DIFFUSION POSTERIOR SAMPLING WITH LOCAL DENOISER RESPONSES

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion split-Gibbs samplers alternate observation updates with prior-conditional draws, but a full reverse diffusion solve at every sweep limits the number of sweeps available under a forward-evaluation budget. We replace that solve with a local-response update requiring two denoiser evaluations. The denoiser is evaluated at the current input and a Gaussian-perturbed copy; their scaled finite-step difference is injected directly alongside independent noise from an offline residual covariance table. This retains an input-dependent perturbation without computing a dense covariance matrix or differentiating through the network. Our analysis characterizes the ideal residual representation and bounds stationary Gaussian output covariance error at fixed admissible coupling after an exact final observation/prior sweep. With fewer online forward evaluations than our full-diffusion reference, the implemented sampler improves held-out inverse-scattering reconstruction with comparable rank calibration. On a held-out natural-image prefix, reconstruction and energy score improve for super-resolution, box inpainting, and random masking, while deblurring favors the reference. Selection controls support retaining the raw response as a whole, without resolving an independent residual benefit. The results support allocating a limited forward budget to inexpensive input-adaptive conditionals and more outer sweeps, with benefits that depend on the inverse problem.

1 INTRODUCTION

Recovering a signal from noisy, incomplete measurements requires more than a plausible reconstruction: different signals can explain the same observation. Pretrained priors generate signals through learned reverse transitions (Ho et al., 2020), noise-conditioned Langevin updates (Song & Ermon, 2019), or reverse stochastic differential equations (Song et al., 2021). Split-Gibbs methods separate the observation model from the prior, alternating a measurement update with a diffusion denoising conditional (Coeurdoux et al., 2024; Xu & Chi, 2024; Wu et al., 2024). Sampling that conditional through an entire reverse trajectory can make each sweep expensive.

The computational choice is to spend evaluations on accurate conditional draws or take more, cheaper sweeps. A denoiser approximates the conditional mean, but uncertainty around it depends on the input; a static covariance cannot adapt at a fixed noise level. We ask how to retain useful local uncertainty directions with a single additional forward evaluation.

We evaluate the denoiser at the current input and a Gaussian-perturbed copy, then add the scaled finite-step difference to its base prediction. This retains the response direction. Independent noise from an offline residual table approximates an additional covariance correction. This local-response update defines moment-matched split Gibbs (MM-SG), replacing the costly prior draw within an inherited split interface with finite-coupling observation compensation (Wang et al., 2026, Appendix A.5). A fixed prior-and-operator diagnostic selects this update or a static one-call alternative. Observation updates, annealing, and stochastic final denoising descent complete the sampler (Section 3.3; Figure 1).

Held-out inverse scattering and natural-image restoration demonstrate the tradeoff. In scattering, the local response improves reconstruction over our full-diffusion split-Gibbs reference with fewer online evaluations and comparable rank calibration; the static alternative is worse on both metrics. On the held-out natural-image prefix, reconstruction and energy score improve over the reference

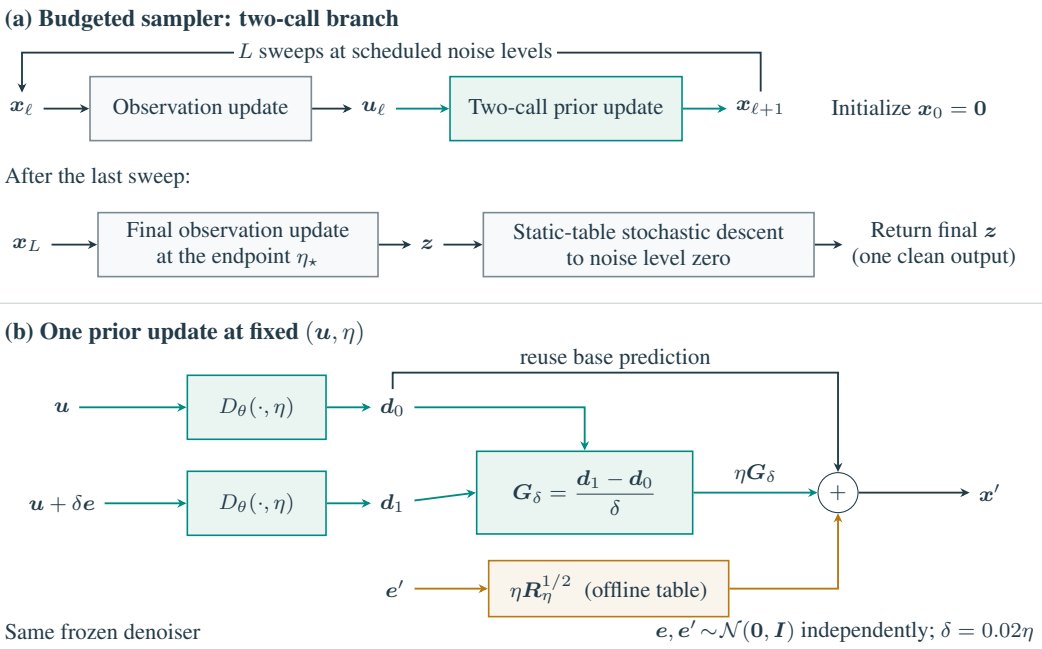


Figure 1: **Local-response sampling (schematic).** The ordinary loop repeats both observation and prior updates; the final observation update precedes static-table stochastic descent. Inside the prior update, the same frozen denoiser produces $\boldsymbol{d}_0$ and $\boldsymbol{d}_1$. Their finite difference retains a local response direction, and an independent probe adds residual-table noise. The base prediction is reused; no matrix Jacobian is formed.

for super-resolution, box inpainting, and random masking, while deblurring favors the reference (Sections 5.1 and 5.3). Selection ablations support retaining the raw response but do not resolve a separate image benefit from the residual.

With the true Bayes denoiser and an exact directional response, response covariance is proportional to the square of conditional covariance. The residual can be easier to represent in a fixed basis. We characterize this representation and bound stationary Gaussian output covariance error at a fixed admissible coupling after an exact final sweep (Section 4). Exact-reference and learned-prior controls show why better conditional moments or more sweeps need not improve finite-budget output. Our contribution is the inexpensive local-response construction, its residual analysis, and the demonstrated accuracy–budget tradeoff, not moment matching or split Gibbs in general.

2 Background

Let $\boldsymbol{x} \in \mathbb{R}^d$ be a clean signal and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon} \in \mathbb{R}^m$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma_y^2 \boldsymbol{I})$, $\sigma_y > 0$, and $\boldsymbol{A} \in \mathbb{R}^{m \times d}$ is known. The target is $p(\boldsymbol{x} \mid \boldsymbol{y}) \propto p(\boldsymbol{x}) \exp[-\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 / (2\sigma_y^2)]$. A pretrained diffusion denoiser $D_\theta(\boldsymbol{u}, \eta)$ approximates the Bayes mean $\boldsymbol{m}_\eta(\boldsymbol{u}) = \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{x} + \eta \boldsymbol{e} = \boldsymbol{u}]$, with $\boldsymbol{e} \sim \mathcal{N}(0, \boldsymbol{I})$. We distinguish this ideal mean from the learned network throughout.

Conditioning a denoising prior. Plug-and-play optimization decouples forward-model inversion from denoising (Venkatakrishnan et al., 2013). Sampling alternatives use denoiser-based Langevin updates (Laumont et al., 2022) or stochastic updates constrained by noiseless linear measurements (Kadkhodaie & Simoncelli, 2021). Within reverse diffusion, observations can enter through resampling known pixels at their noise level (Lugmayr et al., 2022) or measurement gradients through the denoised estimate (Chung et al., 2022). Particle methods instead correct proposals with importance weights and resampling (Wu et al., 2023), including a filtering formulation for linear inverse problems (Dou & Song, 2024).

The conditional sampling interface. Split Gibbs introduces an auxiliary noisy signal $\boldsymbol{u}$, then alternates an observation update and a prior conditional $p_\eta(\boldsymbol{x} \mid \boldsymbol{u}) \propto p(\boldsymbol{x})\mathcal{N}(\boldsymbol{u}; \boldsymbol{x}, \eta^2 \boldsymbol{I})$ (Vono et al., 2019; Coeurdoux et al., 2024). We use the finite-coupling covariance compensation described in Wang et al. (2026, Appendix A.5) for the linear-Gaussian observation update. It defines an exact augmented target when $\sigma_y^2 \boldsymbol{I} - \eta^2 \boldsymbol{A}\boldsymbol{A}^\top \succ 0$; outside that range, a noise-matching anneal is used.

Conditional covariance and related approximations. Writing $\boldsymbol{J}_\eta(\boldsymbol{u}) = \nabla_{\boldsymbol{u}} \boldsymbol{m}_\eta(\boldsymbol{u})$, the denoising identities give (Efron, 2011; Zhang et al., 2023)

$$\mathbb{E}[\boldsymbol{x} \mid \boldsymbol{u}] = \boldsymbol{m}_\eta(\boldsymbol{u}), \qquad \mathrm{Cov}(\boldsymbol{x} \mid \boldsymbol{u}) = \eta^2 \boldsymbol{J}_\eta(\boldsymbol{u}). \tag{1}$$

Moment-matching denoising Gibbs sampling uses these moments to construct an input-dependent Gaussian conditional and, for images, estimates the diagonal Hessian stochastically (Zhang et al., 2023). TMPD and MMPS instead approximate the noisy measurement likelihood with a Gaussian to form a likelihood-guided score (Boys et al., 2024; Rozet et al., 2024). Our update draws a prior conditional rather than modifying that score.

Covariance representations also differ. Reverse variances can be learned during training (Nichol & Dhariwal, 2021) or estimated from offline score statistics (Bao et al., 2022). Posterior covariance can be calibrated from offline denoising error under an isotropic model or learned in an orthonormal basis (Peng et al., 2024); trajectory-based estimators instead add low-rank curvature updates (Rissanen et al., 2025). Finite-difference denoiser actions can estimate posterior principal directions by subspace iteration (Manor & Michaeli, 2024). Covariance-aware sampling uses transformed probe–response products to rescale the same transformed probe (Schioppa & Salimans, 2026, Section 2.3 and Appendix C); Lanczos sampling approximates covariance square-root actions through repeated matrix–vector products (Akhtar et al., 2026). We instead inject the raw directional response and add independent noise from a fixed offline residual table. We do not report matched comparisons with these methods.

3 METHOD

Given $(\boldsymbol{y}, \boldsymbol{A}, \sigma_y)$, a fixed denoiser $D_\theta : \mathbb{R}^d \times (0, \infty) \to \mathbb{R}^d$, fixed covariance tables, branch choice, schedules, and an online forward-evaluation budget B, we return one clean sample. At sweep ℓ, an observation update forms $\boldsymbol{u}_\ell \in \mathbb{R}^d$ from the current clean state $\boldsymbol{x}_\ell \in \mathbb{R}^d$ and $\boldsymbol{y}$; a prior update then produces $\boldsymbol{x}_{\ell+1}$. We specify that prior update for fixed $(\boldsymbol{u}, \eta)$ before describing the repetition and readout. For images, D_θ is a frozen learned denoiser; for analytic mixtures it is the true Bayes mean, which alone does not make the Gaussian prior update exact. Neither the denoiser nor the tables are fitted to $\boldsymbol{y}$. All Gaussian noises in the chain are fresh and mutually independent.

3.1 A STATIC COVARIANCE REFERENCE

Let $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_d] \in \mathbb{R}^{d \times d}$ be a fixed orthonormal basis and let $\hat{j}_k(\eta)$ estimate the average conditional variance along $\boldsymbol{w}_k$, divided by η^2. A calibration pair consists of a prior or unconditional model sample $\boldsymbol{x}_n$ and $\boldsymbol{u}_n = \boldsymbol{x}_n + \eta \boldsymbol{e}_n$, with $\boldsymbol{e}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. For N pairs, the basic denoising-error estimate is

$$\hat{j}_k(\eta) \approx \frac{1}{N\eta^2} \sum_{n=1}^{N} \left[\boldsymbol{w}_k^\top (D_\theta(\boldsymbol{u}_n, \eta) - \boldsymbol{x}_n)\right]^2, \quad \boldsymbol{S}_\eta = \boldsymbol{W}\,\mathrm{diag}(\hat{\boldsymbol{j}}(\eta))\boldsymbol{W}^\top. \tag{2}$$

With the exact Bayes mean and true prior samples, the population estimate equals $\mathbb{E}_{p_\eta}[\boldsymbol{w}_k^\top \boldsymbol{J}_\eta(\boldsymbol{u})\boldsymbol{w}_k]$, where p_η is the noisy-prior law. The resulting one-call update is $\boldsymbol{x}' = D_\theta(\boldsymbol{u}, \eta) + \eta \boldsymbol{S}_\eta^{1/2} \boldsymbol{e}$. It assigns the same calibrated covariance at every input rather than adapting it to the current $\boldsymbol{u}$. Appendix B.2 specifies pooling, variance-reduction, and clipping in the evaluated variants.

3.2 RETAINING THE RAW LOCAL RESPONSE

Instead of assigning all uncertainty from the table, the two-call update probes the denoiser at the current input:

$$\boldsymbol{G}_\delta(\boldsymbol{u}, \boldsymbol{e}; \eta) = \frac{D_\theta(\boldsymbol{u} + \delta \boldsymbol{e}, \eta) - D_\theta(\boldsymbol{u}, \eta)}{\delta}, \qquad \delta = 0.02\eta, \tag{3}$$

$$\boldsymbol{x}' = D_\theta(\boldsymbol{u}, \eta) + \eta\big[\boldsymbol{G}_\delta(\boldsymbol{u}, \boldsymbol{e}; \eta) + \boldsymbol{R}_\eta^{1/2}\boldsymbol{e}'\big]. \tag{4}$$

Here $\boldsymbol{e}, \boldsymbol{e}' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ are independent. We calibrate the response power $\hat{q}_k(\eta) \approx \mathbb{E}[(\boldsymbol{w}_k^\top \boldsymbol{G}_\delta)^2]$, averaging over noisy calibration samples and independent probes. At each response-calibration level, we set

$$r_k(\eta) = \mathrm{clip}\big(\hat{j}_k(\eta) - \hat{q}_k(\eta), 0, 1\big), \qquad \boldsymbol{R}_\eta = \boldsymbol{W}\,\mathrm{diag}(\boldsymbol{r}(\eta))\boldsymbol{W}^\top. \tag{5}$$

Here $\mathrm{clip}(\cdot, 0, 1)$ limits each coefficient to $[0, 1]$. The image sampler interpolates the stored r_k between calibration levels; Appendix B.2 specifies this order and the vector variants. The response is injected directly; it is not replaced by noise with the same pooled power. This is the input-dependent part of the construction. The residual provides an independent positive-semidefinite covariance correction. The base prediction in Equation (3) is reused, so the update requires two forward evaluations, batched together in the image runners, and no automatic differentiation.

With the exact Bayes mean and response $\boldsymbol{J}_\eta(\boldsymbol{u})\boldsymbol{e}$, Equation (4) has covariance $\eta^2(\boldsymbol{J}_\eta^2 + \boldsymbol{R}_\eta)$. Its error divided by η^2 is

$$(\boldsymbol{J}_\eta^2 + \boldsymbol{R}_\eta) - \boldsymbol{J}_\eta = \boldsymbol{R}_\eta - \boldsymbol{\Phi}_\eta(\boldsymbol{u}), \qquad \boldsymbol{\Phi}_\eta(\boldsymbol{u}) = \boldsymbol{J}_\eta(\boldsymbol{u})(\boldsymbol{I} - \boldsymbol{J}_\eta(\boldsymbol{u})). \tag{6}$$

Exact covariance matching in this idealization would require $\boldsymbol{R}_\eta = \boldsymbol{\Phi}_\eta(\boldsymbol{u})$. The fixed table approximates this input-dependent residual rather than the full $\boldsymbol{J}_\eta$. A finite forward difference can be non-Gaussian and have nonzero probe mean, so its calibrated power is not necessarily its variance. A learned denoiser also need not have the symmetric positive-semidefinite Jacobian of the Bayes mean. Appendix A.3 separates these effects from the ideal identity.

3.3 OBSERVATION UPDATES, SCHEDULE, AND CLEAN READOUT

Use the reduced singular-value decomposition $\boldsymbol{A} = \boldsymbol{U}\,\mathrm{diag}(s_i)\boldsymbol{V}^\top$, with rank $r_A \geq 1$, $\boldsymbol{U} \in \mathbb{R}^{m \times r_A}$, $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{r_A}] \in \mathbb{R}^{d \times r_A}$, and $s_i > 0$. Set $\tilde{y}_i = (\boldsymbol{U}^\top \boldsymbol{y})_i / s_i$ and $\tau_i = \sigma_y / s_i$. For fresh $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\xi_i = \boldsymbol{v}_i^\top \boldsymbol{\xi}$, the observation update is

$$\boldsymbol{v}_i^\top \boldsymbol{u} = \begin{cases} (1 - g_i)\boldsymbol{v}_i^\top \boldsymbol{x} + g_i \tilde{y}_i + \eta\sqrt{1 - g_i}\,\xi_i, & \eta \leq \tau_i, \quad g_i = \eta^2/\tau_i^2, \\ \tilde{y}_i + \sqrt{\eta^2 - \tau_i^2}\,\xi_i, & \eta > \tau_i. \end{cases} \tag{7}$$

In the unmeasured subspace, $\boldsymbol{P}_\perp \boldsymbol{u} = \boldsymbol{P}_\perp(\boldsymbol{x} + \eta\boldsymbol{\xi})$, with $\boldsymbol{P}_\perp = \boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^\top$. The first branch is the compensated conditional; the second is a high-noise matching rule related to spectral diffusion solvers (Kawar et al., 2022). Exact-target invariance is asserted only below $\tau_{\min} = \min_i \tau_i$.

Fix the calls per prior update $c \in \{1, 2\}$ using the prior/operator heuristic in Appendix B.3, without using $\boldsymbol{y}$. Resolve the endpoint $\eta_\star$ from $0.95\tau_{\min}$ using the applicable grid rule, retaining $\eta_\star < \tau_{\min}$. The fixed terminal grid $\eta_\star = a_0 > \cdots > a_{n_F} = 0$ requires n_F calls, leaving $L = \lfloor (B - n_F)/c \rfloor$ ordinary sweeps; the reported budgets permit $L \geq 2$. The coupling levels $\eta_0, \ldots, \eta_{L-1}$ follow a geometric decay with a final 10% hold. Appendix B.1 specifies the grids, snapping, and node conventions.

Initialize $\boldsymbol{x}_0 = \boldsymbol{0}$. For $\ell = 0, \ldots, L - 1$, draw $\boldsymbol{u}_\ell$ by Equation (7) at η_ℓ conditional on $\boldsymbol{x}_\ell$, then obtain $\boldsymbol{x}_{\ell+1}$ from the static update in Section 3.1 if $c = 1$, or Equation (4) if $c = 2$. After these sweeps, draw one additional observation state $\boldsymbol{z}$ at $\eta_\star$ conditional on $\boldsymbol{x}_L$. Now $\boldsymbol{z}$ is a state at the current noise level. For each consecutive terminal pair $a > b \geq 0$, set $r = b^2/a^2$ and update

$$\boldsymbol{z} \leftarrow r\boldsymbol{z} + (1 - r)D_\theta(\boldsymbol{z}, a) + \sqrt{a^2 - b^2}\,[r\boldsymbol{I} + (1 - r)\boldsymbol{S}_a]^{1/2}\boldsymbol{e}. \tag{8}$$

This Gaussian transition substitutes D_θ and $\boldsymbol{S}_a$ into the reverse conditional moments under a Brownian-noise coupling (Appendix A). Return the final $\boldsymbol{z}$, without iterate averaging or an additional denoising prediction. The last transition retains noise: $\boldsymbol{z} \leftarrow D_\theta(\boldsymbol{z}, a) + a\boldsymbol{S}_a^{1/2}\boldsymbol{e}$. The descent

uses the static table even after two-call sweeps. Stopping uses $cL + n_F \leq B$, not a convergence test; independent restarts give multiple outputs. Algorithm 1 states the complete ordering. There are $L + 1$ observation updates. The forward-call count includes probes and descent but excludes calibration, routing, and operator preprocessing (Appendix B.2); it is not runtime.

4 What the covariance approximation controls

The analysis concerns the exact Bayes mean unless stated otherwise. It explains the approximation and a restricted stationary effect; it does not establish convergence of the learned, annealed implementation. Full proofs are in Appendix A.

Proposition 1 (Residual covariance)**.** *Fix $\eta > 0$ and $\boldsymbol{R}_\eta \succeq 0$, and use the Bayes mean and exact response $\boldsymbol{J}_\eta(\boldsymbol{u})\boldsymbol{e}$ in Equation* (4)*. Conditional on $\boldsymbol{u}$, the update has mean $\boldsymbol{m}_\eta(\boldsymbol{u})$ and covariance $\eta^2(\boldsymbol{J}_\eta(\boldsymbol{u})^2 + \boldsymbol{R}_\eta)$. With population calibration at this level, without clipping or coefficient pooling, $\boldsymbol{R}_\eta = \operatorname{diag}_{\boldsymbol{W}} \mathbb{E}_{p_\eta} \boldsymbol{\Phi}_\eta$, where $\operatorname{diag}_{\boldsymbol{W}}$ retains the diagonal in basis $\boldsymbol{W}$. If this residual is positive semidefinite, the conditional covariances averaged over $\boldsymbol{u} \sim p_\eta$ match Equation* (1) *on that diagonal. Under the same calibration, a Gaussian prior gives the exact denoising conditional if and only if $\boldsymbol{\Phi}_\eta$ is diagonal in $\boldsymbol{W}$.*

For a Gaussian prior, the denoiser is affine, so its forward difference is exact for any positive δ. The residual can be easier to represent than the full covariance: a rotated two-dimensional Gaussian Jacobian with eigenvalues α and $1-\alpha$, $0 < \alpha < 1/2$, has $\boldsymbol{\Phi}_\eta = \alpha(1-\alpha)\boldsymbol{I}$. The two-call conditional is then exact even when the static diagonal table misses off-diagonal entries of $\boldsymbol{J}_\eta$ (Appendix A.3). This is a representation advantage at one coupling, not a finite-budget guarantee. Exactness of every static descent transition instead requires its table to recover $\boldsymbol{J}_a$ at each queried level; diagonal Gaussian prior covariance in $\boldsymbol{W}$ with population calibration is sufficient. Pooling additionally requires equal coefficients within each pool.

Eigenvalues near zero or one give a small absolute residual, not necessarily small relative error or divergence in directions of tiny conditional variance. For multimodal priors, Jacobian eigenvalues can exceed one; the residual can then be indefinite, and clipping changes the matching identity.

Theorem 1 (Stationary Gaussian output bound)**.** *Fix $\boldsymbol{y}$, let $p = \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \succ 0$, and fix $0 < \eta < \tau_{\min}$. Alternate the compensated observation update with $\boldsymbol{x}' \mid \boldsymbol{u} \sim \mathcal{N}(\boldsymbol{m}_\eta(\boldsymbol{u}), \eta^2 \boldsymbol{S})$, using fixed $\boldsymbol{S} \succeq 0$ and fresh independent noise. Put $\boldsymbol{J} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma} + \eta^2 \boldsymbol{I})^{-1}$ and $\boldsymbol{C}_{\text{post}} = (\boldsymbol{\Sigma}^{-1} + \sigma_y^{-2}\boldsymbol{A}^\top \boldsymbol{A})^{-1}$. The chain has a unique stationary law whose mean is the true posterior mean. If $-\epsilon_- \boldsymbol{J} \preceq \boldsymbol{S} - \boldsymbol{J} \preceq \epsilon_+ \boldsymbol{J}$, with $\epsilon_\pm \geq 0$, then one additional exact observation/prior sweep applied to that law gives an output covariance $\boldsymbol{C}_{\text{out}}$ satisfying*

$$-\tfrac{1}{2}\epsilon_- \boldsymbol{C}_{\text{post}} \preceq \boldsymbol{C}_{\text{out}} - \boldsymbol{C}_{\text{post}} \preceq \tfrac{1}{2}\epsilon_+ \boldsymbol{C}_{\text{post}}. \tag{9}$$

The factor $1/2$ is sharp.

For a Gaussian prior and its Bayes mean, the two-call update has $\boldsymbol{S} = \boldsymbol{J}^2 + \boldsymbol{R}_\eta$, so the theorem's hypothesis measures residual error $\boldsymbol{S} - \boldsymbol{J} = \boldsymbol{R}_\eta - \boldsymbol{\Phi}_\eta$. The exact final sweep remains essential. Static-table descent supplies it when the Bayes mean and $\boldsymbol{S}_a = \boldsymbol{J}_a$ are used at every queried positive level of this Gaussian prior. The theorem neither bounds finite-schedule mixing error nor removes learned-denoiser error. Spending twice as much on a sweep can improve its conditional covariance while making the resulting finite-budget chain worse.

5 Experiments

We test three questions: whether the local response improves uncertainty at fixed online cost, whether the resulting sampler improves held-out inverse-problem results, and when conditional accuracy fails to improve the final samples. Each output is a separately initialized run. Denoiser evaluations (NFE) count forward calls per output, including the terminal descent; they do not measure wall time or backpropagation cost. Unless specified otherwise, intervals are paired, target-level percentile bootstrap intervals with 20,000 resamples and nominal 95% coverage. Appendix B separates selection evidence from final evaluation.

Table 1: **Held-out inverse scattering.** Means across 20 targets; 25 outputs per target. PSNR (dB) evaluates the posterior mean. All NFE include the terminal descent. ICE averages rank-calibration errors over four projection families.

Receivers	Method	NFE ↓	PSNR ↑	ICE ↓
180	**MM-SG**	1,000	40.28	0.0555
180	Full-diffusion SG	3,020	40.10	0.0570
180	Euler SG	1,000	40.23	0.0649
180	Annealed Langevin	1,000	40.31	0.0549
180	Static one-call	1,000	40.06	0.1256
360	**MM-SG**	1,000	41.73	0.0513
360	Full-diffusion SG	3,026	41.53	0.0528
360	Euler SG	1,000	41.64	0.0562

Comparators and metrics. The full-diffusion split-Gibbs reference runs a reverse solver for its prior conditional (Appendix B.5). It is a local reference in the family of Wu et al. (2024), not a reproduction of that paper's full protocol. Euler split Gibbs and annealed Langevin provide cheaper alternatives. Our internal ablations share the observation update and terminal descent. We report ensemble-mean peak signal-to-noise ratio (PSNR; higher is better), projected-rank integrated calibration error (ICE; lower is better), and raw ensemble energy score (ES; lower is better). Energy score evaluates both location and dispersion (Gneiting & Raftery, 2007); ensemble-mean PSNR alone can favor concentrated ensembles at small sample counts. The rank statistic probes specified linear functionals, not the full joint posterior. Appendix B.7 defines all estimators and their limitations.

5.1 Held-out inverse scattering: reconstruction and uncertainty

We use the linear inverse-scattering task from InverseBench (Zheng et al., 2025): 128×128 permittivity images, a pretrained EDM prior (Karras et al., 2022), and noisy measurements from 20 transmitters. The final cohort contains 20 held-out targets with 25 outputs per target. The 180-receiver configuration is followed by transfer to 360 receivers without retuning. The finite-difference update and its calibration are frozen before these final measurements.

Table 1 shows that MM-SG improves PSNR over the full-diffusion reference by 0.180 dB $[0.111, 0.255]$ at 180 receivers while using 1,000 rather than 3,020 evaluations. The paired ICE difference is -0.0014 $[-0.0079, 0.0050]$: calibration is comparable, not demonstrably better. Against the static one-call update at the same budget, the gain is 0.222 dB $[0.193, 0.253]$ and ICE decreases by 0.0701 $[0.0610, 0.0790]$. The latter comparison directly supports the value of the local response within this sampler.

At 360 receivers, the PSNR advantage over the full-diffusion reference is 0.197 dB $[0.127, 0.280]$, again with comparable ICE (difference -0.0015 $[-0.0073, 0.0054]$). The result survives this operator change without output-based retuning. It is not an overall dominance result: at 180 receivers, annealed Langevin has slightly higher average PSNR and slightly lower ICE, with the PSNR difference from MM-SG unresolved by the paired interval. Comparisons between these sampler families also change their state representation or update placement; internal ablations are the cleaner evidence about the covariance construction.

5.2 Which part of the update matters?

We compare the static update, the full two-call update, an internal response-power control, and the raw response without residual noise. The power control replaces the response by an independent centered Gaussian probe shaped to its radial-bin power, retaining the same residual table. It removes directional information and the response's possible conditional mean shift, while changing higher-order structure. Thus it tests the finite response as a whole, not direction alone, and is not an implementation of Schioppa & Salimans (2026).

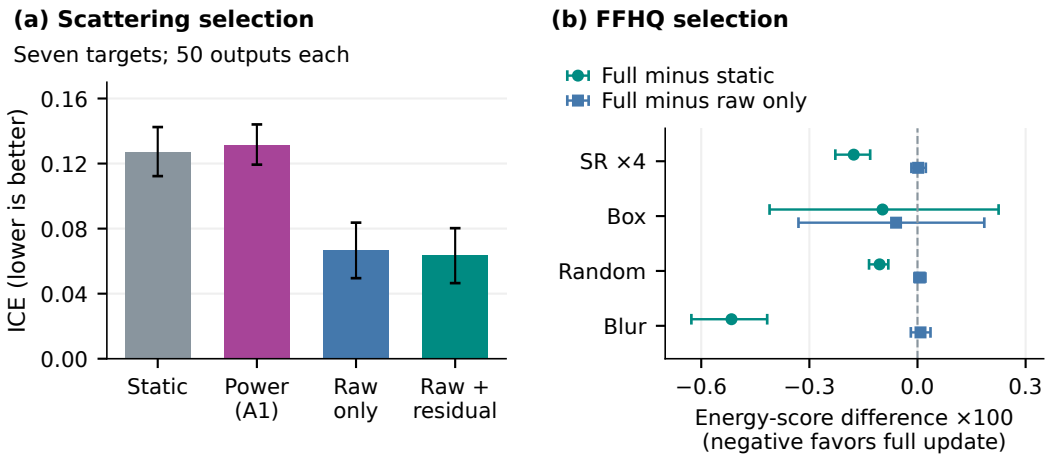


Figure 2: **Raw-response and residual ablations at fixed budgets.** Left: scattering selection cohort, seven targets and 50 outputs per target; bars show mean ICE with target-bootstrap intervals. The scattering variants use 1,000 forward evaluations. A1 is the internal response-power control. Retaining the raw response reduces ICE relative to the static and response-power controls. Right: FFHQ selection cohort, 15 images and four outputs per image; paired differences in raw energy score, multiplied by 100. Negative favors the full update. The residual correction has no resolved advantage over the raw response alone. Intervals are descriptive 95% intervals, with no multiple-comparison correction.

On scattering selection targets, ICE falls from 0.1267 for the static update and 0.1315 for the power control to 0.0636 for the full update (Figure 2); raw-only gives 0.0665. The residual is motivated by covariance completion and exact ideal conditionals in the representable Gaussian case (Proposition 1), not a demonstrated need or advantage over raw-only. Raw-only retains static-table terminal descent; held-out results test the full construction, not the residual separately.

Exact-reference controls explain why an extra call need not help (Appendix C.3, Figure 9). For a heteroscedastic Gaussian mixture with a rotated observation operator, the maximum reference-corrected band log-variance error decreases from 0.0673 to 0.0106. A matched-half-schedule exact-conditional control gives 0.0115. With milder covariance heterogeneity, however, the static update is already accurate: its error is 0.0085, compared with 0.0160 for the two-call update. In a separated mixture, the two-call update also worsens the responsibility discrepancy; the corresponding exact-conditional control with fewer sweeps exhibits similar deterioration. These controls support a conditional-accuracy versus mixing tradeoff, not monotonic improvement from additional computation within a sweep.

5.3 Held-out natural images and selection ablations

We evaluate four restoration tasks on FFHQ (Karras et al., 2019) with a pretrained 256×256 ADM denoiser (Dhariwal & Nichol, 2021) and measurement-noise standard deviation 0.05 on the $[-1, 1]$ scale. Per-output perceptual error uses VGG LPIPS (Zhang et al., 2018). Baseline settings are frozen after tuning on 15 separate images. The held-out evaluation uses the same 60 image IDs across four tasks, with four independent outputs per image and method. This is the largest complete ID prefix at the fixed 26 September 2026 noon KST cutoff, selected by record existence rather than scores. It is not the full planned 100-image confirmation (Appendix B.6).

Table 2 compares MM-SG with full-diffusion SG and the six fast adaptations DPS, DAPS, RED-diff, DiffPIR, DDNM+, and pseudoinverse-guided diffusion (Chung et al., 2023; Zhang et al., 2025; Mardani et al., 2024; Zhu et al., 2023; Wang et al., 2023; Song et al., 2023). At 999 evaluations, MM-SG improves PSNR over full-diffusion SG by 0.205, 1.260, and 0.511 dB on super-resolution, box inpainting, and random masking, respectively; energy score and per-sample perceptual error also improve. On deblurring, all three metrics favor full-diffusion SG. The paired intervals exclude zero in these directions (Figure 3; Table 3). Faster methods remain competitive: RED-diff has the

Table 2: **Held-out FFHQ: all eight methods on the 60-image prefix.** Cells report ensemble-mean PSNR ↑ / raw energy score ↓ multiplied by 100. SR: 4× super-resolution; box: central 128 × 128 inpainting; random: 70% missing pixels; blur: Gaussian deblurring. NFE are actual forward evaluations per output, including descent; †: 3,016 for SR and 3,000 otherwise. Offline calibration and backward-pass costs are excluded. Baseline adaptations are in Appendix B.5; perceptual scores are in Table 4.

Method	NFE	SR	Box	Random	Blur
MM-SG	999	29.44 / 2.42	25.76 / 3.59	32.88 / 1.52	29.22 / 2.66
Full-diffusion SG	3,000†	29.23 / 2.46	24.50 / 4.10	32.36 / 1.61	29.29 / 2.58
DPS	1,000	27.44 / 2.99	25.29 / 3.76	32.33 / 1.68	27.51 / 2.89
DAPS-1K	1,000	29.32 / 2.57	25.57 / 3.81	31.73 / 1.78	29.26 / 2.69
RED-diff	1,000	27.77 / 3.40	25.83 / 4.43	32.45 / 1.87	28.36 / 3.35
DiffPIR	100	28.02 / 2.69	23.73 / 4.36	27.75 / 2.72	28.69 / 2.59
DDNM+	100	27.32 / 3.98	25.53 / 4.11	33.09 / 1.88	29.50 / 2.73
ΠGDM	100	28.88 / 2.51	25.61 / 3.63	32.50 / 1.62	28.86 / 2.50

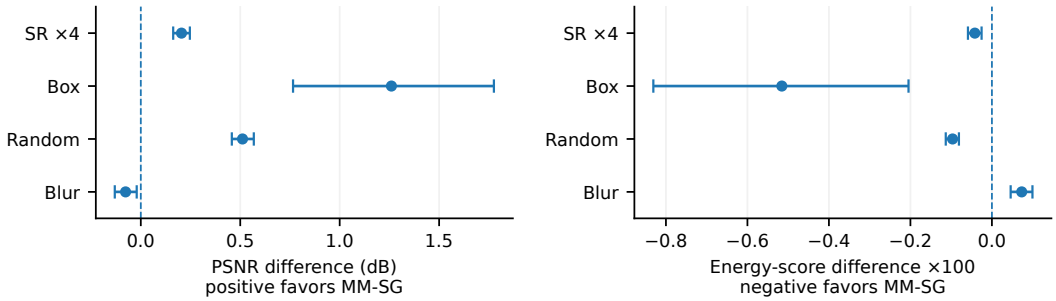


Figure 3: **Held-out paired effects versus full-diffusion SG.** Each point is a mean difference across 60 shared image IDs; bars are central 95% paired bootstrap intervals (20,000 resamples). Positive PSNR (dB) and negative energy-score differences (multiplied by 100) favor MM-SG. Both improve on three tasks and worsen on blur. Tasks are analyzed separately, without multiplicity adjustment; four tasks do not create 240 independent image targets. Appendix Figures 5–6 show all comparators.

highest mean box PSNR, DDNM+ leads random-mask and blur PSNR, and pseudoinverse-guided diffusion has the lowest blur energy score.

The internally specified noninferiority criteria pass on this prefix, including the deblurring PSNR margin; they do not imply equivalence or superiority on every task. The energy-score rule uses a mean of per-image ratios against the lowest-mean-score fast comparator on the same cohort, not a ratio of the table means. The whitened measurement-residual diagnostic ranges from 0.998 to 1.026. It is distinct from the update's additive covariance residual and does not establish full posterior calibration (Appendix B.6).

Selection ablations and budgets. The separate 15-image selection cohort remains the evidence for component attribution and budget dependence (Table 6; Figure 2). Against the static one-call control, MM-SG improves energy score on super-resolution, random inpainting, and deblurring, with paired intervals excluding zero; box inpainting is unresolved. The PSNR differences are -0.02, -0.46, -0.21, and $+0.41$ dB in task order. Removing the residual yields nearly identical scores, with every paired interval containing zero. These controls are absent from the held-out prefix. Increasing MM-SG from 249 to 999 evaluations improves selection random-inpainting energy score but worsens deblurring (Appendix C). These results support inexpensive conditionals, not monotonic budget gains or state-of-the-art restoration.

5.4 Failures that delimit the claim

The final Gaussian suite fails its joint criterion: at 3,000 evaluations, the one-call branch passes five spectra, short of six of seven (Appendix C.4). One failure is consistent with a basis-approximation

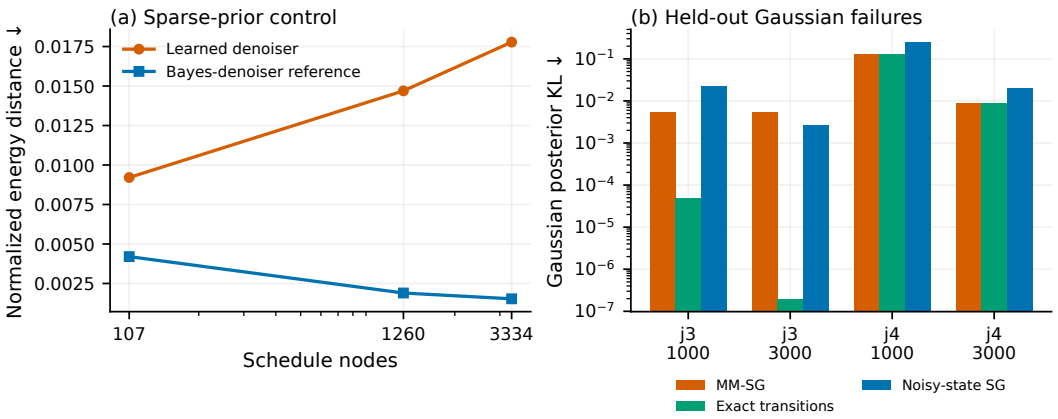


Figure 4: Two distinct failure mechanisms. Left: a development control with a learned sparse-signal denoiser versus a lattice Bayes-denoiser reference, using the same sampler family but separately calibrated tables (Appendix C). Recorded normalized energy distances are averages over two run seeds at 107 and 1,260 schedule nodes and one at 3,334; no confidence intervals are inferred from these few repeats. Right: analytic Gaussian output divergences, averaged over measurements, for two final spectra. j3 exposes an approximation floor; j4 also fails with exact conditional draws. The panels use different priors, metrics, and abscissae and should not be compared numerically.

floor; another persists with exact conditionals and is dominated by finite-schedule error. With the Bernoulli–Laplace prior of Zach et al. (2026), more evaluations worsen learned-denoiser normalized energy distance but improve the numerical Bayes reference (Figure 4). More sweeps cannot be assumed to repair a misspecified denoiser. The frozen prior/operator diagnostic selects two calls on all image tasks but is not a reliable winner predictor across controlled cells: it remains a deployment heuristic, not a posterior-accuracy certificate.

6 Discussion

The local-response update supplies input-dependent perturbations without a full reverse solve at every sweep. With fewer online forward evaluations than our full-diffusion reference, it improves held-out scattering reconstruction with comparable rank calibration, and natural-image reconstruction and energy score except in deblurring. Selection controls support retaining the raw response as a whole, not a separate residual benefit; the table approximates, rather than exactly recovers, the missing covariance.

The ideal covariance interpretation and the implemented finite-step response have different scopes. Fixed development probes show disagreement with automatic-differentiation responses at the implemented step size (Appendix C.6); finite differences can also shift the conditional mean (Appendix A.3). Pooling and clipping also change population matching. The stationary Gaussian bound requires a fixed admissible coupling and an exact final observation/prior sweep. Static descent realizes that sweep in the representable Gaussian setting with the true mean and correct static covariance at every queried positive level. The finite-schedule Gaussian failures and adverse learned-prior budget curve delimit the empirical benefit separately from this theorem.

The natural-image evidence covers the fixed-cutoff complete prefix, not the planned full confirmation, and its component ablations remain selection evidence. Energy scores and projected ranks assess particular functionals; the measurement-residual diagnostic does not establish full posterior calibration. A faithful matched-budget comparison with the closest covariance-aware and matrix-free square-root samplers is absent. The forward-call accounting also excludes offline calibration and is not a measured-runtime comparison.

Retaining a local response supplies input-adaptive perturbations at low conditional forward cost, leaving more outer sweeps than a full reverse solve.

ETHICS STATEMENT

The experiments concern synthetic inverse problems and restoration of images from an existing face dataset. The method produces plausible samples conditioned on a learned prior; these samples must not be interpreted as verified details of an individual or a physical scene. Face restoration inherits representation biases and privacy concerns from its training data. The reported uncertainty diagnostics do not establish reliability for clinical, forensic, or other consequential decisions.

REPRODUCIBILITY STATEMENT

Appendices A–C provide proofs, the complete update ordering, calibration details, benchmark protocols, budget accounting, and additional results. The accompanying manuscript source includes the numerical aggregates, plotting scripts for quantitative figures and tables, and editable vector source for the overview. The experiment package provides sampler code and per-target records. The supplied snapshot does not bundle the external pretrained checkpoints, original image datasets, or calibration arrays; regenerating all sample ensembles therefore requires those dependencies. Rebuilding the manuscript and its plots does not require them.

AI USE STATEMENT

During the preparation of this manuscript, the authors used AI tools to assist with language editing, including grammar, clarity, and readability, as well as with the preparation and refinement of selected figures, including figure layout and visual organization. All AI-assisted outputs were carefully reviewed, verified, and edited by the authors. AI tools were not used to generate or analyze research data, derive results, or formulate scientific interpretations or conclusions.

REFERENCES

Md Sahil Akhtar, Aymane El Gadarri, Vivek F. Farias, and Adam D. Jozefiak. The value of covariance matching in Gaussian DDPMs and the Lanczos sampler. *arXiv preprint arXiv:2605.22723*, 2026.

Fan Bao, Chongxuan Li, Jun Zhu, and Bo Zhang. Analytic-DPM: An analytic estimate of the optimal reverse variance in diffusion probabilistic models. In *International Conference on Learning Representations*, 2022.

Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and O. Deniz Akyildiz. Tweedie moment projected diffusions for inverse problems. *Transactions on Machine Learning Research*, 2024.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. In *Advances in Neural Information Processing Systems*, 2022.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Florentin Coeurdoux, Nicolas Dobigeon, and Pierre Chainais. Plug-and-play split Gibbs sampler: Embedding deep generative priors in Bayesian inference. *IEEE Transactions on Image Processing*, 33:3496–3507, 2024.

Prafulla Dhariwal and Alex Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, 2021.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *International Conference on Learning Representations*, 2024.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Zahra Kadkhodaie and Eero P. Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. In *Advances in Neural Information Processing Systems*, 2021.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, 2022.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, 2022.

Rémi Laumont, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian imaging using plug & play priors: When Langevin meets Tweedie. *SIAM Journal on Imaging Sciences*, 15(2):701–737, 2022.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using denoising diffusion probabilistic models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Hila Manor and Tomer Michaeli. On the posterior distribution in denoising: Application to uncertainty quantification. In *International Conference on Learning Representations*, 2024.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *International Conference on Learning Representations*, 2024.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, 2021.

Xinyu Peng, Ziyang Zheng, Wenrui Dai, Nuoqian Xiao, Chenglin Li, Junni Zou, and Hongkai Xiong. Improving diffusion models for inverse problems using optimal posterior covariance. In *International Conference on Machine Learning*, 2024.

Severi Rissanen, Markus Heinonen, and Arno Solin. Free Hunch: Denoiser covariance estimation for diffusion models without extra costs. In *International Conference on Learning Representations*, 2025.

François Rozet, Gérôme Andry, François Lanusse, and Gilles Louppe. Learning diffusion priors from observations by expectation maximization. In *Advances in Neural Information Processing Systems*, 2024.

Andrea Schioppa and Tim Salimans. Covariance-aware sampling for diffusion models. *arXiv preprint arXiv:2605.13910*, 2026.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, 2019.

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.

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

Maxime Vono, Nicolas Dobigeon, and Pierre Chainais. Split-and-augmented Gibbs sampler—application to large-scale inference problems. *IEEE Transactions on Signal Processing*, 67(6): 1648–1661, 2019.

Weizhou Wang, Jonathan Weare, and Aaron R. Dinner. Composing diffusion priors with explicit physical context via generative Gibbs sampling. In *Advances in Neural Information Processing Systems*, 2026. To appear.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2023.

Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, 2023.

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L. Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. In *Advances in Neural Information Processing Systems*, 2024.

Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. In *Advances in Neural Information Processing Systems*, 2024.

Martin Zach, Youssef Haouchat, and Michael Unser. A statistical benchmark for diffusion-posterior-sampling algorithms. In *International Conference on Learning Representations*, 2026.

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.

Mingtian Zhang, Alex Hawkins-Hooker, Brooks Paige, and David Barber. Moment matching denoising Gibbs sampling. In *Advances in Neural Information Processing Systems*, 2023.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy T. Feng, Caifeng Zou, Yu Sun, Nikola Kovachki, Zachary E. Ross, Katherine L. Bouman, and Yisong Yue. InverseBench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *International Conference on Learning Representations*, 2025.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (NTIRE)*, 2023.

A Derivations and proofs

Vectors are column vectors, $\preceq$ is the Loewner order, and $\|\cdot\|$ denotes the Euclidean or induced operator norm; $\|\cdot\|_F$ is the Frobenius norm. For a symmetric matrix $\boldsymbol{Y}$, define $\operatorname{diag}_{\boldsymbol{W}} \boldsymbol{Y} = \boldsymbol{W} \operatorname{diag}(\boldsymbol{w}_k^\top \boldsymbol{Y} \boldsymbol{w}_k) \boldsymbol{W}^\top$. The theoretical denoiser is the exact Bayes mean unless a learned denoiser is explicitly considered. The conditional identities below hold for any Borel probability prior on $\mathbb{R}^d$; positive-definite Gaussian covariance is imposed for Theorem 1. All population calibration expectations use the noisy-prior law p_η, not the measurement-conditioned chain law. The basis and tables are fixed when drawing online probes.

A.1 The compensated augmentation

For $0 < \eta < \tau_{\min}$, consider

$$\pi_\eta(d\boldsymbol{x}, d\boldsymbol{u} \mid \boldsymbol{y}) \;\propto\; p(d\boldsymbol{x})\, \mathcal{N}(\boldsymbol{u}; \boldsymbol{x}, \eta^2 \boldsymbol{I})\, \mathcal{N}(\boldsymbol{y}; \boldsymbol{A}\boldsymbol{u}, \sigma_y^2 \boldsymbol{I} - \eta^2 \boldsymbol{A}\boldsymbol{A}^\top)\, d\boldsymbol{u}. \tag{10}$$

The last covariance is positive definite. Integrating out $\boldsymbol{u}$ adds $\eta^2 \boldsymbol{A}\boldsymbol{A}^\top$ to it, recovering $\mathcal{N}(\boldsymbol{y}; \boldsymbol{A}\boldsymbol{x}, \sigma_y^2 \boldsymbol{I})$ and hence the desired marginal posterior. In a measured singular direction, write $x_i = \boldsymbol{v}_i^\top \boldsymbol{x}$ and $u_i = \boldsymbol{v}_i^\top \boldsymbol{u}$. The conditional density in u_i is proportional to

$$\exp\left[-\frac{(u_i - x_i)^2}{2\eta^2} - \frac{(\tilde{y}_i - u_i)^2}{2(\tau_i^2 - \eta^2)} \right].$$

Its precision is $\eta^{-2} + (\tau_i^2 - \eta^2)^{-1}$, its variance is $\eta^2(1 - g_i)$, and its mean is $(1 - g_i)x_i + g_i \tilde{y}_i$, with $g_i = \eta^2/\tau_i^2$. Null-space coordinates retain mean $\boldsymbol{x}$ and variance $\eta^2 \boldsymbol{I}$. Thus

$$\boldsymbol{u} \mid \boldsymbol{x}, \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{F}\boldsymbol{x} + \boldsymbol{b}, \eta^2 \boldsymbol{F}), \quad \boldsymbol{F} = \boldsymbol{I} - \eta^2 \sigma_y^{-2} \boldsymbol{A}^\top \boldsymbol{A}, \quad \boldsymbol{b} = \eta^2 \sigma_y^{-2} \boldsymbol{A}^\top \boldsymbol{y}. \tag{11}$$

Equivalently, $\boldsymbol{F} = \boldsymbol{I} - \sum_i g_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$ and $\boldsymbol{b} = \sum_i g_i \tilde{y}_i \boldsymbol{v}_i$. Here $0 \prec \boldsymbol{F} \preceq \boldsymbol{I}$. The conditional of $\boldsymbol{x}$ given $\boldsymbol{u}$ is proportional to $p(d\boldsymbol{x})\mathcal{N}(\boldsymbol{u}; \boldsymbol{x}, \eta^2 \boldsymbol{I})$ and does not otherwise involve $\boldsymbol{y}$. Exact alternating conditionals therefore preserve the posterior. As in Method, $\boldsymbol{A}$ has at least one measured direction, so $\tau_{\min}$ is defined; unmeasured directions are permitted. At $\eta = \tau_{\min}$ the last covariance in Equation (10) is singular, so the stated density argument uses the strict window. The high-noise matching branch is not justified by this argument. Compensation is inherited, not a new component of the method.

A.2 Denoising moments and the descent

Let $\varphi(\boldsymbol{u}, \boldsymbol{x}) = \exp[-\|\boldsymbol{u} - \boldsymbol{x}\|^2 / (2\eta^2)]$ and $Z(\boldsymbol{u}) = \int \varphi(\boldsymbol{u}, \boldsymbol{x}) p(d\boldsymbol{x}) > 0$. Then $p(d\boldsymbol{x} \mid \boldsymbol{u}) = Z(\boldsymbol{u})^{-1} \varphi(\boldsymbol{u}, \boldsymbol{x}) p(d\boldsymbol{x})$. Every polynomial in $\boldsymbol{x}$ multiplied by φ is bounded for fixed $\boldsymbol{u}$, locally uniformly after differentiation in $\boldsymbol{u}$. Thus all conditional moments are finite, and differentiation under the integral is valid without moments of the unconditioned prior. Since $\partial_{u_j} \varphi = \eta^{-2}(x_j - u_j)\varphi$, the quotient rule gives

$$\begin{aligned} \partial_{u_j} m_{\eta,i}(\boldsymbol{u}) &= \eta^{-2} \big(\mathbb{E}[x_i (x_j - u_j) \mid \boldsymbol{u}] - \mathbb{E}[x_i \mid \boldsymbol{u}] \mathbb{E}[x_j - u_j \mid \boldsymbol{u}] \big) \\ &= \eta^{-2} \big(\mathbb{E}[x_i x_j \mid \boldsymbol{u}] - \mathbb{E}[x_i \mid \boldsymbol{u}] \mathbb{E}[x_j \mid \boldsymbol{u}] \big). \end{aligned} \tag{12}$$

Consequently $\boldsymbol{J}_\eta$ is symmetric positive semidefinite and Equation (1) follows.

For a common Brownian-noise coupling, write $\boldsymbol{x}_a = \boldsymbol{x} + \boldsymbol{B}_{a^2}$ and $\boldsymbol{x}_b = \boldsymbol{x} + \boldsymbol{B}_{b^2}$, with $a > b \geq 0$ and a d-dimensional Brownian motion independent of $\boldsymbol{x}$. Given $(\boldsymbol{x}, \boldsymbol{x}_a = \boldsymbol{z})$, the Brownian bridge has mean $r\boldsymbol{z} + (1 - r)\boldsymbol{x}$ and covariance $b^2(1 - r)\boldsymbol{I}$, where $r = b^2/a^2$. Total expectation and covariance give

$$\mathbb{E}[\boldsymbol{x}_b \mid \boldsymbol{x}_a = \boldsymbol{z}] = r\boldsymbol{z} + (1 - r)\boldsymbol{m}_a(\boldsymbol{z}), \tag{13}$$

$$\begin{aligned} \operatorname{Cov}(\boldsymbol{x}_b \mid \boldsymbol{x}_a = \boldsymbol{z}) &= b^2(1 - r)\boldsymbol{I} + (1 - r)^2 a^2 \boldsymbol{J}_a(\boldsymbol{z}) \\ &= (a^2 - b^2)\big[r\boldsymbol{I} + (1 - r)\boldsymbol{J}_a(\boldsymbol{z}) \big]. \end{aligned} \tag{14}$$

Equation (8) substitutes D_θ and the static table into these moments. For a Gaussian prior, these reverse conditionals are Gaussian, so the substituted transition is exact when both $D_\theta(\cdot, a) = \boldsymbol{m}_a$

and $\boldsymbol{S}_a = \boldsymbol{J}_a$ at the queried level. Diagonal Gaussian prior covariance in $\boldsymbol{W}$ with unpooled population calibration at every queried level is sufficient. Finite-grid interpolation is not itself an exactness argument. The forward Brownian process is Markov, so composing its exact reverse conditionals down to zero gives $p(\boldsymbol{x} \mid \boldsymbol{u})$. At $b = 0$, Equation (14) retains covariance $a^2\boldsymbol{J}_a$, not zero. For non-Gaussian priors, these two moments do not determine the reverse conditional.

A.3 Residual covariance and finite differences

Proof of Proposition 1. Fix $\boldsymbol{u}$. The exact-response update is $\boldsymbol{m}_\eta(\boldsymbol{u}) + \eta[\boldsymbol{J}_\eta(\boldsymbol{u})\boldsymbol{e} + \boldsymbol{R}_\eta^{1/2}\boldsymbol{e}']$. Independence and zero means of $\boldsymbol{e}, \boldsymbol{e}' \sim \mathcal{N}(0, \boldsymbol{I})$ give conditional mean $\boldsymbol{m}_\eta(\boldsymbol{u})$ and covariance $\eta^2(\boldsymbol{J}_\eta\boldsymbol{J}_\eta^\top + \boldsymbol{R}_\eta)$; symmetry gives $\boldsymbol{J}_\eta^2$. The population denoising-error estimate in Equation (2) satisfies

$$\eta^{-2}\mathbb{E}[(\boldsymbol{w}_k^\top(\boldsymbol{m}_\eta(\boldsymbol{u}) - \boldsymbol{x}))^2] = \mathbb{E}_{p_\eta}[\boldsymbol{w}_k^\top \boldsymbol{J}_\eta(\boldsymbol{u})\boldsymbol{w}_k], \tag{15}$$

by conditioning on $\boldsymbol{u} = \boldsymbol{x} + \eta\boldsymbol{e}_0$, where $\boldsymbol{x} \sim p$ and $\boldsymbol{e}_0 \sim \mathcal{N}(0, \boldsymbol{I})$ are independent. The probe $\boldsymbol{e}$ is independent of this corruption. Its power is $\mathbb{E}_{\boldsymbol{e}}[(\boldsymbol{w}_k^\top \boldsymbol{J}_\eta(\boldsymbol{u})\boldsymbol{e})^2] = \boldsymbol{w}_k^\top \boldsymbol{J}_\eta(\boldsymbol{u})^2\boldsymbol{w}_k$. Subtracting the population powers gives $\boldsymbol{R}_\eta = \operatorname{diag}_{\boldsymbol{W}} \mathbb{E}_{p_\eta}(\boldsymbol{J}_\eta - \boldsymbol{J}_\eta^2)$. If this matrix is positive semidefinite, it can be sampled as the independent residual, and

$$\operatorname{diag}_{\boldsymbol{W}} \mathbb{E}_{p_\eta}\big[\boldsymbol{J}_\eta(\boldsymbol{u})^2 + \boldsymbol{R}_\eta\big] = \operatorname{diag}_{\boldsymbol{W}} \mathbb{E}_{p_\eta}\boldsymbol{J}_\eta(\boldsymbol{u}). \tag{16}$$

This is an average of conditional covariances, not a pointwise match or a claim about unconditional moments of $\boldsymbol{x}$.

These averages exist even for priors without finite moments. Conditional on $\boldsymbol{u}$, $\boldsymbol{x} = \boldsymbol{u} - \eta\boldsymbol{e}_0$ implies $\boldsymbol{J}_\eta(\boldsymbol{u}) = \operatorname{Cov}(\boldsymbol{e}_0 \mid \boldsymbol{u})$. Hence

$$\mathbb{E}\boldsymbol{J}_\eta = \boldsymbol{I} - \mathbb{E}\big[\mathbb{E}[\boldsymbol{e}_0 \mid \boldsymbol{u}]\mathbb{E}[\boldsymbol{e}_0 \mid \boldsymbol{u}]^\top\big] \preceq \boldsymbol{I}.$$

Since $\boldsymbol{J}_\eta \succeq 0$, conditional Jensen gives

$$\|\boldsymbol{J}_\eta\|_F^2 \le (\operatorname{tr}\boldsymbol{J}_\eta)^2 \le \big(\mathbb{E}[\|\boldsymbol{e}_0\|^2 \mid \boldsymbol{u}]\big)^2 \le \mathbb{E}[\|\boldsymbol{e}_0\|^4 \mid \boldsymbol{u}].$$

Taking expectations bounds the last quantity by $d(d+2)$, establishing integrability of $\boldsymbol{J}_\eta^2$ and the subtraction above.

For $p = \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$, the mean is $\boldsymbol{m}_\eta(\boldsymbol{u}) = \boldsymbol{m} + \boldsymbol{J}(\boldsymbol{u} - \boldsymbol{m})$, with $\boldsymbol{J} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma} + \eta^2\boldsymbol{I})^{-1}$. Its forward difference is exactly $\boldsymbol{J}\boldsymbol{e}$ for every $\delta > 0$. Eigenvalues j of $\boldsymbol{J}$ lie in $[0, 1)$, and $j(1-j) \in [0, 1/4]$. Thus $\boldsymbol{\Phi} = \boldsymbol{J} - \boldsymbol{J}^2 \succeq 0$ and the ideal population clipping is inactive. The update is Gaussian with the correct mean and covariance $\eta^2[\boldsymbol{J} + \operatorname{diag}_{\boldsymbol{W}}\boldsymbol{\Phi} - \boldsymbol{\Phi}]$. It equals the Gaussian conditional exactly if and only if $\operatorname{diag}_{\boldsymbol{W}}\boldsymbol{\Phi} = \boldsymbol{\Phi}$. This part also allows singular $\boldsymbol{\Sigma} \succeq 0$; the stationary theorem does not. □

A representation advantage at one coupling. In two table coordinates, let $\boldsymbol{Q}_\omega$ be a planar rotation through $0 < \omega < \pi/2$ and set

$$\boldsymbol{J} = \boldsymbol{Q}_\omega \operatorname{diag}(\alpha, 1-\alpha)\boldsymbol{Q}_\omega^\top, \qquad 0 < \alpha < 1/2.$$

Then $J_{12} = (2\alpha - 1)\sin\omega\cos\omega \neq 0$, but $\boldsymbol{\Phi} = \alpha(1-\alpha)\boldsymbol{I}$. The positive-definite covariance $\boldsymbol{\Sigma} = \eta^2\boldsymbol{J}(\boldsymbol{I} - \boldsymbol{J})^{-1}$ realizes this Jacobian, because $\boldsymbol{\Sigma} + \eta^2\boldsymbol{I} = \eta^2(\boldsymbol{I} - \boldsymbol{J})^{-1}$. Population residual calibration is therefore exact, whereas the static table $\operatorname{diag}_{\boldsymbol{W}}\boldsymbol{J}$ is not. Pooling these two equal residual coefficients preserves exactness. The example fixes one coupling: it does not make the static table exact at this or other descent levels, and does not imply improved finite-budget sampling.

Pooling, clipping, and calibration error. Let $\mathcal{G}$ partition the basis indices into nonempty groups. Define $\mathcal{P}_{\mathcal{G}}(\boldsymbol{Y}) = \boldsymbol{W}\operatorname{diag}(t_k)\boldsymbol{W}^\top$, where t_k is the average of $\boldsymbol{w}_j^\top\boldsymbol{Y}\boldsymbol{w}_j$ over the group containing k. Without clipping, exact-response population calibration then yields $\boldsymbol{R}_\eta = \mathcal{P}_{\mathcal{G}}(\mathbb{E}\boldsymbol{\Phi}_\eta)$. Idempotence of $\mathcal{P}_{\mathcal{G}}$ gives

$$\mathcal{P}_{\mathcal{G}}\big(\mathbb{E}[\boldsymbol{J}_\eta^2 + \boldsymbol{R}_\eta]\big) = \mathcal{P}_{\mathcal{G}}(\mathbb{E}\boldsymbol{J}_\eta).$$

Only the pooled averages are matched. For a Gaussian prior, conditional exactness requires $\mathcal{P}_{\mathcal{G}}(\boldsymbol{\Phi}) = \boldsymbol{\Phi}$: no off-diagonal residual entries and equal diagonal residual coefficients within each group. Static-update and static-descent exactness instead require representation of $\boldsymbol{J}$, not merely $\boldsymbol{\Phi}$.

Clipping ensures a valid additive covariance, but cannot subtract excess raw-response covariance. For example, a scalar prior equally supported on $\{-h, h\}$ has $m_\eta(u) = h\tanh(hu/\eta^2)$ and $J_\eta(0) = h^2/\eta^2$. If $h > \eta$, then $J_\eta(0)^2 > J_\eta(0)$ and no nonnegative residual can correct that local excess. This does not contradict the averaged identity under its stated positive-semidefiniteness condition. In the image implementation, both estimated powers are clipped before their nonnegative difference is formed at calibration nodes; pooling, clipping, and interpolation need not preserve the population identity.

There is a separate learned-mean contribution to the denoising-error table. For $\boldsymbol{b}_\eta(\boldsymbol{u}) = D_\theta(\boldsymbol{u}, \eta) - \boldsymbol{m}_\eta(\boldsymbol{u})$ and finite denoising risk under the true prior, conditional centering gives

$$\eta^{-2}\mathbb{E}[(\boldsymbol{w}_k^\top(D_\theta(\boldsymbol{u}, \eta) - \boldsymbol{x}))^2] = \mathbb{E}[\boldsymbol{w}_k^\top \boldsymbol{J}_\eta(\boldsymbol{u})\boldsymbol{w}_k] + \eta^{-2}\mathbb{E}[(\boldsymbol{w}_k^\top \boldsymbol{b}_\eta(\boldsymbol{u}))^2]. \tag{17}$$

Thus even population denoising-error power need not equal Bayes conditional variance for a learned denoiser. If the directional prior variance v_k is finite, estimating $\boldsymbol{w}_k^\top \boldsymbol{x}$ from $\boldsymbol{w}_k^\top \boldsymbol{u}$ by scalar linear shrinkage about its prior mean with coefficient $\beta = v_k/(v_k + \eta^2)$ has risk $(1-\beta)^2 v_k + \beta^2\eta^2 = \eta^2 v_k/(v_k + \eta^2)$. The Bayes estimator using all of $\boldsymbol{u}$ has no larger risk, so $\mathbb{E}[\boldsymbol{w}_k^\top \boldsymbol{J}_\eta(\boldsymbol{u})\boldsymbol{w}_k] \le v_k/(v_k+\eta^2)$. This is an averaged bound, not a bound at every $\boldsymbol{u}$. Using estimated v_k to cap a learned-denoiser table is a stabilization rule. Finite samples, rescaled calibration noise, and model-generated rather than true-prior samples introduce further differences from the ideal calibration experiment.

The implemented forward difference. At fixed $(\boldsymbol{u}, \eta)$, write $\boldsymbol{\mu}_\delta = \mathbb{E}_{\boldsymbol{e}}\boldsymbol{G}_\delta$ and $\boldsymbol{V}_\delta = \mathrm{Cov}_{\boldsymbol{e}}(\boldsymbol{G}_\delta)$ whenever these moments are finite. Independence of the residual probe gives the exact conditional moments

$$\begin{aligned} \mathbb{E}[\boldsymbol{x}' \mid \boldsymbol{u}] &= D_\theta(\boldsymbol{u}, \eta) + \eta\boldsymbol{\mu}_\delta, & \mathrm{Cov}(\boldsymbol{x}' \mid \boldsymbol{u}) &= \eta^2(\boldsymbol{V}_\delta + \boldsymbol{R}_\eta), \\ \mathbb{E}_{\boldsymbol{e}}[\boldsymbol{G}_\delta \boldsymbol{G}_\delta^\top] &= \boldsymbol{V}_\delta + \boldsymbol{\mu}_\delta\boldsymbol{\mu}_\delta^\top. \end{aligned} \tag{18}$$

The response calibration uses this uncentered power, not just $\boldsymbol{V}_\delta$. Even with the Bayes mean, unclipped unpooled population subtraction using the finite difference would give

$$\mathrm{diag}_{\boldsymbol{W}}\, \mathbb{E}_{p_\eta}[\boldsymbol{V}_\delta + \boldsymbol{R}_\eta] = \mathrm{diag}_{\boldsymbol{W}}\, \mathbb{E}_{p_\eta}\boldsymbol{J}_\eta - \mathrm{diag}_{\boldsymbol{W}}\, \mathbb{E}_{p_\eta}[\boldsymbol{\mu}_\delta\boldsymbol{\mu}_\delta^\top],$$

provided the population powers are finite and that residual is positive semidefinite. An independent centered correction cannot remove the mean shift.

A sufficient condition for the following local expansion is that $D_\theta(\cdot, \eta)$ is C^4 and its fourth derivatives grow at most polynomially. Write $f = D_\theta(\cdot, \eta)$ and let $\mathrm{D}^k f$ denote its kth derivative as a multilinear map with its induced operator norm. For fixed $\boldsymbol{u}$ and $0 < \delta \le 1$, Taylor's formula with integral remainder gives

$$\begin{aligned} \boldsymbol{G}_\delta &= \mathrm{D}f(\boldsymbol{u})\boldsymbol{e} + \frac{\delta}{2}\mathrm{D}^2 f(\boldsymbol{u})[\boldsymbol{e}, \boldsymbol{e}] + \frac{\delta^2}{6}\mathrm{D}^3 f(\boldsymbol{u})[\boldsymbol{e}, \boldsymbol{e}, \boldsymbol{e}] + \boldsymbol{T}_\delta(\boldsymbol{e}), \\ \boldsymbol{T}_\delta(\boldsymbol{e}) &= \frac{\delta^3}{6}\int_0^1 (1-t)^3 \mathrm{D}^4 f(\boldsymbol{u} + t\delta\boldsymbol{e})[\boldsymbol{e}, \boldsymbol{e}, \boldsymbol{e}, \boldsymbol{e}]\, dt. \end{aligned}$$

If $\left\|\mathrm{D}^4 f(\boldsymbol{v})\right\| \le C_\eta(1 + \|\boldsymbol{v}\|^\nu)$ for some finite $\nu \ge 0$, the integral yields

$$\mathbb{E}\,\|\boldsymbol{T}_\delta(\boldsymbol{e})\| \le C'_{\boldsymbol{u},\eta}\delta^3\mathbb{E}[(1 + \|\boldsymbol{e}\|^\nu)\,\|\boldsymbol{e}\|^4] = O(\delta^3).$$

The Gaussian moment is finite, and $C'_{\boldsymbol{u},\eta}$ is independent of $\delta \in (0, 1]$. The linear and cubic terms have zero expectation by probe symmetry; the quadratic term contracts with $\mathbb{E}[\boldsymbol{e}\boldsymbol{e}^\top] = \boldsymbol{I}$. Therefore

$$\mathbb{E}_{\boldsymbol{e}}\boldsymbol{G}_\delta(\boldsymbol{u}, \boldsymbol{e}; \eta) = \frac{\delta}{2}\,\Delta D_\theta(\boldsymbol{u}, \eta) + O(\delta^3), \tag{19}$$

where the Laplacian acts componentwise. This is an expansion as $\delta \to 0$ at fixed dimension, input, and noise level, not a uniform error bound for the evaluated network or schedule. When the Laplacian is nonzero, the leading update-mean shift is of order $\eta\delta$. With a learned Jacobian, the linearized covariance is $\eta^2(\nabla D_\theta \nabla D_\theta^\top + \boldsymbol{R}_\eta)$, not necessarily a squared symmetric Bayes Jacobian. Neither discrepancy is removed by Proposition 1.

A.4 STATIONARY GAUSSIAN COVARIANCE

Proof of Theorem 1. Fix the measurement and coupling. Substituting Equation (11) into the affine Gaussian denoiser and adding independent Gaussian prior-update noise gives

$$\begin{aligned} &\boldsymbol{x}' = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{c} + \boldsymbol{\zeta}, && \boldsymbol{M} = \boldsymbol{J}\boldsymbol{F}, && \boldsymbol{c} = (\boldsymbol{I} - \boldsymbol{J})\boldsymbol{m} + \boldsymbol{J}\boldsymbol{b}, \\ &\mathbb{E}\boldsymbol{\zeta} = 0, && \mathrm{Cov}(\boldsymbol{\zeta}) = \boldsymbol{Q}_S = \eta^2\boldsymbol{J}\boldsymbol{F}\boldsymbol{J} + \eta^2\boldsymbol{S}. \end{aligned} \tag{20}$$

The innovations are independent of the current state and across sweeps. Because $0 \prec \boldsymbol{J} \prec \boldsymbol{I}$ and $0 \prec \boldsymbol{F} \preceq \boldsymbol{I}$, $\|\boldsymbol{M}\| \le \|\boldsymbol{J}\|\,\|\boldsymbol{F}\| < 1$, with no commutativity requirement. Iteration from a fixed state gives a decaying initial term, the convergent deterministic sum $\sum_{n\ge 0}\boldsymbol{M}^n\boldsymbol{c}$, and a Gaussian noise sum with absolutely summable covariance. Define

$$\mathcal{L}(\boldsymbol{Y}) = \sum_{n=0}^{\infty} \boldsymbol{M}^n \boldsymbol{Y} (\boldsymbol{M}^\top)^n, \qquad \|\mathcal{L}(\boldsymbol{Y})\| \le \frac{\|\boldsymbol{Y}\|}{1 - \|\boldsymbol{M}\|^2}. \tag{21}$$

The noise sum after n sweeps has the same distribution as $\sum_{k=0}^{n-1}\boldsymbol{M}^k\boldsymbol{\zeta}_k$ for independent innovations $\boldsymbol{\zeta}_k$. This latter series converges in mean square. An independent random initial state also has $\boldsymbol{M}^n\boldsymbol{x}_0 \to 0$ almost surely, so the chain law converges to the same Gaussian. Thus the unique stationary law has mean $(\boldsymbol{I} - \boldsymbol{M})^{-1}\boldsymbol{c}$ and covariance $\mathcal{L}(\boldsymbol{Q}_S)$; $\boldsymbol{S} = 0$ is allowed. Every summand preserves the Loewner order, and shifting the series gives

$$\mathcal{L}(\boldsymbol{Y}) = \boldsymbol{Y} + \boldsymbol{M}\mathcal{L}(\boldsymbol{Y})\boldsymbol{M}^\top, \qquad \mathcal{L}(\boldsymbol{M}\boldsymbol{Y}\boldsymbol{M}^\top) = \boldsymbol{M}\mathcal{L}(\boldsymbol{Y})\boldsymbol{M}^\top.$$

Let $\boldsymbol{C}_{\mathrm{post}}$ be as in the statement and let $\boldsymbol{m}_{\mathrm{post}} = \boldsymbol{C}_{\mathrm{post}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{m} + \sigma_y^{-2}\boldsymbol{A}^\top\boldsymbol{y})$. Using $\boldsymbol{I} - \boldsymbol{J} = \eta^2\boldsymbol{J}\boldsymbol{\Sigma}^{-1}$ and Equation (11),

$$\boldsymbol{I} - \boldsymbol{M} = \eta^2\boldsymbol{J}\boldsymbol{C}_{\mathrm{post}}^{-1}, \qquad \boldsymbol{c} = \eta^2\boldsymbol{J}(\boldsymbol{\Sigma}^{-1}\boldsymbol{m} + \sigma_y^{-2}\boldsymbol{A}^\top\boldsymbol{y}).$$

These identities identify the stationary mean with $\boldsymbol{m}_{\mathrm{post}}$ for every fixed $\boldsymbol{S}$. Setting $\boldsymbol{S} = \boldsymbol{J}$ yields the exact Gibbs conditionals of Equation (10); their invariant posterior and uniqueness imply $\boldsymbol{C}_{\mathrm{post}} = \mathcal{L}(\boldsymbol{Q}_J)$.

One additional exact observation/prior sweep replaces the innovation covariance by $\boldsymbol{Q}_J$ while retaining $\boldsymbol{M}$ and $\boldsymbol{c}$. Its output mean is still $\boldsymbol{m}_{\mathrm{post}}$, and $\boldsymbol{C}_{\mathrm{out}} = \boldsymbol{M}\mathcal{L}(\boldsymbol{Q}_S)\boldsymbol{M}^\top + \boldsymbol{Q}_J$. Subtracting the stationary identity for $\boldsymbol{C}_{\mathrm{post}}$ gives

$$\boldsymbol{C}_{\mathrm{out}} - \boldsymbol{C}_{\mathrm{post}} = \boldsymbol{M}\mathcal{L}\big(\eta^2(\boldsymbol{S} - \boldsymbol{J})\big)\boldsymbol{M}^\top. \tag{22}$$

Positivity of $\mathcal{L}$ and congruence by $\boldsymbol{M}$ bound this signed error between $-\epsilon_-\boldsymbol{K}$ and $\epsilon_+\boldsymbol{K}$, where $\boldsymbol{K} = \boldsymbol{M}\mathcal{L}(\eta^2\boldsymbol{J})\boldsymbol{M}^\top$. Put $\boldsymbol{B} = \eta^2\boldsymbol{J}\boldsymbol{F}\boldsymbol{J}$. Congruence by $\boldsymbol{F}$ of $\boldsymbol{J} \preceq \boldsymbol{I}$ gives $\boldsymbol{F}\boldsymbol{J}\boldsymbol{F} \preceq \boldsymbol{F}^2$, and $\boldsymbol{F}^2 \preceq \boldsymbol{F}$ since its eigenvalues are in $(0, 1]$. Congruence by $\boldsymbol{J}$ then yields

$$\boldsymbol{M}(\eta^2\boldsymbol{J})\boldsymbol{M}^\top = \eta^2\boldsymbol{J}\boldsymbol{F}\boldsymbol{J}\boldsymbol{F}\boldsymbol{J} \preceq \boldsymbol{B}, \qquad \boldsymbol{B} \preceq \eta^2\boldsymbol{J}^2 \preceq \eta^2\boldsymbol{J}. \tag{23}$$

Applying the positive operator and its shift identity,

$$\boldsymbol{K} = \mathcal{L}\big(\boldsymbol{M}(\eta^2\boldsymbol{J})\boldsymbol{M}^\top\big) \preceq \mathcal{L}(\boldsymbol{B}) \preceq \tfrac{1}{2}\big[\mathcal{L}(\boldsymbol{B}) + \mathcal{L}(\eta^2\boldsymbol{J})\big] = \tfrac{1}{2}\boldsymbol{C}_{\mathrm{post}}. \tag{24}$$

The middle inequality follows from $\mathcal{L}(\boldsymbol{B}) \preceq \mathcal{L}(\eta^2\boldsymbol{J})$; the last equality uses $\boldsymbol{Q}_J = \boldsymbol{B} + \eta^2\boldsymbol{J}$. Multiplication by the nonnegative $\epsilon_\pm$, reversing the lower comparison after its minus sign, proves Equation (9).

For sharpness, take $d = 2$, $\boldsymbol{A} = (s, 0)$ with $s > 0$, and $\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_m, \lambda)$ with positive finite variances. Choose $0 < \eta < \sigma_y/s$. The second coordinate is unmeasured, while $\tau_{\min} = \sigma_y/s$ remains defined. Write $j_m = \lambda_m/(\lambda_m + \eta^2)$ and $j = \lambda/(\lambda + \eta^2)$, and take $\boldsymbol{S} = \mathrm{diag}(j_m, (1+\rho)j)$ with $\rho > 0$. In the unmeasured coordinate, the stationary variance is $\eta^2[j^2 + (1+\rho)j]/(1 - j^2)$ and the exact posterior variance is $\lambda = \eta^2 j/(1 - j)$. The exact final sweep multiplies their difference by j^2, so

$$\frac{C_{\mathrm{out}} - \lambda}{\lambda} = \frac{\rho j^2}{1 + j} \;\longrightarrow\; \frac{\rho}{2} \quad \text{as } \lambda/\eta^2 \to \infty. \tag{25}$$

Here $\epsilon_+ = \rho$ and $\epsilon_- = 0$. Replacing $1 + \rho$ by $1 - \rho$, with $0 < \rho \le 1$, gives the negative counterpart and sharpness of the lower factor. The factor is approached by admissible finite-variance priors; equality at an infinite prior variance is not required. □

Scope of the final sweep. In the same scalar direction, omitting the exact final sweep leaves relative stationary error $\rho/(1+j) > \rho/2$ for every $j < 1$. Thus the final-sweep requirement cannot simply be dropped. It is a statement about an exact prior-conditional draw after the final observation update, not an additional observation update alone. A static-table descent supplies that draw when its Gaussian reverse transitions are exact at every queried level, as shown in Equation (14); the single-coupling residual example does not satisfy this static-table requirement. The theorem concerns covariance conditional on fixed $\boldsymbol{y}$ at stationarity. It does not bound the measurement-averaged finite-schedule divergence used in the Gaussian experiments. Any fixed $\boldsymbol{S}$ satisfying the relative-error hypothesis is permitted, including a calibrated or interpolated table at that coupling. The theorem does not certify that its hypotheses or exact-readout condition hold for the learned, finite-difference, annealed implementation. The empirical covariance–mixing comparisons and the learned-denoiser control address different questions from this stationary bound.

B Implementation and experimental protocols

B.1 Complete sampling procedure

Algorithm 1 distinguishes the L ordinary observation/prior sweeps from the terminal readout. With $c \in \{1, 2\}$ calls per prior update and n_F descent transitions, the actual forward count is $cL + n_F$, where $L = \lfloor (B - n_F)/c \rfloor$. The reported budgets permit $L \geq 2$. The FFHQ runner constructs L coupling nodes, all used for prior updates. Scattering and the evaluated vector runners construct $L+1$ nodes, reserving the last for the final observation update. In either convention there are exactly $L + 1$ observation updates and one returned clean state.

Algorithm 1 MM-SG for one independent posterior draw

Require: Fixed D_θ; $\boldsymbol{y}, \boldsymbol{A}, \sigma_y$; tables $\boldsymbol{S}_\eta, \boldsymbol{R}_\eta$; budget B; grids and schedule policy below
Compute the spectral observation quantities in Equation (7)
Resolve $\eta_\star$ from $0.95\tau_{\min}$ using the applicable snapping rule below
Read $c \in \{1, 2\}$ fixed by the prior/operator diagnostic (Appendix B.3)
Form terminal levels $\eta_\star = a_0 > a_1 > \cdots > a_{n_F} = 0$
Set $L = \lfloor (B - n_F)/c \rfloor$ and $\boldsymbol{x}_0 \leftarrow \mathbf{0}$
Form sweep levels $\eta_0, \ldots, \eta_{L-1}$ using the applicable node convention
for $\ell = 0, \ldots, L - 1$ **do**
    Draw $\boldsymbol{u}_\ell$ by Equation (7) at η_ℓ conditional on $\boldsymbol{x}_\ell$, including the null space
    if $c = 1$ **then**
        $\boldsymbol{x}_{\ell+1} \leftarrow D_\theta(\boldsymbol{u}_\ell, \eta_\ell) + \eta_\ell \boldsymbol{S}_{\eta_\ell}^{1/2} \boldsymbol{e}$, with fresh $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
    else
        Draw independent $\boldsymbol{e}, \boldsymbol{e}' \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$; set $\delta = 0.02\eta_\ell$
        Evaluate $\boldsymbol{d}_0 = D_\theta(\boldsymbol{u}_\ell, \eta_\ell)$ and $\boldsymbol{d}_1 = D_\theta(\boldsymbol{u}_\ell + \delta \boldsymbol{e}, \eta_\ell)$
        $\boldsymbol{x}_{\ell+1} \leftarrow \boldsymbol{d}_0 + \eta_\ell [(\boldsymbol{d}_1 - \boldsymbol{d}_0)/\delta + \boldsymbol{R}_{\eta_\ell}^{1/2} \boldsymbol{e}']$
    end if
end for
Draw $\boldsymbol{z}$ by Equation (7) at $\eta_\star$ conditional on $\boldsymbol{x}_L$
for $j = 0, \ldots, n_F - 1$ **do**
    Apply Equation (8) from a_j to a_{j+1} with static $\boldsymbol{S}_{a_j}$ and fresh Gaussian noise
end for
return $\boldsymbol{z}$; restart from zero with independent noises for each additional output

Schedules and grids. Let $N_{\text{sch}} = L$ for FFHQ and $N_{\text{sch}} = L + 1$ otherwise. The schedule uses $H = \max(1, \text{round}(0.1 N_{\text{sch}}))$ terminal hold nodes, with ties rounded to the nearest even integer, and $N_a = N_{\text{sch}} - H$ decay nodes. Before any node snapping, these are $\eta_{\text{start}}(\bar{\eta}/\eta_{\text{start}})^{j/N_a}$ for $j = 1, \ldots, N_a$, followed by H copies of $\bar{\eta}$. Thus the first node is already below the decay anchor, and the endpoint occurs in both the decay and hold. The anchor η_{start} is 10 for scattering and vector controls; for FFHQ it is the checkpoint's largest noise level (approximately 157). Repeated nodes are not removed.

Scattering and analytic-mixture runners use $\bar{\eta} = \eta_\star = 0.95\tau_{\min}$ without snapping. Their terminal grid is the 1,000-point descending sequence

$$s_j = \left[100^{1/7} + \tfrac{j}{999}(10^{-3/7} - 100^{1/7})\right]^7, \qquad j = 0, \ldots, 999.$$

The descent starts at the exact $\eta_\star$, includes every $s_j < \eta_\star$, and appends zero. FFHQ uses $s_t = \sqrt{(1-\bar{\alpha}_t)/\bar{\alpha}_t}$ for $t = 0, \ldots, 999$, with $\bar{\alpha}_t = \prod_{i=0}^{t}(1-\beta_i)$ and $\beta_i = 10^{-4} + (0.02 - 10^{-4})i/999$. It first snaps $0.95\tau_{\min}$ to the nearest grid value (ties upward); if this is not below $\tau_{\min}$, it chooses the largest grid value strictly below it. It then sets $\bar{\eta} = \eta_\star$ and snaps all decay nodes to their nearest values. Terminal descent uses all grid values below $\eta_\star$, through s_0, then zero. These rules require an admissible grid endpoint, as in all reported image settings.

The learned-vector control instead snaps each node downward on its variance-preserving diffusion grid, with the grid index bounded below by one; it forms the geometric decay toward the unsnapped $0.95\tau_{\min}$. Its terminal descent uses the corresponding native grid down to its smallest positive level, then zero. Across variants, the final transition is $\boldsymbol{z} \leftarrow D_\theta(\boldsymbol{z}, a) + a\boldsymbol{S}_a^{1/2}\boldsymbol{e}$: it retains calibrated noise. There is no extra denoising prediction, appended sample, or iterate average after this readout.

Denoiser access and numerical implementation. The image chains use fixed learned denoisers, whereas the analytic mixtures use their true Bayes means. Even with a true mixture mean, the Gaussian prior update is not an exact mixture-conditional draw. The exact-conditional controls draw mixture components as well as within-component noise; their observation-rule differences are stated in Appendix C.3. The sparse-signal reference instead approximates the Bayes mean numerically (Appendix C.5). V3/V3PL run internally in the shared-component covariance eigenbasis, transform the calibrated principal-component basis into those coordinates, and rotate the final state back.

Image states, Gaussian noises, and spectral arithmetic use float64. FFHQ evaluates the network on float32 inputs and casts its predictions back to float64 before differencing. Its wrapper scales an input by $1/\sqrt{1+\eta^2}$, selects the nearest discrete noise index, and returns $\boldsymbol{u} - \eta\hat{\boldsymbol{\epsilon}}$ using only the noise-prediction channels, not the learned variance head. Scattering returns float64 predictions through its pretrained-model wrapper; internal network precision is delegated to that model. Analytic-mixture arithmetic uses float64. No Jacobian is formed, no denoiser backward pass is taken, and the candidate has no adaptive inner trial, online covariance fitting, or convergence-based stop. Variance factors are clamped below at zero before square roots where needed for roundoff; the covariance-table clipping is specified next. Image runners raise an error for a nonfinite final state rather than retrying or substituting a draw. Both image harnesses cast completed draws to float32 before evaluation; FFHQ subsequently stores its sample arrays in float16. Clean chain states are not clipped to the image range; metric-specific clipping is described in Appendix B.7.

B.2 Calibration and amortized cost

Image priors. FFHQ calibration uses 128 unconditional model samples and an orthonormal two-dimensional Fourier transform, with 64 radial bins independently in each color channel. Scattering uses 64 unconditional samples and 48 radial bins. For frequencies (f_x, f_y) in cycles per pixel, the bin is $\min(\lfloor n_b\sqrt{f_x^2 + f_y^2}/\sqrt{1/2}\rfloor, n_b - 1)$, with $n_b = 64$ or 48. Power is averaged over samples and coefficients within a bin. Spectral filters are symmetric and act on real white noise, so inverse transforms return real samples; equivalently, the covariance is diagonal in a real Fourier representation.

The denoising-error estimate uses two independent corruption replicates. Response power uses one corruption and an independent probe, with $\delta = 0.02\eta$ as in inference. Both power tables are clipped to $[0, 1]$. At each response-calibration level, the residual is their nonnegative difference. The static and residual tables are then interpolated separately, linearly in log noise, with constant extension outside their grids. In particular, the image sampler does not subtract interpolated powers anew at each query. Proposition 1 instead treats population calibration at the queried level; clipping and finite-grid interpolation are implementation approximations.

FFHQ uses every fifth ascending checkpoint level plus the maximum for its 201-level static table, and every second entry of that table's grid for its 101-level response table. Scattering takes indices $0, 10, \ldots, 990$ of its descending 1,000-point grid and adds index 999, giving 101 levels for both

tables. Calibration batches contain eight FFHQ or 16 scattering samples; the two response queries are concatenated into a doubled batch. FFHQ prior-sample generation uses 998 reverse transitions down to the second-smallest grid level and one final mean prediction, hence 999 calls per sample. Scattering uses 999 transitions to its smallest grid level and one mean prediction, hence 1,000 calls. FFHQ clips those generated calibration samples to $[-1, 1]$; scattering does not. The nominal FFHQ calibration work is

$$128 \times 999 + 201 \times 2 \times 128 + 101 \times 2 \times 128 = 205{,}184 \tag{26}$$

denoiser evaluations over individual images: prior-sample generation, static calibration, and response calibration, respectively. This is an operation count reconstructed from the configuration, not a measured wall-clock total. The analogous scattering count is $64[1000 + 2(101) + 2(101)] = 89{,}856$. These counts exclude Fourier transforms and table construction. Tables are shared across measurements and observation operators; the prior/operator diagnostic incurs additional work (Appendix B.3). The online NFE comparison excludes calibration and routing: it measures forward-evaluation work after preprocessing, not end-to-end speed. If calibration is reusable, $\lceil 205{,}184/(3{,}000 - 999) \rceil = 103$ individual outputs offset its nominal forward cost relative to the 3,000-call reference; the corresponding ceiling is 102 at 3,016 reference calls. These are outputs, not images with multiple outputs, and exclude routing, operator, and backward work. This arithmetic is neither a runtime break-even nor a comparison with the 100-call methods.

Vector priors. The mixture runners estimate a principal-component basis from prior samples and use componentwise denoising-error power in that basis. V3 and V3PL use 65,536 calibration samples; the heteroscedastic controls use 16,384. The 64-dimensional learned signal priors use 16,384 training signals for the basis and a separately selected set for denoising-error estimation; response power uses 2,048 training signals. Gaussian calibration noise is rescaled to have unit empirical second moment per basis direction. If the estimated prior variance in a direction is v_k, the static estimate is capped at $v_k/(v_k + \eta^2)$, the scalar linear-estimator bound for the average normalized Bayes conditional variance (Appendix A.3). That bound is a population statement, not a pointwise bound in $\boldsymbol{u}$; imposing it on a learned-denoiser error estimate is a stabilization choice. The two-call vector controls interpolate the static and response-power tables separately and clip their difference at the queried level, unlike the image residual-table interpolation. V3/V3PL use 56 log-spaced calibration levels from $\min(10^{-3}, 0.5\eta_\star)$ to 12; the heteroscedastic controls use 48 from 10^{-3} to 12. These variants require prior or training samples. The method does not universally eliminate data access, although the image variants calibrate from generated samples without retraining the denoiser.

Online work and storage. Each output requires $cL + n_F$ denoiser evaluations and $L + 1$ observation updates. For a batch of q image chains, a two-call step evaluates $2q$ inputs together and reuses the first q predictions; batching does not halve the evaluation count or remove the larger activation batch. A static or full two-call prior update uses one forward/inverse Fourier-transform pair to shape its independent noise, as does each descent transition. This adds $O(qd \log d)$ arithmetic per update, beyond the denoiser. The measured observation is transformed once. Each FFHQ observation update applies one forward and one inverse right-singular-vector transform, implemented by masks, block transforms, or Fourier transforms rather than a dense factorization.

The scattering observation implementation uses two projections onto its stored row space and three syntheses from it, plus a high-noise subspace synthesis when that branch is active. These operations cost $O(qdr_A)$ per observation update and require $O(dr_A)$ storage for the right-singular vectors. Their preprocessing, and the full operator cache retained by the harness, are additional to online denoiser work. Vector covariance sampling uses a dense principal-component basis, with $O(qd^2)$ transforms and $O(d^2)$ basis storage. Compact image tables contain $O(Tn_b)$ coefficients per channel for T stored levels; FFHQ additionally caches an expanded d-entry filter at each queried level, keyed by log noise rounded to nine decimal places. For Q cached table-level pairs this is $O(Qd)$ storage, beyond $O(qd)$ chain state and denoiser memory. These operation and storage counts are not wall-time or peak-memory measurements.

Ablations. The scalar control averages static variance over directions. The response-power control (A1 in the records) pools the squared Fourier magnitude of the current response within radial bins, applies the resulting scale to a new independent Gaussian probe, and adds the same residual noise. The raw-only control sets $\boldsymbol{R}_\eta = 0$. All use the same terminal static-table descent. In particular,

raw-only removes the residual from the repeated prior updates, not all static covariance use from the complete algorithm.

B.3 THE FROZEN ROUTING DIAGNOSTIC

The routing statistic compares two sample constructions; it does not certify posterior accuracy. Draw $\boldsymbol{x}_n$ from the prior (or its model sampler), form $\boldsymbol{u}_n = \boldsymbol{x}_n + \eta_\star \boldsymbol{e}_n$, and compare a one-call sample $\boldsymbol{z}_n$ with a 64-step static-covariance descent sample $\boldsymbol{z}_n^{\mathrm{ref}}$ from the same $\boldsymbol{u}_n$. For $s \in \{\eta_\star, 2\eta_\star\}$, define

$$\boldsymbol{c}_s(\boldsymbol{z}) = \frac{s^2 \boldsymbol{u} + \eta_\star^2 (\boldsymbol{z} + s\boldsymbol{e}')}{s^2 + \eta_\star^2}, \qquad s_c = \frac{s\eta_\star}{\sqrt{s^2 + \eta_\star^2}}, \tag{27}$$

$$T(s) = \frac{\sum_n \left\| \boldsymbol{F}[\boldsymbol{z}_n - D_\theta(\boldsymbol{c}_s(\boldsymbol{z}_n), s_c)] \right\|^2}{\sum_n \left\| \boldsymbol{F}[\boldsymbol{z}_n^{\mathrm{ref}} - D_\theta(\boldsymbol{c}_s(\boldsymbol{z}_n^{\mathrm{ref}}), s_c)] \right\|^2}, \tag{28}$$

Here $\boldsymbol{F}$ is from Equation (11) at $\eta_\star$. At each s, the same fresh $\boldsymbol{e}'_n$ is used for the numerator and reference, independently across n and the two values of s. Use one call if $|T(\eta_\star) - 1| \le 0.05$ and $|T(2\eta_\star) - 1| \le 0.08$; otherwise use two. Image diagnostics use 48 prior samples. The reference uses 64 positive geometric levels from $\eta_\star$ to $\min(s_{\min}, \eta_\star)$, followed by zero, giving 64 transitions; $s_{\min}$ is the smallest level of the corresponding diffusion grid. No measurement value enters the calculation. The reference uses the same denoiser and an approximate descent, so agreement need not indicate agreement with the true prior conditional.

The image diagnostic is evaluated once per distinct endpoint and reused across measurements. Including its auxiliary raw-response and prior-sample checks, the implemented script uses $48(2 + 64 + 8) = 3{,}552$ denoiser evaluations per endpoint group: two base/perturbed queries, 64 reference transitions, and two test levels for each of four sample constructions. Prior-sample generation and calibration are excluded from this count. Tasks sharing an endpoint reuse these model evaluations and apply their own operator weighting. Scattering uses its earlier development calibration for this diagnostic, separately from the final inference table. Gaussian diagnostics evaluate the corresponding moments analytically rather than sampling these constructions. The evaluated branch is fixed in the runner configuration; the diagnostic is not rerun for each observation.

The second noise level, operator weighting, reference construction, and sample counts were chosen after development statistics had been inspected, then frozen for subsequent evaluation. They are not a universal pre-data design. The diagnostic selects one call in the final Gaussian suite and two on all image tasks. Its reported winner-prediction audit agrees in 13 of 20 metric comparisons; these comparisons share priors and operators and are not 20 independent validations. The method's scope is therefore not defined by passing this statistic.

B.4 SCATTERING DATA AND SPLITS

The InverseBench task uses 20 transmitters, sensor radius 1.6, wave parameter 6, and complex measurement-noise scale 10^{-4}. Images represent 128×128 permittivity contrast. The prior is the benchmark's pretrained EDM model (Karras et al., 2022; Zheng et al., 2025). In physical coordinates $\boldsymbol{x}_{\mathrm{phys}} = (\boldsymbol{x} + \mathbf{1})/2$; the real-stacked transformed measurement is $\boldsymbol{y}' = 2\,\mathrm{stack}(\boldsymbol{y}) - \boldsymbol{A}\mathbf{1}$. The resulting model is $\boldsymbol{y}' = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}'$ with covariance $2 \times 10^{-8} \boldsymbol{I}$. For 180 receivers, the explicit real operator is 7200×16384 with numerical rank 6549. The evaluated solver uses the cached singular vectors and values.

Development uses official validation targets 0–2. Selection uses validation targets 3–9, with 50 outputs per method and target; some of these targets carry earlier pilot exposure. Final evaluation uses official test targets 0–19, with 25 independent outputs per method, on both receiver configurations. Before the final freeze, test target 0 had a mechanics-only shape/range inspection, and no scientific metric; the provenance explicitly records it. Targets 20–99 are not used here. The 360-receiver sampler outputs were first evaluated after the final freeze, without retuning. The final runner's measurement-noise seed convention is shared with earlier validation runs, while the underlying targets are disjoint; this is a target-level holdout, not a claim of entirely new noise realizations. The final cohort is paired across methods.

B.5 FFHQ DATA, OPERATORS, AND BASELINE ADAPTATIONS

The pretrained denoiser is the unconditional FFHQ-256 ADM checkpoint distributed with the DPS implementation, using its native diffusion grid. The four operators are 4×4 average pooling with stride four; a central 128×128 missing box; a fixed 70% random missing-pixel mask; and circular convolution with a 61×61 Gaussian kernel of standard deviation 3. Images are normalized to $[-1, 1]$, with independent observed-space Gaussian noise of standard deviation 0.05. This is not the 0.05 noise convention on $[0, 1]$ used in some published restoration tables.

Image identifiers 69300–69314 are used for tuning (two outputs per image), 69315–69329 for selection (four outputs), and 69330–69339 for development. At most three configurations per fast baseline and task are considered. The frozen selected parameters, in super-resolution/box/random/blur order, are: DPS guidance scale $(0.6, 0.5, 1.0, 0.6)$; DAPS likelihood scale $(0.01, 0.03, 0.01, 0.01)$; RED-diff learning rate 0.01 for every task; DiffPIR regularization $(4, 4, 4, 3.5)$; DDNM noise parameter $(1, 1, 1, 0.85)$; and pseudoinverse-guided diffusion parameter 0.5 throughout. Other algorithm defaults are those of the supplied baseline adapters. Tuning excluded numerically divergent settings under the recorded amendment; no failed runs were imputed as successful outputs.

These comparisons use common data and the same denoiser but include adaptations. DAPS-1K uses 200 outer levels, five denoiser calls and 100 likelihood/proximal Langevin steps per level; its 1,000 NFE exclude 20,000 inner forward/adjoint pairs. The Gaussian likelihood gradient uses the tuned scale listed above, not the measurement-noise scale 0.05. DDNM+ disables time-travel resampling. For super-resolution and masks, it uses the scalar noisy-data blend with regularized inverse coefficients $s_i/(s_i^2 + \sigma_y^2)$. Its selected blur branch instead solves $(\boldsymbol{A}^\top \boldsymbol{A} + \rho \boldsymbol{I})\hat{\boldsymbol{x}} = \boldsymbol{A}^\top \boldsymbol{y} + \rho D_\theta(\boldsymbol{x}, a)$, with $\rho = \sigma_y^2 / \max(b^2, 10^{-12})$ and consecutive diffusion levels $a > b \geq 0$. This per-coordinate data step replaces the scalar blend; it is a local adaptation, not the native DDNM+ update (Wang et al., 2023, Appendix I). DPS and pseudoinverse-guided diffusion backpropagate through the denoiser; their NFE do not include that backward cost. DiffPIR, DDNM+, and pseudoinverse-guided diffusion use 100 forward calls. The full-diffusion SG reference changes the expensive prior-conditional solver within the local harness and is not labeled native PnP-DM. RED-diff optimizes a point estimate with Adam, rather than taking ancestral posterior draws (Mardani et al., 2024). The local implementation starts each estimate at zero and returns the final optimized image; its reported ensemble statistics summarize independently randomized optimization runs. Ensemble-mean PSNR estimates posterior-mean quality for the sampling methods but is not a posterior-mean guarantee for every comparator. We make no numerical comparison with published tables under different masks, noise conventions, tuning budgets, or readouts.

Full-diffusion split-Gibbs reference. Both image references initialize the clean state at zero and alternate an observation update with a reverse diffusion path. For each inner pair $a > b \geq 0$, the selected transition is

$$\boldsymbol{z} \leftarrow r\boldsymbol{z} + (1 - r) D_\theta(\boldsymbol{z}, a) + \sqrt{a^2 - b^2}\, \boldsymbol{e}, \qquad r = b^2/a^2, \quad \boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}),$$

with fresh independent noise, except at the FFHQ terminal transition specified below. The variance factor is $a^2 - b^2$, not the optional bridge factor $r(a^2 - b^2)$. Each transition costs one denoiser evaluation. Every outer observation update treats the previous path's output as level zero; the final output is the last path state, without another denoiser call or iterate averaging. These local operating points are not a native PnP-DM reproduction.

For FFHQ, the observation update uses Equation (7), including high-noise matching. The outer schedule follows the geometric decay and 10% hold in Appendix B.1, with starting parameter $s_{999} \approx 157.4073$ and the same strictly admissible snapped endpoint. Its first actual node is below that starting parameter. Ordinary inner paths use the native levels nearest a 100-point grid uniform in the seventh root of noise, between s_{999} and s_0; merging duplicate levels leaves 99 positive points. Each path starts at the exact snapped outer level, traverses all strictly smaller grid levels above s_0, and ends at s_0. The last outer stage instead uses the full native grid. The transition onto s_0 omits noise, unlike the candidate's stochastic descent to zero. The outer count is the first integer at least two whose total reaches 3,000 evaluations. Super-resolution uses 63 stages at endpoint 0.18899239544636204: 2,962 calls in ordinary stages plus 54 in the final stage, totaling 3,016. Box inpainting, random masking, and deblurring each use 75 stages at endpoint 0.04688549021144125: 2,990 plus 10 calls, totaling 3,000.

For scattering, the observation update is compensated below each directional threshold and pins higher-noise measured coordinates to $\widetilde{y}_i$, rather than using FFHQ's matching rule. The outer starting parameter is 10, the endpoint is $0.95\tau_{\min}$ without snapping, and the endpoint hold is 10%. Ordinary inner grids have 100 positive levels uniform in the seventh root of noise from 100 to 10^{-3}; the last stage uses 1,000 such levels. Every path starts at its exact outer level, includes all smaller grid levels, and ends at zero, retaining the last Gaussian noise. The stage count is the smallest meeting 3,000 calls: 101 stages and 3,020 evaluations for 180 receivers, versus 105 stages and 3,026 for 360 receivers. These counts include the final fine path, not an additional candidate-style terminal sweep.

B.6 HELD-OUT FFHQ PREFIX AND INTERNAL PROTOCOL

The internal final protocol specifies FFHQ IDs 69000–69099, distinct from tuning IDs 69300–69314, selection IDs 69315–69329, and development IDs 69330–69339. The same image IDs are used for super-resolution, box inpainting, random masking, and deblurring. Each method produces four independently initialized outputs per image; image-level metrics are then averaged separately within each task. The eight methods and their settings are frozen before final evaluation. MM-SG uses the raw finite-difference response plus the offline residual, with $\delta = 0.02\eta$ and the self-calibration based on 128 unconditional model samples. The one-time calibration cost is 205,184 nominal offline calls (Appendix B.2); none is included in the online NFE column.

Internal amendment A2, dated 25 September 2026, fixes the primary analysis at noon KST on 26 September 2026. The protocol and amendment are internal records, not public preregistrations. At that cutoff, it uses the largest prefix of $n \in \{20, 30, \dots, 100\}$ consecutive IDs for which every task and all eight methods have complete records. The amendment records that this rule was set before inspecting the affected scores; some outputs already existed, and completed records were retained when in-flight jobs were stopped and execution was requeued. Eligibility depends only on record existence, not measured performance. The frozen analysis contains IDs 69000–69059: 1,920 per-image records, but only 60 distinct image targets, each reused across tasks. The full 100-image confirmation is not available. The static one-call, raw-only, residual-only, and response-power controls are absent from this prefix, as is a faithful published covariance-aware comparator.

Actual forward evaluations per output are 999 for MM-SG, 3,016 for full-diffusion SG on super-resolution and 3,000 on the other tasks, 1,000 for DPS/DAPS-1K/RED-diff, and 100 for DiffPIR/DDNM+/pseudoinverse-guided diffusion. These counts include the final descent but exclude backward work and one-time calibration. Appendix B.1 specifies the stage-dependent high-noise rules, and Appendix B.5 details the baseline adaptations.

Frozen criteria and descriptive intervals. The primary decision requires the one-sided 95% lower bound on the paired PSNR difference to be at least -0.3 dB on at least three tasks, separately against full-diffusion SG, DAPS-1K, and RED-diff. Its energy-score criterion uses the fast comparator with the lowest mean energy score on this cohort for each task. For task t, let $E_{i,t}^m$ be method m's raw energy score on image i, let $\mathcal{F}$ contain the six fast methods, and use the superscript MM for MM-SG. The comparator and statistic are

$$m_t^* = \arg\min_{m \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} E_{i,t}^m, \qquad R_t^{\mathrm{ES}} = \frac{1}{n} \sum_{i=1}^{n} \frac{E_{i,t}^{\mathrm{MM}}}{E_{i,t}^{m_t^*}}. \tag{29}$$

The comparator is held fixed during bootstrap resampling. Thus R_t^{ES} is a mean of per-image ratios, not the ratio of the arithmetic mean scores in Table 2; its bounds do not adjust for choosing the comparator on this cohort. The primary criterion requires the one-sided upper bound on R_t^{ES} to be at most 1.10 on at least three tasks. A separate secondary criterion requires this bound to be below one on at least two tasks. These rules use 10,000 paired image-level resamples with PCG64 seed 8622001. By contrast, the differences in Figure 3 and Table 3 use 20,000 resamples with PCG64 seed 20260926 and central 2.5%/97.5% endpoints. Both analyses resample images within tasks, never outputs or pooled task–image pairs.

The remaining primary criterion requires the mean whitened measurement residual to lie in $[0.9, 1.1]$ on every task. In the measured coordinates of Section 3, let r_A be the number of nonzero measure-

ment singular values. The per-image diagnostic is

$$Q = \frac{1}{K r_A} \sum_{k=1}^{K} \sum_{j=1}^{r_A} \left(\frac{\widetilde{y}_j - \boldsymbol{v}_j^\top \boldsymbol{x}^{(k)}}{\tau_j} \right)^2 . \tag{30}$$

Here $\boldsymbol{v}_j$ is the corresponding right singular vector and τ_j the whitened noise standard deviation. The evaluator computes Q from unclipped native-coordinate outputs and then averages over images. This diagnostic measures residual size relative to measurement noise; it is not the additive covariance residual $\boldsymbol{R}_\eta$ and does not characterize full posterior uncertainty. All primary criteria and the secondary energy-score criterion pass on the prefix (Table 5). This internal decision coexists with worse deblurring reconstruction, energy score, and perceptual error than full-diffusion SG.

B.7 Metric definitions and uncertainty intervals

Reconstruction and perceptual error. Let $\bar{\boldsymbol{x}} = K^{-1} \sum_k \boldsymbol{x}^{(k)}$. PSNR is computed from $\bar{\boldsymbol{x}}$ against the ground truth in unit-range physical/image coordinates; scattering outputs are not clipped. For FFHQ, the mean is formed in native coordinates before mapping and clipping to $[0, 1]$ for ensemble-mean PSNR; energy score instead clips each mapped sample individually. Per-sample LPIPS uses the VGG backend on unclipped outputs in the model's native $[-1, 1]$ units (Zhang et al., 2018). Averaging per-sample PSNR or LPIPS is different from evaluating the ensemble mean; tables label the distinction.

Energy score. For unit-range image vectors, the reported estimator is

$$\widehat{\mathrm{ES}} = \frac{1}{K\sqrt{d}} \sum_{k=1}^{K} \left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}_{\mathrm{true}} \right\| - \frac{1}{2K(K-1)\sqrt{d}} \sum_{k \neq l} \left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}^{(l)} \right\| . \tag{31}$$

This pairwise estimator is unbiased for the ensemble distribution's population energy score when outputs are independent. The population score is proper; a finite cohort with four outputs cannot establish all aspects of high-dimensional posterior dependence. The paper reports arithmetic means of raw per-image scores. The original decision rules also used paired per-image ratios; those ratios are retained as decision records, not substituted for the raw proper score in the main table.

Projected ranks. For each fixed projection, let $L_<$ be the number of samples strictly below the truth and $L_=$ the number tied. The randomized rank is $U = (L_< + V(L_= + 1))/(K + 1)$ with a reproducible $V \sim \mathrm{Uniform}(0, 1)$ shared across methods. For each of four projection families, 64 projections estimate the proportion C_α within $[(1-\alpha)/2, (1+\alpha)/2]$, for $\alpha \in \{0.50, 0.80, 0.90, 0.95\}$. ICE averages $|C_\alpha - \alpha|$ over these levels and families. The families comprise three measured-space bands and the numerical null complement. With a separate diagnostic threshold η_{band}, the measured bands are $\tau_i < \eta_{\mathrm{band}}$; $s_i > 10^{-3}$ outside that set; and $s_i \leq 10^{-3}$. The threshold solves $\sum_{\tau_i < \eta_{\mathrm{band}}} \log(\eta_{\mathrm{band}}/\tau_i) = 10$ on the operator spectrum; it is approximately 0.01919 for 180 receivers and is recomputed by the same rule for 360. It is distinct from the sampler endpoint $\eta_\star$. Measured-space functionals are scaled by directional noise levels. Projections of one target are dependent. We resample targets, not the 256 projections, for all uncertainty intervals. This summary is a restricted calibration diagnostic, not a full simulation-based calibration study or an exact-reference distance.

Reference diagnostics. The mixture responsibility score $\widehat{D}$ is an unbiased pairwise estimate of squared discrepancy between the mean sampled responsibilities and the analytic posterior weights. It can be slightly negative due to finite-sample debiasing and does not measure the whole posterior. Band log-variance errors are centered by a matched exact-reference sample statistic before taking absolute values and their maximum. Normalized energy distance compares sampler outputs with independent reference draws and divides by the reference pairwise distance; its sample estimator also has a finite-sample noise floor. Analytic Gaussian comparisons report $\mathbb{E}_{\boldsymbol{y}} \mathrm{KL}(q_{\mathrm{out}}(\cdot \mid \boldsymbol{y}) \| p(\cdot \mid \boldsymbol{y}))$, averaging over the prior-predictive measurement law. Output means and covariances are propagated analytically, not fitted to samples.

Intervals. Descriptive intervals resample target-level paired differences 20,000 times with seed 20260926 and use central percentile endpoints. They are descriptive, without multiplicity adjustment. Frozen accept/reject criteria use their original one-sided bounds and resample counts. The difference between those bounds and a central 95% interval is intentional. Projection, sample, and task counts are never pooled as independent target replications.

C Additional results and diagnostic controls

C.1 Held-out image effects and selection comparisons

Table 3 gives the paired held-out differences behind Figure 3. Table 4 reports mean per-sample perceptual error for all eight methods; this is not perceptual error of the ensemble mean. The frozen internal analysis in Table 5 uses different interval tails and the per-image ratio statistic defined in Equation (29). The same images are paired within each task, without treating tasks as independent target replications.

Table 3: Held-out FFHQ: MM-SG minus full-diffusion SG. Each entry is the mean difference [central 95% paired interval], over 60 images. Energy-score differences are multiplied by 100. Positive PSNR and negative energy score favor MM-SG; the observed signs reverse on deblurring. No multiplicity adjustment is applied.

Task	ΔPSNR (dB)	ΔES ×100
SR ×4	$+0.205$ [$+0.163, +0.247$]	-0.042 [$-0.059, -0.025$]
Box	$+1.260$ [$+0.765, +1.775$]	-0.516 [$-0.831, -0.205$]
Random 70%	$+0.511$ [$+0.458, +0.568$]	-0.097 [$-0.113, -0.081$]
Blur	-0.077 [$-0.131, -0.021$]	$+0.073$ [$+0.046, +0.100$]

Table 4: Held-out FFHQ mean per-sample LPIPS ↓ (60 images, four outputs each). The MM-SG minus full-diffusion SG differences are -0.0142 [$-0.0167, -0.0116$], -0.0174 [$-0.0211, -0.0138$], -0.0175 [$-0.0194, -0.0155$], and $+0.0059$ [$+0.0029, +0.0090$] in task order, with central 95% paired intervals.

Method	SR	Box	Random	Blur
MM-SG	0.212	0.134	0.125	0.242
Full-diffusion SG	0.226	0.151	0.142	0.236
DPS	0.266	0.218	0.158	0.245
DAPS-1K	0.246	0.201	0.190	0.252
RED-diff	0.327	0.153	0.179	0.300
DiffPIR	0.246	0.175	0.257	0.225
DDNM+	0.433	0.220	0.152	0.280
ΠGDM	0.245	0.170	0.134	0.223

Table 5: Internally specified held-out FFHQ criteria. PSNR columns are one-sided 95% lower bounds for MM-SG minus each comparator (dB); the margin is -0.3. The energy-score ratio is the mean per-image ratio [one-sided 95% upper bound] against pseudoinverse-guided diffusion, the lowest-mean-ES fast comparator on each task. $\bar{Q}$ is the mean whitened measurement residual, not a covariance or full-posterior calibration metric. Bounds use 10,000 resamples.

Task	Full SG	DAPS	RED-diff	ES ratio [upper]	$\bar{Q}$
SR ×4	+0.170	+0.084	+1.577	0.957 [0.964]	1.021
Box	+0.848	-0.016	-0.250	0.986 [1.021]	0.998
Random 70%	+0.466	+1.039	+0.370	0.942 [0.951]	1.026
Blur	-0.122	-0.080	+0.791	1.061 [1.071]	1.003

Complete paired comparisons. Figures 5 and 6 extend the main-text effect plots to every comparator in Table 2. They use the same stored target-paired differences, not new runs or a different estimator. Forward-evaluation budgets differ across methods as reported in that table.

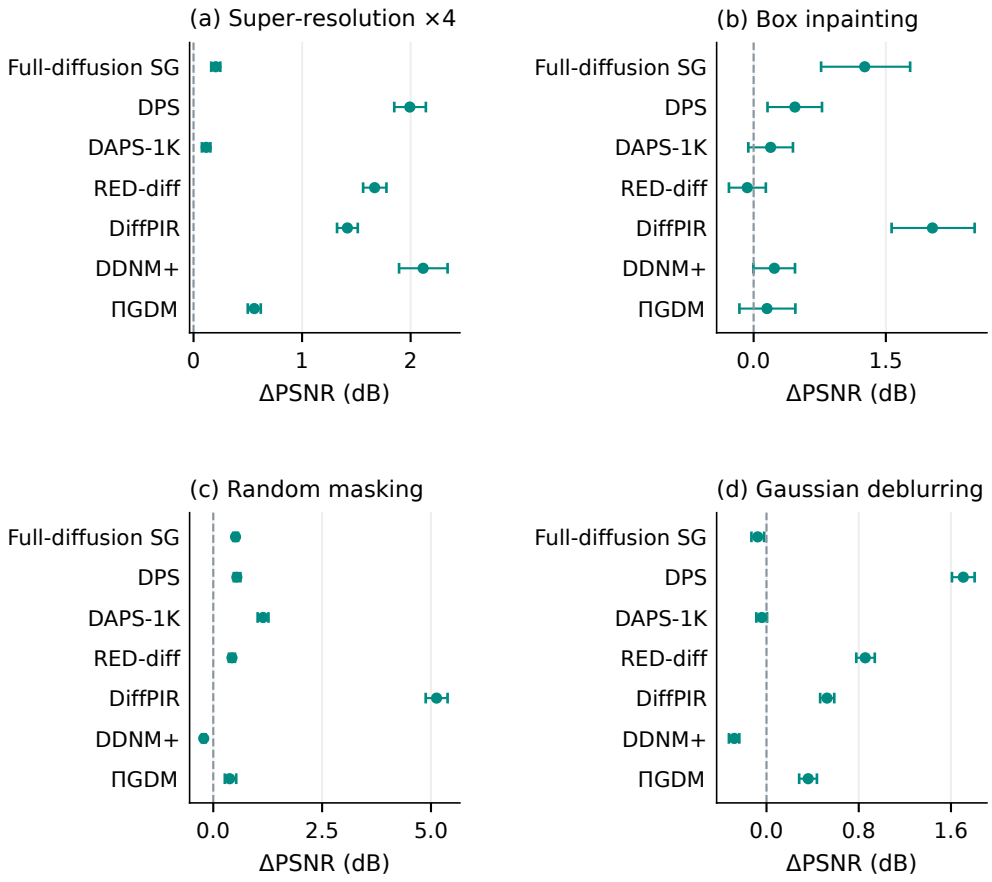


Figure 5: **Held-out FFHQ reconstruction differences against every comparator.** Points show mean ensemble-mean PSNR differences in dB, MM-SG minus the named comparator; positive favors MM-SG. Panels use $4\times$ super-resolution, central-box inpainting, 70% random masking, and Gaussian deblurring. Each comparison pairs the same 60 image IDs within its task, with four outputs per method and image. Horizontal bars are central 95% target-bootstrap intervals (20,000 resamples, no multiplicity adjustment). Axis ranges differ by task; NFE differ as in Table 2. These are the fixed-cutoff held-out prefix, not selection ablations or 240 independent targets.

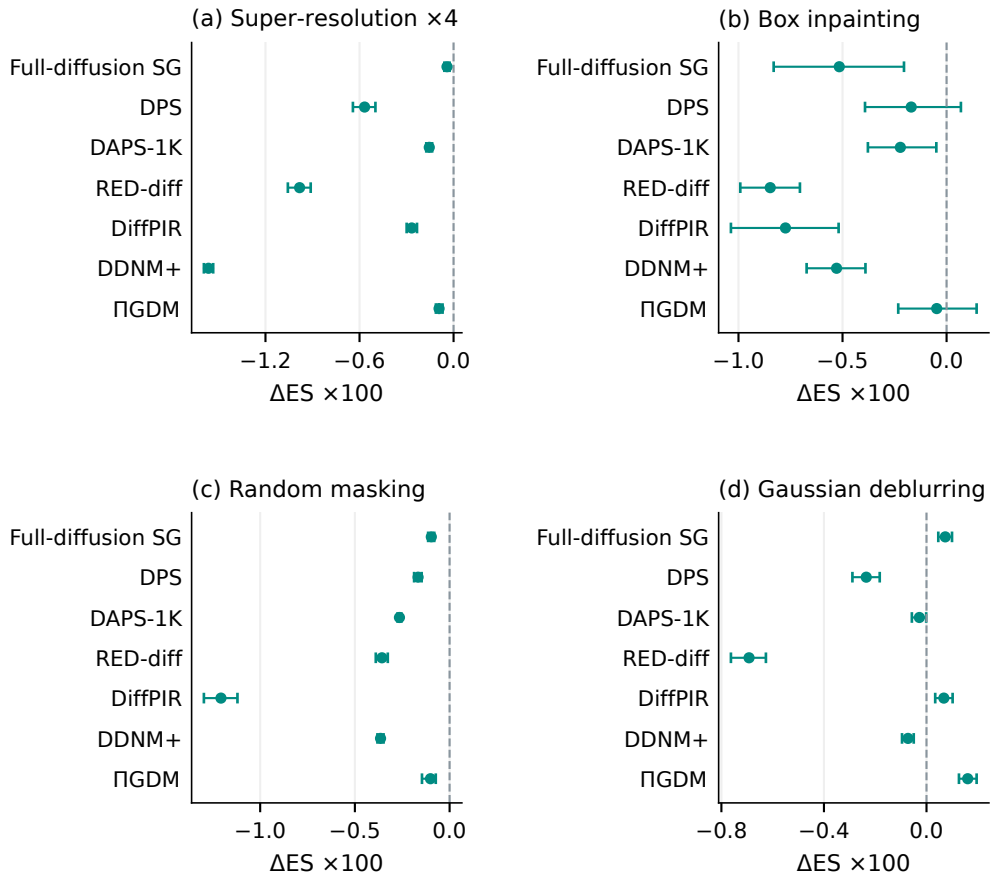


Figure 6: **Held-out FFHQ energy-score differences against every comparator.** Points show mean raw ensemble energy-score differences multiplied by 100, MM-SG minus the named comparator; negative favors MM-SG. Tasks, 60 paired image IDs, four outputs per method and image, and central 95% interval procedure are as in Figure 5. Axis ranges differ by task. These are differences of scores, not the mean per-image ratios used by the frozen internal criterion. All seven comparators are retained, including the favorable deblurring results of the full-diffusion reference and pseudoinverse-guided diffusion.

The broader comparison in Table 6 uses the separate 15-image selection cohort. It includes lower budgets and additional one-call chains that were not run on the held-out prefix. All selection comparisons below retain that cohort, including the component ablations, budget curve, and qualitative panel.

Table 6: **FFHQ selection results.** Each cell is ensemble-mean PSNR ↑ / raw energy score ↓ multiplied by 100. SR: $4\times$ super-resolution; box: central 128×128 inpainting; random: 70% missing pixels; blur: Gaussian deblurring. Budgets are actual forward evaluations, not runtime. Adaptations are detailed in Appendix B.5; Table 9 reports perceptual scores.

Method	NFE	SR	Box	Random	Blur
MM-SG	999	27.90 / 2.93	23.70 / 4.60	31.51 / 1.76	27.69 / 3.21
MM-SG (250)	249	27.90 / 2.93	23.47 / 4.63	31.27 / 1.82	27.90 / 3.10
Full-diffusion SG	3000–3016	27.73 / 2.98	22.35 / 5.45	30.88 / 1.92	27.77 / 3.14
Euler SG	1000	27.32 / 2.88	15.79 / 11.69	31.48 / 1.79	27.17 / 3.50
Annealed Langevin	1000	27.75 / 2.87	15.82 / 11.48	31.57 / 1.77	27.56 / 3.44
Noisy-state SG	1000	27.90 / 2.95	23.90 / 4.99	31.77 / 1.80	27.57 / 3.53
DPS	1000	25.72 / 3.75	23.51 / 4.57	30.83 / 2.02	25.89 / 3.56
DAPS-1K	1000	27.79 / 3.15	23.70 / 4.95	30.58 / 2.04	27.65 / 3.34
RED-diff	1000	26.22 / 4.20	23.60 / 5.99	31.10 / 2.23	26.75 / 4.17
DiffPIR	100	26.53 / 3.24	22.12 / 5.35	26.21 / 3.31	27.11 / 3.15
DDNM+	100	26.33 / 4.45	23.40 / 5.71	31.74 / 2.16	27.93 / 3.30
ΠGDM	100	27.48 / 2.97	23.30 / 4.96	31.24 / 1.88	27.40 / 2.97

C.2 Image ablation values

Table 7 reports all internal scattering controls on the selection cohort. Table 8 reports paired FFHQ differences with target-level intervals. The response-power control has a worse mean raw energy score on each task, but the absence of a faithful published covariance-aware baseline limits the interpretation to the present implementation. Box inpainting illustrates the importance of the aggregation rule: its raw-score difference from the response-power control has a central interval just below zero, while the original paired-ratio criterion did not resolve a gain. We retain both observations and do not revise the frozen decision using the newer aggregation.

Table 7: Scattering selection ablations: seven targets, 50 outputs each, 1,000 forward evaluations per output. PSNR (dB) of the posterior mean and average per-sample PSNR are different estimands.

Update	Mean PSNR ↑	Sample PSNR ↑	ICE ↓
Scalar one-call	38.099	36.579	0.1351
Static one-call	38.079	36.569	0.1267
Pooled response (A1)	38.211	36.655	0.1315
Raw response only	38.262	36.097	0.0665
MM-SG	38.263	36.077	0.0636

Table 8: FFHQ selection: full update minus each internal control; ΔPSNR is in dB. ΔES and its paired 95% interval are multiplied by 100; negative favors the full update. No multiple-comparison correction is applied.

Task	Control	ΔPSNR	ΔES	95% interval for ΔES
SR ×4	Static	-0.021	-0.177	[-0.228, -0.131]
SR ×4	Pooled response	-0.025	-0.177	[-0.225, -0.134]
SR ×4	Raw only	-0.022	+0.002	[-0.017, +0.023]
Box	Static	-0.455	-0.097	[-0.410, +0.225]
Box	Pooled response	-0.217	-0.225	[-0.465, -0.015]
Box	Raw only	+0.030	-0.059	[-0.330, +0.185]
Random 70%	Static	-0.213	-0.105	[-0.134, -0.081]
Random 70%	Pooled response	-0.295	-0.089	[-0.109, -0.071]
Random 70%	Raw only	-0.058	+0.006	[-0.009, +0.022]
Blur	Static	+0.412	-0.516	[-0.627, -0.417]
Blur	Pooled response	+0.261	-0.461	[-0.557, -0.374]
Blur	Raw only	-0.013	+0.008	[-0.018, +0.036]

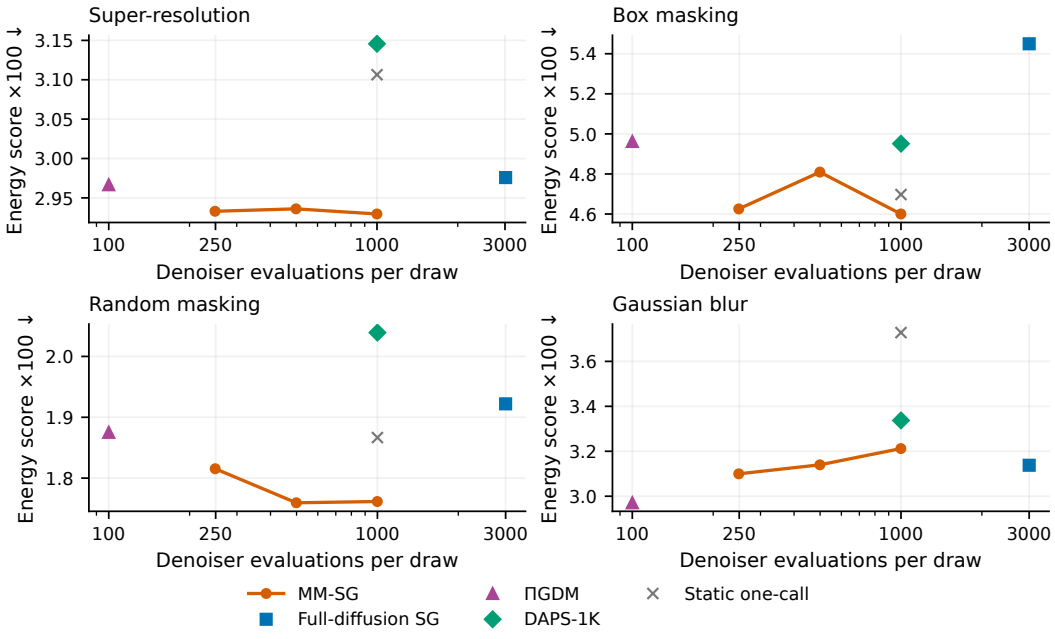


Figure 7: FFHQ selection energy score versus actual online forward evaluations. Each point averages the same 15 images, with four outputs per image. Lines connect MM-SG at 249, 499, and 999 evaluations; other markers are single operating points, not full frontiers. Axis ranges differ by task. More evaluations are not uniformly better, and the 100-evaluation pseudoinverse-guided baseline is competitive. Backward-pass and offline calibration costs are excluded.

Table 9: FFHQ selection average per-sample VGG LPIPS ↓, across four outputs and 15 images per task. These are perceptual sample-quality scores, not posterior calibration measures.

Method	SR	Box	Random	Blur
MM-SG	0.235	0.148	0.127	0.264
MM-SG (500)	0.235	0.147	0.128	0.261
MM-SG (250)	0.235	0.150	0.130	0.256
Static one-call	0.242	0.151	0.127	0.294
Pooled response (A1)	0.243	0.151	0.126	0.291
Raw response only	0.233	0.152	0.126	0.268
Full-diffusion SG	0.245	0.166	0.147	0.257
Euler SG	0.323	0.355	0.159	0.327
Annealed Langevin	0.240	0.358	0.133	0.276
Noisy-state SG	0.229	0.146	0.120	0.275
DPS	0.315	0.249	0.177	0.285
DAPS-1K	0.262	0.205	0.185	0.275
RED-diff	0.360	0.167	0.185	0.331
DiffPIR	0.283	0.195	0.292	0.259
DDNM+	0.424	0.219	0.155	0.302
ΠGDM	0.262	0.190	0.144	0.246

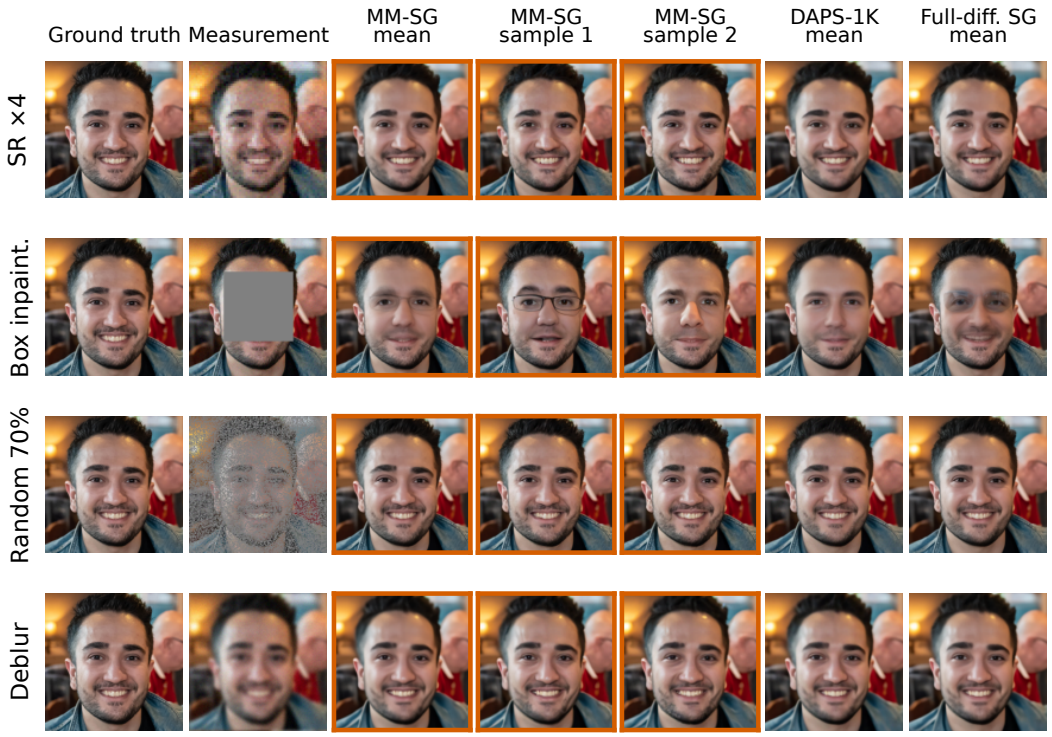


Figure 8: Illustrative FFHQ example from the recorded experiment artifacts. Rows correspond to super-resolution, box inpainting, random inpainting, and Gaussian blur. The panel shows the ground truth, observation, MM-SG posterior mean and two outputs, and comparator ensemble means. Plausible visual variation does not establish posterior correctness.

C.3 EXACT-REFERENCE MIXTURE CONTROLS

V3 and V3PL are 1,024-dimensional Gaussian mixtures with analytic denoisers and posterior samplers. The observation operator measures 640 orthogonal directions using directional noise levels derived from scattering-spectrum quantiles. V3 uses a flat shared within-component spectrum; V3PL uses a power-law spectrum. Component separation includes unmeasured and weakly measured directions. The final cohorts contain 600 and 300 observations, respectively, with 50 outputs per method and observation and a 2,000-draw exact-reference pool. These are internal controlled testbeds, not external image benchmarks. Table 10 reports the final one-call branch selected by the diagnostic; the two-call controls of Figure 9 use the earlier selection observations and must not be interpreted as this final cohort. The V3/V3PL exact-conditional references fix measured coordinates at $\tilde{y}_i$ when $\eta > \tau_i$, whereas MM-SG uses the noise-matching branch of Equation (7). Full- and reduced-schedule exact chains share this reference update, so their comparison tests sweep count without changing the conditional sampler.

Table 10: Final analytic-mixture comparisons. Columns report the debiased responsibility discrepancy $\widehat{D}$ and normalized energy distance multiplied by 10^3. Lower is better. Exact reference draws establish the Monte Carlo floor; exact conditional chains retain finite-schedule error.

Method / budget	V3 $\widehat{D}$	V3 nED	V3PL $\widehat{D}$	V3PL nED
Exact reference draws	-0.00026	-0.011	-0.00056	-0.040
MM-SG / 1000	0.00440	0.112	0.00256	0.204
MM-SG / 3000	0.00157	0.045	0.00284	0.162
Exact conditional / 1000	0.00469	0.136	0.00115	0.097
Exact conditional / 3000	0.00087	0.082	0.00084	0.074
Noisy-state SG / 1000	0.00663	0.193	0.00665	0.490
Noisy-state SG / 3000	0.00121	0.013	0.00550	0.359
Annealed Langevin / 1000	0.01003	0.271	0.00782	0.651
Annealed Langevin / 3000	0.00194	0.116	0.00674	0.480
Full-diffusion SG / 3000	0.03629	1.029	0.01150	1.407

HET and HET-M are 64-dimensional mixtures with six components. Their covariances share a random orthonormal basis; means have independent normal coordinates scaled by 0.3. The observation operator is a separately rotated 40-dimensional subspace with noise levels geometrically spaced from 0.02 to 0.5. HET component spectra are $k^{-\alpha_c}$ for $\alpha_c \in \{0.5, 1, 1.5, 2, 2.5, 3\}$, whereas HET-M spectra are $s_c k^{-1.5}$ with six geometrically spaced scales between $3^{-1/2}$ and $3^{1/2}$. Thus their distinction is covariance heterogeneity. Each has 100 observations and 100 outputs per observation. Figure 9 centers mean log-variance statistics by an equal-size exact-reference ensemble, then takes the largest absolute difference over top, middle, and low spectral bands. All compared HET/HET-M chains use a shared observation-update variant: when $\eta > \tau_i$, the measured coordinate is fixed at $\tilde{y}_i$, without the noise term in Equation (7); the compensated branch is unchanged. The static and two-call chains use 1,000 evaluations; exact controls use matched schedules, not fictitious neural compute costs.

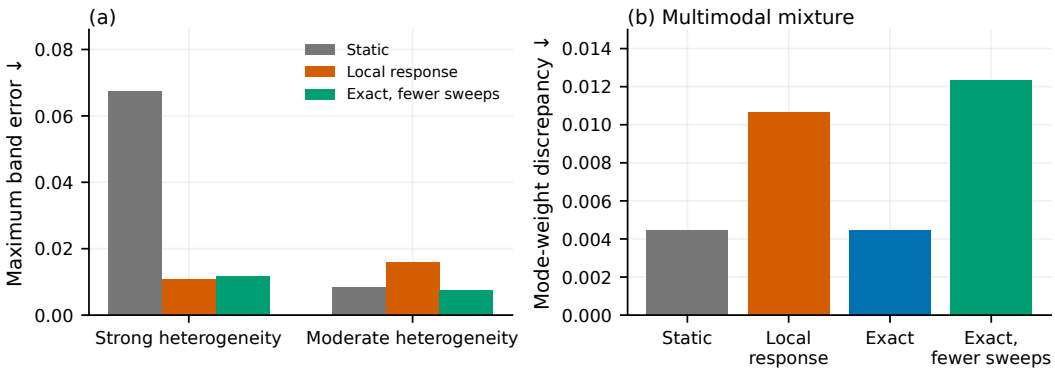


Figure 9: **Exact-reference controls separate covariance error from sweep count.** Left: maximum reference-corrected band log-variance error for heteroscedastic mixtures with stronger (HET) and milder (HET-M) covariance heterogeneity. Right: debiased responsibility discrepancy for a separated mixture (V3). "Exact, fewer sweeps" uses exact conditional draws with the reduced sweep schedule of the two-call method; Appendix C.3 specifies the observation-update variants. Bars summarize the control cohorts (100 targets and 100 outputs per target in HET/HET-M; 600 targets and 50 outputs in V3); they are diagnostic comparisons, not held-out image results. Lower is better on both panels.

C.4 Final Gaussian failure analysis

The seven final Gaussian spectra are fresh perturbations of predeclared spectral families in dimension 1,024. For $k = 1, \ldots, 1024$, their unperturbed shapes are $3k^{-1.2}$, $15k^{-1.6}$, $10k^{-2}$, $10\exp[-(k-1)/150]$, a broken power law with breakpoint 50 and exponents 1 and 3, $8k^{-1.8}$, and $10k^{-1.4}$. Scale, exponent, decay-length, or breakpoint factors are drawn independently log-uniformly between 0.8 and 1.25 as applicable, with eigenvalues floored at 3×10^{-5}. One fresh random basis is shared across the seven rows. Gaussian output laws are propagated analytically conditional on their calibrated tables, so the reported divergences do not depend on fitting moments to sampler outputs.

At each nominal forward budget a row passes the frozen criterion if its divergence is strictly below the isotropic calibration and no larger than the best of Euler SG, noisy-state SG, annealed Langevin, and the full-diffusion reference plus slack. Slack is 0.01 at 1,000 evaluations and 0.002 at 3,000. At least six of seven rows must pass at both budgets. Two image-derived Gaussian rows are additionally required. Their tested static covariance uses a two-dimensional cosine basis, whereas the seven synthetic rows use a calibrated principal-component basis. The image rows pass but are replications with fresh calibration draws, not new covariance spectra. Table 11 shows the synthetic rows. The conjunction fails because only five pass at 3,000 evaluations. This negative result is retained even though other parts of the final evaluation succeed.

The Gaussian exact-conditional reference also fixes measured coordinates at $\tilde{y}_i$ for $\eta > \tau_i$, unlike the noise-matching MM-SG branch. Thus it is a finite-schedule reference, not a control changing only the prior covariance. Spectrum j3 has nearly unchanged MM-SG divergence (0.00543 to 0.00540), while the exact chain drops from about 5.0×10^{-5} to 2.0×10^{-7}. This is consistent with a static-basis floor. For j4, the divergences at 3,000 evaluations are 0.00911 for MM-SG and 0.00912 for the exact chain; both remain limited by the schedule. These mechanisms require different interpretations, and neither supports declaring all finite-budget errors to be covariance errors.

C.5 Learned-prior control and reproducible evidence

The sparse-signal control uses length-64 cumulative Bernoulli–Laplace increments with slab probability 0.1 and unit Laplace scale, from the statistical diffusion-posterior benchmark (Zach et al., 2026). The imputation operator measures 33 coordinates. Four development targets (50, 60, 86, 31) receive 300 independent outputs per run. The reference denoiser uses deterministic forward-backward inference on a windowed lattice; it approximates the Bayes mean up to discretization

Table 11: Final Gaussian posterior divergence, averaged over measurements. "Best ref." is the minimum over the four fixed external sampler-family controls in the criterion; "Exact" uses exact prior conditionals on the same schedule but the high-noise observation rule specified below. The rows test the routed static branch. Budget is nominal; the full-diffusion reference can use extra forward calls. Values are rounded to five decimal places.

Spectrum	Budget	MM-SG	Isotropic	Best ref.	Exact	Pass
j1	1000	0.00017	0.00406	0.00835	0.00006	yes
j2	1000	0.00135	0.06748	0.01973	0.00006	yes
j3	1000	0.00543	0.32963	0.02205	0.00005	yes
j4	1000	0.12985	0.12221	0.13751	0.12984	no
j5	1000	0.00637	2.37741	0.03161	0.00011	yes
j6	1000	0.00115	0.05025	0.01944	0.00005	yes
j7	1000	0.00037	0.00238	0.00375	0.00034	yes
j1	3000	0.00010	0.00252	0.00159	0.00000	yes
j2	3000	0.00130	0.06054	0.00192	0.00000	yes
j3	3000	0.00540	0.31871	0.00270	0.00000	no
j4	3000	0.00911	0.00904	0.00962	0.00912	no
j5	3000	0.00622	2.30763	0.01797	0.00000	yes
j6	3000	0.00109	0.04506	0.00196	0.00000	yes
j7	3000	0.00003	0.00063	0.00093	0.00000	yes

error, assessed by refinement in the experiment artifacts. It is not a Gibbs denoising estimate. Reference posterior pools for energy-distance evaluation are separate.

The L1/O1 curves in Figure 4 change both the denoiser and its calibrated table. Existing crossed-table controls help separate these effects: at 1,260 schedule nodes, the learned denoiser with the reference table gives normalized energy distance 0.01535, while the reference denoiser with the learned table gives 0.00253. Each crossed-table value uses run seed 0; the corresponding original-table values at that same seed are 0.01465 and 0.00234. The original-table curves average seeds 0 and 1 at this node count, giving 0.01469 and 0.00190. Thus exchanging the table alone does not reproduce the reference-denoiser gain in this control. These four development targets and few run repeats support a diagnostic interpretation, not a general ranking. Schedule-node counts 107, 1,260, and 3,334 include the final observation update; the numbers of ordinary sweeps are one smaller.

The registry and plotting scripts reproduce the quantitative figures and tables offline without inference; the overview has a separate editable vector source. Recreating ensembles requires the experiment code, external checkpoints and datasets, and regenerated calibration tables. The qualitative panel is a recorded artifact.

C.6 DEVELOPMENT CHECK OF FINITE-STEP RESPONSE FIDELITY

A fixed development-only gate compared Equation (3) with automatic-differentiation Jacobian–vector products for the frozen FFHQ denoiser. Images 69330 and 69331 were tested at ascending native-grid indices 100, 500, and 900 ($\eta \approx 0.3423, 3.443, 60.82$), with one deterministically seeded corruption of each clean image and an independent Gaussian probe per case. These are development corruptions, not states visited by the sampler. Noisy states were float64, denoiser inputs float32, and $\delta = 0.02\eta$. Initial differentiation failed in legacy activation checkpointing; a validation-only bypass preserved every checked base prediction bitwise without changing sampling code.

This gate assessed a proposed common-interface adaptation of the ConvDCT construction of Schioppa & Salimans (2026, Appendix C), not a native end-to-end comparator. Both the response and the adapter's same-probe noise increment had to agree with their automatic-differentiation counterparts within relative Euclidean error 0.10 (absolute error 10^{-8} for reference norms at most 10^{-8}). All twelve checks failed; response errors ranged from 0.150 to 4.211. The adaptation was closed with zero main-comparison trajectories. Exact cases, seeds, norms, transform settings, and source hashes are in the numerical registry.

A precision diagnostic repeated only image 69330 at index 500. With TF32 disabled, forward-mode and reverse-over-reverse products agreed to 5.62×10^{-6} relative error, while the forward-difference discrepancy remained 0.266 for batched and separate queries. This does not identify an exclusive cause or measure conditional mean, covariance, posterior, or reconstruction error. No sampler, step size, tolerance, or benchmark result changed; native-comparator performance remains unknown.