Elucidating Post-Training Conditioning in Pretrained Generative Flows

Anonymous Authors

Abstract

Conditional adaptation of pretrained generators has developed along seemingly distinct lines: some methods preserve the Gaussian source and add conditioning adapters or modules, bridges instead learn observation-centered transports from scratch, and recent work constructs such bridges from pretrained generators. We show that these approaches are parameterizations of a common post-training design space, determined by source information, explicit conditioning, and whether posterior randomness enters at the source or inside the transport. We then isolate the consequences of each choice. First, informative sources reduce the correction demanded of finite adapters, while vanishing source noise raises the expansion burden, producing a capacity–training crossover verified on GMMs and pretrained FLUX and SiT models. Second, source-noise ODEs and bridge-noise SDEs reach comparable conditional Energy, but require 4 and 64 evaluations, respectively. Third, with an informative source, removing the explicit condition arm largely preserves point fidelity despite a remaining posterior cost. These findings favor two practical recipes: an informative source-only ODE as the simplest efficient default, and a velocity-derived SDE that turns one source-only point-source checkpoint into both a deterministic restorer and an inference-time fidelity–diversity controller.

1 Introduction

Scaling has turned diffusion and flow models into powerful but expensive image priors [2, 6, 12, 15, 18]. Reusing them is therefore increasingly attractive: for a new conditional task, we would rather adapt an existing generator than relearn the image distribution from scratch.

Image inverse problems expose the central difficulty. Given a degraded observation Y, the target is the posterior $p(X \mid Y)$. Conventional adaptation preserves the pretrained Gaussian source and injects Y through guidance, LoRA, ControlNet, or another conditioning pathway [5, 8, 11, 22, 25]. This reuses the prior but retains the complete Gaussian-to-clean transport.

Observation-centered bridges demonstrate that a different transport can be much more efficient. By starting at, or near, the degraded image, they exploit structure already present in the observation and reach clean data in few steps [9, 13, 23, 27]. Canonical bridges train this transport specifically for the restoration task. Recent methods combine observation-centered paths with pretrained diffusion or flow priors [7, 16, 21], confirming that prior reuse and bridge geometry are compatible. Each method, however, commits to a particular source, path, conditioning interface, and sampler.

We expose these choices as independent coordinates of one post-training design space. Starting from a pretrained Gaussian-to-data flow, we continuously move its source toward the observation and adapt the resulting transport with lightweight modules. Gaussian-source conditioning, stochastic observation-centered adaptation, and deterministic residual restoration become points on one source-information axis.

This axis controls finite-resource risk. Gaussian-source adaptation begins with broad support and must contract toward the posterior. Observation-centered adaptation begins near a point estimate and must expand toward it. The former is coverage-first; the latter is fidelity-first. Provided sufficient randomness remains, both can represent the same conditional law in the expressive limit, but differ under limited adapter rank, training time, data, and NFE. We derive the expansion required near a concentrated source and empirically identify the resulting capacity–time crossover.

We then separate source information from randomness placement. Randomness may enter through the source and be transported by an ODE, or enter through a stochastic bridge and be sampled by an SDE. Source

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.

randomness is more NFE-efficient in our experiments, while bridge randomness makes diversity controllable at inference.

Finally, we separate statistical generality from architectural practicality. Explicit conditioning preserves access to Y throughout transport and is the general route to $p(X \mid Y)$, but requires a new conditioning interface. We therefore study an architecture-free alternative, which we call *source-only adaptation*: the observation enters through the initial state but is not provided separately to the learned field. This simplicity comes at the cost of losing a universal posterior guarantee. We derive its oracle point-estimation cost and show that it vanishes near the observation endpoint. Empirically, high-information sources retain nearly all point fidelity without the condition arm, although posterior quality can still degrade.

This result motivates a *velocity-derived SDE* that adds inference-time diversity to the same source-only adapter. Its score is obtained analytically from the learned velocity rather than from a second network. One model therefore supports deterministic few-step restoration and a continuous family of stochastic samplers without introducing a new image-conditioning architecture. Our framework unifies purpose-built restoration bridges with modular adaptation of pretrained generators while exposing the statistical, architectural, and computational trade-offs.

Our contributions are the following:

- We unify Gaussian-source adaptation and observation-started restoration within one post-training framework. The framework separates source information, source randomness, bridge randomness, explicit conditioning, and inference cost.
- We characterize the resulting finite-resource trade-off. Simple bounds quantify the expansion required near a concentrated source and the oracle point-risk cost of removing the conditioning arm. GMM experiments and natural-image inverse problems with pretrained flows support the predicted fidelity–distribution frontier and its rank–training–NFE crossover.
- We compare source and bridge randomness under matched adaptation. Both provide valid posterior mechanisms in the exact-field limit, but source randomness supports substantially more efficient ODE sampling, whereas bridge randomness provides direct stochastic control at inference.
- We introduce a velocity-derived SDE sampler that requires no new conditioning architecture and moves its principal stochastic choices to inference. We derive an exact drift-error identity and validate its fidelity–distribution control with a pretrained FLUX model on image inverse problems.

2 Related work

Restoration transports. Flow matching learns a velocity field by simulation-free regression [12, 19], while stochastic interpolants extend it to flexible paths and diffusions [1, 2, 15]. Observation-centered restoration paths underlie I^2SB, ResShift, FlowIE, and RFMSR [9, 13, 23, 27]. ECSI systematizes bridge design choices [26], and DiSI combines separately trained regression and generation fields at inference [14]. Unlike these purpose-built transports, we study the post-training adaptation of an existing Gaussian-source flow.

Adapting generative priors. LoRA and ControlNet provide adaptation interfaces for pretrained models [8, 25]; StableSR, DiffBIR, and FluxIR use task encoders, control branches, or adapters for restoration [5, 11, 22]. Training-free solvers instead retain the unconditional prior and impose the known operator at inference [24].

Restoration paths from pretrained priors. IRBridge, E-Bridge, and bridge-aligned prompting combine pretrained priors with task-specific restoration paths [7, 16, 21]. We instead factor this design into source information, randomness placement, explicit conditioning, adapter capacity, and inference cost. Our fidelity analysis is related to the perception–distortion trade-off [4], but concerns conditional sampling during post-training.

Table 1: **Post-training design space.** Rows vary explicit conditioning; columns locate posterior randomness.

	Source randomness ($\alpha < 1$)	**Bridge randomness** ($\gamma_t > 0$)
Condition arm	Conditional ODE	Conditional SDE
Source only	Source-only ODE	**Velocity-derived SDE**

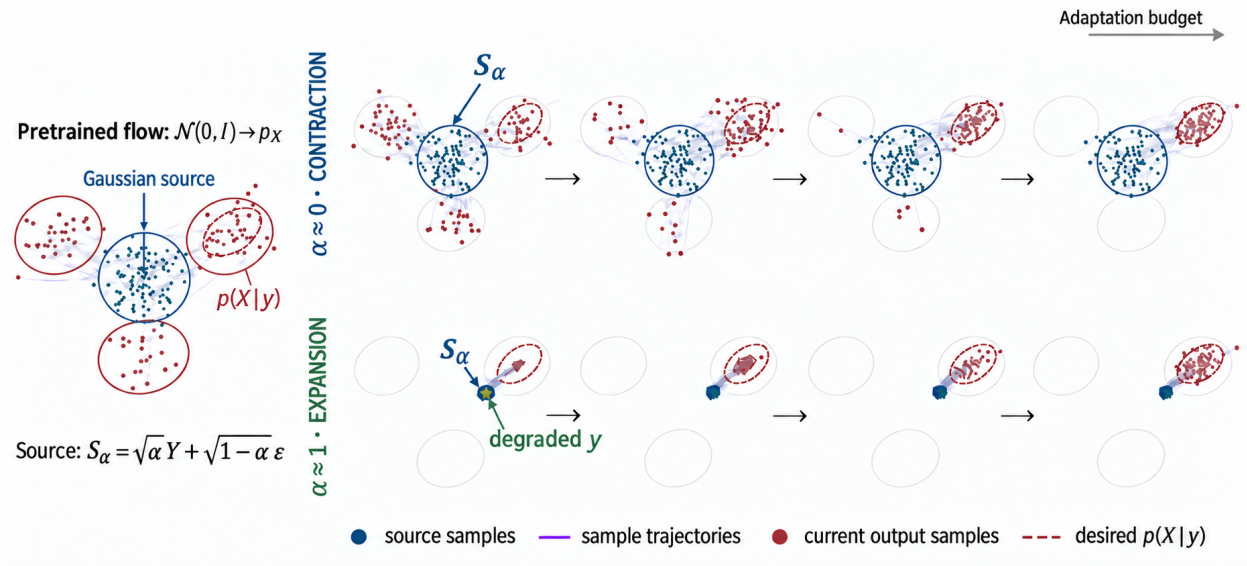

Figure 1: **Two adaptation routes.** Gaussian-source adaptation contracts broad support, whereas observation-centered adaptation expands from an anchored source. Adaptation budget denotes training steps and adapter capacity.

3 A unified framework for post-training conditioning

Let $(X, Y) \sim p_{X,Y}$ be a clean target and its degraded observation in a common normalized state space. Draw independent $\epsilon, \xi \sim \mathcal{N}(0, I_d)$ and define

$$S_\alpha = \sqrt{\alpha}\,Y + \sqrt{1-\alpha}\,\epsilon, \qquad Z_t = a_t S_\alpha + b_t X + \gamma_t \xi, \tag{1}$$

where $a_0 = 1$, $b_0 = \gamma_0 = 0$, and $a_1 = \gamma_1 = 0$, $b_1 = 1$. The parameter α controls *source information*: $\alpha = 0$ recovers the pretrained Gaussian source, $0 < \alpha < 1$ retains both observation information and source randomness, and $\alpha = 1$ starts from the observation. The interior scale γ_t controls *bridge randomness*. With an explicit condition arm, flow matching learns

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}\|v_\theta(Z_t, t, Y) - \dot{Z}_t\|^2, \qquad v^*(z, t, y) = \mathbb{E}[\dot{Z}_t \mid Z_t = z, Y = y]. \tag{2}$$

Without that arm, we instead learn the *source-only velocity* $\bar{v}(z, t) = \mathbb{E}[\dot{Z}_t \mid Z_t = z]$. The learned marginals can be sampled by the probability-flow ODE $dZ_t = v(Z_t, t)dt$. In the explicitly conditioned case, the Gaussian path in Eq. (1) allows the conditional score $s_t(z \mid y) = \nabla_z \log p_t(z \mid y)$ to be recovered analytically from the conditional velocity. The same conditional marginals are therefore preserved by the SDE

$$dZ_t = \{v(Z_t, t, y) + w_t s_t(Z_t \mid y)\}\,dt + \sqrt{2w_t}\,dW_t, \tag{3}$$

for any diffusion gauge $w_t > 0$ [2, 15]. We write $w_t = \kappa \bar{w}_t$ where κ controls inference-time stochasticity.

The endpoint $\alpha = 1$ is qualitatively different. Conditional on $Y = y$, the source is the single point $S_1 = y$, which a deterministic ODE can map only to another single point. Its generated conditional law is therefore a Dirac measure and cannot represent a non-degenerate posterior. Every $\alpha < 1$, by contrast, retains a full-dimensional Gaussian component and therefore has sufficient support for deterministic posterior transport

in principle. This is a support fact, not a guarantee that a finite adapter learns the required map. At $\alpha = 1$, posterior sampling requires another source of randomness, such as the SDE in Eq. (3).

This limitation concerns posterior sampling, not point estimation. The Bayes-optimal deterministic endpoint under squared error is the conditional mean $\mu_y = \mathbb{E}[X \mid y]$. For any restoration $\widehat{X}$ independent of the held-out target given y, the standard conditional bias–variance decomposition gives, with $V_y = \operatorname{tr}\operatorname{Cov}(X \mid y)$,

$$\mathbb{E}[\|\widehat{X} - X\|^2 \mid y] = V_y + \|\mathbb{E}[\widehat{X} \mid y] - \mu_y\|^2 + \operatorname{tr}\operatorname{Cov}(\widehat{X} \mid y). \tag{4}$$

The conditional mean has MSE V_y, whereas an independent exact posterior draw has MSE $2V_y$. This familiar identity is the squared-error instance of the perception–distortion trade-off, with posterior correctness playing the role of perfect perception [4]. Accordingly, our experiments report both single-draw fidelity and conditional Energy.

4 Finite-resource analysis

4.1 Source information versus expansion

We first isolate source randomness by setting $\gamma_t = 0$ and sampling the adapted field with its deterministic ODE. Conditional on y, all output randomness then comes from S_α.

Figure 1 summarizes the resulting two-sided adaptation geometry. The Gaussian route starts with broad pretrained support and must contract it toward $p(X \mid y)$; the informative route quickly provides a conservative point restoration but must expand its residual noise to recover posterior spread. With sufficient adaptation, both stochastic routes can reach the same conditional law from opposite directions.

Increasing α moves the source toward the observation. When Y is informative, this shortens the conditional transport and reduces the correction demanded of a finite adapter (Appendix A.5). It also reduces the conditional variance of the flow-matching velocity target and hence its contribution to stochastic-gradient variance (Appendix A.6). However, reducing source noise also has a cost: the flow must expand a more concentrated source to recover posterior variance. Let T_y be the resulting endpoint map, let K_y be its Lipschitz constant, and recall that $V_y = \operatorname{tr}\operatorname{Cov}(X \mid y)$. Applying the Gaussian Poincaré inequality coordinate-wise gives the following necessary bound [10].

Proposition 1 (Necessary expansion bound)**.** *(Proof in Appendix A.2.) For S_α defined in Eq.* (1)*,*

$$\operatorname{tr}\operatorname{Cov}(T_y(S_\alpha) \mid y) \le (1-\alpha) d K_y^2. \tag{5}$$

Matching posterior variance V_y therefore requires $K_y \ge \sqrt{V_y/[d(1-\alpha)]}$.

Thus the sensitivity required for posterior sampling diverges as $\alpha \to 1$. This is a necessary condition on the endpoint map, not a guarantee about optimization or posterior calibration. It nevertheless exposes the opposing effects of source information: larger α reduces the initial mismatch but increases the sensitivity needed to recover spread. As an illustration, for standardized jointly Gaussian (X, Y) with correlation $\rho \in [0, 1]$, the source and posterior conditional laws match at $\alpha = \rho^2$ (Appendix A.3). This calibrated example is not a universal prescription for choosing α. In a similarly controlled setup (Appendix A.4), a low-rank adapter can correct only a few directions in which the pretrained flow is wrong. An informative source makes many predictable directions approximately correct from the outset, preserving the adapter's limited rank. As rank increases, the adapter can repair more mismatched directions and source anchoring becomes less advantageous.

Together, these results motivate a capacity–time crossover: **informative sources should dominate with low-rank or undertrained adapters, while Gaussian-source adaptation should catch up as capacity and training increase.** At the other extreme, α too close to one remains difficult because the adapter and solver must realize the expansion in Eq. (5). Section 5.1 tests these predictions in controlled GMMs and pretrained image flows (Table 2 and Figures 2–4).

4.2 Value of the condition arm

An informative source provides a particularly simple conditioning mechanism: the observation enters through the model's existing state input. This is especially attractive when a pretrained generator has no native observation input, since it requires neither a dedicated conditioning branch nor an architectural change. We therefore ask how much point fidelity is lost when the model receives only S_α. The following result shows that the oracle excess MSE is exactly the uncertainty about the conditional mean that remains after observing the noisy source.

Proposition 2 (Oracle point-risk cost of removing the condition arm)**.** *Let $Y \in \mathbb{R}^d$ and $X \in \mathbb{R}^p$, and let $S_\alpha = \sqrt{\alpha}Y + \sqrt{1-\alpha}\,\epsilon$ with $\epsilon \sim \mathcal{N}(0, I_d)$ independent of (X, Y), and suppose that $m(y) = \mathbb{E}[X \mid Y = y]$ is L-Lipschitz. For $\alpha > 0$, the excess MSE induced by removing the conditioning arm is*

$$\begin{aligned} \Delta_\alpha &:= \mathbb{E}\|X - \mathbb{E}[X \mid S_\alpha]\|^2 - \mathbb{E}\|X - m(Y)\|^2 \\ &= \mathbb{E}\|m(Y) - \mathbb{E}[m(Y) \mid S_\alpha]\|^2 \\ &\le L^2 d \frac{1-\alpha}{\alpha}. \end{aligned} \tag{6}$$

The equality gives a direct interpretation: Δ_α is the price of predicting from the noisy source rather than from the original observation. It vanishes as $\alpha \to 1$, when $S_1 = Y$. A high-information source should therefore make the condition branch largely redundant for point fidelity, but this is deliberately an oracle point-estimation statement. It neither guarantees that a finite adapter attains the Bayes rule nor that a source-only transport reproduces $p(X \mid Y)$. Section 5.3 exposes this distinction empirically: point fidelity can be preserved even when posterior quality worsens (Table 4).

4.3 Source-only posterior sampling

Equation (1) already covers both locations of posterior randomness. Setting $\gamma_t = 0$ leaves randomness in the source, whereas $\gamma_t > 0$ injects it inside the transport; choosing $\alpha < 1$ and $\gamma_t > 0$ combines both. In particular, bridge randomness allows the point source $\alpha = 1$ to recover a non-degenerate posterior in principle, but requires SDE sampling; its probability-flow ODE remains a point predictor. For the derivation below, we use $a_t = 1 - t$, $b_t = t$, and $\gamma_t = c\sqrt{2t(1-t)}$ in Eq. (1), with $c > 0$. The SDE requires the conditional score $s_y^*(z, t) = \nabla_z \log p_t(z \mid y)$. With an explicit condition arm, this score is recovered exactly from the conditional velocity:

$$s_y^*(z, t) = \frac{t v^*(z, t, y) - z + y}{c^2 t}. \tag{7}$$

The conditioned bridge therefore defines a valid posterior sampler [2, 15].

Without the condition arm, the network learns only the source-only velocity $\bar{v}(z, t) = \mathbb{E}[\dot{Z}_t \mid Z_t = z]$. We therefore approximate the unavailable conditional velocity by $\bar{v}$. Unlike a path starting from independent Gaussian noise, this bridge starts from the observation itself; its state may therefore remain informative about the missing condition. Substituting $\bar{v}$ for v^* in Eq. (7) gives

$$\widetilde{s}_y(z, t) = \frac{t\bar{v}(z, t) - z + y}{c^2 t}, \qquad 0 < t < 1. \tag{8}$$

Here y is used only by the analytic sampler, as the known bridge anchor; it is never passed through the learned network. The resulting $\widetilde{s}_y$ supplies the score needed for SDE sampling. The following algebraic identity makes the approximation cost explicit.

Proposition 3 (Velocity-derived drift-error identity)**.** *(Proof in Appendix D.) Let $e_y(z, t) = \bar{v}(z, t) - v^*(z, t, y)$ be the velocity error introduced by removing the conditioning arm. For $w_t = \kappa c^2$ with $\kappa \ge 0$, define the exact and velocity-derived SDE drifts by $b_y^* = v^* + \kappa c^2 s_y^*$ and $\widetilde{b}_y = \bar{v} + \kappa c^2 \widetilde{s}_y$. Then*

$$\widetilde{s}_y - s_y^* = \frac{1}{c^2} e_y, \qquad \widetilde{b}_y - b_y^* = (1 + \kappa) e_y. \tag{9}$$

Table 2: **Capacity accelerates Gaussian-source catch-up.** Earliest update count at which $\alpha = 0$ is within 10% of the informative reference; parentheses give the loss ratio. Results use five paired seeds.

Adapter rank	Energy vs. $\alpha = .9$	Single-draw MSE vs. $\alpha = .999$
2	$>$ 50k (1.60)	$>$ 50k (2.55)
4	50k (0.93)	$>$ 50k (1.25)
8	5k (0.63)	50k (0.88)

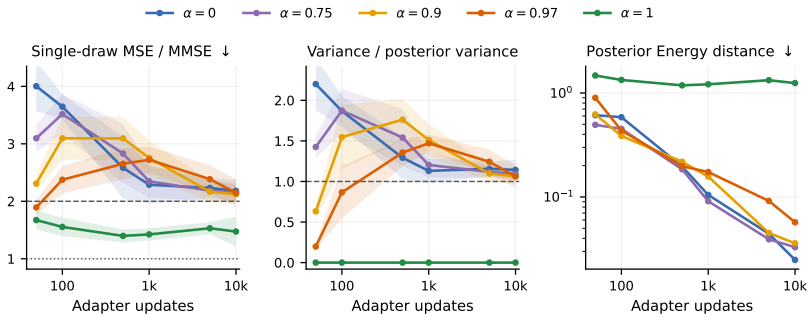

Figure 2: **Two-sided adaptation in a controlled GMM.** The Gaussian source contracts from over-dispersion, whereas informative stochastic sources expand from under-dispersion and may briefly overshoot. Horizontal lines mark conditional-mean and posterior-draw MSE.

The proposition isolates the drift-level effect of the approximation: the SDE drift error is exactly $(1 + \kappa)$ times the velocity mismatch. Increasing κ adds inference-time stochasticity but amplifies this mismatch linearly. The identity does not characterize the accumulated distributional error or make the approximate sampler exact. It nevertheless yields a practical architecture-free sampler: train one source-only velocity model with the observation as its source, then select κ and NFE at test time to trade fidelity for diversity. Section 5.2 first compares exact source and bridge randomness (Table 3); Section 5.4 then evaluates the source-only approximation (Table 5); Appendix Figure 12 audits the drift-error identity.

5 Experiments

We evaluate our findings in two settings: a controlled GMM with exact posteriors and real-image experiments using FLUX.2 on FFHQ and SiT-XL/2 on ImageNet. In every case, the pretrained backbone is frozen and only adapters are trained. Full architectural, training, and evaluation details, together with additional replications, are given in Appendix B. We report PSNR or MSE for single-draw fidelity and conditional Energy and sample spread for posterior quality; these quantities need not improve together, as Eq. (4) makes explicit.

5.1 Where should the observation enter?

GMM. We pretrain an MLP on a symmetric four-component mixture and let the condition select one component. The exact posterior makes both point risk and distributional error measurable. We then freeze the backbone and adapt rank-2/4/8 bottlenecks from either the Gaussian source or informative sources.

Table 2 tests the capacity prediction directly. At rank 2, the Gaussian source remains worse than the informative references after 50k updates. It matches the Energy reference at 50k updates with rank 4 and by 5k with rank 8; single-draw fidelity catches up later, at 50k updates with rank 8. Thus increasing adapter capacity accelerates the Gaussian source's catch-up, without implying a universal ordering over α.

Figure 2 explains why fidelity and distributional metrics must be read together. Informative stochastic sources initially act like low-variance point restorers and then expand; as they recover uncertainty, single-draw MSE can rise toward posterior-draw risk even while Energy improves (Eq. (4)). The Gaussian source instead

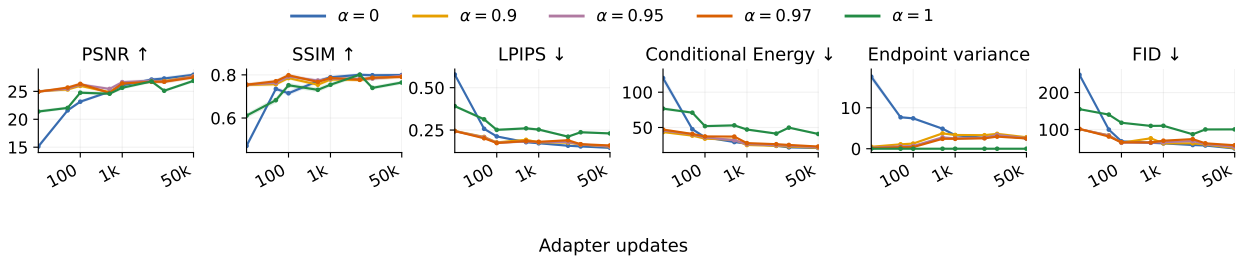

Figure 3: **Two-sided adaptation with FLUX.2.** On FFHQ SR×4, informative sources expand from under-dispersion while the Gaussian source contracts from over-dispersion. Their fidelity and conditional Energy converge with training. Endpoint variance is shown in units of 10^{-3}.

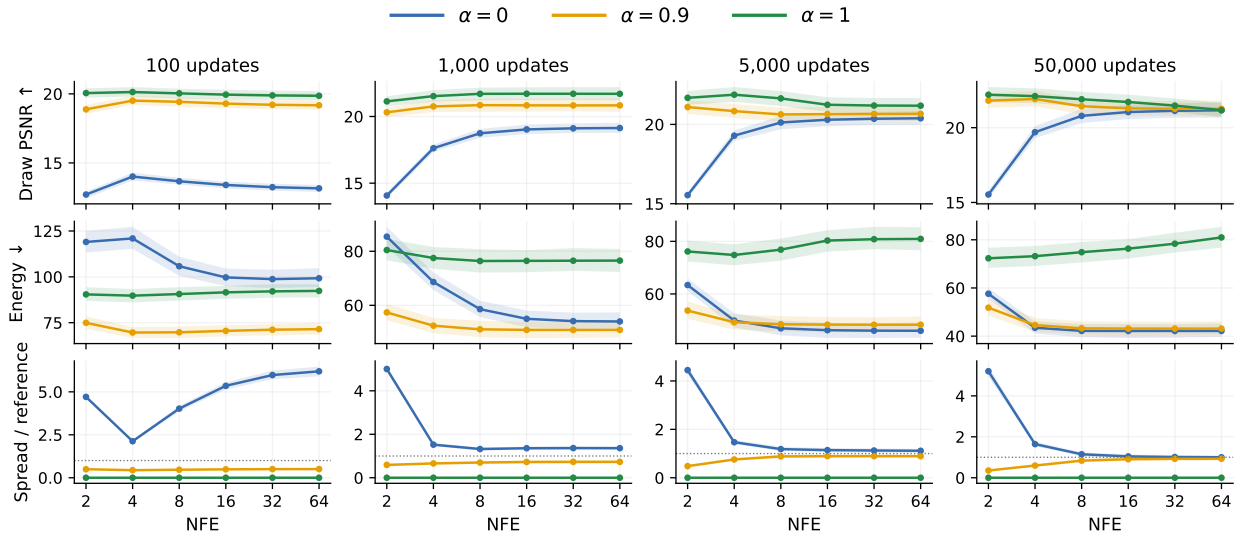

Figure 4: **Source information changes the useful inference budget.** On ImageNet SR with SiT, the Gaussian source benefits from longer integration, whereas the informative source reaches peak fidelity early and later recovers spread. The deterministic source cannot expand. Spread is normalized by the 50k-update, 64-NFE Gaussian arm; bands are 95% bootstrap intervals.

begins over-dispersed and contracts. The size of the informative route's transient overshoot is not universal: Appendix 8 repeats the same fixed-rank protocol for one-, two-, and four-mode posteriors. The two-mode case overshoots only mildly, and the four-mode case expands almost monotonically.

Pretrained image flows. Figure 3 shows that the same dynamics persist when adapting FLUX.2 to FFHQ SR×4. The mechanism is particularly visible in endpoint spread: the Gaussian source contracts while informative sources expand. At 100 updates, $\alpha = .9$ improves PSNR from 23.40 to 26.36 dB over $\alpha = 0$; by 50k updates, their Energy scores have converged to 19.81 and 19.60. The same early informative-source advantage disappears with training on Gaussian blur and inpainting (Appendix C), showing that the finite-training crossover is not specific to super-resolution.

Inference budget. Source choice also determines how much of the learned field must be integrated. At 50k updates, the PSNR-optimal NFE is 64/4/2 for $\alpha = 0/.9/1$, respectively (Figure 4). The informative $\alpha = .9$ source reaches peak fidelity in four evaluations.

Practical restoration quality. The four-NFE FLUX adapters are also competitive with published DAPS baselines on FFHQ restoration (Appendix Table 6). Figure 5 shows a representative SR comparison.

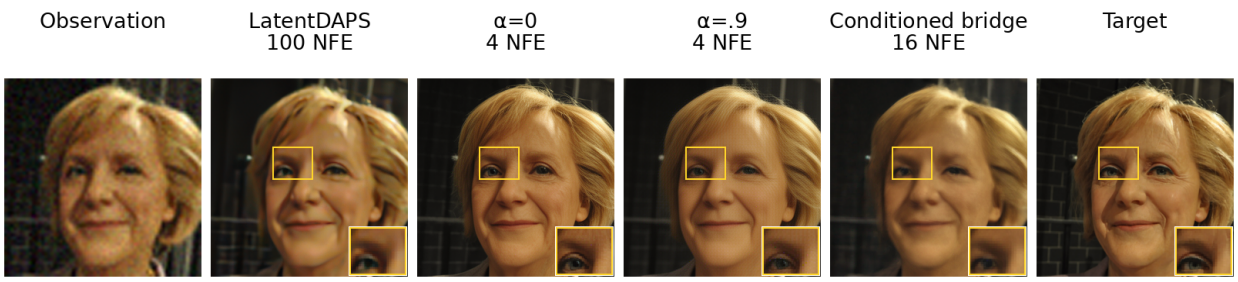

Figure 5: **FFHQ SR×4 comparison.** Outputs and NFE are labeled above each column; boxes enlarge the eye region.

Table 3: **Where randomness enters.** Matched rank-64 conditioned FLUX.2 adapters on FFHQ SR×4 after 50k updates.

Randomness	Sampler	NFE	PSNR ↑	Energy ↓	Spread $\times 10^3$
source ($\alpha = .9$)	ODE	4	28.12	20.66	2.18
source + bridge ($\alpha = .9$)	SDE	4	28.12	20.28	2.30
none ($\alpha = 1$)	ODE	4	28.44	33.98	0.00
bridge ($\alpha = 1$)	SDE	4	27.75	23.96	1.66
bridge ($\alpha = 1$)	SDE	16	28.06	21.28	2.12
bridge ($\alpha = 1$)	SDE	64	27.30	20.76	3.40

5.2 Where should posterior randomness enter?

We train matched stochastic interpolants with the condition arm present and compare source noise sampled by an ODE with bridge noise from a deterministic observation source sampled by an SDE. We also include the deterministic ODE from the point source to make the failure mode explicit: bridge noise during training does not produce posterior samples unless the corresponding SDE is used.

Both stochastic routes attain similar Energy at different computational costs: the four-NFE source-noise ODE matches the 64-NFE point-source bridge SDE. The ODE from the point source has zero spread and much worse Energy. Thus both placements recover comparable distributional quality, but source noise is substantially more NFE-efficient in this matched comparison.

5.3 How much does removing the conditioning arm cost?

We train matched adapters with the degraded-image condition tokens either present or omitted throughout training and sampling; the source-only model receives Y only through S_α. At $\alpha = 0$, it therefore receives no observation information and restoration collapses. At high α, point fidelity is nearly unchanged, consistent with the oracle prediction in Proposition 2.

At $\alpha = .9$, branch removal changes blur PSNR by only $-.04$ dB and improves SR PSNR by .78 dB. Energy remains worse, especially for blur. This is precisely the distinction not covered by the oracle bound: source injection can preserve point fidelity without preserving the full conditional law. The condition arm is therefore largely redundant for point restoration in this regime, but still improves posterior structure.

5.4 Does the velocity-derived approximation work?

We apply the velocity-derived SDE to one point-source checkpoint trained without a condition arm; only κ changes at inference. Comparison with an explicitly conditioned bridge at matched κ measures the joint cost of removing the arm and deriving its unavailable score from the source-only velocity. We select NFE for each row by minimum Energy. A conditioned $\alpha = .9$ source-noise ODE provides a posterior reference; its source-only counterpart appears in Table 4.

Table 4: **Cost of removing the condition arm.** Conditioned and source-only FLUX.2 adapters on FFHQ inverse problems; "Source" omits image tokens throughout adaptation and sampling.

Task	α	$\text{PSNR}_{\text{cond}} \uparrow$	$\text{PSNR}_{\text{source}} \uparrow$	ΔPSNR	$\text{Energy}_{\text{cond}} \downarrow$	$\text{Energy}_{\text{source}} \downarrow$
SR×4	0	28.31	11.68	−16.63	19.59	198.23
	.5	28.23	26.47	−1.77	19.26	21.09
	.9	28.03	**28.81**	+0.78	**19.81**	20.47
Gaussian blur	0	28.10	11.73	−16.38	20.11	193.64
	.5	28.08	25.72	−2.36	20.02	23.52
	.9	28.06	28.02	−0.04	**20.84**	23.16

Table 5: **Velocity-derived stochastic sampling.** FLUX.2 on FFHQ SR×4. NFE minimizes Energy for each row; Δ compares the velocity-derived SDE with the conditioned bridge at matched κ.

field / sampler	κ	NFE*	Energy $\downarrow$	ΔE	PSNR $\uparrow$	ΔP	LPIPS $\downarrow$	spread $\times 10^3$
source-only / ODE	0	4	29.04	—	**29.84**	—	0.153	0.00
source-only / derived SDE	.1	4	26.39	+0.94	29.40	+1.81	**0.134**	0.18
conditioned / SDE	.1	64	25.45	ref.	27.59	ref.	0.190	1.54
source-only / derived SDE	.3	4	25.12	+2.61	29.11	+1.57	0.137	0.44
conditioned / SDE	.3	64	22.51	ref.	27.53	ref.	0.182	2.44
source-only / derived SDE	.5	4	24.82	+3.28	28.82	+1.36	0.142	0.65
conditioned / SDE	.5	64	21.54	ref.	27.46	ref.	0.181	2.87
conditioned source / ODE	—	32	**18.46**	—	27.65	—	0.156	3.72

The velocity-derived SDE turns one deterministic source-only predictor into a test-time fidelity–diversity slider (Table 5). Raising κ from 0 to .5 lowers Energy from 29.04 to 24.82 while reducing PSNR by 1.02 dB. Exact conditioning recovers more spread and lowers Energy by 0.94–3.28, but selects 64 rather than four NFE. The approximation is therefore conservative and compute-efficient; its main cost is under-dispersion rather than point fidelity. Appendix Figure 14 provides a qualitative comparison.

The paired field audit in Appendix Figure 12 verifies the error identity. The velocity mismatch is largest near the Brownian-bridge entrance and falls sharply thereafter; the measured drift error follows $(1 + \kappa)E_v$ as predicted.

A small validation sweep over κ and NFE is therefore the practical recommendation when no image-conditioning interface is available. The appendix reports the full slider, repeated samples, and Gaussian-blur validation.

6 Conclusion

We unified Gaussian-source adaptation, observation-centered bridges, and pretrained restoration bridges as parameterizations of one post-training design space. This view isolates three practical trade-offs. Informative sources are safer under limited adaptation, whereas Gaussian sources catch up with capacity and training; source randomness achieves comparable posterior quality with fewer evaluations than bridge randomness; and removing the condition arm near an informative source largely preserves point fidelity but can still degrade posterior structure. These findings favor two simple operating points. An informative source-only ODE is our efficient default. A velocity-derived SDE instead turns one source-only point-source checkpoint into both a deterministic restorer and an inference-time fidelity–diversity controller. Explicit conditioning remains preferable when posterior accuracy justifies the additional architecture.

AI use statement

We have used generative AI tools (LLMs) in this work for coding, including training and inference scripts as well as the evaluation and metric pipelines. We used it for correcting and editing the paper (spelling, rephrasing, summarizing, paraphrasing). We have also use it for related work, for exploration of potential missing references and related research papers.

We did not used generative AI tools to help us in the research of ideas or the design of methods or in the experimental protocol. We did not use it either to generate experimental results. We did not conduct human-subject study, so disclosures related to participants are not applicable to this work. We have reviewed all AI-assisted work. Every related work suggested by an AI tool was investigated and studied. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

Not applicable.

Reproducibility statement

Code will be released upon publication.

References

[1] Michael S. Albergo, Mark Goldstein, Nicholas M. Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. In *International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 921–937. PMLR, 2024.

[2] Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

[3] Giulio Biroli, Tony Bonnaire, Valentin de Bortoli, and Marc Mézard. Dynamical regimes of diffusion models. *Nature Communications*, 15:9957, 2024. doi: 10.1038/s41467-024-54281-3.

[4] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[5] Junyuan Deng, Xinyi Wu, Yongxing Yang, Congchao Zhu, Song Wang, and Zhenyao Wu. Acquire and then adapt: Squeezing out text-to-image model for image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23195–23206, 2025.

[6] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

[7] Jinhui Hou, Zhiyu Zhu, and Junhui Hou. Energy-oriented diffusion bridge for image restoration with foundational diffusion models. *arXiv preprint arXiv:2604.10983*, 2026.

[8] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

[9] Shuwei Huang, Tianyao Luo, Jicheng Liu, and Pan Zhou. RFMSR: Residual flow matching for image super-resolution. *arXiv preprint arXiv:2607.12753*, 2026.

[10] Michel Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.

[11] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. DiffBIR: Toward blind image restoration with generative diffusion prior. In *European Conference on Computer Vision*, 2024.

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

[13] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A. Theodorou, Weili Nie, and Anima Anandkumar. I^2sb: Image-to-image schrödinger bridge. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 22042–22062. PMLR, 2023.

[14] Yi Liu, Jia Ma, Wengen Li, Jihong Guan, Shuigeng Zhou, and Yichao Zhang. Disentangling generation and regression in stochastic interpolants for controllable image restoration. *arXiv preprint arXiv:2605.21381*, 2026.

[15] Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, 2024.

[16] Sudarshan Rajagopalan and Vishal M. Patel. Your pre-trained diffusion model secretly knows restoration. *arXiv preprint arXiv:2604.04924*, 2026.

[17] Gabriel Raya and Luca Ambrogioni. Spontaneous symmetry breaking in generative diffusion models. In *Advances in Neural Information Processing Systems*, 2023.

[18] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

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

[20] Alexander Tyurin. Same flow, different paths: Variance reduction in flow matching. *arXiv preprint arXiv:2609.17287*, 2026.

[21] Hanting Wang, Tao Jin, Wang Lin, Shulei Wang, Hai Huang, Shengpeng Ji, and Zhou Zhao. IRBridge: Solving image restoration bridge with pre-trained generative diffusion models. In *International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 63266–63289. PMLR, 2025.

[22] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C. K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 2024. doi: 10.1007/s11263-024-02168-7.

[23] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. ResShift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems*, 2023.

[24] Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

[25] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *IEEE International Conference on Computer Vision*, 2023.

[26] Shaorong Zhang, Yuanbin Cheng, and Greg Ver Steeg. Exploring the design space of diffusion bridge models. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

[27] Yixuan Zhu, Wenliang Zhao, Ao Li, Yansong Tang, Jie Zhou, and Jiwen Lu. FlowIE: Efficient image enhancement via rectified flow. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

A Proofs and statistical assumptions

All expectations below concern the specified joint law $p(X, Y)$. Conditional on Y, the source noise and a held-out target draw are independent. We assume finite second moments; differentiability and integrability are imposed where a Jacobian is used. The transport coordinates are Euclidean. In latent image experiments, a conditional-mean identity in latent space does not imply that its nonlinear decoding is the RGB mean.

A.1 Risk and point-source support

Support at the source endpoints. Conditioned on $Y = y$ and for every $\alpha < 1$, $S_\alpha \sim \mathcal{N}(\sqrt{\alpha}y, (1-\alpha)I_d)$ is non-atomic and has a strictly positive density. It can therefore be transported to any regular conditional law; for smooth positive source and target densities, this map can be chosen as a regular flow map. At $\alpha = 1$, the conditional source is δ_y, and the pushforward of a Dirac mass by any deterministic endpoint map is another Dirac mass. It cannot equal a non-degenerate posterior.

Derivation of Eq. (4). Write $\widehat{\mu}_y = \mathbb{E}[\widehat{X} \mid y]$ and expand $\widehat{X} - X = (\widehat{X} - \widehat{\mu}_y) + (\widehat{\mu}_y - \mu_y) + (\mu_y - X)$. The cross terms vanish by conditional centering and independence, giving Eq. (4). A correct independent posterior draw has $\widehat{\mu}_y = \mu_y$ and variance trace V_y, hence risk $2V_y$. The mean of M conditionally independent correct draws has variance V_y/M, hence risk $(1 + 1/M)V_y$. This is an MSE statement, not a factor-of-two PSNR statement. With MMSE $= \mathbb{E}_Y V_Y$, the corresponding integrated risks are MMSE, 2MMSE, and $(1 + 1/M)$MMSE.

Proof of Proposition 2. At the source, the one-step endpoint associated with a source-only population velocity is $S_\alpha + \mathbb{E}[X - S_\alpha \mid S_\alpha] = \mathbb{E}[X \mid S_\alpha]$. At $\alpha = 1$, $S_1 = Y$, so the conditional means coincide. At $\alpha = 0$, $S_0 = \epsilon$ is independent of (X, Y), giving $\mathbb{E}[X \mid S_0] = \mathbb{E}[X]$. Finally, S_α is generated from Y using independent Gaussian noise, so $X \to Y \to S_\alpha$ is a Markov chain. The data-processing inequality yields $I(X; S_\alpha) \le I(X; Y)$. These statements concern information and source-time point prediction. They do not claim that a deterministic source-only ODE samples $p(X \mid Y)$; at $\alpha = 1$ it cannot produce a non-degenerate conditional law.

Quantitative cost of removing the conditioning arm. Assume $Y \in \mathbb{R}^d$, $\epsilon \sim \mathcal{N}(0, I_d)$ is independent, $\alpha > 0$, and $m(y) = \mathbb{E}[X \mid Y = y]$ is L-Lipschitz. The Markov chain $X \to Y \to S_\alpha$ gives $\mathbb{E}[X \mid S_\alpha] = \mathbb{E}[m(Y) \mid S_\alpha]$. Conditional-expectation orthogonality then yields the exact projection identity

$$\mathbb{E}\|X - \mathbb{E}[X \mid S_\alpha]\|^2 - \mathbb{E}\|X - m(Y)\|^2 \tag{10}$$

$$= \mathbb{E}\|m(Y) - \mathbb{E}[m(Y) \mid S_\alpha]\|^2. \tag{11}$$

The conditional mean $\mathbb{E}[m(Y) \mid S_\alpha]$ is the best square-integrable function of S_α. Comparing it with the admissible estimator $m(S_\alpha/\sqrt{\alpha})$ therefore gives

$$\mathbb{E}\|m(Y) - \mathbb{E}[m(Y) \mid S_\alpha]\|^2 \le \mathbb{E}\|m(Y) - m(S_\alpha/\sqrt{\alpha})\|^2 \tag{12}$$

$$\le L^2 \frac{1-\alpha}{\alpha} \mathbb{E}\|\epsilon\|^2 = L^2 d \frac{1-\alpha}{\alpha}. \tag{13}$$

This proves Eq. (6). The bound formalizes a plug-and-play advantage of high α: the source itself becomes an increasingly accurate carrier of the observation. It controls optimal point risk only. It neither guarantees that a finite adapter reaches the Bayes rule nor that a source-only transport has the correct conditional distribution.

A.2 Proof of the expansion price

Here $\gamma_t = 0$ and sampling follows the deterministic ODE, so conditional output randomness is inherited entirely from S_α. Condition on $Y = y$ and write $F(\epsilon) = T_y(\sqrt{\alpha}y + \sqrt{1-\alpha}\epsilon)$. The classical Gaussian Poincaré

inequality [10], applied to each output coordinate, gives

$$\operatorname{tr}\operatorname{Cov}(F(\epsilon)) \le \mathbb{E}\|J_\epsilon F(\epsilon)\|_F^2 \tag{14}$$

$$= (1-\alpha)\mathbb{E}\|J_s T_y(S_\alpha)\|_F^2 \le (1-\alpha)dK_y^2. \tag{15}$$

For any coupling of two square-integrable random vectors, the L^2 triangle inequality on their centered versions gives

$$\sqrt{\mathbb{E}\|\widehat{X}-X\|^2} \ge \left|\sqrt{\operatorname{tr}\operatorname{Cov}(X)} - \sqrt{\operatorname{tr}\operatorname{Cov}(\widehat{X})}\right|. \tag{16}$$

Minimizing over couplings and applying the variance upper bound gives an explicit Wasserstein lower bound whenever the output variance is mismatched. If the ODE field has integrated spatial Lipschitz constant $L_y = \int_0^1 \operatorname{Lip}_z(v_\theta(\cdot,t,y))dt$, Grönwall gives $K_y \le e^{L_y}$. Exact variance matching therefore requires $L_y \ge \frac{1}{2}\log[V_y/(d(1-\alpha))]$ whenever the right side is positive. This is a necessary expansion requirement, not a sufficient calibration test.

A.3 Gaussian case - alpha optimum

In the main Gaussian model, with $\rho \in [0,1]$, $p(X \mid y) = \mathcal{N}(\rho y, (1-\rho^2)I)$ and $p(S_\alpha \mid y) = \mathcal{N}(\sqrt{\alpha}y, (1-\alpha)I)$. The Gaussian Wasserstein formula gives

$$\mathbb{E}_Y W_2^2/d = (\sqrt{\alpha}-\rho)^2 + (\sqrt{1-\alpha}-\sqrt{1-\rho^2})^2. \tag{17}$$

The two vectors $(\sqrt{\alpha}, \sqrt{1-\alpha})$ and $(\rho, \sqrt{1-\rho^2})$ have unit norm. Their squared distance is minimized uniquely at $\alpha = \rho^2$. Since $I(X;Y) = -\frac{d}{2}\log(1-\rho^2)$, this is also $\alpha^* = 1 - e^{-2I(X;Y)/d}$. The information-to-noise formula is specific to this calibrated isotropic model, not a rule for every joint law with the same mutual information.

To relate this to the pretrained field, set $D_t = (1-t)^2 + t^2$ and $q_t = (2t-1)/D_t$. Independent Gaussian-to-Gaussian pretraining gives $v_{\text{pre}}(z,t) = q_t z$. At $\alpha = \rho^2$, the conditional source and target are independent draws from the same Gaussian, and

$$v^*(z,t,y) = q_t(z-\rho y) = v_{\text{pre}}(z,t) - q_t\rho y. \tag{18}$$

The spatial slope is unchanged: a shift-only adapter is sufficient. Although the endpoints have the same law, independent-pair FM still contracts and re-expands between them; zero endpoint Wasserstein distance does not make this particular field identically zero.

A.4 A study of low-rank adaptation burden

Fix t and y, center the state as $\widetilde{Z}_t = Z_t - \mathbb{E}[Z_t \mid y]$, and suppose that the exact conditional field differs from the frozen linear field by $\Delta_{\alpha,t}\widetilde{Z}_t$ after its conditional bias has been fit. Thus $\Delta_{\alpha,t}$ is the matrix-valued spatial correction that an unlimited adapter would reproduce. Let A_t be the correction implemented by a rank-r linear adapter. With $\Sigma_{\alpha,t} = \operatorname{Cov}(Z_t \mid y) \succ 0$, and with all expectations in this paragraph taken over $Z_t \mid Y = y$,

$$\mathbb{E}\|(A_t - \Delta_{\alpha,t})\widetilde{Z}_t\|^2 = \|(A_t - \Delta_{\alpha,t})\Sigma_{\alpha,t}^{1/2}\|_F^2. \tag{19}$$

Eckart–Young gives

$$\inf_{\operatorname{rank}(A_t)\le r} \mathbb{E}\|(A_t - \Delta_{\alpha,t})\widetilde{Z}_t\|^2 = \sum_{j>r} \sigma_j^2(\Delta_{\alpha,t}\Sigma_{\alpha,t}^{1/2}), \tag{20}$$

where $\sigma_j(M)$ is the jth singular value of M, ordered from largest to smallest. The left side is the smallest expected squared error achievable by any rank-r linear correction at this time and condition. The right side is the weighted residual field energy outside the r directions the adapter can represent. Integrating over t gives the pathwise linear-Gaussian approximation burden; we use it only as intuition for nonlinear LoRA.

In the anisotropic Gaussian case the source dependence is explicit. Let $X_i = \rho_i Y_i + \sqrt{B_i}\eta_i$, where $B_i = 1 - \rho_i^2$, and put $A = 1 - \alpha$. The centered conditional FM slope in coordinate i is

$$q_{A,B_i}(t) = \frac{tB_i - (1-t)A}{(1-t)^2 A + t^2 B_i}. \tag{21}$$

The unit-Gaussian pretrained slope is $q_{1,1}(t)$. Hence $\Delta_{\alpha,t}$ is diagonal with entries $q_{A,B_i}(t) - q_{1,1}(t)$, and coordinate i needs no spatial correction when $A = B_i$, equivalently $\alpha = \rho_i^2$. A rank-r adapter can cancel at most r independent diagonal mismatches; the remaining cost is exactly the spectral tail above. If most coordinates are predictable (B_i small), a weak adapter minimizes its tail by choosing A small, which in our case means a large α. With more rank it can correct those predictable coordinates while choosing a larger A. This explains why weak adapters can benefit from a more informative source.

A.5 A shorter suffix and approximate marginal overlap

Let $G \sim \mathcal{N}(0, I_d)$ be the original Gaussian source, paired independently with X, define $W_t = (1-t)G + tX$, and construct the synthetic observation $Y_\tau = (1-\tau)G + \tau X$ from that same pair. Then, pathwise,

$$(1-u)Y_\tau + uX = W_{\tau+(1-\tau)u}, \qquad X - Y_\tau = (1-\tau)(X - G). \tag{22}$$

The squared endpoint displacement decreases by exactly $(1-\tau)^2$. This example uses the induced endpoint coupling; it is not obtained by independently redrawing G after observing Y_τ. Its conditional marginal velocity must also retain the conditioning information.

For a less exact source, let q_u be a conditional Wasserstein displacement geodesic from q_0 to the target $p(\cdot \mid y)$, and let ν_y be the proposed anchored source. The triangle inequality gives

$$W_2(\nu_y, p(\cdot \mid y)) \le W_2(\nu_y, q_\tau) + (1-\tau) W_2(q_0, p(\cdot \mid y)). \tag{23}$$

The first term is the price of imperfect intermediate-state matching; the second is the remaining ideal transport. This formalizes the approximate shortcut intuition without asserting that a real degradation lies on a specific pretrained characteristic. Actual NFE also depends on field error, curvature, and the solver, but Diffusion regime analyses [3, 17] motivate looking for dispensable early transport.

A.6 Velocity ambiguity and stochastic gradients

Conditioning on y, remove deterministic means. With independent $e \sim \mathcal{N}(0, AI)$ and $r \sim \mathcal{N}(0, BI)$, $Z_t - \mathbb{E}[Z_t \mid y] = (1-t)e + tr$. For the same FM velocity label $U = X - S_\alpha$ defined in Eq. (1), this gives $U - \mathbb{E}[U \mid y] = r - e$. Gaussian conditioning gives

$$\mathrm{Cov}(U \mid Z_t, Y) = \frac{AB}{(1-t)^2 A + t^2 B} I. \tag{24}$$

Completing the square or substituting $r = t/(1-t)$ into its integral gives

$$\int_0^1 \mathbb{E}\,\mathrm{tr}\,\mathrm{Cov}(U \mid Z_t, t, Y)\, dt = \frac{\pi d}{2}\sqrt{AB}. \tag{25}$$

At $A = 0$, the interior ambiguity is zero because the state reveals the target. This is compatible with nonzero uncertainty at the source itself.

For half-squared FM loss, let $H = (Z_t, t, Y)$ and $J_\theta(H) = \partial v_\theta(H)/\partial\theta$. The per-example gradient is $G = J_\theta^T(v_\theta - U)$. Applying total covariance conditional on H gives the exact decomposition

$$\mathrm{Cov}(G) = \mathrm{Cov}\left[J_\theta^T(v_\theta - v^*)\right] + \mathbb{E}\left[J_\theta^T\,\mathrm{Cov}(U \mid H) J_\theta\right]. \tag{26}$$

The second term is gradient noise due to ambiguous regression labels. The first also depends on field error, the input distribution, and adapter geometry. Thus Eq. (25) identifies an optimization mechanism without imposing monotonicity on every trained gradient statistic. This also distinguishes our question from changing the path coupling while preserving the same marginal field [20].

B Evaluation definitions and protocols

B.1 Architectures and implementation details

Controlled GMM. The clean marginal is

$$p_X = \frac{1}{4} \sum_{s \in \{-1,+1\}^2} \mathcal{N}(s \odot (\mu_1, \mu_2), \sigma I_2), \qquad \mu_1 = \mu_2 = .9, \quad \sigma = .19. \tag{27}$$

Here σ is the component variance, so every marginal coordinate has unit variance: $\mu_j^2 + \sigma = .9^2 + .19 = 1$. The condition selects one mixture component and therefore gives an analytic conditional target. The frozen unconditional field is a three-layer MLP of width 64 (4,674 parameters). Each adapted linear map receives a zero-initialized residual bottleneck; ranks 2, 4, 8, 16, and 64 contain respectively 528, 1,056, 2,112, 4,224, and 16,896 trainable parameters. We use Adam, learning rate .002, batch size 256, uniform $t \in [0, 1]$, and five paired pretraining/adaptation seeds 501–505. Training samples exact (X, Y) pairs and evaluation samples the analytic conditional GMM independently. The main capacity table uses ranks 2/4/8; rank 64 is only a saturation control because it is no longer a parameter-efficient modification of this small network.

FLUX.2. We use the official FLUX.2 Klein Base 4B checkpoint and its frozen VAE at 256^2 resolution. The evolving latent and degraded latent are packed as separate token sequences with distinct position identifiers and concatenated before every transformer call; only the state tokens are integrated. Ordinary residual LoRA factors replace the frozen `qkv`, `proj`, `linear1`, and `linear2` projections in attention and MLP blocks. Unless stated otherwise, rank is 64, giving 70,778,880 trainable parameters. LoRA weights use FP32 AdamW master parameters with BF16 backbone compute, AdamW learning rate 10^{-4}, betas $(.9, .95)$, weight decay .01, batch size one without accumulation, unit gradient clipping, and uniform physical flow time. Source-only experiments zero the degraded token sequence but still construct S_α from the known observation. Thus they change information available to the field, not the data or source.

SiT. We use the official class-conditional SiT-XL/2 256^2 checkpoint with its frozen latent codec and the null class token. A condition encoder applies the frozen SiT patch embedding to the degraded latent. At every transformer block, a trainable bottleneck projects those condition features down to the stated rank and back to the hidden width, with a zero-initialized output so the model is exactly the pretrained field at initialization. All original SiT weights remain frozen. The NFE study uses rank 1024 (132,149,248 trainable parameters), AdamW at 10^{-4} with zero weight decay, batch size one, uniform physical time, one training seed, 128 held-out ImageNet identities, eight draws, and Heun integration at actual NFE 2/4/8/16/32/64.

Distributional metrics. For independent $A, A' \sim Q$ and $X, X' \sim P$, Energy distance is

$$\mathcal{E}(Q, P) = 2\mathbb{E}\|A - X\| - \mathbb{E}\|A - A'\| - \mathbb{E}\|X - X'\|. \tag{28}$$

It vanishes exactly when the laws agree (finite first moments). For GMMs we sample the exact conditional target and use the off-diagonal U-statistic estimator. It can be slightly negative at finite sample size. Whitened Energy applies the same invertible oracle posterior whitening to predictions and targets; it measures discrepancy in units of posterior uncertainty. Raw Energy instead retains the original coordinate scale. The main GMM row uses whitened Energy.

With one paired image reference, use the proper Energy score

$$\widehat{\mathrm{ES}}(Q, x) = \frac{1}{M} \sum_i \|\widehat{x}_i - x\| - \frac{1}{2M(M-1)} \sum_{i \neq j} \|\widehat{x}_i - \widehat{x}_j\|. \tag{29}$$

Averaged over held-out pairs, this equals half the conditional Energy distance plus a target-only constant. Repeated real posterior samples for each image are not necessary for estimating this average proper score. Its absolute value need not be zero for the correct posterior. Its finite image-set estimate is not by itself a certificate of conditional calibration. For a deterministic law $Q = \delta_{m(y)}$, including the $\alpha = 1$ ODE arm, the second term is exactly zero and $\widehat{\mathrm{ES}} = \|m(y) - x\|$. The score is therefore meaningful and finite, but the point predictor receives no reward for posterior diversity.

Individual draws and the model mean. For PSNR, SSIM, and LPIPS, "individual draw" means computing the metric for each generated image and then averaging. "Ten-draw mean" first averages the ten images pixelwise and then computes the metric. The latter estimates the model's conditional mean and is expected to favor distortion metrics; it is neither a posterior sample nor a free one-evaluation output. For FID, we follow the pinned DAPS implementation: generated images undergo an 8-bit PNG round trip, then bilinear resizing and ImageNet normalization before torchvision Inception-v3 prelogit features and `piq.FID`. We report both one fixed draw per identity and the ten-draw mean, using 100 identities. This reproduces the published DAPS protocol exactly. It is too small to be treated as a high-precision population FID: in 2048 feature dimensions its sample covariance has rank at most 99. We therefore use it for direct baseline comparison and do not interpret small FID differences as posterior evidence.

For toys we additionally compute $\|\widehat{\Sigma}_y - \Sigma_y\|_F / \operatorname{tr} \Sigma_y$, mode-mass error, sliced Wasserstein discrepancy, and interval coverage. GMM single-draw risk in the learning row is evaluated using the known target moments: $1 + \text{bias/MMSE} + \text{spread/MMSE}$, with the Monte Carlo contribution to squared estimated mean bias removed. It is not the risk of a mean image. The stored CSV also contains direct paired-draw MSE.

Conditional spread. For M independent samples at one fixed condition y, the reported conditional endpoint variance is

$$\widehat{V}(y) = \frac{1}{M-1} \sum_{m=1}^{M} \frac{\|\widehat{x}^{(m)} - \overline{x}\|_2^2}{d}, \qquad \overline{x} = \frac{1}{M} \sum_{m=1}^{M} \widehat{x}^{(m)}. \tag{30}$$

We average $\widehat{V}(y)$ over held-out conditions. This is variance across random seeds *at fixed* y, not variance across different images. It is zero for a deterministic $\alpha = 1$ ODE. Matching total spread alone does not imply the correct covariance, mode weights, or realism, so we report Energy as the complementary distributional metric.

Toy and image learning curves. The GMM learning curves use five paired seeds, 24 conditions, 64 draws, and separate validation and test streams. The main curve fixes rank 16 and 16 NFE; the capacity table uses ranks 2/4/8 and extends selected arms to 50k updates. Five-seed bands use Student-t intervals. The FLUX study fixes rank 64 and one training seed with standard uniform-time flow matching. Evaluation uses 16 held-out FFHQ identities, ten draws, two inference replicates, and 95% identity-bootstrap intervals.

Benchmark protocol. The FFHQ benchmark uses the official first 100 test identities, disjoint training data (59k faces), and a disjoint validation pool for NFE selection. The frozen observations are produced by pinned DAPS operators with noise standard deviation .05 in $[-1, 1]$: cubic SR×4, Gaussian blur with a 61×61 kernel and standard deviation 3, and a random 128×128 inpainting box. A bicubic lift for the SR condition is distinct from the forward operator. Measurements and masks are fixed when comparing our models. Published baseline numbers use the same benchmark protocol, not necessarily our exact realized noise.

PSNR/SSIM/LPIPS use DAPS's `piq` protocol on clipped $[0, 1]$ images with replace-pooling LPIPS. FID uses the pinned official Inception-prelogit implementation, PNG-quantized predictions, and the same 100 reference images; this is not a conventional 50k-sample FID evaluation. Our standard-FM benchmark numbers average the prescribed first draw across five inference replicates; ten-draw ensembles and best-of-ten values are reported separately. DAPS's published linear-task configuration uses $250 \times 4 = 1000$ denoiser evaluations; LatentDAPS uses the same schedule plus inner latent updates. Its newer 100-NFE configurations must not be assigned the 1000-NFE table's scores.

C Complementary experiments

C.1 Practical restoration quality across inverse problems

Table 6 compares the resulting adapters with published DAPS baselines under the 100-image FFHQ protocol. The comparison establishes that the mechanisms studied in the main paper yield competitive restorers; it is not intended as the paper's primary contribution.

Table 6: **Restoration quality and inference cost.** Results follow the 100-image FFHQ protocol of Zhang et al. [24].

Task	Method	NFE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
SR ×4	LatentDAPS	1000	27.48	.801	.182	59.62
	Pixel DAPS	1000	29.07	.818	.177	51.44
	Ours ($\alpha = 0$)	**4**	28.17	.809	.141	**45.52**
Gaussian blur	LatentDAPS	1000	27.93	.764	.234	64.52
	Pixel DAPS	1000	**29.19**	**.817**	.165	53.33
	Ours ($\alpha = 0$)	**4**	27.96	.801	**.134**	**44.99**
Box inpainting	LatentDAPS	1000	23.99	.802	.194	46.52
	Pixel DAPS	1000	24.07	.814	.133	43.10
	Ours ($\alpha = 0$)	**4**	**25.12**	**.847**	**.098**	**36.30**

C.2 Posterior proximity and integration cost

We separate two related effects that should not be conflated. First, source proximity changes which α is easiest to adapt. We keep a unit-variance four-mode marginal and one revealed component, but vary the component variance σ. Thus $\mu_1 = \mu_2 = \sqrt{1-\sigma}$ and the source whose conditional variance matches the posterior has $\alpha_{\text{var}} = 1 - \sigma$. Table 7 reports a three-seed rank-16 experiment evaluated at a fixed 64 NFE. The early selected source becomes more informative as the posterior narrows, but remains below α_{var} for the narrowest posterior. This is consistent with a finite model balancing proximity against expansion and frozen-field correction costs; it is not evidence that variance matching alone determines the optimum.

Table 7: **The preferred early source follows posterior geometry.** Three-seed GMM results at 50 updates and 64 NFE. $\widehat{\alpha}$ is selected on validation; because posterior scale varies, Energy is comparable only within each row.

σ	$\mu_1 = \mu_2$	$\alpha_{\text{var}} = 1 - \sigma$	$\widehat{\alpha}$ at 50	Test Energy ↓
.75	.50	.25	.25	.092
.19	.90	.81	.75	.200
.0199	.99	.9801	.81	.024

Second, holding the baseline posterior ($\sigma = .19$) fixed, we sweep NFE on the same five rank-16 models. Table 8 exposes the predicted integration effect. The required standard-deviation scale is $\sqrt{\sigma/(1-\alpha)}$ from Eq. (31). At 10k updates, four NFE resolve the contracting and approximately variance-matched routes, but under-expand the $\alpha = .97$ route. More steps are needed to recover its posterior variance. The exact solver error decreases monotonically with NFE, whereas Energy need not: a discretization error can accidentally offset a learned-field error. We therefore report fixed-NFE variance and a numerical convergence criterion, rather than selecting NFE by the noisy minimum of a downstream metric or claiming a universal monotone NFE law.

For one conditional Gaussian component, with source variance $A = 1 - \alpha$ and posterior variance σ, independent linear coupling gives

$$c_\alpha(t) = A(1-t)^2 + \sigma t^2, \qquad v_t(z) = \frac{\sigma t - A(1-t)}{c_\alpha(t)} z. \tag{31}$$

The endpoint standard-deviation ratio is therefore $\sqrt{\sigma/A}$. This calculation isolates solver stiffness from learned field error; it does not predict that every downstream metric improves monotonically with NFE.

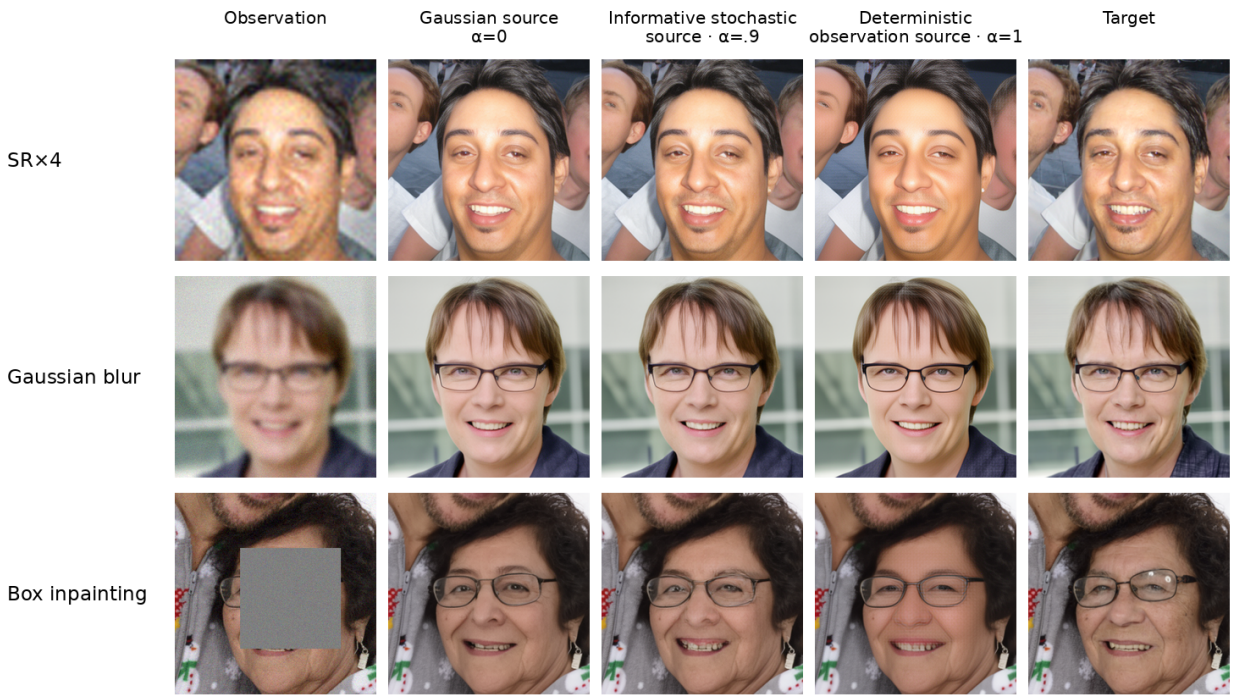

Figure 6: **Source randomness preserves fine detail.** Four-step FLUX.2 restorations across super-resolution, Gaussian deblurring, and box inpainting. The informative stochastic source avoids the over-smoothed deterministic prediction.

Table 8: **Near-deterministic sources require finer integration.** One-mode GMM at 10k updates. Variance ratio targets one; "Scale" is the exact endpoint standard-deviation ratio, with values above one requiring expansion.

α	Scale $\sqrt{.19/(1-\alpha)}$	Var.@4 NFE	Var.@16 NFE	5%-spread NFE
0	.44	1.05	1.14	8
.75	.87	.96	1.09	8
.9	1.38	.85	1.06	8
.97	2.52	.58	1.06	16

C.3 Additional exact-Gaussian evidence

C.4 Additional learning dynamics

The transient expansion seen in the one-mode GMM is not universal. Figure 8 repeats the same rank-16, 16-NFE protocol for one-, two-, and four-mode conditional posteriors. The overshoot is mild for two modes and absent for four modes, while Energy continues to detect distributional error after total variance is approximately correct.

The same finite-training crossover appears under Gaussian deblurring, providing a degradation-level check beyond the SR experiment in the main paper.

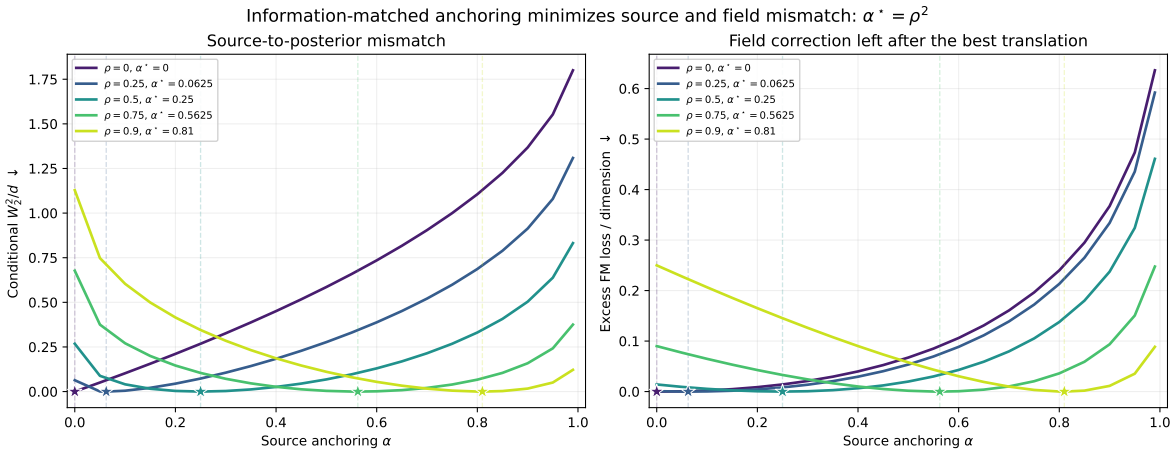

Figure 7: **Information matching in the exact Gaussian model.** Left: source–posterior distance. Right: residual field error after the best condition-dependent translation. Stars mark $\alpha^* = \rho^2$.

D Interior noise and matched SDE sampling

We use the stochastic-interpolant construction of Albergo et al. [2] and its flow-transformer formulation in SiT [15]. For completeness, the path used in our experiments is

$$Z_t = (1-t)\sqrt{\alpha}\, Y + tX + h(t)\xi, \qquad h(t)^2 = (1-\alpha)(1-t)^2 + 2c^2 t(1-t), \tag{32}$$

where $\xi \sim \mathcal{N}(0, I_d)$ is independent. The parameter c controls training-path noise and the bridge term vanishes at both endpoints. Write $m = \sqrt{\alpha} y$. Differentiating the interpolation gives target velocity $U_t = X - m + h'(t)\xi$. Gaussian integration by parts gives $s^*(z,t,y) = -\mathbb{E}[\xi \mid z,t,y]/h(t)$, with no requirement that $p(X \mid y)$ itself be Gaussian. Eliminating $\mathbb{E}[X \mid z,t,y]$ between the state and velocity identities gives

$$tv^* - z + m = (th' - h)\mathbb{E}[\xi \mid z,t,y] = (h^2 - thh')s^*. \tag{33}$$

For Eq. (32), $h^2 - thh' = (1-\alpha)(1-t) + c^2 t$. This proves the velocity-to-score formula where the denominator is nonzero. It works for a deterministic known anchor plus Gaussian perturbations, without knowing the degradation operator. If both source and bridge noise are arbitrary and unknown, velocity alone generally does not determine the score.

Explicitly, the conditional score and the marginal-preserving SDE are

$$s^*(z,t,y) = \frac{tv^*(z,t,y) - z + \sqrt{\alpha} y}{(1-\alpha)(1-t) + c^2 t}, \qquad dZ_t = [v^* + w(t)s^*]dt + \sqrt{2w(t)}\, dB_t. \tag{34}$$

Here c defines the path on which the field is learned, whereas $w(t)$ is the sampling diffusion gauge; they are distinct choices. Setting $w = 0$ recovers the probability-flow ODE.

The Fokker–Planck equation for drift $v^* + ws^*$ and diffusion $\sqrt{2w}I$ is

$$\partial_t p = -\nabla\cdot(pv^*) - w\nabla\cdot(ps^*) + w\Delta p = -\nabla\cdot(pv^*). \tag{35}$$

This verifies the same smooth marginals when initialized correctly. It does not select a singular point-source entrance by itself. With $Y = y$ retained as an explicit condition and a valid entrance at $\alpha = 1$, the exact SDE therefore ends at $p(X \mid Y = y)$. The associated ODE, however, maps the one initial state to one endpoint and cannot be a posterior sampler.

An exact-field GMM isolates this distinction without optimization error. We start from the deterministic $\alpha = 1$ source and use $c = .25$ toward a four-mode posterior. The ODE remains deterministic for every NFE. The matched SDE creates the missing variance, occupies all four modes, and converges to the target covariance as its discretization is refined. With $\kappa = 1$ and 16 NFE, its variance ratio is .990, mode-mass TV

is .084, and relative covariance error is .014; the corresponding ODE has variance ratio zero, mode-mass TV .75, and relative covariance error one (Figure 10).

Learned FLUX fields show the same hierarchy. We train rank-64 adapters for 50k updates on FFHQ SR×4 with $c = .25$, then sample each checkpoint with its ODE and matched SDE. At $\alpha = 1$, only the SDE creates endpoint spread. At $\alpha = .9$, the samplers are nearly indistinguishable because the source already supplies randomness, as shown in Figure 11 and summarized in Table 3.

What changes without the condition branch? The identity above is conditional: both $v^*(z, t, y)$ and the known anchor $m = \sqrt{\alpha}y$ enter the score. If the field sees only Z_t, regression instead returns the source-only velocity $\bar{v}(z, t) = \mathbb{E}[U_t \mid Z_t = z]$. Define $D_t = (1 - \alpha)(1 - t) + c^2 t$ and substitute $\bar{v}$ into the exact conversion:

$$\widetilde{s}_y(z, t) = \frac{t\bar{v}(z, t) - z + \sqrt{\alpha}y}{D_t}. \tag{36}$$

Writing $s^*_y = \nabla_z \log p_t(z \mid y)$, $e_y = \bar{v} - v^*(\cdot, \cdot, y)$, and defining the exact and velocity-derived drifts as $b^*_\kappa = v^* + w_\kappa s^*_y$ and $\widetilde{b}_\kappa = \bar{v} + w_\kappa \widetilde{s}_y$ gives

$$\widetilde{s}_y - s^*_y = \frac{t}{D_t}(\bar{v} - v^*), \qquad \widetilde{b}_\kappa - b^*_\kappa = \left(1 + \frac{w_\kappa t}{D_t}\right)(\bar{v} - v^*). \tag{37}$$

At $\alpha = 1$, $Z_0 = Y$ makes the explicit branch redundant at the source endpoint. Bridge noise and overlapping paths can make Y non-measurable from Z_t later, but their entire effect on the velocity-derived dynamics is exposed by the pooled-to-conditional velocity gap in Eq. (37).

This identity also yields a terminal stability guarantee. Couple the exact conditional and velocity-derived SDEs with the same Brownian path and suppose the exact conditional drift is L_t-Lipschitz. Then the standard synchronous-coupling argument gives

$$\mathbb{E}\|\widetilde{Z}_1 - Z^*_1\|^2 \le C_L \int_0^1 \mathbb{E}\left[\left(1 + \frac{w_\kappa(t)t}{D_t}\right)^2 \|\bar{v} - v^*\|^2\right] dt, \quad C_L = \exp\int_0^1 (2L_t + 1)dt, \tag{38}$$

apart from numerical discretization error. For the point source and $w_\kappa = \kappa c^2$, the multiplier inside the integral is simply $(1 + \kappa)^2$. Thus the score conversion introduces no singular drift error: more training reduces the statistical term, more NFE reduces the numerical term, and κ can be selected after training. The construction is exact whenever $\bar{v} = v^*(\cdot, \cdot, y)$ on the sampled path, including cases in which Y is recoverable from Z_t.

A source-only stochastic bridge is exact only when its pooled and conditional velocities agree along the sampled paths. Matching the endpoint marginal alone does not determine the observation–target coupling; the drift-error identity above makes this limitation explicit.

We next evaluate the velocity-derived construction itself. One source-only $\alpha = 1$ velocity checkpoint is reused throughout; only the inference gauge changes. The deterministic endpoint has zero spread. Low and intermediate diffusion create nonzero spread and improve conditional Energy without retraining. A broader pilot sweep confirms that excessive diffusion eventually overshoots the useful frontier. Relative to the ODE, $\kappa = .25$ and .5 lower Energy by 2.61 and 2.95, respectively; their identity-paired 95% intervals are $[2.49, 2.72]$ and $[2.81, 3.08]$.

Paired field-gap audit. For Figure 12, we compare source-only and conditioned adapters at every common training checkpoint on exactly the same validation states $Z_t = (1 - t)y + tX + c\sqrt{2t(1 - t)}\xi$. The audit uses 32 identities, four fixed ξ draws per identity, and seven path times. Uncertainty is computed over identity means after averaging bridge draws. Since the population conditional velocity is unavailable for images, v^{cond}_θ is the operational reference rather than an oracle. The right panel is an implementation check of the analytic identity, not an independently discovered linear law: after measuring $e_y = \bar{v}_\theta - v^{\text{cond}}_\theta$, both outputs are inserted into the same reverse-SDE drift and the measured ratio agrees with $1 + \kappa$ to within 3×10^{-8} over the reported gauges.

This single-checkpoint sweep is the intended plug-and-play use case. $\kappa = .1$ gives the best LPIPS and $\kappa = .3$ nearly reaches the best Energy while retaining more fidelity than $\kappa = .5$. Thus $\kappa \in [.1, .3]$ is

a useful conservative starting range; broader values can then be explored when diversity is the primary objective. Increasing NFE refines the numerical solution, whereas longer training reduces the distinct velocity-approximation term in Eq. (38).

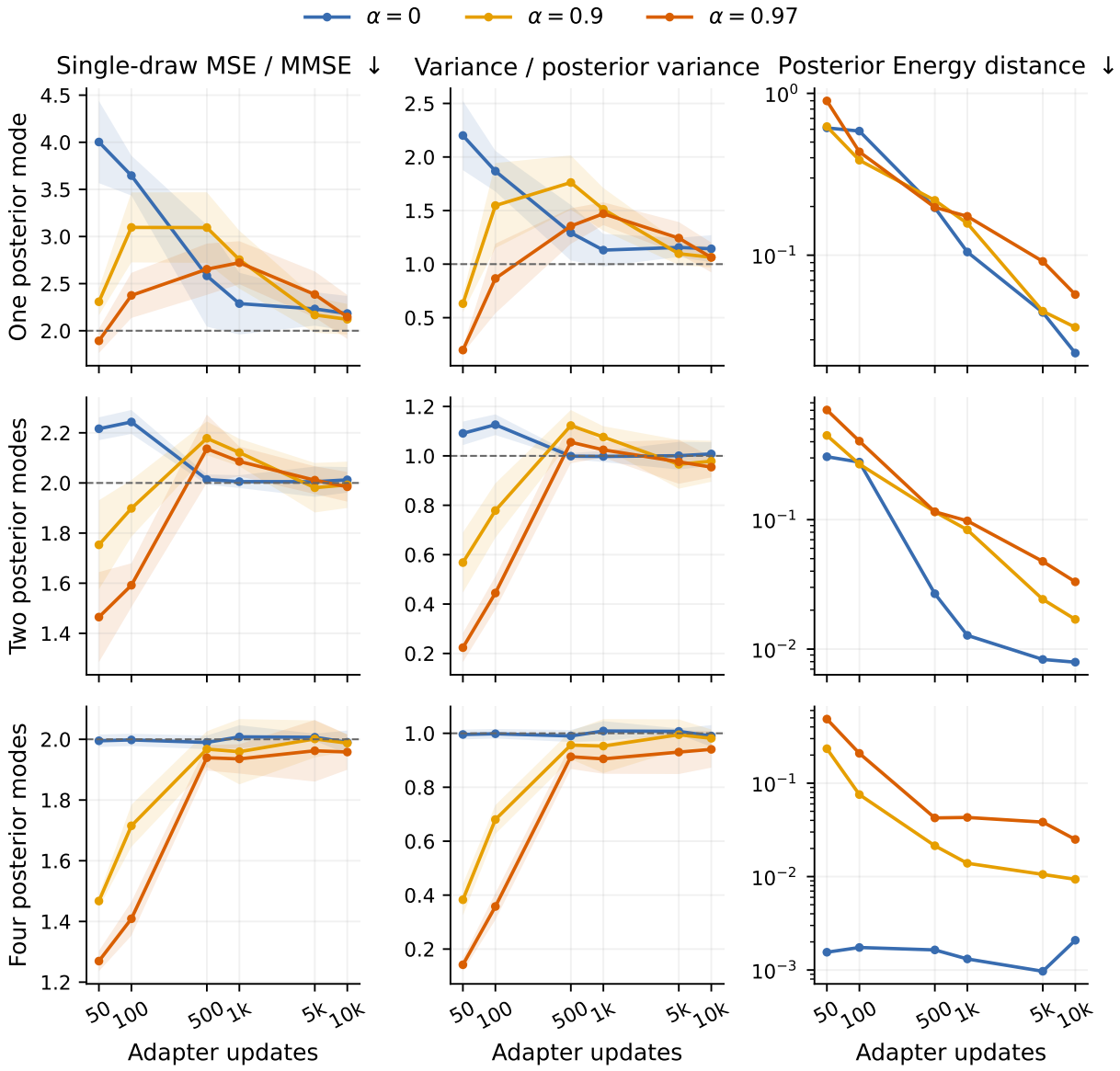

Figure 8: **Source dynamics across GMM geometries.** Informative sources expand from low variance, whereas the Gaussian source contracts. Curves average five seeds.

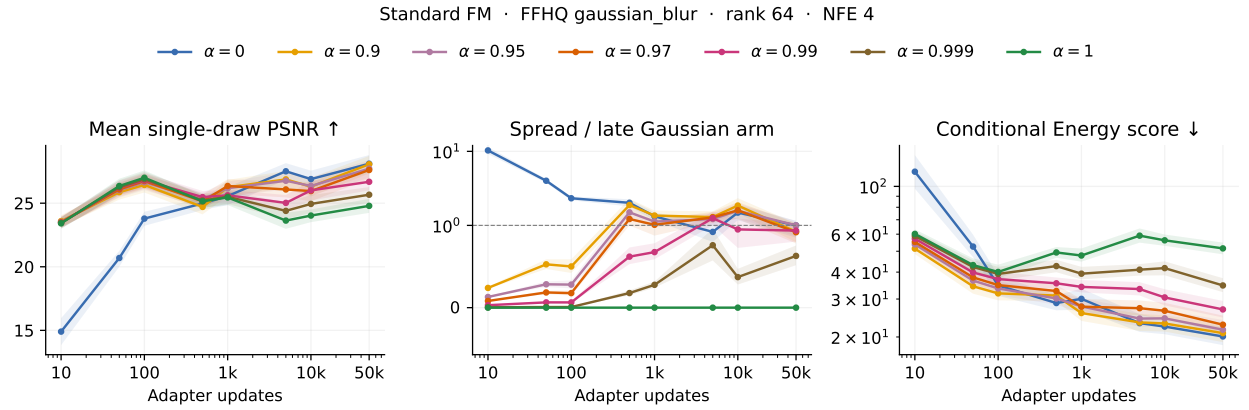

Figure 9: **FLUX.2 learning dynamics under Gaussian blur.** Rank-64 adapters with four-step sampling reproduce the crossover observed for super-resolution.

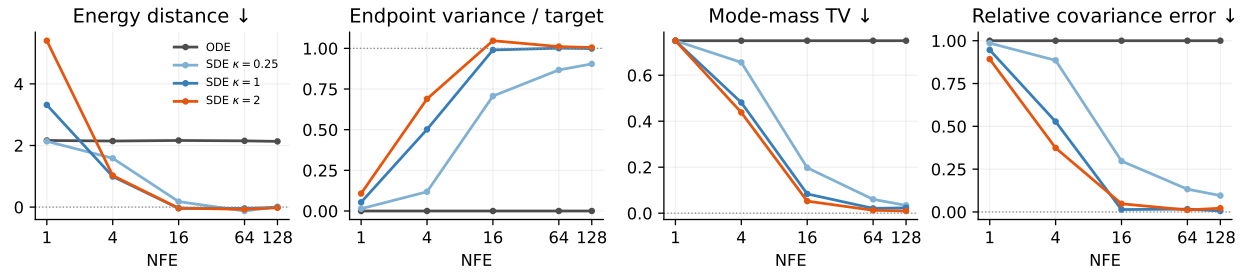

Figure 10: **A point-source bridge requires stochastic sampling.** For an exact four-mode GMM with $\alpha = 1$, the ODE remains deterministic, whereas matched SDEs recover posterior statistics as NFE increases.

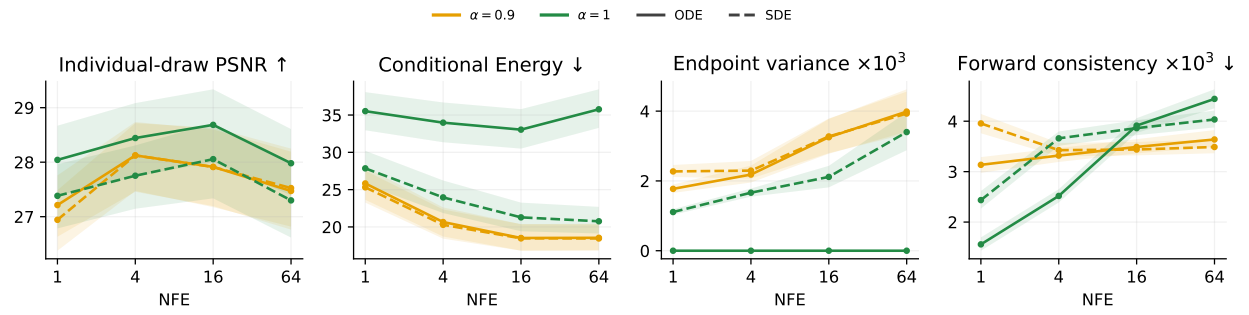

Figure 11: **Source and bridge randomness with FLUX.2.** At $\alpha = 1$, only the SDE creates endpoint spread; at $\alpha = .9$, source noise makes the ODE and SDE nearly coincide. Bands are 95% bootstrap intervals over 16 FFHQ identities.

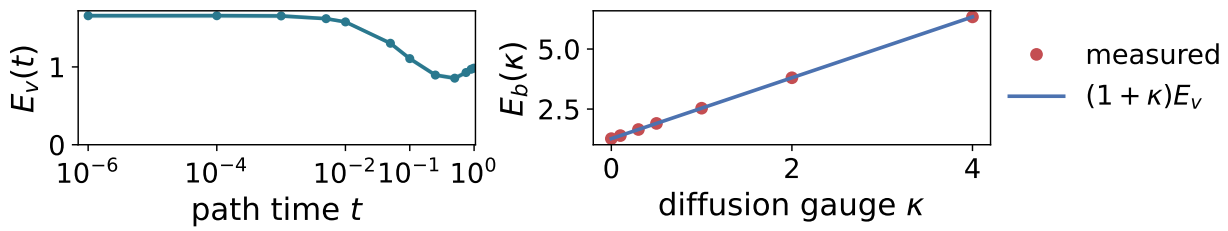

Figure 12: **Velocity error and drift amplification.** Left: source-only velocity mismatch along the bridge. Right: drift error follows the predicted $1 + \kappa$ factor.

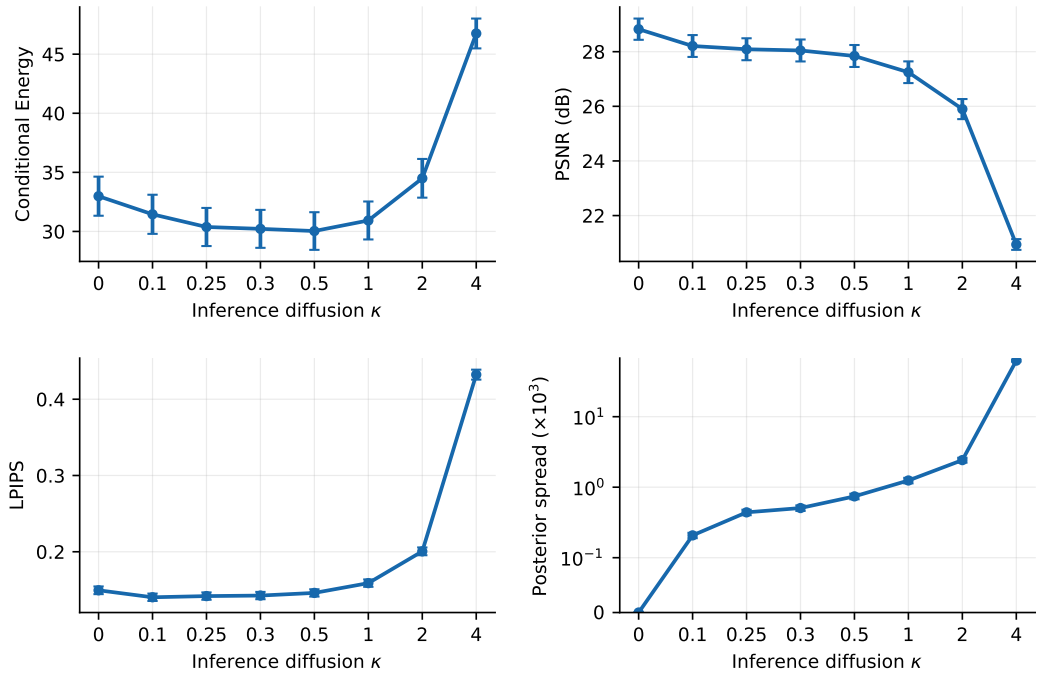

Figure 13: **Inference-time stochasticity control.** Every point reuses one source-only FLUX.2 checkpoint; only κ changes. Moderate diffusion improves distributional quality and creates spread, while excessive diffusion reduces fidelity. Error bars are 95% intervals over identities.

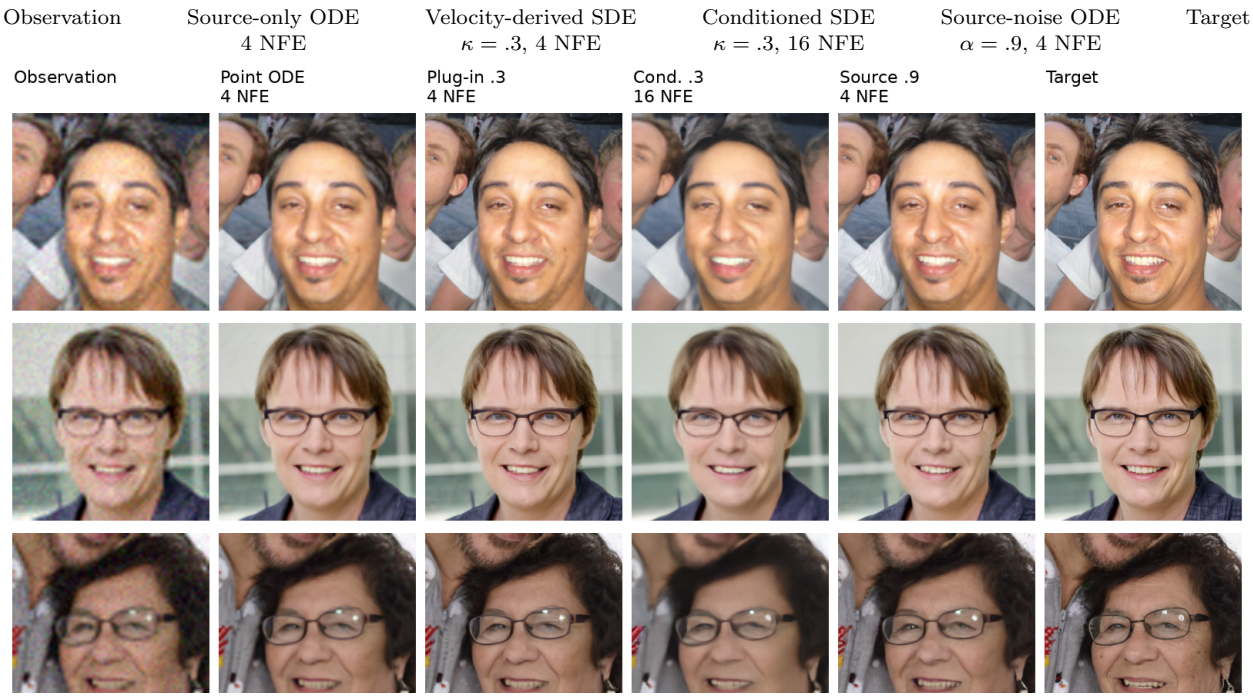

Figure 14: **Sampling strategies on FFHQ SR×4.** Deterministic, velocity-derived, explicitly conditioned, and source-noise outputs.

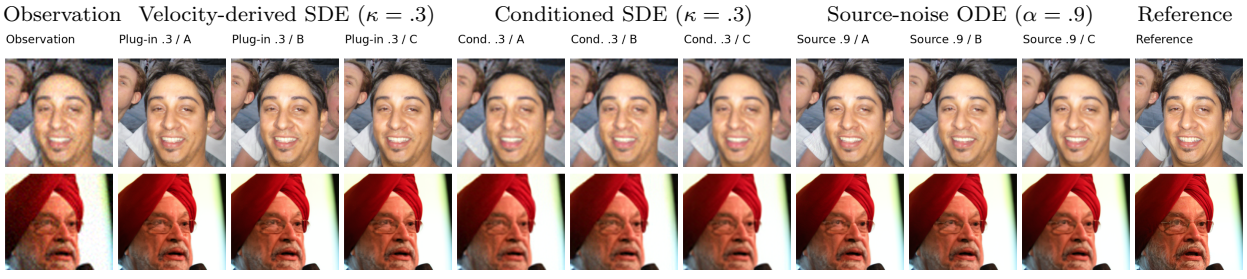

Figure 15: **Repeated samples reveal the remaining posterior gap.** Three draws for two held-out FFHQ identities. The velocity-derived SDE remains closest to its deterministic prediction, while explicit conditioning and source noise produce greater variation. Samples are fixed, not metric-selected.