right singular vector $\boldsymbol{v}_i$ of $\boldsymbol{A}$. For any $\boldsymbol{x}_1$ with finite second moments,

$$\mathbb{E}\big[(\boldsymbol{v}_i^\top(\boldsymbol{x}_1 - \widetilde{\boldsymbol{x}}_0))^2\big] - \mathbb{E}\big[(\boldsymbol{v}_i^\top(\boldsymbol{x}_1 - \boldsymbol{x}_0))^2\big] = (1 - w_i^2)E_i + \tau^2(1-w_i)^2 \;\ge\; 0, \tag{23}$$

with equality if and only if $s_i = 0$. Indeed, writing $a_i := \boldsymbol{v}_i^\top \boldsymbol{x}_1$ and $n_i := \boldsymbol{u}_i^\top \boldsymbol{n}$ (with $n_i := 0$ if $s_i = 0$), the Tikhonov center gives $\boldsymbol{v}_i^\top(\boldsymbol{x}_1 - \boldsymbol{x}_0) = w_i a_i - \frac{s_i}{s_i^2+\lambda} n_i - \boldsymbol{v}_i^\top \boldsymbol{\eta}$, a sum of uncorrelated terms with second moments $w_i^2 E_i$, $\tau^2 w_i(1-w_i)$ and $\tau^2 w_i$, whereas the isotropic source gives $E_i + \tau^2$.

Evaluation. Figure 15 evaluates $\mathcal{E}_t$ for $q_\tau(\cdot \mid \boldsymbol{y})$ and $\mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I})$ in two settings where it is available exactly. On the two-dimensional mixture of App. B.7 at $y = 2.5$, the conditional covariance is computed by quadrature against the closed-form posterior. The two sources coincide in the null direction there, so the gap comes entirely from the observed direction. On CelebA Gaussian deblurring, gap is evaluated under a stationary Gaussian prior with the empirical power spectrum, for which $\boldsymbol{A}$, $\boldsymbol{W}$ and the posterior covariance are simultaneously diagonal in the Fourier basis. There the gap is present at every t and peaks at 56%, since the graded Fourier weights w_i leave the source something to exploit in every direction.

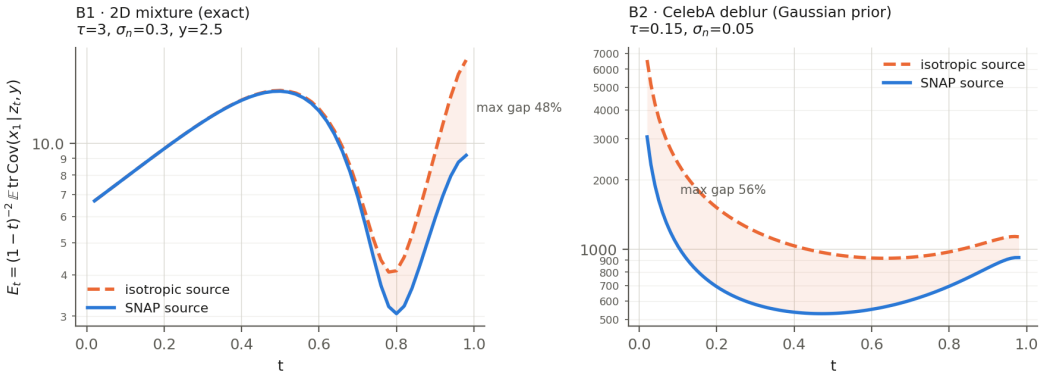

Figure 15: **Irreducible velocity-regression error** $\mathcal{E}_t$ of Eq. (21) under the SNaP source $q_\tau(\cdot \mid \boldsymbol{y})$ and the isotropic source $\mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I})$. **Left:** two-dimensional mixture of App. B.7 ($\tau = 3$, $\sigma_{\mathrm{n}} = 0.3$, $y = 2.5$), exact. **Right:** CelebA Gaussian deblurring ($\tau = 0.15$, $\sigma_{\mathrm{n}} = 0.05$) under a stationary Gaussian prior. Lemma 2 guarantees the ordering; the shaded region is the gap.

Scope. Three limits apply. First, the lemma concerns the $r = t$ part of the objective. For $r < t$ the target Eq. (14) also contains the network's own derivatives, so the lemma does not transfer verbatim to the interior of the (r, t) range; a fraction p_{ratio} of training pairs have $r = t$ (Table 9). Second, like the displacement in Eq. (7), $\mathcal{E}_t$ decreases as the source covariance shrinks, so it cannot be used to choose $\boldsymbol{\Sigma}$, and the lemma only compares q_τ with sources that are at least as spread. A narrower source can have a smaller $\mathcal{E}_t$: on the mixture of App. B.7, where $\tau = 3$, the source $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ has a smaller $\mathcal{E}_t$ than q_τ at every t, because it is narrower in the null direction. Third, the lemma compares regression problems, not trained samplers; Table 5 and App. B.7 compare the samplers directly, with everything but the source held fixed.

App. B.6 shows that, for any posterior with finite second moments, the velocity-regression error that no network can remove is smaller under $q_\tau(\cdot \mid \boldsymbol{y})$ than under any isotropic source $\mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$ with $\sigma \geq \tau$ (Lemma 2); Table 5 and App. B.7 measure the effect on the learned map.

B.7 Verification on a Gaussian-mixture inverse problem

We verify the construction on a problem where the posterior is available in closed form. With a four-component Gaussian-mixture prior on $\mathbb{R}^2$ and observation of the first coordinate ($\boldsymbol{A} = \boldsymbol{e}_0^\top$, $\sigma_{\mathrm{n}} = 0.3$), the posterior is itself a Gaussian mixture, and for $y \approx 2.5$ it is bimodal over $\boldsymbol{x}[1] \in \{\pm 2.5\}$ with the MMSE estimate falling between the modes. The setting therefore separates posterior sampling from posterior-mean estimation: a point estimator lands in the empty valley between the modes, and a mode-collapsed sampler reports the component variance $\varsigma^2 = 0.12$ in the null direction rather than the posterior variance 6.36.

Setup. The clean data $\boldsymbol{x}_1 \in \mathbb{R}^2$ is drawn from a four-component equal-weight Gaussian mixture with isotropic component covariance $\varsigma^2 \boldsymbol{I}$ ($\varsigma = 0.35$) and means at $(\pm a, \pm a)$, $a = 2.5$. The forward operator observes only the first coordinate, so $m = 1$, $n = 2$ and $\mathrm{null}(\boldsymbol{A}) = \mathrm{span}(\boldsymbol{e}_1)$. The working-prior scale is $\tau = 3$.

Closed-form posterior. For each mixture component k,

$$\widetilde{\boldsymbol{\Sigma}}_k = \big(\boldsymbol{\Sigma}_k^{-1} + \boldsymbol{A}^\top \boldsymbol{A}/\sigma_{\mathrm{n}}^2\big)^{-1}, \quad \widetilde{\boldsymbol{\mu}}_k = \widetilde{\boldsymbol{\Sigma}}_k\big(\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k + \boldsymbol{A}^\top \boldsymbol{y}/\sigma_{\mathrm{n}}^2\big), \tag{24}$$

and

$$\widetilde{\pi}_k \propto \pi_k \, \mathcal{N}\big(\boldsymbol{y}; \boldsymbol{A}\boldsymbol{\mu}_k, \boldsymbol{A}\boldsymbol{\Sigma}_k\boldsymbol{A}^\top + \sigma_{\mathrm{n}}^2\big), \qquad p(\boldsymbol{x}_1 \mid \boldsymbol{y}) = \sum_k \widetilde{\pi}_k \mathcal{N}(\widetilde{\boldsymbol{\mu}}_k, \widetilde{\boldsymbol{\Sigma}}_k). \tag{25}$$

All quantities labeled "posterior" below are computed from this expression.

Network and optimization. An $\boldsymbol{x}$-prediction MLP (width 256, depth 4, SiLU) with random-Fourier embeddings (dimension 64) for r, t and σ_{n}, conditioned on $\boldsymbol{y}$, with $\boldsymbol{u}^{\boldsymbol{\theta}}_{r,t} = \big(\mathrm{net}(\boldsymbol{z}, r, t, \boldsymbol{y}, \sigma_{\mathrm{n}}) - \boldsymbol{z}\big)/(1-r)$. We use the objective Eq. (15) with $p = 1$ and $c = 10^{-3}$; t is logit-normal ($\mu = 0$, $\sigma = 1$) and $r = t$ with probability 0.75, else $r \sim \mathcal{U}(0, t)$. AdamW, learning rate 2×10^{-4}, batch size 4096, 20,000 steps, EMA decay 0.999. One-step sampling is $\widehat{\boldsymbol{x}}_1 = \mathrm{net}(\boldsymbol{x}_0, r{=}0, t{=}1, \boldsymbol{y}, \sigma_{\mathrm{n}})$; few-step sampling composes $\boldsymbol{u}_{r,t}$ over a uniform partition of $[0, 1]$.

The source is exact. Drawing 2×10^5 samples from the perturb-and-solve construction in Eq. (13) at $y = 2.5$ and comparing against the closed form reproduces $\mathcal{N}(\boldsymbol{m}_{\boldsymbol{y}}, \tau^2 \boldsymbol{W})$ to Monte-Carlo accuracy: $\|\Delta \boldsymbol{m}_{\boldsymbol{y}}\| = 3.8 \times 10^{-3}$ and $\|\Delta \mathrm{cov}\|_F = 4.8 \times 10^{-3}$ (Table 15). The anisotropy predicted by Eq. (10) is visible directly: the observed direction is pinned at variance 0.089 while the null direction carries the full working-prior variance $\tau^2 = 9$.

Posterior recovery. Figure 16 shows the source, the closed-form posterior, and 2×10^4 one-step samples for three in-distribution measurements. The flow transports the vertically-spread source onto both posterior modes at $\boldsymbol{x}[1] = \pm 2.5$ while pinning $\boldsymbol{x}[0] \approx y$, and the MMSE estimate falls in the empty valley between them. In the null direction the one-step sampler produces variance 6.66 against the closed-form 6.36; a mode-collapsed sampler would report $\varsigma^2 = 0.12$. Table 13 reports mean, per-direction variance and MMD^2 at three measurements.

Per-direction calibration. Projecting onto the right singular vectors of $\boldsymbol{A}$ (here $\boldsymbol{e}_0$ with $s = 1$ and $\boldsymbol{e}_1$ with $s = 0$), the flow assigns almost all of the null direction's freedom (6.66 against posterior 6.36, source 9.0) and contracts the well-measured direction from the source value 0.089 toward the posterior 0.053, reaching 0.079 (Table 16).

Gaussian ground truth. With a single Gaussian prior matched to the working prior ($\boldsymbol{\Sigma} = \tau^2 \boldsymbol{I}$, $\tau = \varsigma = 0.35$), the source *is* the posterior and the optimal map is the identity. Retraining reproduces the source to $\|\Delta \boldsymbol{m}_{\boldsymbol{y}}\| = 5.6 \times 10^{-4}$ and matches the Gaussian posterior with mean errors ≤ 0.05 and MMD$^2 \leq 3 \times 10^{-3}$ (Table 14).

Table 13: Posterior match at three in-distribution measurements.

y	mean (post)	mean (SNaP)	$\|\Delta\text{mean}\|$	Var (observed, null) post $\to$ SNaP	MMD^2
2.5	$(2.5, 0)$	$(2.474, 0.065)$	0.069	$(0.052, 6.37) \to (0.079, 6.66)$	$4.0{\times}10^{-3}$
2.0	$(2.212, 0)$	$(2.045, 0.051)$	0.174	$(0.052, 6.37) \to (0.084, 6.64)$	$1.1{\times}10^{-2}$
-2.5	$(-2.5, 0)$	$(-2.444, 0.108)$	0.121	$(0.052, 6.37) \to (0.081, 6.59)$	$2.6{\times}10^{-3}$

C Implementation Details

C.1 Datasets and evaluation protocol

CelebA. Official aligned images with the canonical partition of 162,770/19,867/19,962 for the train/validation/test images, center-cropped to 178×178 and resized to 128×128. We report results on the first 100 images of the test set.

Table 14: Single-Gaussian ground-truth check. The map reduces to the Wiener estimate, with the posterior mean recovered to within 0.05 and the posterior variances over-estimated by 37–39%.

y	mean (post)	mean (SNaP)	$\|\Delta\text{mean}\|$	Var (observed, null) post → SNaP	MMD^2
0	$(0, 0)$	$(-0.040, 0.026)$	0.048	$(0.052, 0.123) \to (0.071, 0.171)$	2.7×10^{-3}
1	$(0.576, 0)$	$(0.553, 0.012)$	0.026	$(0.052, 0.123) \to (0.069, 0.160)$	1.7×10^{-3}

Table 15: The square-root-free sampler Eq. (13) reproduces $\mathcal{N}(\boldsymbol{m_y}, \tau^2\boldsymbol{W})$ to Monte-Carlo accuracy at $y = 2.5$.

	$\boldsymbol{m_y}$	$\tau^2\boldsymbol{W}$ (diag)	error
closed form	$[2.4752,\ 0]$	$[0.0891,\ 9.000]$	—
RTO Eq. (13)	$[2.4752,\ 0.0038]$	$[0.0889,\ 8.995]$	$\|\Delta\boldsymbol{m_y}\| = 3.8\times10^{-3}$, $\|\Delta\text{cov}\|_F = 4.8\times10^{-3}$

AFHQ-Cat. $5{,}153/500/100$ train/validation/test images, resized to 256×256. AFHQ-Cat has no official validation split. We followed Martin et al. (2025) for train/val/test split.

fastMRI brain. The AXT2 subset, complex multi-coil k-space at 320×320. We split at the volume (patient) level into $615/77/77$ train/val/test volumes, giving $4{,}305/539/539$ slices after discarding the first four and last five slices of every volume, which contain little or no anatomy. Coil handling is full multi-coil SENSE, not RSS or an emulated single-coil reduction. The ground truth is the ESPIRiT coil-combined complex image $\boldsymbol{x}_1 = \sum_i \bar{\boldsymbol{S}}_i \,\text{crop}(\boldsymbol{F}^H \boldsymbol{k}_i)$ on a 320×320 grid, retained as two real channels (real/imaginary) and normalized per slice by $\max|\boldsymbol{x}_1|$; phase is modeled throughout and magnitude is taken only when computing PSNR/SSIM. The forward operator is $\boldsymbol{A} = \boldsymbol{MFS}$ with the same per-slice ESPIRiT maps at train and test time, zero-padded to 20 coils. Because the coils couple, the measurement-space Gram $\boldsymbol{AA}^H$ is not diagonal and the posterior-source solve admits no closed form, unlike the inpainting, super-resolution, and deblurring operators, where $\boldsymbol{A}^H\boldsymbol{A}$ is diagonal in the pixel or Fourier basis and the solve costs $O(n)$ or a single FFT pair. We therefore solve it with conjugate gradient in the algebraically equivalent image-space form $\boldsymbol{A}^H(\boldsymbol{AA}^H + \lambda\boldsymbol{I})^{-1} = (\boldsymbol{A}^H\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^H$, whose spectrum lies in $[\lambda, \approx1]$ and whose right-hand side already lies in $\text{range}(\boldsymbol{A}^H)$; the measurement-space form is rank-deficient by the coil count and, at the same iteration budget, returns an anchor worse than the plain adjoint.

C.2 Solutions of the source construction

The source Eq. (13) requires one application of $\boldsymbol{K} = \boldsymbol{A}^\top(\boldsymbol{AA}^\top + \lambda\boldsymbol{I})^{-1}$ per measurement, with $\lambda = \sigma_\text{n}^2/\tau^2$. Writing $\boldsymbol{A} = \boldsymbol{U\Sigma V}^\top$, the covariance is $\tau^2\boldsymbol{W}$ with $\boldsymbol{W} = \boldsymbol{V}\,\text{diag}(w_i)\boldsymbol{V}^\top$ and $w_i = \lambda/(s_i^2 + \lambda)$. We summarise the cases used in this work.

Image deblurring. For deblurring $\boldsymbol{A} = \mathbf{H}$ is a spatially invariant blur. Under circular boundary conditions $\mathbf{H} = \boldsymbol{F}^H\boldsymbol{\Lambda F}$ with $\boldsymbol{\Lambda}$ the Fourier coefficients $\hat{h}$ of the kernel, so $\mathbf{HH}^H = \boldsymbol{F}^H|\boldsymbol{\Lambda}|^2\boldsymbol{F}$ and

$$\boldsymbol{m_y} = \boldsymbol{F}^H\left[\frac{\overline{\boldsymbol{\Lambda}}\,\boldsymbol{Fy}}{|\boldsymbol{\Lambda}|^2 + \lambda}\right], \qquad w_i = \frac{\lambda}{\left|\hat{h}_i\right|^2 + \lambda}, \tag{26}$$

elementwise in the Fourier domain, at the cost of two FFTs. Because $\left|\hat{h}_i\right| > 0$ for every i, no w_i reaches 1 and no direction is unmeasured; the spectrum instead decays smoothly.

Image inpainting. For inpainting $\boldsymbol{A} = \boldsymbol{M}$ is a row-selection matrix that keeps the observed coordinates, so $\boldsymbol{MM}^\top = \boldsymbol{I}_m$ and the solve is elementwise:

$$(\boldsymbol{m_y})_i = \begin{cases} \dfrac{y_i}{1+\lambda}, & i \in \Omega, \\ 0, & i \notin \Omega, \end{cases} \qquad w_i = \begin{cases} \dfrac{\lambda}{1+\lambda}, & i \in \Omega, \\ 1, & i \notin \Omega, \end{cases} \tag{27}$$

where Ω is the set of observed pixels. Unobserved coordinates carry the full working-prior variance τ^2, while observed coordinates are shrunk toward the prior mean by the factor $1/(1+\lambda)$ and retain

Table 16: Per-direction variance at $y = 2.5$ (Eq. (10)). The source over-disperses both directions relative to the posterior; the trained flow moves both toward it, almost exactly in the null direction and partially in the observed one.

direction	Var posterior	Var one-step SNaP	Var source
$\boldsymbol{e}_0$ (observed, $s = 1$)	0.0518	0.0791	0.0891
$\boldsymbol{e}_1$ (null, $s = 0$)	6.372	6.656	9.000

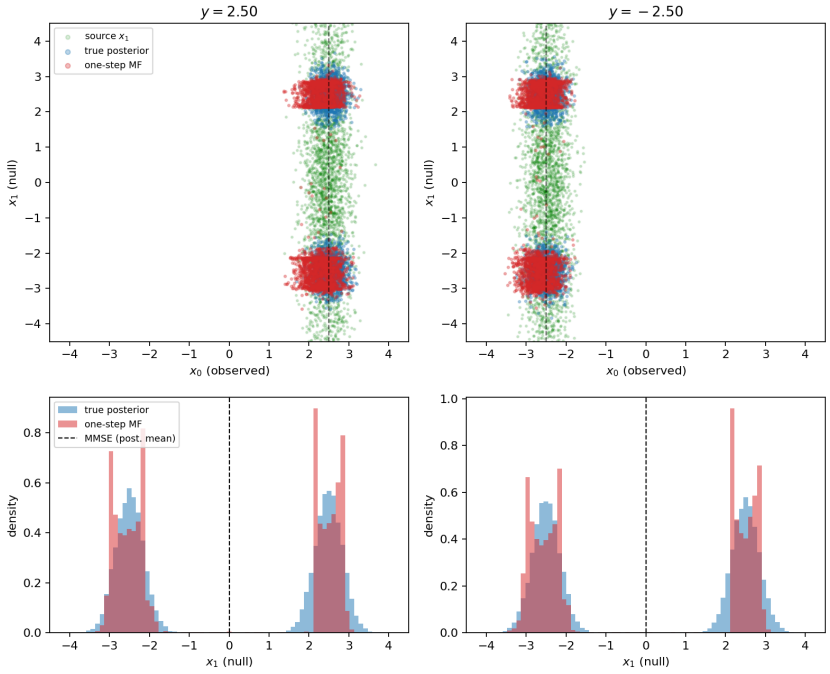

Figure 16: Posterior recovery for observations $y = 2.5$ (left) and $y = -2.5$ (right). The top row compares samples from the source distribution (green), the closed-form posterior (blue), and the one-step model (red). Dashed vertical lines indicate the observed values. The bottom row shows the corresponding marginals along the unobserved coordinate x_1; the dashed line marks the posterior mean. The one-step model captures both posterior modes, although it exhibits some within-mode distortion relative to the exact posterior.

variance $\tau^2\lambda/(1+\lambda)$. Box and random inpainting share this structure and differ only in the geometry of Ω.

Image super-resolution. We use box-decimated super-resolution by implementing strided subsampling with no anti-alias prefilter, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}[\ldots, :: s, :: s]$, so $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}$, $\boldsymbol{A}^\top$ is zero-filled upsampling, and Eq. (27) applies with Ω the sampled grid locations.

Multi-coil MRI. With $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{F}\boldsymbol{S}$ the sensitivity maps $\boldsymbol{S}$ are not unitary, so $\boldsymbol{A}\boldsymbol{A}^\top$ is not diagonal and the solve with $\boldsymbol{A}\boldsymbol{A}^\top + \lambda\boldsymbol{I}$ has no closed form. We solve it with conjugate gradients, warm-started at the zero-filled adjoint. Measured iteration counts and cost are reported in Table 17.

General linear operators. For a general $\boldsymbol{A}$ with no exploitable structure, Eq. (13) is solved by conjugate gradients. All operators used in this paper fall under the cases above. The solve is required once per draw and does not depend on t, r, or the network. For a fixed operator, the solve can be reduced to a diagonal inverse for diagonal (and Fourier diagonal) forward models or use a short conjugate gradient run for most others.

Table 17: Cost of the source construction Eq. (13). Larger λ (noisier data) improves conditioning and accelerates CG.

Problem	$\boldsymbol{A}\boldsymbol{A}^\top$ structure	Cost	Notes
Inpainting (random)	$\boldsymbol{I}_m$	free	$\boldsymbol{W}$ diagonal
Inpainting (box)	$\boldsymbol{I}_m$	free	identical structure
MRI, single-coil	$\boldsymbol{I}_m$	free	diagonal in k-space
SR, average-pool	$d^{-2}\boldsymbol{I}$	free	disjoint rows of norm $1/d$
Deblurring, Gaussian	diag. in Fourier	2 FFTs	$\boldsymbol{A}\boldsymbol{A}^\top = \lvert\hat{h}\rvert^2$
Deblurring, uniform	diag. in Fourier	2 FFTs	$\hat{h}$ has exact zeros
SR, bicubic	block-Fourier	$O(n \log n)$	polyphase folding identity
MRI, multi-coil	not diagonal	CG, 10–20 it. typical	standard SENSE

Algorithm 1 SNaP training step.

1: $\boldsymbol{x}_1 \sim p, \quad \boldsymbol{y} \leftarrow \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{n}$ ▷ training pair, Eq. (1)
2: $t, r \leftarrow \textsc{SampleTimes}()$ ▷ t logit-normal; $r = t$ with probability p_{ratio}
3: $\boldsymbol{x}_0 \leftarrow \textsc{SampleSource}(\boldsymbol{y}, \sigma_{\text{n}})$ ▷ perturb-and-solve, Eq. (13)
4: $\boldsymbol{z}_r \leftarrow (1-r)\boldsymbol{x}_0 + r\,\boldsymbol{x}_1; \quad \boldsymbol{v} \leftarrow \boldsymbol{x}_1 - \boldsymbol{x}_0$
5: $\boldsymbol{u},\ \mathrm{d}_r\boldsymbol{u} \leftarrow \textsc{Jvp}\big(\boldsymbol{u}^{\boldsymbol{\theta}};\ (\boldsymbol{z}_r, r, t),\ (\boldsymbol{v}, 1, 0)\big)$ ▷ one forward-mode pass
6: $\boldsymbol{u}_{\text{tgt}} \leftarrow \boldsymbol{v} + (t-r)\,\mathrm{sg}[\mathrm{d}_r\boldsymbol{u}]$ ▷ MeanFlow target, Eq. (14)
7: $\ell \leftarrow \|\boldsymbol{u} - \boldsymbol{u}_{\text{tgt}}\|_2^2 / n$
8: $\mathcal{L} \leftarrow \mathrm{sg}[\omega(\ell)]\ \ell$ ▷ weighted objective, Eq. (15)
9: $\boldsymbol{\theta} \leftarrow \textsc{AdamW}\big(\boldsymbol{\theta}, \nabla_{\boldsymbol{\theta}}\mathcal{L}\big)$ ▷ optimizer step

C.3 Training and implementation details

Architecture. The backbone is a DDPM-style residual U-Net (`GroupNorm`$_{32}$, Swish, variance-scaled initialization, zero-initialized output and residual-branch convolutions) with three total scalar conditioning inputs. The time t is embedded by a sinusoidal positional encoding followed by a two-layer MLP into a $4\,\mathrm{ch}$-dimensional vector; the second MeanFlow time r and the noise level σ_{n} each get their own identical embedding, and the three are *summed* before entering the body, where each residual block adds a per-channel affine projection of the sum to its first activation. The σ_{n} embedding takes $10\,\sigma_{\text{n}}$ as input so that the small noise levels used here occupy a useful part of the sinusoidal range. Conditioning on $\boldsymbol{y}$ is by *concatenation* rather than FiLM, with input $[\,\boldsymbol{z}_t,\ \mathrm{anchor}]$ of $2C$ channels, where the anchor is the adjoint $\boldsymbol{A}^\top\boldsymbol{y}$.

Configurations: CelebA 128×128 uses $\mathrm{ch} = 64$, $\mathrm{ch_mult} = (1,2,2,2)$, two residual blocks per level and attention at 16^2 (9.1M parameters); AFHQ-Cat 256×256 uses $\mathrm{ch} = 64$, $\mathrm{ch_mult} = (1,2,4,4,8,8)$ down to an 8×8 bottleneck, two residual blocks, attention at 32^2 and 16^2 and dropout 0.1 (110.5M); fastMRI brain 320×320 uses $\mathrm{ch} = 32$, $\mathrm{ch_mult} = (1,2,4,8,8)$, four residual blocks and attention at 40^2 (40.7M). Attention is a hand-written block of 1×1 convolutional Q, K, V projections, a batched matrix product scaled by $C^{-1/2}$, and a `float32` softmax, since `torch.func.jvp` has no forward-mode rule for the fused scaled-dot-product kernels. The same restriction rules out `torch.utils.checkpoint`, so activation memory is a hard ceiling and width is added only at the two deepest levels.

Training. Given a clean image $\boldsymbol{x}_1$ we form $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{n}$, draw the source $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{m}_{\boldsymbol{y}}, \tau^2\boldsymbol{W})$ by Eq. (13), and set $\boldsymbol{z}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$ and $\boldsymbol{v} = \boldsymbol{x}_1 - \boldsymbol{x}_0$. Times are drawn as $t = \sigma(\mu + s\,\epsilon)$ with $\epsilon \sim \mathcal{N}(0,1)$, $\mu = 0$ and $s = 1$ (logit-normal on $(0,1)$); with probability $p_{\text{ratio}} = 0.5$ we set $r = t$, and otherwise $r \sim \mathcal{U}(0, t)$. A further 10% of the $r \neq t$ samples, or 5% of the batch, is redrawn at the corner $r \sim \mathcal{U}(0, 0.1)$, $t \sim \mathcal{U}(0.9, 1)$, because the base schedule reaches the slice at which one-step sampling is scored with probability only 7.3×10^{-4}.

The network is an $\boldsymbol{x}_1$-predictor, so $\boldsymbol{u}(\boldsymbol{z}_r, r, 1) = \big(f_{\boldsymbol{\theta}}(\boldsymbol{z}_r, r, 1) - \boldsymbol{z}_r\big)/(1-r)$. The total derivative is one forward-mode pass, with the differentiated arguments $(\boldsymbol{z}, r, t)$ carrying tangent $\boldsymbol{v}$ on $\boldsymbol{z}$, 1 on r and 0 on t, while the anchor, the coverage map and σ_{n} are closed over and carry zero tangent. The regression target is $\boldsymbol{u}_{\text{tgt}} = \boldsymbol{v} + (t-r)\,\mathrm{sg}[\mathrm{d}\boldsymbol{u}/\mathrm{d}r]$ and the loss is Eq. (15) with $\boldsymbol{\Delta} = \boldsymbol{u}^{\boldsymbol{\theta}} - \boldsymbol{u}_{\text{tgt}}$,

$p = 1$ and $c = 10^{-3}$. The stop-gradient is on the weight only, and a second stop-gradient is applied to the derivative term. On CelebA and AFHQ, ℓ is first divided by its all-reduced running mean (decay 0.99) so that c is dimensionless and the loss shape is stable across τ and across training; the MRI runs use the unnormalized form. No endpoint $\|\hat{\boldsymbol{x}}_1 - \boldsymbol{x}_1\|_2^2$ term is used, since its population minimizer is the posterior mean and would invalidate every calibration metric.

Optimization is AdamW with weight decay 10^{-4}, cosine annealing to 10^{-6} over the full epoch budget, no warm-up and gradient-norm clipping at 1.0. The learning rate is 2×10^{-4} on CelebA and AFHQ and 1×10^{-4} on MRI. Everything is trained in `float32`, since forward-mode AD is not autocast-safe. The effective batch is 32 throughout (CelebA 8×4 GPUs; AFHQ 2×4 accumulation steps $\times 4$ GPUs; MRI 4×4 GPUs). Budgets are 200 epochs ($50{,}000$ steps) on CelebA, 120 epochs ($15{,}000$ steps) for AFHQ super-resolution and 80 epochs ($10{,}000$ steps) for the remaining AFHQ operators, and 80 epochs ($43{,}040$ steps) for brain MRI, where an epoch is a fixed-size draw with replacement rather than a pass over the data. EMA is implemented but disabled in all reported runs, having never beaten the raw weights.

Choice of τ. We set τ^2 to the mean posterior variance left along the directions the measurement leaves open, estimated offline on 256 training images as $\tau^2 = \langle \boldsymbol{d}, \boldsymbol{W}\boldsymbol{d} \rangle / \operatorname{tr} \boldsymbol{W}$, where $\boldsymbol{d} = \boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{y})$ is the residual of the linear-MMSE (Wiener) estimate under the empirical image power spectrum. Since $\boldsymbol{W}$ depends on τ, we solve this by fixed-point iteration. The linear estimate is suboptimal, so its residual upper-bounds the posterior variance and τ errs toward over-dispersion. This gives $\tau = 0.1$ for $\times 2$ super-resolution and random inpainting on CelebA, 0.15 for deblurring and 0.3 for box inpainting; $\tau = 0.1$ for $\times 4$ super-resolution, random inpainting and deblurring on AFHQ and 0.5 for box inpainting; and $\tau = 0.25$ for brain MRI. The ablation in App. B shows that reconstruction quality changes little across this range. Dataset splits and preprocessing are given in App. C.1.

Source solves. The draw Eq. (13) needs only $\boldsymbol{A}$, $\boldsymbol{A}^\top$, and one solve with $(\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I})^{-1}$, where $\lambda = \sigma_{\mathrm{n}}^2 / \tau^2$ is floored at 10^{-8}, and for the operators used here that solve is a diagonal inverse, except for multi-coil MRI.

Circular Gaussian deblurring is Fourier-diagonal, with $\boldsymbol{A}\boldsymbol{A}^\top = |H(\omega)|^2$ and one division per mode. Box-decimation super-resolution has $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}$ on the low-resolution grid, so the solve is the scalar division by $1 + \lambda$. Inpainting has $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{M}$ for a $0/1$ mask, giving a pixel-diagonal division by $\boldsymbol{M} + \lambda$. Super-resolution and inpainting are $\mathcal{O}(n)$ and Gaussian deblurring is $\mathcal{O}(n \log n)$ from the FFT pair; none add any measurable time to a training step.

Multi-coil MRI is the only iterative case, and we solve it in image space using $\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I})^{-1} = (\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}^\top$, running conjugate gradient on $\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}$, whose spectrum lies in $[\lambda, \approx 1]$ and whose right-hand side already lies in $\operatorname{range}(\boldsymbol{A}^\top)$, with a cap of 200 iterations and relative-residual tolerance 10^{-6}, which is typically reached in 10–20 iterations and per-sample reductions so that batched slices do not share a step size. The measurement-space route is implemented but not used, since $\boldsymbol{A}\boldsymbol{A}^\top$ acts on k-space while factoring through the image and is therefore rank-deficient by the coil count on top of the mask; a short CG there leaves the anchor over-scaled at 6.5 dB, worse than the plain adjoint at 24.4 dB, against 26.0 dB for the image-space form at equal cost. Requesting the anchor $\boldsymbol{m}_{\boldsymbol{y}}$ inside the sampler splits the single solve into two by linearity, which is free for the diagonal operators and doubles the CG only for MRI.

Cost. All models were trained on four A6000 GPUs with manual data-parallel gradient all-reduce. Measured end-to-end wall-clock: AFHQ-Cat super-resolution 6 h 07 m for 120 epochs; AFHQ random inpainting, box inpainting and deblurring 4 h 04 m, 4 h 00 m and 4 h 02 m for 80 epochs each; CelebA box inpainting 3 h 29 m for 200 epochs; and brain MRI at $\times 8$ acceleration 19 h 05 m for 80 epochs. Throughput on the 256-pixel AFHQ configuration is 6.6 images/s/GPU at 26.1 GB peak memory with batch size 2; the 110.5M model costs 18% more memory and 21% less throughput than a 28.0M baseline, because memory is activation- rather than parameter-dominated and the added width sits at the 16^2 and 8^2 levels. At inference an M-sample posterior costs M network evaluations and M source draws. Each draw is $O(n)$ for super-resolution and inpainting, $O(n \log n)$ for Gaussian deblurring, and a CG solve for multi-coil MRI.

Baseline budgets. The NFE column of each table counts network evaluations per reported output. SNaP and the MSE regressor use one evaluation per draw, so an M-draw row costs M evaluations.

On natural images PnP-GS uses 23, DDRM 20, DiffPIR 100, OT-ODE 180 and DPS 1000; on fastMRI, DiffPIR uses 100 and CSGM, DPS, DAPS and PnP-DM use 1000. The subscript on PnP-Flow1, PnP-Flow5, Flower1-OT and Flower5-OT is N_{Avg}, the number of network evaluations averaged at each time step, so a run of N steps costs $N \cdot N_{\mathrm{Avg}}$ evaluations; we use the values tabulated by Pourya et al. (2026). On CelebA both methods use $N = 100$ for every operator, so Tab. 1 reports the five-evaluation variants at 500 evaluations, and the single-evaluation variant Flower1-OT of Tab. 2 at 100. On AFHQ-Cat the step counts depend on the operator: PnP-Flow uses $N = 500$ for deblurring and super-resolution, 200 for random inpainting and 100 for box inpainting, while Flower uses $N = 100$, 500, 200 and 100 on those same four operators. The budgets behind the NFE range quoted in Tab. 6 are therefore 2500, 2500, 1000 and 500 evaluations for PnP-Flow5, and 500, 2500, 1000 and 500 for Flower5-OT. Every SNaP entry in Tab. 2 is a single draw (M=1); averaged draws are reported only in the distortion tables and in the ablations of App. B.

C.4 Metric details

For natural images, PSNR is computed on $[0, 1]$ with data range 1, SSIM on the clamped $[0, 1]$ image, and LPIPS with an AlexNet backbone on the clamped $[-1, 1]$ image. Images are normalized to $[-1, 1]$ for training and mapped back to $[0, 1]$ for scoring. For MRI, PSNR and SSIM are computed on magnitude images, per slice, with the data range set to the maximum intensity of the ground-truth slice. We do not report LPIPS on MRI: its backbone is trained on RGB natural images and its perceptual judgments do not transfer to grayscale magnitude reconstructions.

Calibration ratio. We check whether the spread of a sampler's draws has the right size for a sampler of $p(\boldsymbol{x} \mid \boldsymbol{y})$. Let $p_{\mathrm{s}}(\boldsymbol{x} \mid \boldsymbol{y})$ be the distribution of the draws and $\boldsymbol{\mu}_{\mathrm{s}}(\boldsymbol{y})$ its mean, and let $(\boldsymbol{x}_1, \boldsymbol{y})$ be a ground-truth image and its measurement. Define

$$B := \mathbb{E}_{\boldsymbol{x}_1,\boldsymbol{y}}\big[\, \|\boldsymbol{\mu}_{\mathrm{s}}(\boldsymbol{y}) - \boldsymbol{x}_1\|_2^2 \,\big], \qquad V := \mathbb{E}_{\boldsymbol{y}}\big[\, \mathrm{tr}\, \mathrm{Cov}_{\mathrm{s}}(\boldsymbol{x} \mid \boldsymbol{y}) \big].$$

If $p_{\mathrm{s}} = p$, then $\boldsymbol{\mu}_{\mathrm{s}}(\boldsymbol{y}) = \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{y}]$ and

$$B = \mathbb{E}_{\boldsymbol{y}}\Big[\mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{y}}\big[\, \|\mathbb{E}[\boldsymbol{x} \mid \boldsymbol{y}] - \boldsymbol{x}_1\|_2^2 \,\big]\Big] = \mathbb{E}_{\boldsymbol{y}}\big[\, \mathrm{tr}\, \mathrm{Cov}(\boldsymbol{x} \mid \boldsymbol{y}) \big] = V.$$

Hence $V/B = 1$ is a necessary condition for $p_{\mathrm{s}} = p$, and we report the calibration ratio

$$\Delta_{\mathrm{cal}} := \frac{V}{B}. \tag{28}$$

Values below 1 indicate draws with too little spread, collapsed toward a point estimate, and values above 1 indicate over-dispersed draws.

Estimation. For each test pair $(\boldsymbol{x}_1^{(i)}, \boldsymbol{y}^{(i)})$, $i = 1, \ldots, N$, we draw $\boldsymbol{s}_{(1)}(\boldsymbol{y}^{(i)}), \ldots, \boldsymbol{s}_{(M)}(\boldsymbol{y}^{(i)})$ with mean $\bar{\boldsymbol{s}}(\boldsymbol{y}^{(i)})$ and use

$$\hat{V} := \frac{1}{N(M-1)} \sum_{i=1}^{N} \sum_{j=1}^{M} \Big\| \boldsymbol{s}_{(j)}(\boldsymbol{y}^{(i)}) - \bar{\boldsymbol{s}}(\boldsymbol{y}^{(i)}) \Big\|_2^2,$$

$$\hat{B} := \frac{1}{N} \sum_{i=1}^{N} \Big\| \bar{\boldsymbol{s}}(\boldsymbol{y}^{(i)}) - \boldsymbol{x}_1^{(i)} \Big\|_2^2 - \frac{\hat{V}}{M},$$

which are unbiased for V and B; the correction $\hat{V}/M$ removes the variance that remains in the mean of M draws. We use $M = 100$. Δ_{cal} is related to the spread-skill ratio used in ensemble forecasting, which compares the standard deviation of the ensemble with the root-mean-squared error of its mean and therefore corresponds to $\sqrt{V/B}$.

FID. FID is computed with InceptionV3 pool3 features on 2048 single-draw reconstructions, quantised to uint8, against reference statistics from 2048 ground-truth images.